



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

**Υποστήριξη Δημιουργικότητας στο
πλαίσιο της Διερευνητικής Αναζήτησης
Πληροφορίας**

Διδακτορική Διατριβή

Μαρία Χ. Ταραμίγκου

Αθήνα, Μάιος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Υποστήριξη Δημιουργικότητας στο πλαίσιο της Διερευνητικής Αναζήτησης (Leveraging Creativity with Exploratory Search)

Μαρία Χ. Ταραμίγκου
Διδακτορική Διατριβή

Τριμελής Συμβουλευτική Επιτροπή:

Γρηγόριος Μέντζας, Καθηγητής Ε.Μ.Π. (επιβλέπων)

Ιωάννης Ψαρράς, Καθηγητής Ε.Μ.Π.

Δημήτριος Ασκούνης, Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 16^η Ιουνίου 2020

Επταμελής Εξεταστική Επιτροπή:

Γρηγόριος Μέντζας
Καθηγητής Ε.Μ.Π.

Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Βασίλειος
Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

Ανδρέας-Γεώργιος
Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Δημήτριος Αποστόλου
Αν. Καθηγητής
Πανεπιστημίου Πειραιώς

Χρυσόστομος Δούκας
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Μάιος 2020

.....

Μαρία Χ. Ταραμίγκου

Διδάκτωρ Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών ΕΜΠ

Copyright © **Μαρία Χ. Ταραμίγκου, 2020**

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η διδακτορική διατριβή τοποθετείται στον χώρο της *Ζήτησης Πληροφορίας* προτείνοντας το μοντέλο της *Διερευνητικής Αναζήτησης* πληροφορίας για την υποστήριξη της *Δημιουργικότητας* με ψηφιακά πληροφοριακά συστήματα

Η επισκόπηση της βιβλιογραφίας σε μοντέλα και συστήματα ζήτησης πληροφορίας προέκρινε τους παράγοντες της *εύνοιας των τυχαίων ανακαλύψεων*, του *πλαισίου αλληλεπίδρασης* αναζητητή-συστήματος και της *προσωπικότητας του αναζητητή* με σκοπό να μελετηθεί η επίδρασή τους στην *δημιουργικότητα* στη διάρκεια και μετά από μία συνεδρία *διερευνητικής αναζήτησης*.

Γι' αυτό το σκοπό προτείναμε ένα διαδραστικό μοντέλο διερεύνησης και μία μεθοδολογία παραγωγής προτάσεων, τα οποία υλοποιήσαμε σε πληροφοριακό σύστημα διερεύνησης πληροφορίας με απώτερο στόχο να αποδείξουμε ότι μπορούν να υποστηρίξουν το χρήστη σε εργασίες για τις οποίες δεν υπάρχει μία και καλύτερη απάντηση, αλλά απαιτούν δημιουργικότητα, όπως η παραγωγή ιδεών ή η συγγραφή μίας ερευνητικής πρότασης.

Η προσέγγισή μας αξιολογήθηκε, υπό το πρίσμα της συνεισφοράς της στη δημιουργικότητα, από έναν οργανισμό κατά τη διάρκεια δημιουργικών ενασχολήσεων, καθώς και σε ελεγχόμενες συνθήκες από φοιτητές και ερευνητές υπό το πρίσμα ανάθεσης δημιουργικής εργασίας. Τα αποτελέσματα της αξιολόγησης συνίστανται στα παρακάτω:

- Η σχεδιαστική πρότασή μας αναφορικά με την *ενίσχυση της εύνοιας τυχαίων ανακαλύψεων* υποστηρίζει τη *δημιουργική διαδικασία* και το *δημιουργικό αποτέλεσμα*.
- Η προτεινόμενη μεθοδολογία που αξιοποιεί το *πλαίσιο αλληλεπίδρασης* επιδρά θετικά στο *δημιουργικό αποτέλεσμα*.
- Κάθε ένας από τους παράγοντες της *εύνοιας τυχαίων ανακαλύψεων*, το *πλαίσιο αλληλεπίδρασης* και η *προσωπικότητα* καθώς και ο συνδυασμός όλων μαζί σε μία προσέγγιση ενθαρρύνουν τη διερευνητική αναζήτηση.

Λέξεις Κλειδιά: Διερευνητική Αναζήτηση, Ζήτηση Πληροφορίας, Δημιουργικότητα, Εύνοια Τυχαίων Ανακαλύψεων, Πλαίσιο Αλληλεπίδρασης, Προσωπικότητα, Κρυφά Μαρκοβιανά Μοντέλα

Abstract

The doctoral thesis is positioned in the field of *Information Seeking* and specifically proposes the *Exploratory Search* model to support *Creativity* with digital information systems.

An overview of the literature on information seeking models and systems has identified the factors of *serendipity*, *interactional context*, and *personality* in order to study their impact on *creativity* during and after an *exploratory search* session.

To this end, we have proposed an interactive exploration model and a recommendation generation methodology, which we have implemented in an information exploration system having as an ultimate goal to prove that they support the user in tasks that are not necessarily solved with a “best match” result, but instead require *creativity*, such as idea generation or the writing of a research proposal.

Our approach has been evaluated by an organization during creative engagements (see training workshops) as well as by students and researchers in the context of creative work assignments. The results of the evaluation consist in the following points:

- Our design proposal for enhancing *serendipity* supports the *creative process* and *creative outcome*.
- The proposed methodology utilizing *interactional context* has a positive effect on creative outcome.
- Each of the factors: *serendipity*, *interactional context*, and *personality* as well as all of them combined in a single approach encourage *exploratory process*.

Keywords: Exploratory Search, Information Seeking, Creativity, Serendipity, Interactional Context, Personality, Hidden Markov Models

Ευχαριστίες

Η παρούσα Διδακτορική Διατριβή σηματοδοτεί το τέλος μιας προσπάθειας που πραγματοποιήθηκε στα πλαίσια του προγράμματος μεταπτυχιακών σπουδών της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Η συγκεκριμένη προσπάθεια δε θα μπορούσε να έχει ολοκληρωθεί χωρίς την υποστήριξη κάποιων ανθρώπων πολύ σημαντικών για εμένα για τους οποίους θα νοιώθω πάντοτε μεγάλη ευγνωμοσύνη.

Καταρχήν θα ήθελα να εκφράσω τη βαθιά ευγνωμοσύνη μου για τον επιβλέποντα Καθηγητή μου κ. Γ. Μέντζα, για τη διαρκή συνδρομή, την καθοδήγηση και την ενθάρρυνση που μου προσέφερε. Τον ευχαριστώ ιδιαίτερα για την ευκαιρία που μου έδωσε, καθώς και για την κατανόηση που έδειξε σε ιδιαίτερες συγκυρίες της ζωής μου. Χωρίς τη δική του υποστήριξη η συγκεκριμένη ερευνητική προσπάθεια δε θα είχε ολοκληρωθεί. Θέλω ακόμα να ευχαριστήσω τον Αν. Καθηγητή Δ. Αποστόλου, για το χρόνο που διέθεσε, τις συμβουλές και την επικοινωνιακή του κριτική, τα οποία συνετέλεσαν στην ολοκλήρωση της συγκεκριμένης ερευνητικής εργασίας.

Θα ήθελα να ευχαριστήσω τα άλλα δύο μέλη της τριμελούς συμβουλευτικής επιτροπής, τους Καθηγητές κ. Ι. Ψαρρά και κ. Δ. Ασκούνη, καθώς και τους Καθηγητές κ. Δ. Ασημακόπουλο, Δ. Αποστόλου Α. Σταφυλοπάτη και τον Αν. Καθηγητή κ. Χ. Δούκα για την τιμή που μου έκαναν να συμμετέχουν στην επταμελή επιτροπή εξέτασης της διατριβής.

Θέλω επίσης να ευχαριστήσω ιδιαίτερώς τους συνεργάτες μου Ευθύμιο Μπόθο, Φώτη Παρασκευόπουλο, Κώστα Χρηστίδη, καθώς και τα υπόλοιπα μέλη της ερευνητικής ομάδας Διοίκησης Πληροφοριακών Συστημάτων. Η συνεργασία μαζί τους, οι προτάσεις, η ανταλλαγή ιδεών, αποτέλεσαν βασική πηγή υποστήριξης και έμπνευσης για την αντιμετώπιση των ερευνητικών και τεχνολογικών προκλήσεων της παρούσας διατριβής.

Ένα πολύ μεγάλο ευχαριστώ οφείλω στους γονείς μου, Χαράλαμπο και Άννα, καθώς και στην αδερφή μου Εβελίνα για την απεριόριστη αγάπη και υποστήριξή τους

όλα αυτά τα χρόνια. Είναι τα πρόσωπα που πάντοτε με ενθάρρυναν να διεκδικώ τους στόχους και τα όνειρά μου.

Το πιο μεγάλο ευχαριστώ όμως αναλογεί στους ανθρώπους που στάθηκαν δίπλα μου όλο αυτό το χρονικό διάστημα: στο σύζυγό μου Χρήστο και στην κόρη μας Άννα. Σας ευχαριστώ για τη συμπαράσταση, την υπομονή σας, μα πάνω από όλα για την αγάπη σας, η οποία είναι κινητήριοις δύναμη για μένα ώστε να προχωρώ.

Θέλω ακόμα να ευχαριστήσω τις παλιές συμφοιτήτριες και αδερφικές μου φίλες πια, Σοφία και Στέλλα, για το γεγονός ότι είναι πάντα διαθέσιμες για εμένα για να μοιραστώ τις ανησυχίες ή τις επιτυχίες μου είτε βρίσκονται κοντά είτε μακριά.

Κλείνοντας θα ήθελα να ευχαριστήσω τον πατέρα του συζύγου μου, Δημήτρη, για τις πολύτιμες συμβουλές του κατά την πορεία συγγραφής της Διατριβής.

Η παρούσα Διατριβή είναι αφιερωμένη στους γονείς μου και ιδιαίτερα στον πατέρα μου, μέντορα της ζωής μου, ο οποίος μας παρακολουθεί από «κάπου ψηλά».

Πίνακας Περιεχομένων

Περίληψη	7
Abstract.....	9
Ευχαριστίες	11
Πίνακας Περιεχομένων.....	13
Κατάλογος Εικόνων.....	18
Κατάλογος Πινάκων	23
Ευρετήριο Όρων	26
1 Εισαγωγή	29
1.1 Ερευνητικό Περιβάλλον	29
1.2 Προκλήσεις.....	30
1.3 Στόχοι	32
1.4 Συνεισφορά.....	33
1.5 Σχέση με τις Δημοσιεύσεις.....	34
1.6 Σχέση με Ερευνητικά Έργα.....	35
1.7 Δομή της Διδακτορικής Διατριβής.....	37
2 Επισκόπηση Βιβλιογραφίας	39
2.1 Εισαγωγή.....	39
2.2 Θεωρίες και Μοντέλα Ζήτησης Πληροφορίας	39
2.2.1 Μοντέλο Γρήγορης Αναζήτησης.....	39
2.2.2 Εννοιολογικά Μοντέλα Περιήγησης, Ζήτησης και Ανάκτησης Πληροφορίας με επίκεντρο το Χρήστη.....	41
2.3 Διερευνητική Αναζήτηση	61
2.3.1 Τι είναι η Διερευνητική Αναζήτηση.....	61
2.3.2 Σχεδίαση Συστημάτων Διερευνητικής Αναζήτησης.....	66
2.3.3 Εύνοια Τυχαίων Ανακαλύψεων	67
2.3.4 Εξατομίκευση	68

2.3.5	Επισκόπηση Συστημάτων Διερευνητικής Αναζήτησης	74
2.3.6	Μετρικές Αξιολόγησης	84
2.4	Δημιουργικότητα	92
2.4.1	Ορισμός Δημιουργικότητας	93
2.4.2	Δημιουργική Διαδικασία	95
2.5	Ο ρόλος της Ζήτησης Πληροφορίας στη Δημιουργικότητα.....	99
2.5.1	Αρχές Σχεδίασης συστημάτων Αναζήτησης για την υποστήριξη της Δημιουργικότητας.....	100
2.5.2	Επισκόπηση Συστημάτων Αναζήτησης για την Υποστήριξη της Δημιουργικότητας.....	103
3	Η Πρόταση της Διατριβής.....	105
3.1	Διαμόρφωση του Προβλήματος και Ερευνητικές Προκλήσεις	105
3.2	Ερευνητικά Ερωτήματα.....	109
3.2.1	1 ^ο Ερευνητικό Ερώτημα: Πώς μπορεί να υποστηριχθεί η εύνοια τυχαίων ανακαλύψεων σε σύστημα διερευνητικής αναζήτησης;	111
3.2.2	2 ^ο Ερευνητικό Ερώτημα: Ποιος είναι ο ρόλος της εύνοιας τυχαίων ανακαλύψεων στην υποστήριξη της δημιουργικότητας στο πλαίσιο της διερευνητικής αναζήτησης;	112
3.2.3	3 ^ο Ερευνητικό Ερώτημα: Ποιος είναι ο ρόλος της υιοθέτησης του πλαισίου αλληλεπίδρασης στην υποστήριξη της Διερευνητικής Αναζήτησης και της Δημιουργικότητας;;	112
3.2.4	4 ^ο Ερευνητικό Ερώτημα: Ποιος είναι ο ρόλος της αξιοποίησης της προσωπικότητας του αναζητητή στην υποστήριξη της Διερευνητικής Αναζήτησης και της Δημιουργικότητας;	113
3.3	Πρόταση και Συνεισφορά Διατριβής	114
3.3.1	Ενίσχυση της Εύνοιας Τυχαίων Ανακαλύψεων	115
3.3.2	Υποστήριξη της Δημιουργικότητας μέσω της Εύνοιας Τυχαίων Ανακαλύψεων στο πλαίσιο της Διερευνητικής Αναζήτησης.....	116
3.3.3	Υποστήριξη Διερευνητικής Αναζήτησης και Δημιουργικότητας μέσω ενσωμάτωσης Πλαισίου Αλληλεπίδρασης	117

3.3.4	Υποστήριξη Διερευνητικής Αναζήτησης και Δημιουργικότητας μέσω ενσωμάτωσης Χαρακτηριστικών Προσωπικότητας.....	118
4	Η Επίδραση της Εύνοιας των Τυχαίων Ανακαλύψεων	123
4.1	Εισαγωγή.....	123
4.2	Σχετικές Εργασίες.....	126
4.2.1	Εύνοια Τυχαίων Ανακαλύψεων	126
4.2.2	Ανθρώπινοι Παράγοντες που επηρεάζουν την Εύνοια Τυχαίων Ανακαλύψεων	127
4.2.3	Μελέτες και Μοντέλα Εισαγωγής Εύνοιας Τυχαίων Ανακαλύψεων στην Ανάκτηση Πληροφορίας.....	128
4.2.4	Χαρακτηριστικά Πληροφοριακών Συστημάτων που υποστηρίζουν ευχάριστες και αναπάντεχες ανακαλύψεις.....	130
4.2.5	Ενεργοποιητές Εύνοιας Τυχαίων Ανακαλύψεων στο Σχεδιασμό Διεπαφής Αναζήτησης	135
4.2.6	Ενεργοποιητές Εύνοιας Τυχαίων Ανακαλύψεων στην Παρουσίαση Αποτελεσμάτων Αναζήτησης	137
4.2.7	Ενεργοποιητές Εύνοιας Τυχαίων Ανακαλύψεων στο περιεχόμενο Αποτελεσμάτων Αναζήτησης	139
4.3	Ερευνητική Προσέγγιση.....	142
4.3.1	Διαδραστικό Μοντέλο Διερεύνησης.....	143
4.3.2	Δημιουργία και Απεικόνιση Ενδείξεων (Clues).....	146
4.3.3	Χώρος Αναζήτησης και Διαφοροποίηση των Αποτελεσμάτων ...	148
4.3.4	Παραδίδοντας Απροσδόκητα Αποτελέσματα.....	152
4.4	Υλοποίηση	155
4.4.1	Αρχιτεκτονική Σχεδίαση	155
4.4.2	Σχεδίαση Υπηρεσίας Προτάσεων	158
4.4.3	Παρουσίαση Εργαλείου & Σενάρια Χρήσης	159
4.5	Αξιολόγηση.....	165
4.5.1	Μετρικές Δημιουργικότητας και Εύνοιας Τυχαίων Ανακαλύψεων	165

4.5.2	Εργαστηριακές Μελέτες.....	167
4.6	Αποτελέσματα.....	172
4.6.1	Πρώτη Εργαστηριακή Μελέτη	172
4.6.2	Δεύτερη Εργαστηριακή Μελέτη.....	178
4.6.3	Πιλοτική Μελέτη	186
4.7	Συμπεράσματα.....	191
5	Η Επίδραση του Πλαισίου Αλληλεπίδρασης και της Προσωπικότητας .	193
5.1	Εισαγωγή.....	193
5.2	Σχετικές Εργασίες.....	195
5.2.1	Μοντελοποίηση Πλαισίου Αλληλεπίδρασης.....	195
5.2.2	Μοντελοποίηση Συμπεριφοράς στην Αναζήτηση	197
5.2.3	Ο ρόλος της Προσωπικότητας στη Συμπεριφορά Ζήτησης Πληροφορίας	199
5.3	Ερευνητική Προσέγγιση.....	201
5.3.1	Μοντελοποίηση Πλαισίου Αλληλεπίδρασης.....	202
5.3.2	Παραγωγή Προβλέψεων	207
5.3.3	Σχεδίαση Προτάσεων	208
5.4	Υλοποίηση	210
5.4.1	Συλλογή Δεδομένων Πλαισίου Αλληλεπίδρασης και Προσωπικότητας.....	214
5.4.2	Εκπαίδευση Μοντέλων	223
5.4.3	Παραγωγή Προτάσεων.....	230
5.4.4	Αρχιτεκτονική Σχεδίαση & Παραδείγματα Προτάσεων.....	233
5.5	Αξιολόγηση.....	236
5.5.1	Υποθέσεις	236
5.5.2	Σχεδιασμός Μελέτης.....	238
5.5.3	Πρωτόκολλο Μελέτης	239
5.6	Αποτελέσματα.....	240

5.6.1	Η Επίδραση των Προτάσεων στην Απόδοση της Διερευνητικής Αναζήτησης	240
5.6.2	Η Επίδραση των Προτάσεων Στο Αποτέλεσμα της Εργασίας	243
5.6.3	Επίδραση των Προτάσεων στη Δέσμευση του Χρήστη	246
5.6.4	Αξιολόγηση των Χαρακτηριστικών των Προτάσεων.....	247
5.7	Συμπεράσματα	249
6	Συμπεράσματα και Επόμενα Βήματα	253
6.1	Συμπεράσματα	253
6.2	Μελλοντική Έρευνα	254
	Δημοσιεύσεις και Ανακοινώσεις	257
	Βιβλιογραφία	259
	Παράρτημα 1: Ερωτηματολόγιο εστιασμένο στην υπηρεσία παροχής προτάσεων από το Κοινωνικό Δίκτυο του Χρήστη (Εργαστηριακό Πείραμα)	283
	Παράρτημα 2: Ερωτηματολόγιο Αθροιστικής Αξιολόγησης	284
	Παράρτημα 3: Ερωτηματολόγιο Αξιολόγησης Κεφαλαίου 5	285
	Παράρτημα 4: Επισκόπηση των διαθέσιμων εργαλείων με επίκεντρο τη δημιουργικότητα, την έμπνευση και την ανακάλυψη από πόρους Ιστού ...	287
	Παράρτημα 5: Κώδικας Matlab Αλγορίθμου Προώθησης.....	295

Κατάλογος Εικόνων

Εικόνα 1 Το παραδοσιακό μοντέλο «Γρήγορης αναζήτησης (lookup search)»	40
Εικόνα 2 Γνωσιακό μοντέλο Αλληλεπίδρασης στην Ανάκτηση Πληροφορίας - Peter Ingwersen 1992.....	45
Εικόνα 3 Το γενικευμένο μοντέλο του Wilson στη Συμπεριφορά Πληροφορίας (1996).....	47
Εικόνα 4 ISP: Μοντέλο Διαδικασίας Αναζήτησης Πληροφορίας κατά Kuhlthau ..	48
Εικόνα 5 Information Seeking Model Marchionini 1995	50
Εικόνα 6 Θεωρία Επεξήγησης Brenda Dervin	56
Εικόνα 7 Διερευνητική Αναζήτηση (από Marchionini 2006a).....	63
Εικόνα 8 Διάγραμμα Venn το οποίο τοποθετεί τη Διερευνητική Αναζήτηση σε σχέση με άλλες ερευνητικές επιστήμες (από White & Roth 2009: Exploratory Search Beyond the Query–Response Paradigm)	64
Εικόνα 9 Το Μοντέλο Δημιουργικής Διαδικασίας του Wallas	96
Εικόνα 10 Creative Problem Solving Model (CPS)	97
Εικόνα 11 Δημιουργικές ιδέες που παράγονται από μία ομάδα (Warr & O'Neill, 2005)	98
Εικόνα 12 Απεικόνιση των φάσεων του μοντέλου Genex (Shneiderman, 1999) .	98
Εικόνα 13 Στόχος της Διατριβή είναι η υποστήριξη της διερεύνησης, ανακάλυψης και συγκέντρωσης πληροφορίας με την προτεινόμενη προσέγγιση διερεύνησης πληροφορίας για τη σύνθεση δημιουργικών ιδεών.	106

Εικόνα 14 Τομείς που στοχεύει να υποστηρίξει η Πρόταση της Διατριβής	108
Εικόνα 15 Τα ερευνητικά ερωτήματα και οι αντίστοιχες ερευνητικές εργασίες που πραγματοποιούνται στη Διατριβή.....	111
Εικόνα 16 Οι παράγοντες που μελετά η Διατριβή για την υποστήριξη της Δημιουργικότητας και της Διαδικασίας Διερεύνησης	114
Εικόνα 17 Διαδραστικό Μοντέλο Διερεύνησης Πληροφορίας	144
Εικόνα 18 Υλοποίηση Αλγορίθμου n-Grams	147
Εικόνα 19 Υλοποίηση Αλγόριθμου Διαφοροποίησης	152
Εικόνα 20 Αρχιτεκτονική CRUISE	157
Εικόνα 21 Επισκόπηση Υπηρεσίας Προτάσεων.....	159
Εικόνα 22 Το Cruise με αναδυόμενα παράθυρα που εμφανίζουν επιπλέον πληροφορία όταν ο χρήστης κάνει κλικ είτε σε όρους είτε σε video.	160
Εικόνα 23 Διεπαφή Χρήστη CRUISE.....	162
Εικόνα 24 Ο χρήστης διατυπώνει ένα νέο μονοπάτι αναζήτησης και εξετάζει τα σχετικά αποτελέσματα. Δημιουργείται ένα νέο σύνολο προτάσεων.....	163
Εικόνα 25 Ο χρήστης λαμβάνει εικόνες και λέξεις στο νέφος αντικειμένων σχετικών με το ερώτημά του αλλά και προτάσεις από tweets, εικόνες και άρθρα Ιστού.....	165
Εικόνα 26 Η απόδοση των αναζητητών κατά τη διάρκεια της Διερεύνησης	174
Εικόνα 27 Νέφος Λέξεων που απεικονίζει τις λέξεις και τη συχνότητα στο σύνολο των ερωτημάτων στο CRUISE	174

Εικόνα 28 Νέφος Λέξεων που απεικονίζει τις λέξεις και τη συχνότητα στο σύνολο των ερωτημάτων στο σύστημα αναφοράς.	174
Εικόνα 29 Συγκριτική Απεικόνιση Εύνοιας Τυχαίων Ανακαλύψεων στα δύο υπό αξιολόγηση συστήματα	176
Εικόνα 30 Οι απαντήσεις των συμμετεχόντων σχετικά με το αν η διαδικασία τους βοήθησε να σχηματίσουν ιδέες κατά το τρίτο βήμα της ανάθεσης.....	183
Εικόνα 31 Οι παράγοντες που συνεισφέρουν στο Δείκτη Υποστήριξης Δημιουργικότητας.....	190
Εικόνα 32 Το σύστημα CRUISE με προτάσεις. Σε αυτό το παράδειγμα η προβλεπόμενη επόμενη ενέργεια είναι ένα «Νέο ερώτημα» και το σημασιολογικό πλαίσιο είναι η «Αρχιτεκτονική» που οδηγεί στο προτεινόμενο ερώτημα 'architecture miami beach', που εμφανίζεται στην κορυφή της οθόνης.....	203
Εικόνα 33 Οι αλληλεπιδράσεις χρήστη-συστήματος σε μια συνεδρία, αναπαρίστανται ως σειριακές μεταβάσεις από μία ενέργεια σε άλλη. Κάθε ενέργεια είναι μια παρατήρηση που προκύπτει από μια συγκεκριμένη κρυφή κατάσταση.	204
Εικόνα 34 Κρυφό Μαρκοβιανό Μοντέλο	206
Εικόνα 35 Οι πέντε Διστάσεις της Προσωπικότητας σύμφωνα με το μοντέλο Big Five	213
Εικόνα 36 Το διάγραμμα ράβδων απεικονίζει το εύρος τιμών κάθε χαρακτηριστικού προσωπικότητας ανά συστάδα. (Ανοιχτό μπλε για τη συστάδα 1, μαύρο για τη συστάδα 2, πράσινο για τη συστάδα 3, πορτοκαλί για τη συστάδα 4 και μοβ για τη συστάδα 5.).....	218

Εικόνα 37 Το διάγραμμα Spiderweb περιγράφει τα κεντροειδή κάθε συστάδας. Τα διανύσματα κεντροειδών αποτελούνται από τις μέσες τιμές κάθε χαρακτηριστικού προσωπικότητας στην αντίστοιχη συστάδα.....	218
Εικόνα 38 Άθροισμα της εκτίμησης της Λογαριθμικής Πιθανότητας των παραμέτρων του HMM.....	225
Εικόνα 39 Διαγράμματα Πιθανοτήτων Εκπομπών για HMM 4 καταστάσεων με προσεγγιστικές $1/N$, $1/M$ αρχικές παραμέτρους.	226
Εικόνα 40 Άθροισμα της Λογαριθμικής Πιθανότητας των HMM παραμέτρων.	228
Εικόνα 41 Οι υψηλότερες τιμές Πιθανοτήτων Εκπομπών για κάθε κατάσταση και συστάδα.....	228
Εικόνα 42 Παραγωγή Προτάσεων βάσει Γενικευμένου Μοντέλου	231
Εικόνα 43 Αλγόριθμος Προώθησης (Forward Algorithm).....	231
Εικόνα 44 Αρχιτεκτονική Σχεδίαση Υπηρεσίας Παραγωγής Προτάσεων βάσει Πλαισίου Αλληλεπίδρασης.....	233
Εικόνα 45 Συνδυάζοντας την προβλεπόμενη ενέργεια "Νέο ερώτημα" και την προτίμηση του χρήστη που συνάγεται από τη συλλογή, το σύστημα προτείνει στο χρήστη να αναζητήσει "Βιώσιμη Ενέργεια"	235
Εικόνα 46 Συνδυάζοντας την προβλεπόμενη ενέργεια "Add Video to collection" και το σημασιολογικό πλαίσιο "Ανανεώσιμη Ενέργεια και Ωκεανός", το σύστημα προτείνει ένα σχετικό βίντεο.....	235
Εικόνα 47 Διάγραμμα που συγκρίνει τους μέσους όρους του αριθμού των διερευνητικών ενεργειών «Προσθήκης Ευρήματος στη Συλλογή» και «Διερεύνησης Συλλογής»	243

Εικόνα 48 Απεικόνιση των μέσων όρων των βαθμολογιών αναφορικά με τα κριτήρια αξιολόγησης ενός δημιουργικού αποτελέσματος ή ιδέας στα τρία διαφορετικά εργαλεία244

Κατάλογος Πινάκων

Πίνακας 1 Σύνοψη Μοντέλων Ζήτησης Πληροφορίας.....	57
Πίνακας 2 Σύνοψη Συστημάτων Διερευνητικής Αναζήτησης	81
Πίνακας 3 Ερευνητικά Ερωτήματα	110
Πίνακας 4 Σύνοψη των ερευνητικών ερωτημάτων που στοχεύει να απαντήσει η Διατριβή, των προτεινόμενων λύσεων, του τρόπου αξιολόγησης και αναφορά της ενότητας που αναλύονται	121
Πίνακας 5 Αρχές που Υποστηρίζουν την Εύνοια Τυχαίων Ανακαλύψεων	133
Πίνακας 6 Περιγραφή πρώτου Πειράματος.....	168
Πίνακας 7 Μέσος Όρος, Τυπική Απόκλιση, και Πηλίκο των απαντήσεων του ResQue βάσει κλίμακας από το 1 (Διαφωνώ Απόλυτα) έως το 5 (Συμφωνώ Απόλυτα).....	173
Πίνακας 8 Επισκόπηση των μέσων των βαθμολογιών των επιστρεφόμενων αποτελεσμάτων για το CRUISE και το σύστημα αναφοράς	175
Πίνακας 9 Ποσοτική εκτίμηση της Εύνοιας Τυχαίων Ανακαλύψεων στα επιστρεφόμενα αποτελέσματα για το CRUISE και το Σύστημα Αναφοράς βάσει της εξίσωσης που παρατέθηκε στην ενότητα 4.5.1	176
Πίνακας 10 Ανάλυση των προτάσεων διδακτορικών διατριβών από δύο καθηγητές ως προς: α) την ποιότητα και β) την πρωτοτυπία.....	177
Πίνακας 11 Συμφωνία μεταξύ των καθηγητών στην αξιολόγηση της ποιότητας και της πρωτοτυπίας.....	178
Πίνακας 12 Δημοτικότητα Ερωτήσεων Μεταξύ Συμμετεχόντων.....	178

Πίνακας 13 Ο μέσος όρος, η τυπική απόκλιση και το mode αποκρίσεων για το CRUISE με βάση μια κλίμακα Likert πέντε βαθμών από το 1 (διαφωνώ απόλυτα) μέχρι το 5 (συμφωνώ απόλυτα). Η αντιληπτή δυνατότητα για εύνοια τυχαίων ανακαλύψεων κανονικοποιήθηκε από το 1 έως το 5	179
Πίνακας 14 Ανάλυση των ιδεών αναφορικά με την πρωτοτυπία, ποιότητα, ποσότητα και ποικιλία με βάση τις βαθμολογίες των εμπειρογνομόνων.....	181
Πίνακας 15 Η συνεισφορά των λειτουργιών του CRUISE (δηλ. νέφος λέξεων, είδος πηγών, προτάσεις από το twitter) στην έμπνευση, όπως έγινε αντιληπτή από τους χρήστες.....	182
Πίνακας 16 Ανάλυση συντελεστή συσχέτισης pearson	185
Πίνακας 17 Συμφωνία μεταξύ καθηγητών για την αξιολόγηση τόσο της ποιότητας όσο και της πρωτοτυπίας	186
Πίνακας 18 Αντιληπτή Χρησιμότητα του CRUISE	189
Πίνακας 19 Κωδικοποίηση και Περιγραφή Αλληλεπιδράσεων Χρήστη με το CRUISE	205
Πίνακας 20 Περιγραφή Amazon Mechanical Turk HIT για τη Συλλογή Δεδομένων Αλληλοεπίδρασης.....	212
Πίνακας 21 Χαρακτηριστικά Συμμετεχόντων	214
Πίνακας 22 Ανάλυση ενδεικτικών χαρακτηριστικών συνεδρίας – Ο χρόνος Δέσμευσης μετρίεται σε δευτερόλεπτα.....	214
Πίνακας 23 Κατανομή Ενεργειών Συμμετεχόντων	215
Πίνακας 24 Συσχέτιση χαρακτηριστικών προσωπικότητας των συμμετεχόντων και χαρακτηριστικών συνεδρίας.....	216

Πίνακας 25 Ανάλυση χαρακτηριστικών συνεδρίας ανά συστάδα χρηστών.	220
Πίνακας 26 Μέσες τιμές χαρακτηριστικών προσωπικότητας ανά σύμπλεγμα (κεντροειδή).....	220
Πίνακας 27 Συσχέτιση Spearman’s ή Pearson χαρακτηριστικών προσωπικότητας με χαρακτηριστικά συνεδρίας ανά συστάδα	222
Πίνακας 28 Απεικόνιση των υψηλότερων τιμών Πιθανοτήτων Εκπομπών για κάθε κατάσταση.	226
Πίνακας 29 Περιγραφή Πειράματος.....	240
Πίνακας 30 Αξιολόγηση της συμπεριφοράς του χρήστη ως προς την αναζήτηση στα τρία διαφορετικά εργαλεία	241
Πίνακας 31 Αξιολόγηση των Δημιουργικών Αποτελεσμάτων από τους ειδήμονες	244
Πίνακας 32 Συμφωνία μεταξύ καθηγητών για την αξιολόγηση τόσο της ποιότητας όσο και της πρωτοτυπίας	245
Πίνακας 33 Αξιολόγηση εργαλείου με την Κλίμακα Δέσμευσης Χρήστη	246
Πίνακας 34 Αξιολόγηση των προτάσεων βάσει κλίμακας βαθμολογίας από το 1 ως το 5 για τα συστήματα προτάσεων που βασίζονται στο Γενικευμένο HMM και στο HMM που βασίζεται και στη προσωπικότητα.	248

Ευρετήριο Όρων

Αγγλικός Όρος	Ελληνικός όρος
Artifacts	Τεχνουργήματα
Clues	Ενδείξεις
Clustering	Συσταδοποίηση
Concept designer	Σχεδιαστής εννοιών
Context modelling	Μοντελοποίηση πλαισίου
Coverage	Κάλυψη
Crowdsourcing	Πληθοπορισμός
Diversification	Διαφοροποίηση
Dynamic query systems	Συστήματα δυναμικών ερωτημάτων
Exploratory search	Διερευνητική αναζήτηση
Fact-finding tasks	Εργασίες εύρεσης γεγονότων
Formative evaluation	Διαμορφωτική αξιολόγηση
Hidden Markov models	Κρυφά Μαρκοβιανά Μοντέλα
Hover	πέραςμα κέρσορα πάνω από αντικείμενα
Human Computer Information Retrieval	Ανθρωπο-Υπολογιστική Ανάκτηση Πληροφορίας
Human-Computer Interaction	Αλληλεπίδραση Ανθρώπου-Υπολογιστή
Information Behavior	Συμπεριφορά στην Πληροφορία
Interactional context	Πλαίσιο αλληλεπίδρασης
Information Foraging	Συλλογή Πληροφορίας
Information seeking	Ζήτηση πληροφορίας
Knowledge intensive tasks	Εργασίες έντασης γνώσης
Knowledge worker	Εργάτης γνώσης
Linked Open Data	Ανοιχτά Διασυνδεδεμένα Δεδομένα
Lookup search	Γρήγορη αναζήτηση

Machine learning	Μηχανική μάθηση
Matrix Factorization	Παραγοντοποίηση Μήτρας
Personal computing	Προσωπική πληροφορική
Pointing devices	Συσκευές κατάδειξης
Question-answering	Εργασίες ερωτήσεων-απαντήσεων
Query disambiguation	Αποσαφήνιση ερωτήματος
Randomness	μη προβλεψιμότητα
Recommender systems	Συστήματα προτάσεων
Reinforcement learning	Ενίσχυση μάθησης
Sensemaking	Επεξήγηση
Search engine	Μηχανή αναζήτησης
Serendipity	Εύνοια των τυχαίων ανακαλύψεων
Serendipitous discoveries	Ευχάριστες και αναπάντεχες ανακαλύψεις
Situationalists	Περιπτωσιολόγοι
Social bookmarking site	Κοινωνικός ιστότοπος καταγραφής σελιδοδεικτών
Social chatter	Κοινωνική συζήτηση
Summative evaluation	Αθροιστική αξιολόγηση
Tag cloud	Νέφος επισημάνσεων
Word cloud	Νέφος λέξεων

1 Εισαγωγή

Το παρόν αποτελεί την διδακτορική διατριβή, η οποία τοποθετείται στην τομή των χώρων της Ζήτησης Πληροφορίας και της Δημιουργικότητας.

1.1 Ερευνητικό Περιβάλλον

Η αναζήτηση πληροφορίας είναι μία θεμελιώδης δραστηριότητα της ζωής για την κάλυψη ανθρώπινων αναγκών όπως: α) εύρεση στοιχείων, β) λήψη αποφάσεων ή ακόμα και για γ) διασκέδαση και αναψυχή. Ιστορικά οι άνθρωποι έψαχναν πληροφορίες από την αλληλεπίδραση με άλλους ανθρώπους είτε μέσω της προφορικής ή της γραπτής επικοινωνίας. Από τα τέλη του προηγούμενου αιώνα η αναζήτηση πληροφορίας μπορούσε πλέον να πραγματοποιείται και μέσω αυτοματοποιημένης διαδικασίας, δηλαδή μέσω των μηχανών αναζήτησης Ιστού.

Οι εξελίξεις στα πεδία της αλληλεπίδρασης ανθρώπου-υπολογιστή (HCI) και της ανάκτησης πληροφοριών (IR) ακολούθησαν παράλληλες ροές αμφότερες επιτυγχάνοντας σημαντική επίδραση στις αρχές του 21ου αιώνα. Με τη διασταύρωση αυτών των δύο κλάδων ασχολείται μια ενεργή κοινότητα ερευνητών που έχουν επηρεάσει τις διεπαφές χρήστη για ιστοσελίδες και μηχανές αναζήτησης του Παγκόσμιου Ιστού (World Wide Web) (Marchionini, 2006a, 2006b). Ηγέτης στο χώρο ο Shneiderman, οποίος όχι μόνο χρησιμεύει ως ηγέτης στην κατανόηση του τρόπου με τον οποίο οι άνθρωποι αναζητούν και χρησιμοποιούν ψηφιακούς πόρους, αλλά και αποτελεί πρωτοπόρο στην ανάπτυξη νέων διεπαφών χρήστη που υποστηρίζουν τη διαδικασία αναζήτησης πληροφοριών. Ο Shneiderman βασίστηκε στο υπόβαθρο που είχε στις βάσεις δεδομένων για την αντιμετώπιση ζητημάτων Ανάκτησης Πληροφορίας. Σε συνεργασία με συναδέλφους στην ψυχολογία, πραγματοποιήθηκαν πολλές μελέτες καταλόγων βιβλιοθηκών που τελικά οδήγησαν σε ένα πρόγραμμα εγκατάστασης και δοκιμών σε σταθμούς εργασίας touch panel στη

Βιβλιοθήκη του Κογκρέσου. Η εμφάνιση του WWW- Παγκόσμιου Ιστού πυροδότησε νέες προκλήσεις στο σχεδιασμό και οδήγησε στην έννοια των διεπαφών δυναμικών ερωτημάτων στη δεκαετία του 1990 που προέβλεπε μερικές από τις εξαιρετικά διαδραστικές τεχνικές AJAX που χρησιμοποιούνται στα συστήματα αναζήτησης των πρώτων χρόνων του νέου αιώνα.

Η αποτελεσματική αλληλεπίδραση με αυτοματοποιημένα συστήματα αναζήτησης είναι κρίσιμος παράγοντας στην αναζήτηση πληροφορίας. Οι αλληλεπιδράσεις αυτές μπορεί να διαφέρουν στην πολυπλοκότητα καθώς εκτείνονται από τη βασική εισαγωγή ερωτήματος και επιλογή αποτελεσμάτων μέχρι χειρονομίες και προφορικές διαλογικές αλληλεπιδράσεις. Τα συστήματα αναζήτησης γίνονται πιο έξυπνα και ενήμερα για τους σκοπούς και τα ενδιαφέροντα του χρήστη, καθώς επίσης και του περιβάλλοντός του όταν πραγματοποιεί αναζήτηση. Αυτό καθιστά εφικτό στα συστήματα αυτά να μπορούν να περιμένουν άρα και να προβλέπουν τις ανάγκες του πιο αποτελεσματικά και να βοηθούν το χρήστη στην ολοκλήρωση της εργασίας.

Οι εξελίξεις στη μοντελοποίηση του χρήστη, η σημασιολογική κατανόηση ερωτημάτων και εγγράφων, οι αλγόριθμοι μηχανικής μάθησης, και η τεχνητή νοημοσύνη αλλάζουν τον τρόπο που οι μηχανές αναζήτησης υποστηρίζουν τους χρήστες αλλά και τη συμπεριφορά των χρηστών, το τι προσμένουν, το οποίο επηρεάζει το πώς συλλέγονται τα δεδομένα του χρήστη και τον τύπο της υποστήριξης που του παρέχεται, καθώς έχουμε πλέον οδηγηθεί στα συστήματα αναζήτησης επόμενης γενιάς.

1.2 Προκλήσεις

Η αφθονία των πληροφοριών που υπάρχουν στα ηλεκτρονικά πληροφοριακά συστήματα έχει προκαλέσει την ανάγκη για αποτελεσματικούς τρόπους διευκόλυνσης της ανακάλυψης χρήσιμων πληροφοριών από τους χρήστες. Η αναζήτηση πληροφοριών ακολουθεί κατά κανόνα ένα μοντέλο διαδικασίας

αναζήτησης που περιλαμβάνει την αρχική αναγνώριση και προσδιορισμό μιας ανάγκης πληροφόρησης, ακολουθούμενη από την εξέταση των αποτελεσμάτων αναζήτησης και την επανάληψη του κύκλου έως ότου εντοπιστεί ένα ικανοποιητικό σύνολο αποτελεσμάτων (Marchionini & White, 2007; Shneiderman, Byrd & Croft, 1998).

Ειδικότερα όμως σε εργασίες όπως η παραγωγή καινοτόμων ιδεών ή η λήψη αποφάσεων, η ανακάλυψη παίζει πολύ σημαντικό ρόλο στην έμπνευση και τη δημιουργικότητα που με τη σειρά τους απαιτούνται για την υποστήριξη τέτοιων εργασιών. Σε τέτοιες περιπτώσεις η αναζήτηση πληροφορίας είναι το μέσο στο οποίο προσφεύγει ο χρήστης προκειμένου να τον υποστηρίξει σε ανακαλύψεις που θα βοηθήσουν τη δημιουργικότητά του και κατ' επέκταση στην επίλυση του προβλήματός του με το καλύτερο δυνατό τρόπο. Παρά τον τεράστιο όγκο πληροφορίας που υπάρχει στον Παγκόσμιο και Κοινωνικό Ιστό, ο οποίος είναι δυνητικά ικανός να εμπνεύσει, αλλά και την εμφάνιση εξειδικευμένων μηχανών έμπνευσης, δεν είναι εύκολο για τους χρήστες να ανακαλύψουν πληροφορίες που θα τους εμπνεύσουν. Συνήθως, οι χρήστες αναζητούν την πληροφορία και την έμπνευση στον κατάλογο αποτελεσμάτων των μηχανών που κατά κύριο λόγο ακολουθούν το συμβατικό πρότυπο αναζήτησης. Οι χρήστες γράφουν τα ερωτήματά τους και η μηχανή παρουσιάζει μια λίστα με αποτελέσματα και τότε ο χρήστης μπορεί να διατρέξει τη λίστα και ενδεχομένως να ανακαλύψει έναν πόρο έμπνευσης διαβάζοντας την προεπισκόπηση των αποτελεσμάτων. Σε τέτοιες περιπτώσεις, το μοντέλο της διερευνητικής αναζήτησης - η οποία περιγράφει ένα πλαίσιο προβλημάτων αναζήτησης πληροφοριών που είναι ανοιχτό, επίμονο και πολύπλευρο, όπως αυτό της λήψης αποφάσεων ή της εκμάθησης, και μια διαδικασία αναζητήσεων πληροφοριών που είναι ευκαιριακή, επαναληπτική και πολυτακτική-είναι το καταλληλότερο.

Με βάση τη γνώση μας, ενώ η ζήτηση πληροφορίας καθώς και το μοντέλο της διερευνητικής αναζήτησης έχει αξιολογηθεί από τους ερευνητές σε γενικότερο επίπεδο ικανοποίησης χρήστη, πολύ λίγες εργασίες υπάρχουν αναφορικά με την

αξιολόγηση της δυνατότητας να υποστηρίξει δημιουργικές εργασίες. Σχεδόν σε όλη την έκταση της Διατριβής, γίνεται μνεία και μελετάται η διευκόλυνση της ανακάλυψης χρήσιμου, καινοτόμου και αναπάντεχου περιεχομένου στη διαδικασία αναζήτησης και διερεύνησης – αυτό το φαινόμενο σε όλη την έκταση του κειμένου θα αναφέρεται ως *εύνοια τυχαίων ανακαλύψεων (serendipity)*.

Επιπρόσθετα, στα πλαίσια της εξατομίκευσης της εμπειρίας κατά τη διάρκεια της διερεύνησης του χρήστη αναπαρίσταται η αλληλεπίδρασή του με το σύστημα, άρα προκύπτει το λεγόμενο πλαίσιο αλληλεπίδρασης, προκειμένου να μελετηθεί η επίδρασή του σε Δημιουργικές εργασίες (διαδικασία και αποτέλεσμα). Στο ίδιο πλαίσιο αναπαρίσταται ο αναζητητής με βάση τα χαρακτηριστικά της προσωπικότητάς του, προκειμένου να διερευνηθεί εάν η ενσωμάτωση του παράγοντα της προσωπικότητας έχει κάποια επίδραση όσον αφορά τη Διερευνητική διαδικασία και το Δημιουργικό Αποτέλεσμα.

1.3 Στόχοι

Τέσσερις στόχοι καλύπτονται στα πλαίσια της Διατριβής.

Αρχικά εξετάζεται η επιλογή συγκεκριμένων μεθοδολογιών, σε επίπεδο διαδραστικής διεπαφής, σε επίπεδο αποτελεσμάτων, αλλά και παρουσίασής αυτών στο πλαίσιο της διερευνητικής αναζήτησης, ως προς τη δυνατότητά τους να υποστηρίξουν την εύρεση χρήσιμων και απρόσμενων αποτελεσμάτων, δηλαδή την *Εύνοια Τυχαίων Ανακαλύψεων*.

Στη συνέχεια εξετάζεται κατά πόσο η παραπάνω μέθοδος μπορεί να συμβάλλει στην υποστήριξη του δημιουργικού αποτελέσματος, σε εργασίες παραγωγής νέων ιδεών, σε επίλυση προβλημάτων και γενικότερα σε εργασίες έντασης γνώσης καθώς και στην υποστήριξη της ίδιας της δημιουργικής διαδικασίας.

Η Διατριβή εστιάζει ακόμα στη μοντελοποίηση του πλαισίου αλληλεπίδρασης του αναζητητή με το εργαλείο που χρησιμοποιεί για τις διερευνητικές του εργασίες.

Στοχεύει στην ανίχνευση μοτίβων διερεύνησης προκειμένου να αξιοποιηθούν για την πρόβλεψη της επόμενης κίνησης του χρήστη, η οποία προδραστικά του προτείνεται προκειμένου να τον βοηθήσει στη διαδικασία διερεύνησης. Μέσω της προηγούμενης προσέγγισης στοχεύει ακόμα να εξετάσει ποια είναι η επίδραση των προδραστικών προτάσεων, που βασίζονται στο πλαίσιο αλληλεπίδρασης, στην ικανότητα του αναζητητή να παραγάγει δημιουργικές ιδέες αμέσως μετά από μία συνεδρία αναζήτησης.

Στην παρούσα Διατριβή η προηγούμενη μέθοδος χρησιμοποιείται σα βάση και στη συνέχεια επεκτείνεται ενσωματώνοντας την προσωπικότητα του αναζητητή. Οι προτεινόμενες ενέργειες προσαρμόζονται αναλόγως των μοντέλων διερεύνησης που αντιπροσωπεύουν τον αναζητητή βάσει των χαρακτηριστικών της προσωπικότητάς του. Στη συνέχεια, όμοια αξιολογείται η συνεισφορά στη διερευνητική διαδικασία και στο δημιουργικό αποτέλεσμα.

1.4 Συνεισφορά

Η συνεισφορά της παρούσας εργασίας συνοψίζεται στις παρακάτω διαστάσεις:

Η Διατριβή δείχνει ότι μπορεί να υποστηριχθεί η *εύνοια των τυχαίων ανακαλύψεων* σε σύστημα διερευνητικής αναζήτησης με τη χρήση συγκεκριμένων σχεδιαστικών τεχνικών και αλγορίθμων. Γί' αυτόν το σκοπό παρουσιάζει μεθοδολογίες και αλγορίθμους, τους οποίους εφαρμόζει σε σύστημα διερευνητικής αναζήτησης, αποδεικνύοντας τη θετική τους συνεισφορά ως προς την εύρεση χρήσιμων και παράλληλα απρόσμενων αποτελεσμάτων σε σχέση με τυπικό σύστημα αναζήτησης.

Η Διατριβή δείχνει ότι η υποστήριξη της *εύνοιας των τυχαίων ανακαλύψεων* μέσω συγκεκριμένων σχεδιαστικών τεχνικών και αλγορίθμων σε σύστημα διερευνητικής Αναζήτησης μπορεί να υποστηρίξει τη *δημιουργικότητα*. Η Διατριβή έδειξε πειραματικά και σε πιλοτική μελέτη ότι ένα σύστημα διερευνητικής

αναζήτησης - που σχεδιάστηκε με σκοπό την υποστήριξη της εύνοιας των τυχαίων ανακαλύψεων - ευνοεί σημαντικά τα Δημιουργικά Αποτελέσματα ενώ επιδρά θετικά και στη Δημιουργική Διαδικασία.

Η Διατριβή προτείνει μια καινοτόμο προσέγγιση δημιουργίας προτάσεων βάσει πλαισίου αλληλεπίδρασης αναζητητή και συστήματος, η οποία βοηθά τους χρήστες στη διαδικασία διερεύνησής τους με προτάσεις προς ανεξερεύνητες πτυχές του υπό έρευνα θέματος, αλλά και με προτροπές να πραγματοποιήσουν συγκεκριμένη ενέργεια. Ταυτόχρονα μελετά την επίδραση των προτάσεων ως προς το δημιουργικό αποτέλεσμα της εργασίας του χρήστη, αποδεικνύοντας ότι η αξιοποίηση του πλαισίου αλληλεπίδρασης επιδρά θετικά.

Η Διατριβή προτείνει μια καινοτόμο προσέγγιση προτάσεων βάσει πλαισίου αλληλεπίδρασης και προσωπικότητας του αναζητητή, η οποία βοηθά τους χρήστες στη διαδικασία διερεύνησής τους με προτάσεις και προτροπές, ενώ ταυτόχρονα μελετά την επίδραση των προτάσεων ως προς το δημιουργικό αποτέλεσμα της εργασίας του χρήστη, αποδεικνύοντας ότι η υβριδική προσέγγιση ενώ δεν έχει σημαντική επίδραση στο δημιουργικό αποτέλεσμα, υποβοηθά τη διαδικασία διερεύνησης.

1.5 Σχέση με τις Δημοσιεύσεις

Μεγάλο μέρος της έρευνας που διεξήχθη στο πλαίσιο της διατριβής έχει παρουσιαστεί σε επιστημονικά συνέδρια και έχει δημοσιευτεί σε διεθνή περιοδικά. Στη συνέχεια δίνεται μια αντιστοίχιση μεταξύ των δημοσιεύσεων και της έρευνας όπως αυτή παρουσιάζεται στα κεφάλαια της διατριβής. Ο πλήρης κατάλογος των δημοσιεύσεων και ανακοινώσεων δίνεται στο τέλος της διατριβής, στο παράρτημα «Δημοσιεύσεις και Ανακοινώσεις».

Οι δημοσιεύσεις [J2], [J3], [J4], και [C1], [C3], [C4], [C5], [C6] παρουσιάζουν τη βιβλιογραφική επισκόπηση των παραγόντων που υποστηρίζουν την εύνοια τυχαίων

ανακαλύψουν σε ένα διαδικτυακό πληροφοριακό σύστημα που υιοθετεί μεθοδολογίες αναζήτησης, περιήγησης και συστημάτων προτάσεων, την υλοποίησή τους σε σύστημα διερευνητικής αναζήτησης και την αξιολόγησή τους ως προς την ποιότητα της αναζήτησης και το δημιουργικό αποτέλεσμα, όπως περιγράφονται στο Κεφάλαιο 4.

Τέλος, η δημοσίευση [J1] παρουσιάζει τη μελέτη αξιοποίησης του πλαισίου αλληλεπίδρασης χρήστη-συστήματος και της προσωπικότητας του χρήστη στη διερευνητική αναζήτηση, την υλοποίηση της προτεινόμενης προσέγγισης και την αξιολόγησή της σε εργασία παραγωγής ιδεών, όπως περιγράφεται στο Κεφάλαιο 5.

1.6 Σχέση με Ερευνητικά Έργα

Η παρούσα διδακτορική διατριβή έχει εν μέρει υποστηριχθεί από την Ευρωπαϊκή Επιτροπή μέσω του ερευνητικού έργου των Τεχνολογιών της Πληροφορίας και Επικοινωνίας (Information and Communication Technologies) COLLAGE .

Έργο COLLAGE

Το έργο COLLAGE που χρηματοδοτήθηκε από την ΕΕ αποσκοπούσε να παρουσιάσει λύσεις πληροφορικής τελευταίας τεχνολογίας.

Η εστίαση του έργου ήταν η ενίσχυση της Κοινωνικής Δημιουργικότητας σε συνθήκες συνεργατικής μάθησης σε χώρους εργασίας. Το όραμα του έργου ήταν να σχεδιάσει αναπτύξει και επαληθεύσει ένα καινοτόμο σύνολο υπηρεσιών Κοινωνικής Δημιουργικότητας, για την υποστήριξη της συνεργιστικής διασύνδεσης διαδικασιών εκμάθησης με κοινωνικές υπολογιστικές υπηρεσίες για εμπνευσμένους εκπαιδευόμενους, χώρων κοινωνικής συγγένειας για την αξιοποίηση της έκφρασης και της εξερεύνησης και της μηχανικής κοινωνικών παιχνιδιών για την υποστήριξη της κοινωνικής αξιολόγησης και της ανταμοιβής της δημιουργικής συμπεριφοράς. Οι χρήστες ήταν μαθητές ακαδημαϊκών ιδρυμάτων, επιχειρηματικών περιβαλλόντων

που ασχολούνταν με διαφορετικά περιβάλλοντα μάθησης, καθώς και εργαζόμενοι επάνω στη δημιουργία ερευνητικών πρωτοτύπων

Οι κύριοι στόχοι του έργου ήταν να:

- **Αναπτύξει μια νέα θεωρία κοινωνικής δημιουργικότητας και μάθησης**, βασισμένη σε ένα περιγραφικό μοντέλο που ενσωματώνει υπάρχοντα μοντέλα δημιουργικότητας, αντανakλαστικής μάθησης, παιχνιδιών και κοινωνικών υπολογιστών. Η θεωρία και το σχετικό μοντέλο θα παράσχουν το πλαίσιο που θα καθοδηγεί την εργασία και την ολοκλήρωση του έργου.

- **Υποστηρίξει την έκφραση και τη διερεύνηση δημιουργικών ιδεών** σε καινοτόμους χώρους κοινωνικής συγγένειας ως μαθησιακούς χώρους που διευκολύνουν μια κοινή προσπάθεια (ενδιαφέροντα, στόχους και πρακτικές), ενσωματώνοντας το σχετικό περιεχόμενο το οποίο οι μαθητές μπορούν να επιλέξουν να αλληλεπιδρούν, όπου μπορεί να δημιουργηθεί νέο περιεχόμενο και η μάθηση έχει στόχο να είναι κοινωνική και ευχάριστη.

- **Ενεργοποιήσει τους εκπαιδευόμενους και να προωθήσει την κοινωνική αξιολόγηση** και εκτίμηση με την προηγμένη μηχανική κοινωνικών παιχνιδιών που επικεντρώνεται στην παιχνιδοποίηση των διαδικασιών εκμάθησης και την ενθάρρυνση των χρηστών να συμμετέχουν ενεργά σε χώρους συγγένειας. Η μηχανική κοινωνικών παιχνιδιών θα έχει ως στόχο να παρέχει ανταμοιβές και αναγνώριση στους εκπαιδευόμενους, προκειμένου να τους παρακινήσει και να τους εμπλέξει σε περαιτέρω δημιουργικές δραστηριότητες.

- **Ενισχύσει την έμπνευση και την πρωτοτυπία των μαθητών με καινοτόμα κοινωνικά υπολογιστικά συστήματα** που θα αναγνωρίσουν, εξορύξουν, επεξεργαστούν τον Κοινωνικό Ιστό και θα προτείνουν περιεχόμενο από το Κοινωνικό Δίκτυο καθώς και επιχειρησιακούς πόρους, με σκοπό να εμπνέουν τους μαθητές και να ενθαρρύνουν τον πειραματισμό.

- **Ενοποιήσει το σύνολο Υπηρεσιών Νέφους που θα προκύψουν από το COLLAGE** σε μια πρωτοποριακή λύση, κοινωνικής δημιουργικότητας, σε διαδικασίες εκμάθησης, προσφέροντας συνδυασμούς χώρων κοινωνικής συγγένειας, μηχανικής παιχνιδιών και κοινωνικών υπολογιστικών συστημάτων εύκολα χρησιμοποιήσιμων για τη μεγιστοποίηση της δημιουργικής παραγωγής των μαθητών.

- **Δοκιμάσει, αξιολογήσει και παραγάγει αντίκτυπο με την υπηρεσία Κοινωνικής Δημιουργικότητας του COLLAGE** σε πραγματικές περιπτώσεις μαθησιακών διαδικασιών, να διαδώσει τα αποτελέσματα του έργου σε σχετικές κοινότητες και παράγοντες, καθώς και θα εκμεταλλευτεί εμπορικά τα αποτελέσματα του έργου. Η FIAT SEPIN, η οποία παρέχει κατάρτιση στον όμιλο Fiat και στο αυτοκινητικό του οικοσύστημα, το CEDEP, η κοινοπραξία εκπαίδευσης διοικητικών στελεχών, και η WAAG η δημιουργική τεχνολογική κοινότητα έχουν επιλεγεί προσεκτικά και θα αποτελέσουν τις ιδανικές δοκιμαστικές μονάδες για τον έλεγχο και την επικύρωση των αποτελεσμάτων του COLLAGE.

1.7 Δομή της Διδακτορικής Διατριβής

Η διδακτορική διατριβή έχει δομηθεί ως εξής:

Στο κεφάλαιο 2 παρουσιάζονται βιβλιογραφικά τα κυριότερα μοντέλα στο χώρο της ζήτησης πληροφορίας. Γίνεται αναφορά στο πιο κλασσικό μοντέλο της γρήγορης αναζήτησης, καθώς επίσης παρουσιάζονται τα μοντέλα ζήτησης πληροφορίας με επίκεντρο το χρήστη, δηλαδή αυτά τα οποία λαμβάνουν υπόψιν την αλληλεπίδραση χρήστη-συστήματος, αλλά και παράγοντες περιφερειακούς όπως το πλαίσιο της αναζήτησης. Γίνεται ιδιαίτερη μνεία στο μοντέλο Διερευνητικής Αναζήτησης, καθώς θα αποτελέσει το πλαίσιο της συγκεκριμένης Διατριβής, με επισκόπηση συστημάτων και αναφορά στις μετρικές αξιολόγησης τέτοιων συστημάτων. Πραγματεύεται ακόμα την έννοια της Δημιουργικότητας και το ρόλο της αναζήτησης πληροφορίας στη Δημιουργική Διαδικασία.

Στο κεφάλαιο 3 παρουσιάζεται η διαμόρφωση του προβλήματος και η διατύπωση των σχετικών ερευνητικών ερωτημάτων, τα οποία απαντώνται στην παρούσα διατριβή.

Στο κεφάλαιο 4 παρουσιάζεται η ερευνητική μελέτη της επίδρασης του παράγοντα της εύνοιας των τυχαίων ανακαλύψεων και η υλοποίηση της προτεινόμενης ερευνητικής προσέγγισης σε σύστημα διερευνητικής αναζήτησης. Η αξιολόγηση της μεθόδου πραγματοποιήθηκε σε πιλοτική και εργαστηριακή μέτρηση.

Στο κεφάλαιο 5 παρουσιάζεται η ερευνητική μελέτη της επίδρασης του παράγοντα του πλαισίου αλληλεπίδρασης αναζητητή-συστήματος και της προσωπικότητας του αναζητητή και η υλοποίηση της προτεινόμενης ερευνητικής προσέγγισης στο σύστημα που παρουσιάστηκε στο Κεφάλαιο 4.

Στο κεφάλαιο 6 παρουσιάζονται τα συμπεράσματα και η προοπτική για μελλοντική έρευνα.

Στο τέλος της Διατριβής παρατίθενται ο κατάλογος των δημοσιεύσεων και ανακοινώσεων που πραγματοποιήθηκαν στο πλαίσιο της εκπόνησής της, καθώς και ένας πλήρης κατάλογος με τη βιβλιογραφία που χρησιμοποιήθηκε. Τέλος, συμπεριλαμβάνονται τα παραρτήματα.

2 Επισκόπηση Βιβλιογραφίας

2.1 Εισαγωγή

Το μεγαλύτερο μέρος της ανασκόπησης της βιβλιογραφίας στο κεφάλαιο αυτό συζητά μοντέλα που έχουν προταθεί στους τομείς της Αναζήτησης και Ανάκτησης Πληροφορίας, της Επιστήμης Πληροφορίας και Αλληλεπίδρασης Ανθρώπου-Υπολογιστή, για να περιγράψουν το πώς οι άνθρωποι διερευνούν, αναζητούν, συλλέγουν και χρησιμοποιούν την πληροφορία. Δίδεται ιδιαίτερη έμφαση στο μοντέλο της Διερευνητικής Αναζήτησης καθώς στόχος του είναι, εκτός από την ανάγκη διερεύνησης του θέματος που απασχολεί τον αναζητητή, η σύνθεση της πληροφορίας προκειμένου να προάγει το γνωστικό επίπεδό του προκειμένου να προκύψουν προϊόντα γνώσης. Τέλος παρουσιάζεται η έννοια της Δημιουργικότητας και γίνεται συσχέτισή της με την αναζήτηση πληροφορίας ως μέσο για τη διευκόλυνσή της.

2.2 Θεωρίες και Μοντέλα Ζήτησης Πληροφορίας

2.2.1 Μοντέλο Γρήγορης Αναζήτησης

Σε τυπικά συστήματα αναζήτησης οι αλληλεπιδράσεις υιοθετούν το μοντέλο των ερωτημάτων αναζήτησης και επιλογής αποτελεσμάτων. Το παραδοσιακό μοντέλο «γρήγορης αναζήτησης» (lookup search) είναι θεμελιώδες στη σχεδίαση συστημάτων ανάκτησης πληροφορίας. Εστιάζει στη συνάφεια ερωτήματος-αποτελέσματος και χρησιμοποιείται για να αναπαραστήσει την αλληλεπίδραση σε ένα εύρος εφαρμογών όπως η αναζήτηση στον Ιστό στην πιο γενική της μορφή. Η μέθοδος «10 blue links» (Broder et al. 2010) είναι το κυρίαρχο τρέχον μοντέλο αλληλεπίδρασης που χρησιμοποιείται στις μεγαλύτερες εμπορικές μηχανές αναζήτησης (Google, Yahoo, Bing).



Εικόνα 1 Το παραδοσιακό μοντέλο «Γρήγορης αναζήτησης (lookup search)»

Τα συστατικά στοιχεία του συγκεκριμένου μοντέλου είναι α) η συλλογή εγγράφων στην οποία εκτελείται η αναζήτηση (Documents), β) η αναπαράσταση των εγγράφων τα οποία αποθηκεύονται στη συλλογή (document syrrogates), γ) η πληροφοριακή ανάγκη του χρήστη (information need) και δ) ένα ερώτημα αναζήτησης (Query). Στα συστήματα που υιοθετούν αυτό το μοντέλο, ο χρήστης παρέχει τη δήλωση ενός ερωτήματος αναπαριστώντας τις πληροφοριακές ανάγκες και το σύστημα επιστρέφει μία ταξινομημένη λίστα περιλήψεων των εγγράφων μαζί με τους σχετικούς τους τίτλους και τα URLs, μαζί με επιπρόσθετη πληροφορία όπως το μέγεθος της σελίδας, της θεματικής κατηγορίας.

Οι εργασίες γρήγορης αναζήτησης (lookup tasks) συνήθως ταιριάζουν σε αναλυτικές στρατηγικές αναζήτησης που ξεκινούν με συγκεκριμένα ερωτήματα και αποδίδουν ακριβή αποτελέσματα με μηδαμινή ανάγκη για διερεύνηση και εξέταση του σετ αποτελεσμάτων και σύγκρισης εγγράφων, όπως για παράδειγμα είναι οι εργασίες εύρεσης γεγονότων ή ερωτήσεις-απαντήσεις (fact-finding ή question-answering tasks.) Μία κοινή υπόθεση πίσω από αυτό το μοντέλο είναι ότι η πληροφορία υπάρχει στη συλλογή και οι αναζητητές χρειάζεται απλά να σχηματίσουν το κατάλληλο ερώτημα με ή χωρίς βοήθεια από το σύστημα αναζήτησης. Απαντάται στο κλασικό πρόβλημα Ανάκτησης Πληροφορίας (Information Retrieval-IR problem), όπου ο χρήστης αναμένει την «καλύτερη απάντηση» σε μία πληροφοριακή ανάγκη, από μία συλλογή εγγράφων που αποτελούνται από συγκεκριμένες λέξεις. Η «καλύτερη απάντηση», δηλαδή το έγγραφο που επιστρέφεται αποκαλείται συχνά και «καλύτερο ταίριασμα» (best match).

Το μοντέλο διερευνήθηκε για αρκετά μεγάλο χρονικό διάστημα από ερευνητές όπως οι Bates 1986, Belkin 1982, Ellis 1984, Kuhlthau 1993, Marchionini 1995. Το

μοντέλο βρέθηκε ανεπαρκές στο να αναπαραστήσει το πώς οι άνθρωποι αλληλοεπιδρούν με το σύστημα αναζήτησης και την ενδεχόμενη δυναμική των πληροφοριακών αναγκών στη διάρκεια μιας συνεδρίας αναζήτησης, καθώς επίσης δε λαμβάνει υπόψιν σημαντικούς παράγοντες όπως το πλαίσιο της εργασίας, και τη χρήση της πληροφορίας. Οι εργασίες αναζήτησης πληροφορίας πραγματοποιούνται επίσης στη διάρκεια του χρόνου σε μία ακολουθία επεισοδίων αναζήτησης χρησιμοποιώντας διασπαρμένες πηγές ενώ πολλές φορές οι χρήστες μπορεί να εκτελούν εργασίες αναζήτησης ταυτόχρονα.

2.2.2 Εννοιολογικά Μοντέλα Περιήγησης, Ζήτησης και Ανάκτησης Πληροφορίας με επίκεντρο το Χρήστη

Σε αυτό το κεφάλαιο αναλύονται τα πιο γνωστά μοντέλα με επίκεντρο το χρήστη στη ζήτηση πληροφορίας στα οποία βασίζεται η σχεδίαση για την υποστήριξη της αναζήτησης πληροφορίας. Τα μοντέλα συμπεριφοράς στην αναζήτηση τα οποία έχουν προταθεί στη βιβλιογραφία βασίζονται στην παρατήρηση του τρόπου που οι άνθρωποι αναζητούν πληροφορία είτε μόνοι τους είτε αλληλοεπιδρώντας με ενδιαμέσους όπως οι βιβλιοθηκάριοι (το μοντέλο αναφοράς του βιβλιοθηκάριου - ενός ανθρώπου ειδικού που προσπαθεί να καλύψει την ανάγκη πληροφορίας του πελάτη). Πράγματι το προαναφερθέν μοντέλο παραμένει το βασικό μοντέλο αλληλεπίδρασης σε πολλά συστήματα όπως οι μηχανές αναζήτησης Παγκόσμιου Ιστού.

Η συμπεριφορά στην αναζήτηση πληροφορίας αποτελεί μέρος του ευρύτερου πεδίου *Συμπεριφορά στην Πληροφορία (Information Behavior)* και περιλαμβάνει *σκόπιμη αναζήτηση ή αναζήτηση χωρίς συγκεκριμένο σκοπό*. Τα μοντέλα της αναζήτησης πληροφορίας είναι οδηγοί για τη σχεδίαση διεπαφών αναζήτησης, καθώς και μοντέλων συμπεριφοράς (πχ. για πρόβλεψη click, μοντελοποίηση ικανοποίησης) που αντιστοιχούν στις ανάγκες του χρήστη και αντιπροσωπεύουν τον τύπο των ενεργειών που πραγματοποιούν. Η παραπάνω θεώρηση περιλαμβάνει

διαφορετικά μοντέλα αναζήτησης πληροφορίας, αντικατοπτρίζοντας τους διαφορετικούς τρόπους με τους οποίους οι άνθρωποι ψάχνουν για πληροφορίες.

2.2.2.1 Περιήγηση

Μια βασική διαφορά των συστημάτων που απέχουν από το παραδοσιακό μοντέλο γρήγορης αναζήτησης είναι ότι το αντικείμενο της πληροφορίας μπορεί να είναι άγνωστο ή μη παρόν. Ένα σημαντικό κλάσμα δραστηριότητας αναζήτησης πληροφορίας συμβαίνει έξω από τα συστήματα αναζήτησης, είτε όταν ο αναζητητής επιλέγει ένα αποτέλεσμα και πλοηγείται μακριά από το σύστημα ή περιηγούμενος σε συνδεδεμένες συλλογές από έγγραφα όπως στον Παγκόσμιο Ιστό. Η δραστηριότητα λέγεται *Περιήγηση (Browsing)*. Η Περιήγηση παραμένει μία θεμελιώδης δραστηριότητα αναζήτησης πληροφορίας και χρειάζεται να κατανοηθεί και να υποστηριχθεί καλύτερα στα συστήματα αναζήτησης επόμενης γενιάς.

Οι Camel et al. (1992) αναγνώρισαν τρεις τύπους Περιήγησης α) *Περιήγηση προσανατολισμένη στην Αναζήτηση (search oriented browsing)*, που είναι η διαδικασία εντοπισμού πληροφορίας σχετικής με συγκεκριμένη εργασία, β) *Περιήγηση για Επισκόπηση (review browsing)*: σάρωση πηγών για εντοπισμό χρήσιμης πληροφορίας και γ) *Περιήγηση για Σάρωση (scan browsing)*, όπου οι αναζητητές σαρώνουν για να βρουν πληροφορία χωρίς να την αξιολογούν ή να την ενσωματώνουν τελικά.

Ο Marchionini (1995) αναγνώρισε τρεις πανομοιότυπους τύπους Περιήγησης οι οποίοι ποικίλουν βάσει του στόχου της αναζήτησης και των τακτικών που χρησιμοποιούνται α) *κατευθυνόμενη περιήγηση (directed browsing)*, που περιλαμβάνει συστηματική περιήγηση στην αναζήτηση ενός γνωστού αντικειμένου, β) *ημι-κατευθυνόμενη περιήγηση (semi-directed browsing)*, όπου ο στόχος είναι λιγότερο συγκεκριμένος και οι στρατηγικές αναζήτησης που χρησιμοποιούνται είναι λιγότερο συστηματικές, γ) *μη κατευθυνόμενη περιήγηση (undirected browsing)*, όπου δεν υπάρχει αληθινός σκοπός και υπάρχει ελάχιστη συγκέντρωση.

Ο Hughes (1997) ιχνηλατεί την ιστορικότητα των διαφόρων θεωριών που αναπτυχθήκαν για να εξηγήσουν τη διερευνητική συμπεριφορά. Ο Hughes προσδιορίζει την εσωτερική διερεύνηση ως διαδικασία που περιλαμβάνει διερευνητικές ενέργειες, όπου σημασία έχει η επίδοση στις καθαυτές ενέργειες και όχι η επίτευξη κάποιου ιδιαίτερου στόχου. Αυτό έρχεται σε αντίθεση με την εξωτερική αναζήτηση που καθοδηγείται από συγκεκριμένο σκοπό, όπως την ανάγκη να βρεις φαγητό ή να αποφύγεις κάποιο κίνδυνο.

Η Bates (2007) προτείνει ότι η Περιήγηση είναι μία γνωστική και συμπεριφορική έκφραση της διερευνητικής συμπεριφοράς που αποτελείται από τέσσερα στοιχεία: α) *γρήγορη ματιά σε μια σκηνή (glimpse a scene)*, β) *στόχευση ενός στοιχείου μιας σκηνής οπτικά ή φυσικά*, γ) *εξέταση αντικειμένων ενδιαφέροντος* δ) *φυσική ή εννοιολογική απόκτηση ή εγκατάλειψη αντικειμένων*. Αυτή η αλληλουχία επαναλαμβάνεται απεριόριστα όσο οι άνθρωποι διερευνούν για να καλύψουν την ανάγκη. Τα συστήματα που υποστηρίζουν διερευνητικές συμπεριφορές σε συλλογές πληροφορίας θα πρέπει να παρέχουν επισκοπήσεις συλλογής, τη δυνατότητα να διασχίσει κανείς μονοπάτια μέσα από τη συλλογή, το να συμπεριλάβουν εργαλεία για απομόνωση και σύγκριση πολλαπλών αντικειμένων ενδιαφέροντος και υποστήριξη για επιθεώρηση εγγράφων και αποθήκευση.

2.2.2.2 Ζήτηση Πληροφορίας

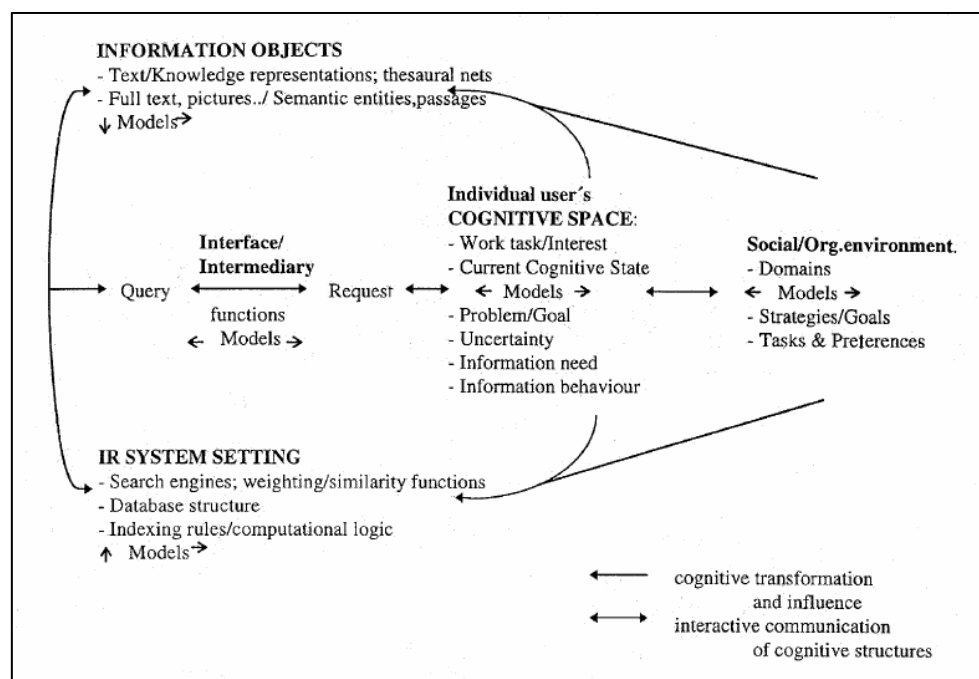
Η ζήτηση πληροφορίας τυπικά υποθέτει ένα μοντέλο της διαδικασίας αναζήτησης, το οποίο περιλαμβάνει αρχικώς αναγνώριση της πληροφοριακής ανάγκης, ακολουθούμενη από την εξέταση των αποτελεσμάτων αναζήτησης και επανάληψη του κύκλου μέχρι να βρεθεί ένα ικανοποιητικό σετ αποτελεσμάτων (Schneiderman et al. 1998; Marchionini and White 2007).

Ένας ικανός αριθμός μοντέλων διαδικασίας ζήτησης πληροφορίας και συμπεριφοράς στη ζήτηση πληροφορίας έχει αναπτυχθεί. Οι Belkin et al. (1982) στο μοντέλο “*Ανώμαλων Καταστάσεων της Γνώσης*” ASK (Anomalous States of Knowledge) για ανάκτηση πληροφορίας δηλώνουν ότι υπάρχει κενό μεταξύ αυτού

που κάποιος γνωρίζει και αυτού που θα ήθελε να γνωρίζει και η ανάγκη γεφύρωσης αυτού του κενού οδηγεί στη ζήτηση και ανάκτηση πληροφορίας. Αυτή η έλλειψη πληροφορίας οδηγεί στον προς επίτευξη στόχο σαν αποτέλεσμα της διαδικασίας αναζήτησης. Το μοντέλο ASK είναι μια κρίσιμη απάντηση σε ένα προηγούμενο (και ακόμα κυρίαρχο) μοντέλο, το οποίο θεωρεί ότι το ερώτημα ενός χρήστη είναι μια άμεση δήλωση της πληροφοριακής του ανάγκης. Σύμφωνα με αυτό το προηγούμενο μοντέλο, η εργασία ανάκτησης πληροφοριών είναι να βρει το έγγραφο ή τα έγγραφα που ταιριάζουν καλύτερα με το ερώτημα, με την έννοια ότι είναι τα πιο σχετικά με αυτό. Η θεωρία των Ανώμαλων Καταστάσεων της Γνώσης (ASK) απαντά αντίθετα ότι, *«ενώ ο χρήστης γνωρίζει ότι υπάρχει κάτι που δεν γνωρίζει αλλά θέλει, δεν γνωρίζει ακριβώς αυτό που δεν γνωρίζει»*. Υπάρχει λοιπόν μια θεμελιώδης αναντιστοιχία μεταξύ εγγράφων και ερωτημάτων. Ένα έγγραφο είναι μια δήλωση του τι γνωρίζει ο συγγραφέας του. Ένα ερώτημα από την άλλη είναι μια δήλωση για το τι ο χρήστης δεν γνωρίζει. Η διάκριση μεταξύ των μοντέλων καλύτερης αντιστοίχισης (best-match) και ASK μπορεί να φαίνεται μόνο θεωρητική, αλλά έχει σημαντικές πρακτικές συνέπειες. Το μοντέλο με την καλύτερη αντιστοίχιση τείνει να απεικονίζει την ανάγκη πληροφόρησης του χρήστη ως στατική και μεταδοτική σε ένα ενιαίο, καλά διαμορφωμένο ερώτημα, μια κατανόηση που ταιριάζει καλά με την ανάκτηση πληροφοριών σε κατάσταση παρτίδας (batch mode). Αντίθετα, το μοντέλο ASK αναγνωρίζει ότι η ανάγκη πληροφόρησης του χρήστη είναι γενικά ασαφής στην αρχή και διευκρινίζεται και επεκτείνεται ως απάντηση στην διερεύνηση του χώρου πληροφοριών. Το μοντέλο, επομένως, παρουσιάζει μια κατανόηση της διαδικασίας ανάκτησης πληροφοριών που είναι εγγενώς επαναληπτική και διαδραστική. Στο μοντέλο του Belkin εντοπίζονται 4 διαστάσεις: α) η *επεξηγηματική περιγραφή* του προβλήματος ASK και όχι ένα ερώτημα το οποίο θα δηλώνει το τι θέλει ο χρήστης να βρει, β) ο *σηματισμός της αναπαράστασης του προβλήματος* από το παραπάνω δεδομένο, γ) ο *κατάλληλος μηχανισμός ανάκτησης*, καθώς ο Belkin θεώρησε ότι για κάθε πρόβλημα δεν είναι κατάλληλος ένας μηχανισμός επιστροφής αποτελεσμάτων, υπάρχουν προβλήματα όπου το «καλύτερο ταίριασμα» ταιριάζει και άλλα που δεν

ταιριάζει και δ) η αξιολόγηση – ανατροφοδότηση από το χρήστη: κάθε ανάκτηση εξαρτάται από την αξιολόγηση των αποτελεσμάτων της προηγούμενης ανάκτησης.

Το μοντέλο του Ingwersen (1992) κατευθύνει τους ερευνητές προς το να λάβουν υπόψιν μεγάλη ποικιλία παραγόντων στη διαδικασία ζήτησης πληροφορίας. Το μοντέλο ενσωματώνει έρευνα σχετικά με την ανάκτηση πληροφορίας και σε επίπεδο συστημάτων και σε γνωστικό επίπεδο. Η προσανατολισμένη προς τα συστήματα έρευνα σχετικά με την ανάκτηση πληροφορίας περιλαμβάνει συλλογές κειμένων, αναπαραστάσεις κειμένων, ερωτήματα και τεχνικές ανάκτησης. Η έρευνα που αφορά

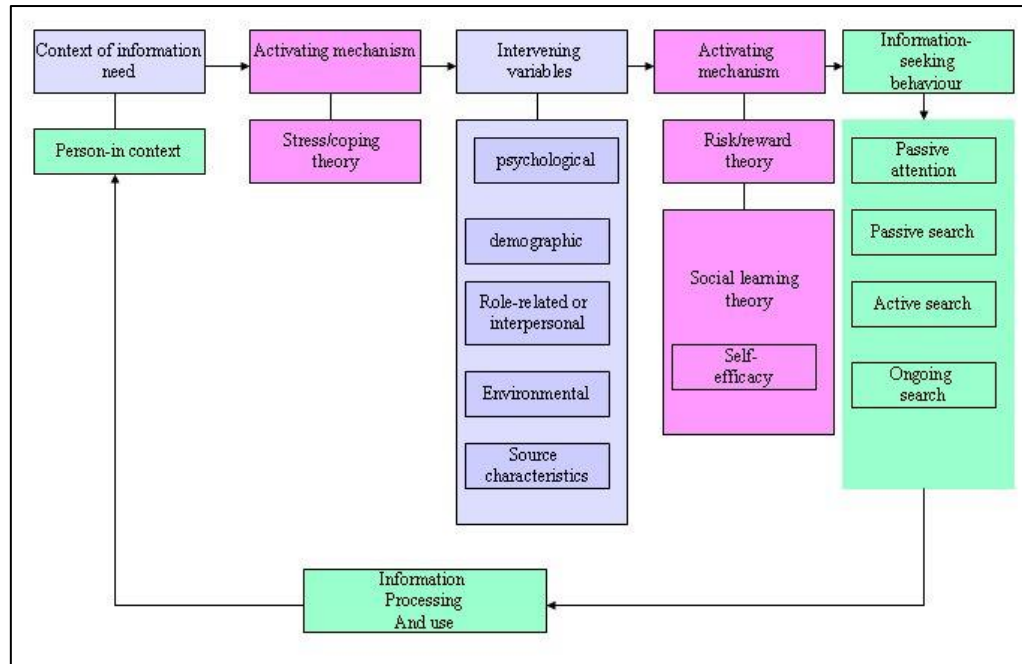


Εικόνα 2 Γνωστικό μοντέλο Αλληλεπίδρασης στην Ανάκτηση Πληροφορίας - Peter Ingwersen 1992

το γνωστικό άξονα της ανάκτησης πληροφορίας περιλαμβάνει το χώρο προβλήματος του αναζητητή, αιτήματα, αλληλεπιδράσεις μεσολαβητών και σχεδίαση διεπαφών. Το μοντέλο του Ingwersen περιλαμβάνει το πλαίσιο ή το κοινωνικο-οργανωτικό περιβάλλον του αναζητητή πληροφορίας. Αργότερα το μοντέλο άλλαξε συμπεριλαμβάνοντας και την αλληλεπίδραση αναζήτησης προσθέτοντας την εργασία, αλλά και την αντίστοιχη κατάσταση όπως την αντιλαμβάνεται ο αναζητητής.

Η γεωγραφική συνάφεια έχει επίσης εφαρμογή στο γνωστικό μοντέλο του Ingwersen στην αλληλεπίδραση μεταξύ εργασιών και αναζητητή.

Ο Wilson πρώτα δημοσίευσε δύο μοντέλα (1981): ένα σχετικό με τη χρήση πληροφορίας και ένα σχετικό με τις πληροφοριακές ανάγκες και την ζήτηση πληροφορίας. Αργότερα αναθεώρησε τα μοντέλα ουσιαστικά και τα ενοποίησε σε ένα γενικευμένο μοντέλο (Wilson 1996,1999). Μία από τις κυριότερες αλλαγές ήταν να εισάγει θεωρίες από άλλα πεδία, ώστε να επεκταθεί η δυναμική των μοντέλων με σκοπό τη δημιουργία υποθέσεων αιτιατών σχέσεων. Ο Wilson υποθέτει ότι μία αντιληπτή πληροφοριακή ανάγκη είναι περιορισμένη σε ένα πλαίσιο. Σύμφωνα με τη θεωρία αντιμετώπισης πίεσης (stress), η αντιληπτή ανάγκη από μόνη της μπορεί να μην ενεργοποιήσει ενέργειες εάν το επίπεδο πίεσης-άγχους δεν είναι αρκετά υψηλό. Παρεμβαίνουσες μεταβλητές όπως δημογραφικές, ψυχολογικές και περιβαλλοντολογικές επηρεάζουν τη συμπεριφορά ή ακόμα μπορεί να παρεμβαίνουν ως εμπόδια στην υποστήριξη δραστηριοτήτων ζήτησης πληροφορίας. Η προσπάθεια αναζήτησης πληροφορίας μπορεί να εξηγηθεί από τη θεωρία ρίσκου/ανταμοιβής και τη θεωρία κοινωνικής μάθησης. Το μοντέλο περιλαμβάνει επίσης ενεργητική και παθητική πληροφοριακή συμπεριφορά. Ο Wilson αναγνώρισε τέσσερις κατηγορίες αναζήτησης και κτήσης πληροφορίας α) *Παθητική Προσοχή (Passive attention)*: το να λαμβάνεις πληροφορία για την οποία δεν είχες σκοπό, για παράδειγμα ακούγοντας ραδιόφωνο ή βλέποντας τηλεόραση, β) *Παθητική Αναζήτηση (Passive search)*: το να βρίσκεις χρήσιμο υλικό που αφορά άλλη εργασία, ψάχνοντας πληροφορία για μια εργασία, γ) *Ενεργητική Αναζήτηση (Active Search)* όταν ενεργά ψάχνεις πληροφορία για να καλύψεις μία συγκεκριμένη ανάγκη και δ) *Συνεχής Αναζήτηση (Ongoing Search)*: η ενεργητική αναζήτηση συνεχίζεται στο χρόνο -κοινή δραστηριότητα για διευρυμένους σκοπούς αναζήτησης. Υπάρχει επίσης ένας βρόγχος ανάδρασης μέσω του οποίου η επεξεργασία της πληροφορίας και η χρήση της αξιολογούνται για να αποφασιστούν επόμενες ενέργειες.

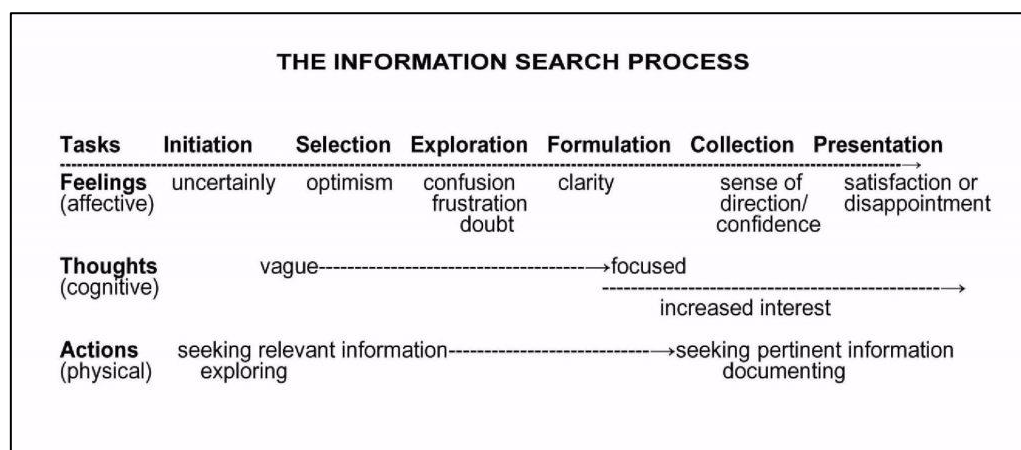


Εικόνα 3 Το γενικευμένο μοντέλο του Wilson στη Συμπεριφορά Πληροφορίας (1996)

Η Kuhlthau (Kuhlthau 1991; 1993) πρότεινε ένα μοντέλο για τη διαδικασία αναζήτησης πληροφορίας (ISP-Information Search Process) το οποίο αναγνωρίζει και δίνει έμφαση στη σημασία των επιμέρους σταδίων που περιλαμβάνονται στις εργασίες εκμάθησης, αλλά και επίλυσης προβλημάτων. Το μοντέλο που ανέπτυξε τονίζει τη διαφορά στα συναισθήματα, τις σκέψεις και ενέργειες που οι αναζητητές βιώνουν στη διάρκεια της αναζήτησης. Οι αλλαγές σε όλα τα παραπάνω εξαρτώνται από το στάδιο της αναζήτησης και κάθε εργασία είναι μοναδική στο στάδιο της διερευνητικής διαδικασίας. Τα στάδια κατά Kuhlthau συνίστανται στα εξής: α) *εκκίνηση (initiation)*: ο χρήστης συνειδητοποιεί την ανάγκη του για πληροφορία όταν παρουσιάζεται το πρόβλημα, β) *επιλογή (selection)*: το γενικό θέμα της αναζήτησης αναγνωρίζεται και επιλέγεται, γ) *διερεύνηση (exploration)*: ζήτηση και διερεύνηση πληροφορίας στο γενικευμένο θέμα, δ) *εστίαση (focus formulation)*: δόμηση του προβλήματος προς επίλυση, ε) *συλλογή (collection)*: συγκέντρωση σχετικής πληροφορίας για το εστιασμένο θέμα και στ) *παρουσίαση (presentation)*: ολοκλήρωση ζήτησης, αναφοράς και χρήσης του αποτελέσματος. Υπάρχει επίσης μία πρόσθετη φάση εκτίμησης στο τέλος της διαδικασίας, στην οποία οι αναζητητές

συλλογίζονται επάνω στα αποτελέσματα της αναζήτησης καθώς επίσης και της διαδικασίας που χρησιμοποίησαν. Μέχρι τη δημοσίευση της εργασίας της, η Kuhlthau ήταν η μόνη ερευνήτρια που είχε εισαγάγει τις ψυχολογικές πτυχές της αναζήτησης στη ζήτηση πληροφορίας.

Αναγνωρίζοντας τη φύση του συναισθήματος στις διαφορετικές στιγμές κατά τη διάρκεια της συνεδρίας αναζήτησης, χρησιμοποιώντας μοντέλα σαν της Kuhlthau επιτρέπεται στο σύστημα να παρέχει υποστήριξη ανάλογη με τη φάση, όπως υποστήριξη στο σχηματισμό ερωτημάτων, ώστε να μειωθεί η αρχική αβεβαιότητα ή καλύτερη υποστήριξη στη διερεύνηση για μείωση της σύγχυσης και της απογοήτευσης.



Εικόνα 4 ISP: Μοντέλο Διαδικασίας Αναζήτησης Πληροφορίας κατά Kuhlthau

Οι Ellis (1989), Ellis et al. (1993) και Ellis & Haugan (1997) πρότειναν ένα μοντέλο αναζήτησης πληροφορίας βασισμένο σε μελέτες που έγιναν στα μοτίβα αναζήτησης κοινωνικών επιστημόνων, ερευνητών φυσικής, χημικών και μηχανικών σε μία βιομηχανική εταιρεία. Η μελέτη του ανέδειξε οχτώ κοινά χαρακτηριστικά των μοτίβων που επιλέχθηκαν: α) *Εκκίνηση*: περιλαμβάνει το μέσο που χρησιμοποιήθηκε για την αναζήτηση πληροφορίας β) *Δημιουργία Αλυσίδων-Chaining*: το να ακολουθείς τους δείκτες ή τις αναφορές από μία αρχική πηγή ή υποσημειώσεις ή παραπομπές της αρχικής πηγής γ) *Περιήγηση*: ημι-δομημένη δραστηριότητα ή ημι-κατευθυνόμενη αναζήτηση σε πιθανούς χώρους πληροφορίας. Μπορεί να οδηγήσει

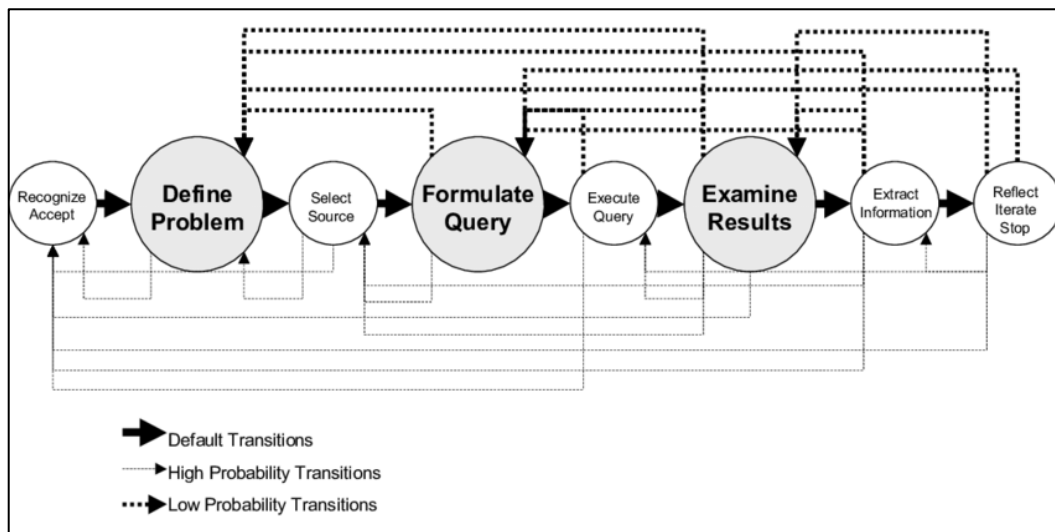
σε πιθανή αποκάλυψη απρόσμενης πληροφορίας, δ) *Διαφοροποίηση*: ο αναζητητής φιλτράρει την πληροφορία διαφοροποιώντας με βάση το περιεχόμενο και την ποιότητα, ε) *Παρακολούθηση*: Συχνά ακολουθεί την εξέλιξη σε συγκεκριμένα πεδία μέσω συγκεκριμένων καναλιών και πηγών, στ) *Εξαγωγή*: επιλεκτική αναγνώριση σχετικού υλικού σε πηγή πληροφοριών, ζ) *Επαλήθευση*: έλεγχος της ακρίβειας της πληροφορίας, η) *Εύρεση*: δραστηριότητες που ολοκληρώνουν τη διαδικασία αναζήτησης.

Οι περισσότερες περιπτώσεις ζήτησης πληροφορίας κατοπτρίζονται επάνω στο μοντέλο του Ellis. Ωστόσο, το μοντέλο δε συλλαμβάνει τις κύριες πτυχές των διεργασιών διερευνητικής αναζήτησης (Marchionini 2006a; White & Roth 2009). Επίσης, το μοντέλο δεν περιλαμβάνει διαδικασίες ανάκτησης και είναι μονής κατεύθυνσης.

Οι Choo et al. (2000) ανέπτυξαν ένα μοντέλο ζήτησης πληροφορίας στο διαδίκτυο το οποίο συνδυάζει περιήγηση και αναζήτηση. Προτείνουν ότι το μεγαλύτερο μέρος του μοντέλου του Ellis είναι ήδη υλοποιημένο από συστατικά στοιχεία τα οποία διατίθενται ήδη από τους εξυπηρετητές Ιστού. Ο χρήστης που αναζητά μπορεί να ξεκινήσει σε μία ιστοσελίδα (εκκίνηση), να ακολουθήσει συνδέσμους σε πόρους πληροφορίας (δημιουργία αλυσίδων), να αποθηκεύσει ιστοσελίδες (διαφοροποίηση), να εγγραφεί σε υπηρεσίες που προσφέρουν ειδοποιήσεις μέσω email (παρακολούθηση) και να αναζητήσει πληροφορία μέσα σε ιστοσελίδες ή πληροφοριακούς πόρους (εξαγωγή).

Ο Marchionini (1995) πρότεινε ένα άλλο μοντέλο κυρίως για ψηφιακά περιβάλλοντα. Στο μοντέλο αυτό η διαδικασία ζήτησης πληροφορίας αποτελείται από οχτώ υπο-διεργασίες, οι οποίες αναπτύσσονται παράλληλα πιθανόν σε διαφορετικούς ρυθμούς: α) *αναγνώριση* και *αποδοχή* ενός πληροφοριακού προβλήματος, β) *προσδιορισμός* και *κατανόηση* του, γ) *επιλογή* ενός *συστήματος αναζήτησης*, δ) *σηματισμός* ενός *αιτήματος*, ε) *πραγματοποίηση αναζήτησης*, στ)

διερεύνηση των αποτελεσμάτων, ζ) εξαγωγή πληροφορίας, η) συλλογισμός/επανάληψη/διακοπή



Εικόνα 5 Information Seeking Model Marchionini 1995

Το μοντέλο είναι πιο κατάλληλο από του Elli's για ψηφιακά και διαδικτυακά περιβάλλοντα. Επίσης, εκφράζει πολύ σημαντικά στοιχεία της ζήτησης πληροφορίας, όπως την πτυχή της συλλογής πληροφορίας στη διαδικασία εξέτασης των αποτελεσμάτων και πτυχές απόκτησης γνώσης κατά την εξαγωγή πληροφορίας. Ωστόσο απουσιάζει η εκμάθηση και κατανόηση, όπως και το πλαίσιο αλλά και η χρήση της πληροφορίας.

Ο Bates (1979;1990) πρότεινε ένα μοντέλο το οποίο περιλαμβάνει τέσσερα στάδια στρατηγικής αναζήτησης: α) *κίνηση*, β) *τακτική*, γ) *στρατήγημα* και δ) *στρατηγική*. Μία *κίνηση* είναι μία δράση του αναζητητή, είτε φυσική είτε πνευματική (απόφαση ή ανάγνωση). Μία *τακτική* είναι συνδυασμός κινήσεων και υπάρχουν άπειροι συνδυασμοί που μπορούν να υποστηρίξουν μία τακτική ανάλογα με την υλοποίηση του συστήματος. *Στρατηγήματα* είναι ένας μεγαλύτερος συνδυασμός κινήσεων και τακτικών, όπως για παράδειγμα το να ψάχνει κανείς παραπομπές ή να ακολουθεί υποσημειώσεις. Οι *στρατηγικές* είναι συνδυασμός κινήσεων, τακτικών και στρατηγημάτων. Παραδείγματα στρατηγικής είναι το να ψάχνει κανείς σχετικές εργασίες για ένα ερευνητικό άρθρο. Τα προηγούμενα μπορεί να αντικατοπτρίζουν

εργασίες σύνθετης αναζήτησης και πολλές φορές να συνδέονται με πολλαπλές συνεδρίες αναζήτησης.

2.2.2.3 Ζήτηση και Ανάκτηση Πληροφορίας

Τα παρακάτω μοντέλα σχετίζονται με το πώς οι άνθρωποι συλλέγουν και τακτοποιούν αντικείμενα στον αληθινό κόσμο. Συνδέονται με τα μοντέλα ζήτησης πληροφορίας ή με διερευνητική συμπεριφορά όπως συνδέονται και με μετρικές κέρδους πληροφορίας.

Συλλογή Πληροφορίας (Information Foraging)

Οι Pirolli και Card (Pirolli & Card, 1999) παρομοίωσαν τη συμπεριφορά των ανθρώπων όταν ψάχνουν πληροφορία και εξετάζουν πηγές, με εκείνη των ζώων όταν εξερευνούν για την τροφή τους. Οι άνθρωποι είναι «πληροφοριοφάγοι» (informavores). Στη διάρκεια αναζήτησης πληροφορίας, οι άνθρωποι συνεχώς λαμβάνουν αποφάσεις για τη φύση τη πληροφορίας που ψάχνουν, όπως για παράδειγμα για το αν θα πρέπει να συνεχίσουν να χρησιμοποιούν μία πηγή και να προσπαθούν να εντοπίσουν περισσότερη πληροφορία ή να την αναζητήσουν σε άλλη πηγή, τι μονοπάτι να ακολουθήσουν ή πότε να λήξουν την αναζήτηση.

Η μεταφορά της συλλογής πληροφορίας τονίζει το πώς οι αναζητητές πληροφορίας μπορεί να χρησιμοποιήσουν στοιχεία από τις αναζητήσεις προηγούμενων επισκεπτών για να βρουν πληροφορία σε μία συλλογή κειμένων. Αυτή η θεωρία απαντάται σε εφαρμογές που αξιοποιούν ιστορικά δεδομένα αναζητήσεων για να υποστηρίξουν τους αναζητητές σε συγκριμένους τομείς.

Μία κεντρική έννοια στη θεωρία «συλλογής πληροφορίας» είναι η «οσμή της πληροφορίας». Οι αναζητητές αξιολογούν την καταλληλότητα μιας συγκεκριμένης διαδρομής θεωρώντας μία αναπαράσταση, συνήθως μία λεκτική περιγραφή όπως τον υπότιτλο ενός αποτελέσματος ή μικρογραφία μιας εικόνας του περιφερικού περιεχομένου. Ο Furnas (1997) πρότεινε ότι ένα αντικείμενο αναπαράστασης κρατάει κατάλοιπο από το αντικείμενο πληροφορίας που αντιπροσωπεύει. Η έννοια του

καταλοίπου εκλεπτύνθηκε περαιτέρω από τον Pirolli (1997) σαν οσμή πληροφορίας και ορίστηκε από τους Card et al. (2001) ως η αντίληψη ενός ατόμου για την αξία, το κόστος ή το μονοπάτι πρόσβασης στις πηγές πληροφορίας κεκτημένων από κεντρικά σημεία όπως οι σύνδεσμοι WWW. Η κερδοφορία μίας πηγής πληροφορίας προσδιορίστηκε ως προς την αξία της πληροφορίας που αποκτήθηκε ανά κόστος μονάδας για την επεξεργασία της πηγής πληροφορίας. Το κόστος προσδιορίζεται από το χρόνο που καταναλώνεται, τις πηγές που χρησιμοποιούνται και τις ευκαιρίες που χάνονται όταν ακολουθείται μία συγκεκριμένη στρατηγική αναζήτησης αντί για άλλες (Russel et al., 1993).

Με τον ίδιο τρόπο που τα ζώα βασίζονται σε χημικές οσμές για να βρουν λεία στην περιοχή τους, αλλά και για να οδηγηθούν και σε άλλα υποσχόμενα μονοπάτια, οι αναζητητές πληροφορίας χρησιμοποιούν ίχνη στο περιβάλλον πληροφορίας που βρίσκονται για να τους οδηγήσουν σε περιεχόμενο ενδιαφέροντος και στις λύσεις των πληροφοριακών προβλημάτων τους. Μέσα στον περιηγητή Ιστού, η οσμή πληροφορίας μπορεί να «ζήσει εκεί» ανάμεσα σε άλλα πράγματα., σε υπερσυνδέσμους, σε λειτουργικότητα αυτοματοποιημένης συμπλήρωσης ερωτημάτων, υπότιτλους αποτελεσμάτων και προηγουμένως επισκεπτόμενους υπερσυνδέσμους που απεικονίζονται με διαφορετικό χρώμα για να αναπαραστήσουν ιστορικότητα.

Μία ενδιαφέρουσα πτυχή του μοντέλου που έχει προταθεί από τους Pirolli και Card (1999) είναι η έννοια του εμπλουτισμού τεμαχίων, όπου οι αναζητητές μπορούν να τροποποιήσουν τα τεμάχια πληροφορίας για να ταιριάξουν στις ανάγκες της τρέχουσας εργασίας αναζήτησης. Μέσα από μια σειρά από εκλεπτύνσεις της πληροφορίας, όπου οι αναζητητές απομονώνουν, αλλά και συνθέτουν σχετική πληροφορία, την τακτοποιούν σε υπο-θέματα και ούτω καθεξής, τα τεμάχια μετασχηματίζονται σε νέα με μεγαλύτερους ρυθμούς επιστροφής σχετικής πληροφορίας και χαμηλότερες επαναλήψεις αναζήτησης, πράγμα το οποίο βοηθά τους συλλέκτες να είναι πιο αποτελεσματικοί στο να βρίσκουν πληροφορία υψηλής ποιότητας.

Υπολογιστικά μοντέλα συμπεριφοράς αναζήτησης μπορούν να χρησιμοποιηθούν σε εργασίες όπως η πρόβλεψη των πηγών που θα επιλεγούν από αναζητητές, δοθέντος ενός πλαισίου συνδέσμων που εξετάζουν. Φυσικά, δοθέντων δεδομένων αλληλεπίδρασης πολύ μεγάλης κλίμακας από μηχανές αναζήτησης, μπορούμε να παρατηρήσουμε μοτίβα πρόσβασης σε πληροφορία χωρίς να χρειάζεται να προσομοιώσουμε τη δραστηριότητα αναζήτησης. Αυτό ωστόσο έχει εφαρμογή μόνο στο περιβάλλον από το οποίο συγκεντρώθηκαν τα δεδομένα.

Σημαντική μελλοντική εργασία στην περιοχή περιλαμβάνει μοντελοποίηση της συμπεριφοράς αναζήτησης με διαφορετικού βαθμού εξειδίκευση, διαφορετικά χαρακτηριστικά του αναζητητή, καθώς και άλλους τύπους αναζήτησης πέρα από την διερεύνηση μιας ιστοσελίδας. Πιο πρόσφατες εργασίες έχουν επικεντρωθεί στην κοινωνική και συνεργατική παραγωγή γνώσης.

«Συλλογή άγριων καρπών» (Berry-picking)

Εκτός από τη συλλογή πληροφορίας και άλλες αναλογίες από τον φυσικό κόσμο έχουν υιοθετηθεί για να μοντελοποιήσουν το πώς οι άνθρωποι ψάχνουν για πληροφορία με μη καθοδηγούμενο τρόπο με έμφαση στην περιήγηση και την διερεύνηση. Ένα από τα πιο γνωστά μοντέλα είναι η α) «*συλλογή άγριων καρπών*» (*berry-picking*) από τον Bates(1989).

Η «συλλογή άγριων καρπών» υποθέτει ότι η ενδιαφέρουσα πληροφορία είναι διασκορπισμένη μέσα σε μία συλλογή εγγράφων για παράδειγμα, όπως τα μούρα στους θάμνους και ότι οι άνθρωποι πρέπει να μαζεύουν τους άγριους καρπούς μόνοι τους. Η προσέγγιση αυτή βλέπει τον αναζητητή να μετακινείται διαμέσου ενός χώρου πληροφορίας μαζεύοντας θραύσματα πληροφορίας ενώ παράλληλα ψάχνει για ίχνη από το τοπικό πλαίσιο που θα καθοδηγήσουν μελλοντικές ή τις αμέσως επόμενες πλοηγήσεις. Σε αντίθεση με τη συλλογή πληροφορίας (information foraging) το μοντέλο εστιάζει στη δυναμική της πληροφοριακής ανάγκης παρά στο συμβιβασμό διερεύνησης-εκμετάλλευσης και στην δράση της συλλογής πληροφορίας. Σύμφωνα με το μοντέλο που συζητάμε το ερώτημα σταθερά

μεταβάλλεται και όσο οι αναζητητές μετακινούνται μεταξύ πηγών η νέα πληροφορία στην οποία εκτίθενται διευκολύνει νέες ιδέες και νέες κατευθύνσεις συμπεριλαμβανομένων νέων αντιλήψεων σχετικά με τις δηλώσεις ερωτημάτων. Σε κάθε στάδιο αναζήτησης, ο αναζητητής μπορεί να αναγνωρίσει χρήσιμες πληροφορίες και πηγές. Η κατανόηση σχετικά με την πληροφοριακή τους ανάγκη βελτιώνεται. Ο αναζητητής αναθεωρεί τη θεωρία σχετικά με το ποια πληροφορία είναι σχετική μέχρι να φθάσει σε ένα τελικό σημείο πλεονασμού – μέχρι δηλαδή είτε η ανάγκη να έχει καλυφθεί είτε να μην έχει πλέον σημασία για τον αναζητητή.

Το berry-picking μοντέλο διαφέρει από το παραδοσιακό μοντέλο γρήγορης αναζήτησης (καλύτερο ταίριασμα ερωτήματος-εγγράφου) στα εξής: α) το ερώτημα εξελίσσεται και β) η φύση της διαδικασίας αναζήτησης ακολουθεί μοτίβο «συλλογής άγριων καρπών» και δεν οδηγεί σε μια μοναδική συλλογή καλύτερου ταιριάσματος.

Το μοντέλο δίνει κατευθύνσεις προς τη σχεδίαση συστημάτων αναζήτησης, με την έννοια ότι τα συστήματα θα πρέπει να προσφέρουν υποστήριξη για αποθήκευση ενδιάμεσων αποτελεσμάτων, ώστε να μπορεί ο αναζητητής να επιστρέψει σε προηγούμενες καταστάσεις εφόσον η κατεύθυνση που έχει πάρει δε φέρει αποτέλεσμα.

2.2.2.4 Ζήτηση, Επεξεργασία και Χρήση Πληροφορίας

Επεξήγηση (Sensemaking)

Ως μεθοδολογία εισάχθηκε τη δεκαετία του 1980 από τους Dervin και Nilan (1986) με σκοπό την εξυπηρέτηση εργασιών στους τομείς της λήψης αποφάσεων και επίλυσης προβλημάτων. Πρόκειται για μία συνεχή προσπάθεια αφομοίωσης γνώσης ή πληροφορίας και δημιουργίας συνδέσεων. Σύμφωνα με τον Dervin παρουσιάζεται σα μία ενεργή αμφίδρομη διαδικασία προσαρμογής δεδομένων σε ένα πλαίσιο ή και πνευματικό μοντέλο, αλλά και προσαρμογής του πλαισίου γύρω από τα δεδομένα.

Οι Dervin (1988) και Nilan (1985) εστίασαν στο «γνωστικό κενό» που βιώνει κανείς όταν προσπαθεί να εξηγήσει δεδομένα. Στο μοντέλο τους, οι αναζητητές οι οποίοι

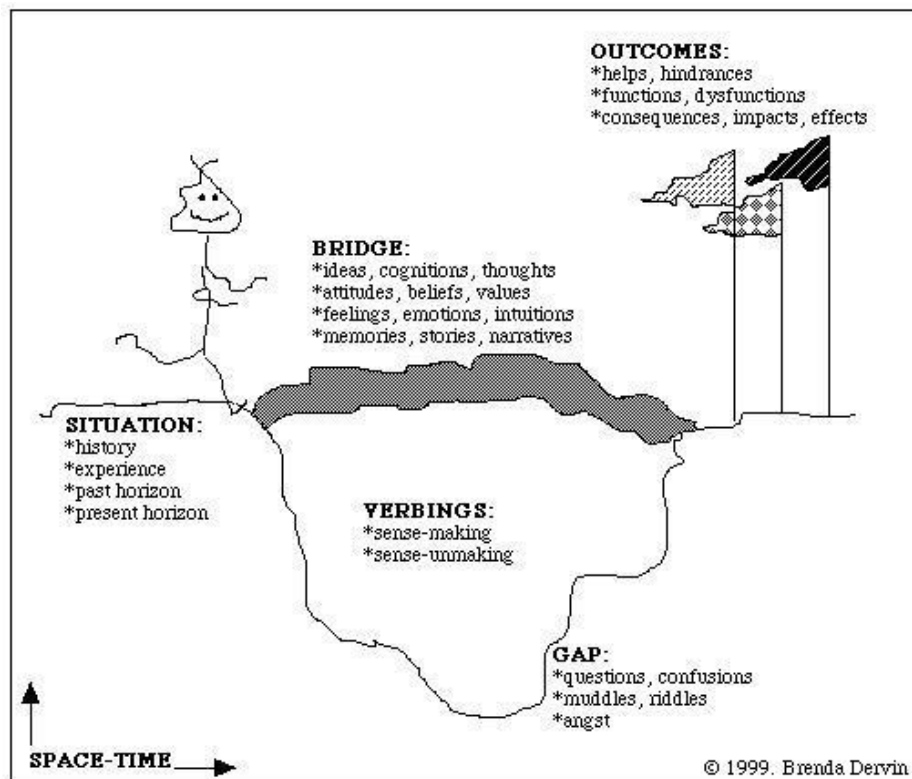
πρέπει να επιλύσουν συγκεκριμένη εργασία αντιμετωπίζουν ένα προβληματικό σημείο ή κενό το οποίο εμποδίζει την πρόοδο. Ο αναζητητής πρέπει να ξεπεράσει το κενό βρίσκοντας βοήθεια ή εξηγώντας την τρέχουσα κατάσταση για να επιτύχει το επιθυμητό αποτέλεσμα. Θεωρούν την ζήτηση πληροφορίας σαν μία διαδικασία εξήγησης της ευαίσθητης κατάστασης. Προτείνουν ότι η ζήτηση πληροφορίας θα έπρεπε να είναι ολιστική και να εστιάζει σε υποκειμενική πληροφορία, η οποία κατασκευάζεται από ανθρώπους, επικοινωνητικούς παράγοντες με εσωτερικές-γνωσιακές συλλήψεις παρά από παθητικούς λήπτες πληροφορίας και καταστάσεις μέσα στις οποίες απλά δρουν.

Ο Russel και οι συνεργάτες του (Russel et al., 1993) του περιέγραψαν την *Επεξήγηση* σαν διαδικασία κωδικοποίησης ανακτημένων πληροφοριών για να απαντήσουν ερωτήσεις εστιασμένες στην εργασία. Όρισαν ένα μοντέλο που απαρτίζεται από 4 κύριες διαδικασίες α) *αναζήτηση* για αναπαραστάσεις που αποτυπώνουν *εμφανή μοτίβα δεδομένων*, β) *δημιουργία στιγμιότυπων αναπαραστάσεων*: κωδικοποίηση πληροφορίας στην αναπαράσταση, γ) *τροποποίηση αναπαράστασης*: όταν τα δεδομένα δεν ταιριάζουν ή είναι ελλιπή, δ) *αξιοποίηση δημιουργημένων αναπαραστάσεων*. Βασικό χαρακτηριστικό του μοντέλου είναι ότι η διαδικασία αποτελείται από πολλές επαναλήψεις έως ότου να θεωρηθεί επιτυχής.

Σύμφωνα με τους Pirolli και Card (2005) το μοντέλο Επεξήγησης περιλαμβάνει τα παρακάτω βήματα: α) *Αναγνώριση* του χάσματος γνώσης, β) *δημιουργία αρχικής δομής* της γνώσης που χρειάζεται για την ολοκλήρωση της εργασίας, γ) *αναζήτηση* πληροφορίας, δ) *ανάλυση* και *σύνθεση* της πληροφορίας για κατανόηση και αντίληψη, ε) *δημιουργία προϊόντος γνώσης* βάσει της προηγούμενης κατανόησης.

Σε ανάλυση γνωστικών εργασιών που διεξήγαγαν αναγνώρισαν δύο βρόχους δραστηριοτήτων κατά την επεξήγηση: α) ένα *βρόχο συλλογής πληροφορίας* που περιλαμβάνει αναζήτηση, διήθηση, διάβασμα και εξαγωγή πληροφορίας σε κάποιο

σχήμα, β) ένα βρόχο επεξήγησης που περιλαμβάνει επαναληπτική ανάπτυξη ενός πνευματικού μοντέλου από το σχήμα που ταιριάζει καλύτερα στις ενδείξεις. Μία



Εικόνα 6 Θεωρία Επεξήγησης Brenda Dervin

ποικιλία εννοιολογικών αλλαγών μπορούν να συμβούν στην πνευματική αναπαράσταση γνώσης όσο ο αναζητητής μαθαίνει για την εργασία, το πρόβλημα ή την κατάσταση.

Οι τυπικές εργασίες που συνήθως απαιτούν ένα βαθμό επεξήγησης συνήθως περιλαμβάνουν: α) νέες καταστάσεις ή προβλήματα, β) σύνθετες λιγότερο δομημένες καταστάσεις ή προβλήματα, γ) ένα νέο πεδίο, δ) ασαφή πληροφοριακή ανάγκη. Υπάρχει επικάλυψη μεταξύ τέτοιων καταστάσεων και της διερευνητικής αναζήτησης, η οποία θα συζητηθεί εκτεταμένα στην επόμενη ενότητα. Οι αναζητητές που εμπλέκονται σε διερευνητικούς και άλλους τύπους σύνθετων εργασιών αναζήτησης σταθερά εμπλέκονται και σε επεξηγηματικές δραστηριότητες όσο κινούνται στο χώρο πληροφορίας. Η κίνηση διακόπτεται όταν αναγνωρίζεται ένα «κενό» το οποίο

απαιτεί πληροφορία για να γεφυρωθεί. Η «επεξήγηση» είναι μία εξατομικευμένη διαδικασία κατασκευής και όχι μία διαδικασία απλής χρήσης πληροφορίας. Ωστόσο η διαδικασία κατασκευής περιλαμβάνει την εφαρμογή της πληροφορίας που έχει βρεθεί.

Διερευνητικές και σύνθετες αναζητήσεις τυπικά περιλαμβάνουν παρατεταμένη εμπλοκή, στην οποία οι αναζητητές επαναληπτικά αναζητούν και μαθαίνουν νέες έννοιες και γεγονότα. Η κτήση πληροφορίας προκαλεί τους αναζητητές δυναμικά να αλλάζουν και να εκλεπτύνουν τους πληροφοριακούς σκοπούς τους, όπως και να ρωτούν πιο στοχευμένα ερωτήματα, τα οποία τους καθοδηγούν βαθύτερα στο πρόβλημα και το χώρο πληροφορίας.

Γενικά η διαδικασία «επεξήγησης» περιλαμβάνει αφομοίωση νέας γνώσης για να ενισχύσει την πρότερη κατάσταση γνώσης ενός ατόμου, παρέχοντας καλύτερη κατανόηση του συγκεκριμένου θέματος. Η διαδικασία αυτή μπορεί επίσης να παραγάγει κατασκευάσματα που μπορούν να διαμοιραστούν μεταξύ ατόμων.

Διερευνητική Αναζήτηση (Exploratory Search)

Στην ίδια κατηγορία με την Επεξήγηση εντάσσεται και το μοντέλο της Διερευνητικής Αναζήτησης για το οποίο επιλέγουμε να αφιερώσουμε ξεχωριστή ενότητα (βλ. ενότητα 2.3) δεδομένου ότι είναι το μοντέλο με το οποίο θα ασχοληθούμε σε όλη την έκταση της Διατριβής.

Στον πίνακα 1 συνοψίζονται τα βασικότερα σημεία που επεξηγούν και διαφοροποιούν τα κυριότερα μοντέλα ζήτησης πληροφορίας μεταξύ τους.

Πίνακας 1 Σύνοψη Μοντέλων Ζήτησης Πληροφορίας

Μοντέλο	Περιγραφή	Ερευνητικός Τομέας
ASK Belkin et al. (1982)	<ul style="list-style-type: none"> Επαναληπτική και διαδραστική διαδικασία ανάκτησης πληροφοριών Το ερώτημα ως επεξηγηματική περιγραφή του προβλήματος ASK 	Ανάκτηση Πληροφορίας με επίκεντρο το χρήστη (Human Centered IR)

Ingwersen (1992)	<ul style="list-style-type: none"> • Γνωσιακό μοντέλο Αλληλεπίδρασης • Περιλαμβάνει το πλαίσιο ή το κοινωνικο-οργανωτικό περιβάλλον & την αλληλεπίδραση αναζήτησης 	Προσανατολισμένη στο σύστημα & Γνωστική Ανάκτηση Πληροφορίας (Systems oriented IR and Cognitive IR)
Wilson (1986;1999)	<ul style="list-style-type: none"> • Πλαίσιο πληροφοριακής ανάγκης • Μηχανισμός ενεργοποίησης π.χ. Stress • Παρεμβαίνουσες μεταβλητές επηρεάζουν τη συμπεριφορά αναζήτησης • Παθητική ή Ενεργητική Συμπεριφορά στη ζήτηση • Περιλαμβάνει βρόγχο αξιολόγησης, επεξεργασίας & χρήσης της πληροφορίας 	Συμπεριφορά στην Πληροφορία & Επιστήμη Πληροφορίας (Information Behaviour & Information Science)
Kuhlthau (1991-93)	<ul style="list-style-type: none"> • Μοντέλο διαδικασίας αναζήτησης πληροφορίας (ISP-Information Search Process) • Δίνει έμφαση στη σημασία των επιμέρους σταδίων που περιλαμβάνονται στις εργασίες εκμάθησης και επίλυσης προβλημάτων • Αναγνώριση σταδίων: Εκκίνηση, επιλογή, διερεύνηση, εστίαση, συλλογή, παρουσίαση • Συνδέει κάθε στάδιο με συναισθήματα σκέψεις και ενέργειες του αναζητητή 	Επιστήμη Πληροφορίας (Information Science)
Elli's (1989;1993;1997)	<ul style="list-style-type: none"> • Μοντέλο αναζήτησης πληροφορίας που βασίζεται σε μελέτες μοτίβων αναζήτησης κοινωνικών επιστημόνων, ερευνητών • Διακρίνει 8 στάδια και ταιριάζει στις περισσότερες περιπτώσεις ζήτησης πληροφορίας • Δεν περιλαμβάνει διαδικασίες ανάκτησης • Δεν περιλαμβάνει βρόχο ανάδρασης 	Συμπεριφορά στην Πληροφορία & Αναζήτηση (Information behavior & Search Behaviour)
Bates (1979;1990)	<ul style="list-style-type: none"> • Αναγνώριση τεσσάρων σταδίων στρατηγικής αναζήτησης: α) κίνηση, β) τακτική, γ) στρατήγημα και δ) στρατηγική 	Επιστήμη της Πληροφορίας & Βιβλιοθηκονομίας (Information & library science)

Information Foraging (Pirolli & Card, 1999)	<ul style="list-style-type: none"> • Οι άνθρωποι είναι πληροφοριοφάγοι (informavores) & ακολουθούν την «οσμή» πληροφορίας • Η οσμή πληροφορίας (information scent) μπορεί να είναι το Url, result caption ένα thumbnail • οι αναζητητές πληροφορίας μπορεί να χρησιμοποιήσουν στοιχεία από τις αναζητήσεις προηγούμενων επισκεπτών για να βρουν πληροφορία σε μία συλλογή κειμένων. • Οι αναζητητές σε κάθε βήμα εκτιμούν το κέρδος που θα έχουν αν ακολουθήσουν ένα μονοπάτι -Συμβιβασμός διερεύνησης εκμετάλλευσης • Η καταγραφή της διαδρομής τροφοδοτεί υπολογιστικά μοντέλα που μπορούν να προβλέψουν τα urls που θα επιλεγθούν ή το μήκος των μονοπατιών που θα ακολουθήσουν 	Διερευνητική Συμπεριφορά & Ζήτηση Πληροφορίας (Exploratory Behaviour & Information Seeking)
Berry-picking (Bates 1989)	<ul style="list-style-type: none"> • εστιάζει στη δυναμική της πληροφοριακής ανάγκης • ο αναζητητής αναθεωρεί ως προς το ποια πληροφορία είναι σχετική μέχρι είτε η ανάγκη να έχει καλυφθεί ή να μην έχει πλέον σημασία • διαφέρει από το παραδοσιακό μοντέλο γρήγορης αναζήτησης (ταιρίασμα ερωτήματος-εγγράφου) στα: α) το ερώτημα εξελίσσεται και β) η φύση της διαδικασίας αναζήτησης ακολουθεί μοτίβο «συλλογής άγριων καρπών» και δεν οδηγεί σε μια μοναδική συλλογή καλύτερου ταιριάσματος • δίνει κατευθύνσεις για τη σχεδίαση (Σ) αναζήτησης που υποστηρίζουν αποθήκευση ενδιάμεσων αποτελεσμάτων, ώστε ο αναζητητής να επιστρέφει σε προηγούμενες καταστάσεις εάν η κατεύθυνση που έχει πάρει δε φέρει αποτέλεσμα 	Διερευνητική Συμπεριφορά (Exploratory Behaviour)

Choo et al. (2000)	<ul style="list-style-type: none"> • Μοντέλο για τον Ιστό • Στηρίζεται στο μοντέλο Elli's model • Αναγνωρίζει 6 γενικά στάδια στη συμπεριφορά ζήτησης πληροφορίας εκκίνηση, δημιουργία αλυσίδων, παρακολούθηση, εξαγωγή, περιήγηση, διαφοροποίηση • Προτείνει Περιήγηση και αναζήτηση 	Συμπεριφορά στην Αναζήτηση Πληροφορίας (Information Search Behaviour)
Marchionini (1995)	<ul style="list-style-type: none"> • Ως μοντέλο-σε σχέση με της Elli's απευθύνεται σε ψηφιακά και διαδικτυακά περιβάλλοντα. • Εκφράζει την πτυχή της συλλογής πληροφορίας στη διαδικασία εξέτασης των αποτελεσμάτων • Εκφράζει πτυχές απόκτησης γνώσης κατά την εξαγωγή πληροφορίας. • Απουσιάζει η εκμάθηση και κατανόηση • Απουσιάζει το πλαίσιο • Απουσιάζει η χρήση της πληροφορίας 	
Sensemaking (Dervin and Nilan 1986; Russel et al., 1993)	<ul style="list-style-type: none"> • Ξεκινά από την ύπαρξη ενός προβληματικού σημείου ή κενού το οποίο εμποδίζει την πρόοδο • Περιλαμβάνει αφομοίωση νέας γνώσης για την ενίσχυση πρότερης κατάστασης γνώσης παρέχοντας καλύτερη κατανόηση του θέματος 	Διερευνητική Συμπεριφορά & Ζήτηση Πληροφορίας (Exploratory Behaviour & Information Seeking)
Exploratory Search (Marchionini 2006a)	<ul style="list-style-type: none"> • Εξειδίκευση στη διερεύνηση πληροφορίας. • Η απάντηση μπορεί να εμφανιστεί έπειτα από ανάλυση της πληροφορίας που συλλέγεται στη διάρκεια των συνεδριών αναζήτησης • Πιθανότητα μεγάλης διάρκειας • Ανάπτυξη διανοητικών δεξιοτήτων ανώτερου επιπέδου, όπως εφαρμογή, σύνθεση, αξιολόγηση • Στόχος είναι η δημιουργία ενός προϊόντος γνώσης • Ομοιότητα με μοντέλα: berry-picking, information foraging, sensemaking 	Διερευνητική Συμπεριφορά & Ζήτηση Πληροφορίας (Exploratory Behaviour & Information Seeking)

	<ul style="list-style-type: none"> • Εστίαση στην εκμάθηση συγκριμένου θέματος – βασική διαφορά από παραδοσιακά μοντέλα επαναληπτικής αναζήτησης (π.χ. επαναληπτική βελτίωση ερωτημάτων Belkin et al. 2011) 	
--	--	--

2.3 Διερευνητική Αναζήτηση

2.3.1 Τι είναι η Διερευνητική Αναζήτηση

Ο όρος Διερευνητική Αναζήτηση μπορεί να χρησιμοποιηθεί για να περιγράψει ένα πρόβλημα ζήτησης πληροφορίας, το οποίο είναι αορίστου χρόνου, επίμονο και πολύπλευρο, αλλά και μία διαδικασία ζήτησης πληροφορίας η οποία είναι ευκαιριακή, επαναληπτική και πολλαπλών τακτικών. Με την έννοια του προβλήματος, η διερευνητική αναζήτηση χρησιμοποιείται ευρέως στην επιστημονική ανακάλυψη, τη μάθηση και τη λήψη αποφάσεων. Με την έννοια της διαδικασίας οι διερευνητικές τακτικές χρησιμοποιούνται σε όλα τα είδη ζήτησης πληροφορίας και αντικατοπτρίζουν τα ενδιαφέροντα και τις προτιμήσεις του αναζητητή όσο και τους στόχους της αναζήτησης (Marchionini 2006a). Οι διερευνητικές αναζητήσεις μπορεί να οδηγήσουν σε πολλαπλά ερωτήματα ή συνεδρίες. Συχνά καθοδηγούνται από περιέργεια και δίνουν έμφαση στην έρευνα, τη μάθηση και γενικότερα την εφαρμογή της γνώσης η οποία αποκτήθηκε. Αν και σχεδόν όλες οι αναζητήσεις είναι κατά μία έννοια διερευνητικές ή εξερευνητικές, δεν είναι μόνο η ενέργεια της διερεύνησης που κάνει την αναζήτηση «διερευνητική», αλλά και το ότι η αναζήτηση θα πρέπει να περιλαμβάνει σύνθετες γνωστικές δραστηριότητες συνδεδεμένες με την απόκτηση γνώσης και την ανάπτυξη γνωσιακών δεξιοτήτων (White και Roth 2009).

Στη διερευνητική αναζήτηση οι χρήστες υποβάλλουν ένα αβέβαιο αίτημα για να πλοηγηθούν σε σχετικά έγγραφα στη συλλογή και έπειτα να διερευνήσουν το περιβάλλον για να κατανοήσουν καλύτερα πώς να το εκμεταλλευτούν αναζητώντας επιλεκτικά και παθητικά αποκτώντας στοιχεία για τα επόμενα βήματά τους (White et

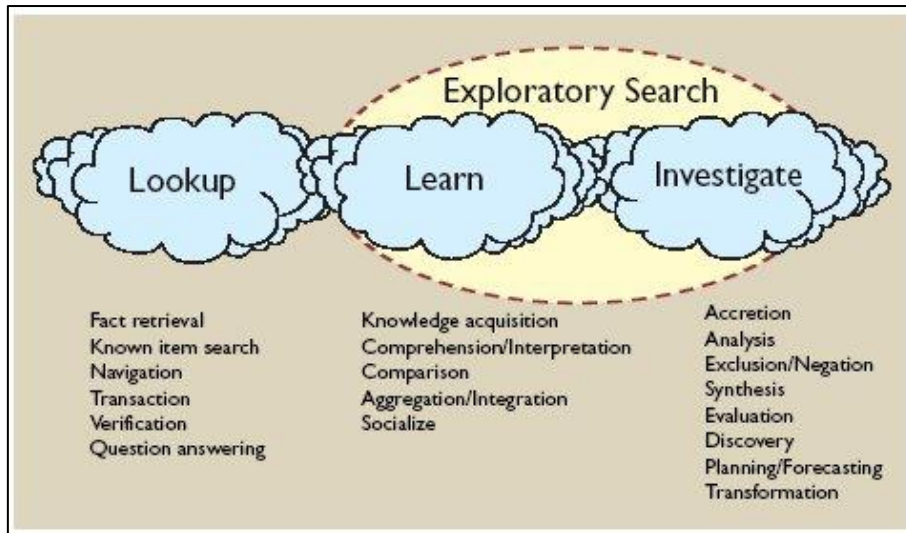
αl. 2006). Η διερευνητική αναζήτηση είναι μια εξειδίκευση της ζήτησης πληροφορίας, η οποία περιγράφει τη δραστηριότητα της προσπάθειας απόκτησης πληροφοριών μέσω ενός συνδυασμού αναζήτησης και περιήγησης. Η διερευνητική αναζήτηση είναι ένας τύπος αναζήτησης πληροφοριών και ένας τύπος επεξήγησης όπως είδαμε στην προηγούμενη ενότητα που επικεντρώνεται στη συλλογή και τη χρήση πληροφοριών για την προώθηση της πνευματικής ανάπτυξης.

Στις διερευνητικές αναζητήσεις η απάντηση μπορεί να εμφανιστεί έπειτα από ανάλυση της πληροφορίας που συλλέγεται στη διάρκεια των συνεδριών αναζήτησης και αυτό μπορεί να κρατήσει ημέρες, εβδομάδες, μήνες. Από την άλλη, οι διερευνητικές αναζητήσεις μπορεί εν δυνάμει να έχουν προφανή επίδραση στην προσωπική ανάπτυξη του αναζητητή, διότι αντικατοπτρίζουν το κυνήγι στόχων εκμάθησης ανώτερου επιπέδου.

Σε αντίθεση με την αναζήτηση πληροφορίας που θα κλείσει ένα κενό στη γνώση κάποιου –στόχος συνήθως κατανοητός και προσδιορισμένος- ο στόχος στις διερευνητικές αναζητήσεις, όπως αναφέραμε παραπάνω, είναι στην αρχή λιγότερο καθαρά ορισμένος, αλλά πάντα σκοπός είναι η ανάπτυξη διανοητικών δεξιοτήτων ανώτερου επιπέδου, όπως εφαρμογή, σύνθεση, αξιολόγηση. Παράδειγμα για το προηγούμενο είναι η δημιουργία ενός προϊόντος γνώσης, όπως ένα ερευνητικό άρθρο ή ο σχηματισμός μίας ενέργειας, όπως η επιλογή μίας ιατρικής θεραπείας.

Ο Marchionini (2006a) πρότεινε ένα σύνολο από δραστηριότητες αναζήτησης συνδεδεμένες με μία διαδικασία διερευνητικής αναζήτησης και ξεχώρισε τις δραστηριότητες που σχετίζονται με διερεύνηση από τις γρήγορες αναζητήσεις που χρησιμοποιούνται από τις σύγχρονες μηχανές αναζήτησης Ιστού (Εικόνα 7).

Αν και οι δραστηριότητες τις γρήγορης αναζήτησης, εκμάθησης και έρευνας φαίνονται ξεχωριστές ενυπάρχει αλληλεπίδραση μεταξύ τους. Για παράδειγμα, η

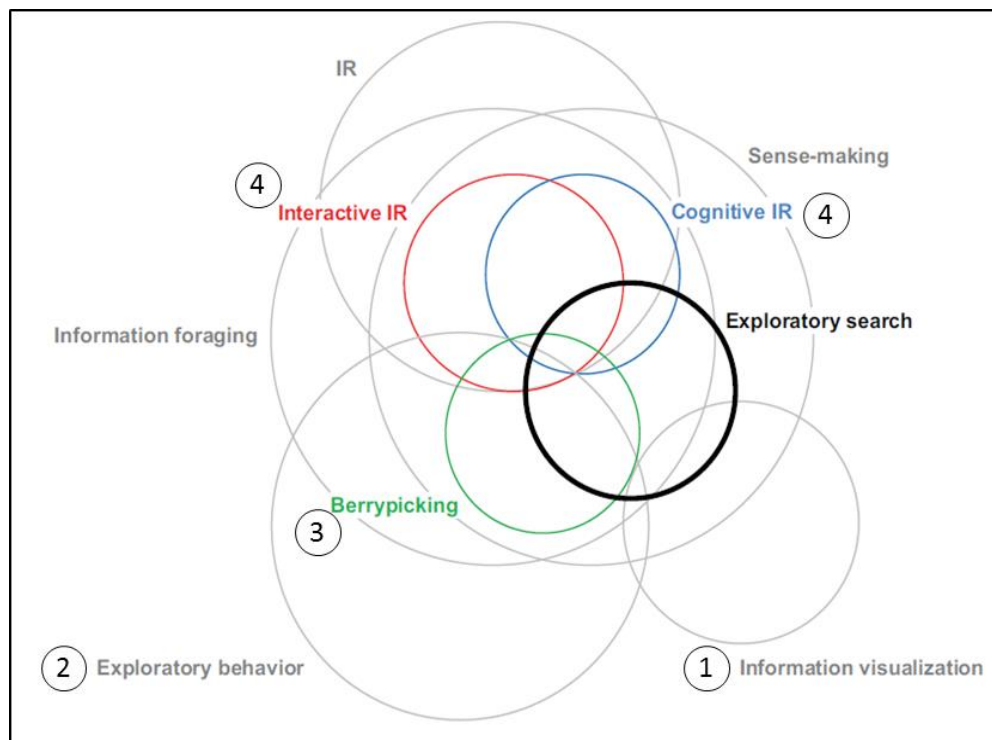


Εικόνα 7 Διερευνητική Αναζήτηση (από Marchionini 2006a)

γρήγορη αναζήτηση είναι ενσωματωμένη στην εκμάθηση ή την έρευνα, η εκμάθηση είναι σημαντικό μέρος της έρευνας. Η γρήγορη αναζήτηση συνήθως περιλαμβάνει την ανάκτηση μοναδικών απαντήσεων, όπως εύρεση γεγονότων, ή ερωτήσεις-απαντήσεις. Σήμερα οι σύγχρονες μηχανές αναζήτησης χειρίζονται τη γρήγορη αναζήτηση πολύ καλά με αλγόριθμους ανάκτησης και ιεράρχησης μεγάλης ακρίβειας. Ωστόσο, οι δραστηριότητες διερευνητικής αναζήτησης απαιτούν μεγαλύτερη συμμετοχή από τον αναζητητή, μεγαλύτερο βαθμό συνέργειας μεταξύ αναζητητή και συστήματος, παροχή περισσότερης λειτουργικότητας από το σύστημα που εκτείνεται πέρα από τον καθορισμό του ερωτήματος και την παρουσίαση αποτελεσμάτων.

Γενικότερα η θέση της Διερευνητικής Αναζήτησης τοποθετείται μεταξύ επιστημών που σχετίζονται με την Ανθρωπο-Υπολογιστική Ανάκτηση Πληροφορίας και την Αλληλεπίδραση Ανθρώπου-Υπολογιστή. Η διερευνητική αναζήτηση παρουσιάζει πολλά κοινά στοιχεία με μοντέλα που παρουσιάστηκαν σε προηγούμενες υπο-ενότητες, αλλά και με άλλες πτυχές της Ζήτησης Πληροφορίας, όπως απεικονίζεται και στην εικόνα 8. Για παράδειγμα, σχετίζεται με 1) την Οπτικοποίηση Πληροφορίας, δηλαδή τον τρόπο διευκόλυνσης να έχει κάποιος μία γενική εποπτεία του χώρου πληροφορίας, 2) τη Διερευνητική Συμπεριφορά, δηλαδή

τον τρόπο διερεύνησης του χώρου πληροφορίας, 3) τα μοντέλα «συλλογής άγριων καρπών» (berry-picking) και συλλογής πληροφορίας (information foraging) τα οποία αποδίδουν τη σπτική της ικανότητας του αναζητητή να ανακαλύψει πληροφορία και 4) με τη Διαδραστική Ανάκτηση Πληροφορίας (IIR) και τη γνωστική Ανάκτηση Πληροφορίας (cognitive IR) -για παράδειγμα το μοντέλο της Επεξήγησης (Sensemaking)-, τα οποία περιγράφουν τα πνευματικά και συμπεριφοριστικά μοντέλα κατά την εύρεση πληροφορίας. Ωστόσο, οδηγείται περισσότερο από περιέργεια παρά από επιθυμία να ικανοποιηθεί μία συγκεκριμένη πληροφοριακή ανάγκη. Επίσης, ο στόχος της αναζήτησης μπορεί να αλλάξει κατά τη διάρκειά της όσο οι πληροφοριακές ανάγκες εξελίσσονται, όπως επίσης είναι πιθανό να μην είχε προσδιοριστεί καλά στην αρχή της έρευνας. Επιπλέον ο στόχος μπορεί να μην είναι η



Εικόνα 8 Διάγραμμα Venn το οποίο τοποθετεί τη Διερευνητική Αναζήτηση σε σχέση με άλλες ερευνητικές επιστήμες (από White & Roth 2009: Exploratory Search Beyond the Query–Response Paradigm)

μεγιστοποίηση του πληροφοριακού κέρδους. Για παράδειγμα συχνά παρατηρείται οι αναζητητές να ακολουθούν ξανά κάποιο μονοπάτι αναζήτησης στον χώρο

πληροφορίας για να επισκεφθούν ξανά ένα συγκεκριμένο «κλαδί» ή να διαβάζουν περιττή πληροφορία προκειμένου να επιβεβαιώσουν ή όχι μία υπόθεση.

Στη διερευνητική αναζήτηση παρατηρείται ένας συνδυασμός περιήγησης και επικεντρωμένης αναζήτησης. Η διερευνητική αναζήτηση διαφέρει ωστόσο από πιο παραδοσιακά μοντέλα επαναληπτικής αναζήτησης (όπως επαναληπτική βελτίωση ερωτημάτων Belkin et al. 2011) καθώς εστιάζει στην εκμάθηση συγκριμένου θέματος, στην κατανόηση του περιεχομένου της συλλογής, όπως και στην αξιοποίηση αναπάντεχων και χρήσιμων ευρημάτων που μπορεί να ωθήσει τους αναζητητές σε άλλες κατευθύνσεις άμεσα ή έμμεσα σχετιζόμενες με το θέμα που θέλουν να γνωρίσουν. Επίσης, διαφέρουν και στο ότι στην απλή επαναληπτική αναζήτηση ο στόχος είναι συνήθως γνωστός και η εργασία αναζήτησης είναι να βρεθεί ο στόχος.

Σε τέτοιες περιπτώσεις χρησιμεύει να προσφέρονται από τα συστήματα προτάσεις για βελτίωση ερωτημάτων, ώστε ο χρήστης να οδηγηθεί στην απάντηση που ψάχνει. Στην περίπτωση της διερευνητικής αναζήτησης η προηγούμενη στρατηγική δεν οδηγεί σε αποκάλυψη νέων κατευθύνσεων για διερεύνηση και ως εκ τούτου παράγοντες που ωφελούν είναι η *καινοτομία*, η *κάλυψη* των διαφορετικών πτυχών της πληροφορίας (coverage) καθώς και η *συνάφεια* με την κατάσταση (Seracevic 2007).

Η διερευνητική αναζήτηση περιγράφει καταρχήν ένα πλαίσιο προβλήματος ζήτησης πληροφορίας, το οποίο είναι ευκαιριακό, επαναληπτικό και πολλαπλών τακτικών. Ως εκ τούτου, η διερευνητική αναζήτηση περιλαμβάνει πολλά από τα τυπικά χαρακτηριστικά της ζήτησης πληροφορίας με την έννοια ότι είναι μία διαδικασία προσπάθειας απόκτησης πληροφορίας. Επιπλέον όμως, είναι ένας τύπος κατανόησης (δηλαδή, μια συνεχιζόμενη και παρακινημένη προσπάθεια να εξηγηθούν οι συνδέσεις μεταξύ ανθρώπων, τόπων και γεγονότων) επικεντρωμένος στη συλλογή και τη χρήση πληροφοριών για την προώθηση της προσωπικής νοητικής ανάπτυξης και της απόκτησης γνώσεων (White & Roth 2009).

2.3.2 Σχεδίαση Συστημάτων Διερευνητικής Αναζήτησης

Τα εργαλεία υποστήριξης διερευνητικής αναζήτησης θα έπρεπε να υποστηρίζουν τους αναζητητές ώστε να προσδιορίσουν καλύτερα το πρόβλημα, να επεξηγήσουν την πληροφορία που συνάντησαν στη διάρκεια της τρέχουσας συνεδρίας, καθώς και να χειριστούν την τυχόν αβεβαιότητα και σύγχυση, παρέχοντας ενημερώσεις προόδου, επεξηγήσεις και περιλήψεις των κυριότερων θεμάτων που δεσπάζουν στην πληροφορία που έχει βρεθεί (Marchionini 2006a; White and Roth 2009).

Οι πληροφοριακές ανάγκες, για παράδειγμα ομάδων έργου που δημιουργούν σύνθετες λύσεις ή προϊόντα, μελετητών ή ερευνητών που μελετούν σύνθετα φαινόμενα, ανθρώπων που πάσχουν από χρόνιες ασθένειες και άλλες περιπτώσεις, καλύπτονται από τις σύγχρονες μηχανές αναζήτησης Ιστού μόνο επιφανειακά. Τα συστήματα διερευνητικής αναζήτησης έρχονται να καλύψουν αυτό το κενό, παρέχοντας λύσεις αναζήτησης οι οποίες ενισχύουν τους ανθρώπους να κινηθούν πέρα από τις γρήγορες αναζητήσεις καλύτερου ταιριάσματος και μοναδικής συνεδρίας.

Η ιδιαίτερη φύση των αναζητήσεων για τις οποίες ο στόχος είναι άγνωστος ή ασαφής, απαιτεί νέα σχεδίαση. Οι White και Roth (2009) πρότειναν τις παρακάτω εφτά γενικές αρχές:

- *Υποστήριξη ερωτημάτων και γρήγορη τελειοποίηση ερωτημάτων:* Ύπαρξη κουμπιών, φίλτρων, δυνατότητα για δυναμικά μεταβαλλόμενα αποτελέσματα.
- *Περιβάλλον/Πλαίσιο αναζήτησης:* διαθέσιμες πληροφορίες σχετικά με την τρέχουσα εργασία διερευνητικής αναζήτησης.
- *Προσφορά Διαφορετικών Οπτικών:* Διήθηση και ομαδοποίηση/κατηγοριοποίηση αποτελεσμάτων με βάση τη θεματική κατηγορία.
- *Απεικονίσεις/Οπτικοποίηση:* Οπτικές αναπαραστάσεις της συλλογής που διερευνάται/αφαιρετικές και γενικευμένες επισκοπήσεις βοηθούν τους ανθρώπους να καταλήξουν σε σχέσεις μεταξύ εννοιών.

- *Υποστήριξη της μάθησης και της κατανόησης*: Παρουσίαση πληροφοριών με τρόπους που να μπορούν να βοηθήσουν τη μάθηση, λαμβάνοντας υπόψιν το τρέχον επίπεδο γνώσεων και δεξιοτήτων του χρήστη.
- *Προσφορά Ιστορικότητας, χώροι εργασίας και ενημερώσεις προόδου*: καταγραφή διαδρομών που ακολούθησαν οι χρήστες, εύκολη επανεπισκεψιμότητα των αποτελεσμάτων, παρακολούθηση της προόδου του χρήστη, διαδικασία συλλογής πληροφοριών και περιλήψεις.
- *Διευκόλυνση Συνεργασίας*: παροχή τρόπων διαμοιρασμού γνώσης

2.3.3 Εύνοια Τυχαίων Ανακαλύψεων

Η εύνοια των τυχαίων ανακαλύψεων μπορεί να διαδραματίσει κάποιο ρόλο στη διαδικασία αναζήτησης πληροφοριών. Όσον αφορά την περιήγηση σε πληροφορία, μια ειδική περίπτωση περιήγησης που οδηγεί σε τυχαία ανακάλυψη είναι η εύνοια τυχαίων ανακαλύψεων (Arted & Choo 1971). Σε αυτό το πλαίσιο, οι άνθρωποι ενδέχεται να συναντούν πληροφορίες χωρίς προμελετημένη ενέργεια, μπορεί να βρίσκουν "πολύτιμες πληροφορίες για το θέμα Β όταν ψάχνουν για το θέμα Α" και αυτό "επιτρέπει σε ένα χρήστη να αναγνωρίζει πληροφορίες αξίας σε άλλα περιβάλλοντα από αυτό που είχαν κατά νου όταν ξεκίνησε η αναζήτηση" (Boyce et al., 1994). Από τη σκοπιά της ανάκτησης πληροφοριών, η εύνοια των τυχαίων ανακαλύψεων υποδηλώνει το σενάριο της ανακάλυψης χρήσιμων πληροφοριών με τυχειότητα και ευχαρίστηση (Thudt et al., 2012).

Η ιδέα της εύνοιας των τυχαίων ανακαλύψεων αναφέρεται από την διεπιστημονική έρευνα ως ένας πρόσθετος παράγοντας για την παροχή βοήθειας στην αναζήτηση πληροφοριών (Foster & Ford 2003) και τη διερευνητική αναζήτηση (White et al., 2006). Σε αντίθεση με τις παραδοσιακές μεθόδους ανάκτησης πληροφοριών, η υποστήριξη της εύνοιας των τυχαίων ανακαλύψεων απαιτεί μια ασαφή προσέγγιση κατά την απάντηση σε ερωτήματα αναζήτησης, αντί να επιστρέψει ακριβή ταιριάσματα (Toms 2000). Με ένα κατάλληλο περιβάλλον και

προσέγγιση, ο σχεδιασμός για την εύνοια των τυχαίων ανακαλύψεων μπορεί να βελτιώσει την εμπειρία των χρηστών (Leong et al., 2012).

Στην έννοια των τυχαίων ανακαλύψεων θα αναφερθούμε ξανά στο Κεφάλαιο 3 και θα ασχοληθούμε εκτεταμένα στο κεφάλαιο 4.

2.3.4 Εξατομίκευση

Η μελέτη της βιβλιογραφίας αποκαλύπτει την επίδραση του παράγοντα της εξατομίκευσης ως καθοριστική στη βελτίωση της εμπειρίας του χρήστη κατά την ανάκτηση πληροφορίας. Για παράδειγμα, πληροφορίες σχετικά με τα προσωπικά ενδιαφέροντα και τη συμπεριφορά μπορούν να χρησιμοποιηθούν για να υποστηρίξουν την εύνοια τυχαίων ανακαλύψεων στα αποτελέσματα αναζήτησης (André et al. 2009α). Άλλοι ερευνητές συνδυάζουν την εξατομίκευση με άλλους παράγοντες που έχουμε εξετάσει ήδη όπως η ποικιλομορφία (Vallet & Castells, 2012). Ήδη από το 1999 ερευνητές όπως οι Liebermann, Campos και Figueiredo δημιουργούν πράκτορες λογισμικού που παρακολουθούν τη δραστηριότητα του χρήστη προκειμένου να προτείνουν ενδιαφέροντες συνδέσμους που μπορεί να διεγείρουν αναπάντεχες συσχετίσεις (Liebermann 1999; Campos and Figueiredo 2001).

Στη διάσταση της εξατομίκευσης ωστόσο, εκτός από τη μοντελοποίηση της συμπεριφοράς αναζήτησης του χρήστη (π.χ. παρακολούθηση ερωτημάτων αναζήτησης και επιλογή συνδέσμων για μεγάλο χρονικό διάστημα) και τη δημιουργία ενός μακροπρόθεσμου προφίλ προτιμήσεων και ενδιαφερόντων, εντάσσονται επίσης ο εντοπισμός του πλαισίου, καθώς και η προσωπικότητα του χρήστη.

Ο εντοπισμός του πλαισίου είναι πολύ κοντινή έννοια με την εξατομίκευση, υπό το πρίσμα ότι προσαρμόζεται στην κατάσταση του χρήστη και όχι στον ίδιο το χρήστη, που σημαίνει ότι δύο χρήστες που εκτελούν το ίδιο ερώτημα και επιλέγουν τον ίδιο σύνδεσμο στη διάρκεια μίας συνεδρίας αναζήτησης, μοιράζονται το ίδιο πλαίσιο. Οποιοσδήποτε με το ίδιο ιστορικό αλληλεπιδράσεων θα έχει την ίδια

προσαρμοσμένη εμπειρία και είναι πιο πιθανό ότι πολλοί άνθρωποι μπορεί να μοιράζονται ένα κοινό ιστορικό αλληλεπιδράσεων παρά ένα όμοιο προφίλ μακροπρόθεσμων προτιμήσεων.

Οι μελέτες έχουν συσχετίσει την προσωπικότητα με την αναζήτηση πληροφοριών (Heinström, 2003, Nie et al., 2014, Marchionini, 2009). Η σχέση μεταξύ της προσωπικότητας και των γνωστικών στρατηγικών επεξεργασίας έχει μελετηθεί και τα χαρακτηριστικά της προσωπικότητας έχουν χρησιμοποιηθεί ως μέσο πρόβλεψης της τάσης ενός ατόμου να επεξεργάζεται πληροφορίες (Higgins, King, Cantor & Kihlstorm, 1981). Ως εκ τούτου, η προσωπικότητα είναι ένα χαρακτηριστικό χρήστη που πρέπει να θεωρηθεί ως ένας παράγοντας που επηρεάζει την συμπεριφορά αναζήτησης και τις τακτικές διερεύνησης πληροφοριών, αφού μπορεί να αντανακλάται στις προτιμήσεις του αναζητητή (Marchionini, 2006a).

2.3.4.1 Εξατομίκευση με Μοντελοποίηση Μακροπρόθεσμων Ενδιαφερόντων

Τα συστήματα αναζήτησης μπορεί να μην είναι αντιπροσωπευτικά των αναγκών ενός μεμονωμένου αναζητητή και η προσωποποίηση μπορεί να βοηθήσει προς αυτήν τη κατεύθυνση (Teevan et al., 2010).

Με βάση τη γνώση μας στο χώρο της Διερευνητικής Αναζήτησης δεν υπάρχουν αρκετές εργασίες που να εστιάζουν στην προσωποποιημένη προσαρμογή της εμπειρίας ή των αποτελεσμάτων αναζήτησης. Ωστόσο, αυτό είναι ένας πολύ σημαντικός παράγοντας που έχει συσχετιστεί με αυξημένη ικανοποίηση χρήστη στα συστήματα αναζήτησης, ο οποίος χρησιμεύει προκειμένου να απομακρύνονται δεδομένα που δε σχετίζονται με την τρέχουσα δραστηριότητα ή τις προτιμήσεις του χρήστη. Παρακάτω παρατίθεται μία σύνοψη των μεθοδολογιών επίτευξης προσωποποίησης στα συστήματα αναζήτησης.

Μέθοδοι για την υποστήριξη της προσωποποίησης αναπτύσσονται συνήθως με τη μακροπρόθεσμη παρακολούθηση των ενεργειών του αναζητητή. Αρκετοί ερευνητές έχουν προτείνει ως προσωποποίηση τη μοντελοποίηση της συμπεριφοράς

του αναζητητή στα πλαίσια μας συνεδρίας (Daoud et al. 2009; Sriram et al. 2004), όμως τα δεδομένα είναι πολύ αραιά και η αναζήτηση -άρα και εργασία- πολύ εξειδικευμένη, ώστε να είναι αρκετά τα δεδομένα για να αναπαραστήσουν επαρκώς τον αναζητητή και να προσαρμόσουν την εμπειρία αναζήτησης (White et al., 2016 p. 267).

Τα μοντέλα προσωποποίησης μπορούν να χτιστούν με βάση τη διαδικτυακή συμπεριφορά του χρήστη ή με βάση τα αρχεία του προσωπικού υπολογιστή του αναζητητή. Η προσωποποίηση με μακροπρόθεσμη παρακολούθηση της συμπεριφοράς του αναζητητή οδηγεί στο σχηματισμό μοναδικών προφίλ. Όπως προτείνεται από τους (White et al. 2013a; Yan et al. 2014), όταν δεν υπάρχει επαρκής πληροφορία για τα ενδιαφέροντα του χρήστη, τα προφίλ μπορούν να εμπλουτιστούν από τα ενδιαφέροντα παρόμοιων χρηστών, όπου το παρόμοιοι χρήστες μπορεί να ορίζεται π.χ. με γεωγραφικά κριτήρια, εμπειρία, προτιμήσεις.

Προκειμένου να δημιουργηθούν προφίλ χρήστη, οι διάφορες προτεινόμενοι μέθοδοι βασίζονται είτε στο περιεχόμενο των αποτελεσμάτων που επιστρέφονται στον υπολογιστή του χρήστη -άρα στην εξαγωγή όρων από τα έγγραφα που είναι σχετικά στον χρήστη και τα οποία αναπαρίστανται ως σταθμισμένα διανύσματα- είτε στο περιεχόμενο των ερωτημάτων του χρήστη και τα κλικ στα αποτελέσματα. Οι Teevan et al. (2005) εξέτασαν την αποτελεσματικότητα μίας προσωποποιημένης μεθόδου χρησιμοποιώντας δεδομένα αποθηκευμένα στον τοπικό υπολογιστή του αναζητητή και προηγούμενες αναζητήσεις (ερωτήματα και επισκέψεις ιστοσελίδων). Οι όροι που εξήχθησαν από αυτά τα αρχεία χρησιμοποιήθηκαν για να ανακατατάξουν τα αποτελέσματα αναζήτησης, καθώς θεωρήθηκε ότι αντιπροσωπεύουν τα προσωπικά ενδιαφέροντα του αναζητητή. Στην εργασία επισκόπησης και αξιολόγησης προσωποποιημένων συστημάτων των Dou et al. (2009), αναλόγως του αν για το σχηματισμό των προφίλ και εν τέλει την εύρεση της συνάφειας των αιτημάτων και των επιλεγμένων αποτελεσμάτων με το προφίλ χρησιμοποιούνται θεματικές κατηγορίες ή «σάκοι λέξεων», οι σχετικές εργασίες χωρίζονται σε αυτές τις δύο κατηγορίες.

Στην κατηγορία των προφίλ που δομούνται ως ιεραρχία θεμάτων, τα ερωτήματα που έχουν πραγματοποιηθεί από το χρήστη και τα έγγραφα που έχουν επιλεγεί από τον χρήστη κατηγοριοποιούνται σε ιεραρχίες θεμάτων που συγκεντρώνονται για τη δημιουργία προφίλ χρήστη. Όταν ο χρήστης πραγματοποιεί ένα ερώτημα, κάθε ένα από τα επιστρεφόμενα έγγραφα κατηγοριοποιείται επίσης. Τα έγγραφα ανακατατάσσονται ανάλογα με το πόσο καλά οι κατηγορίες τους ταιριάζουν με τα ενδιαφέροντα των προφίλ (Pretschner & Gauch, 1999; Chirita et al., 2005; Gauch et al., 2003; Speretta & Gauch, 2005).

Στην κατηγορία των προφίλ που αντιπροσωπεύονται από σάκους λέξεων υπάρχουν αρκετές εργασίες. Οι Sugiyama et al. (2004) αναπαράστησαν τις προτιμήσεις των χρηστών ως διανύσματα ξεχωριστών όρων και τα κατέταξαν συγκεντρώνοντας παλαιότερες προτιμήσεις, συμπεριλαμβανομένων τόσο των μακροπρόθεσμων όσο και των βραχυπρόθεσμων προτιμήσεων. Οι Shen et al. (2005) χρησιμοποίησαν για πρώτη φορά τη γλωσσική μοντελοποίηση για να συλλέξουν πληροφορίες συμφραζομένων από ένα ιστορικό βραχυπρόθεσμης αναζήτησης. Οι Tan et al. (2006) έπειτα χρησιμοποίησαν τη μέθοδο για να συλλέξουν το πλαίσιο από ένα μακροπρόθεσμο ιστορικό αναζήτησης. Οι Teevan et al. (2005) και Chirita et al. (2006) εκμεταλλεύονται πλούσια μοντέλα ενδιαφερόντων χρηστών, που βασίζονται τόσο σε πληροφορίες σχετικές με την αναζήτηση όσο και σε άλλες πληροφορίες σχετικά με το χρήστη, συμπεριλαμβανομένων των εγγράφων και των μηνυμάτων ηλεκτρονικού ταχυδρομείου που έχει διαβάσει και δημιουργήσει ο χρήστης. Στο έργο των Liu et al. (2002;2004), οι λέξεις-κλειδιά συνδέονται με κατηγορίες και, επομένως, τα προφίλ χρηστών αντιπροσωπεύονται από μια ιεραρχική δομή κατηγοριών που βασίζεται σε κατηγορίες λέξεων-κλειδιών.

2.3.4.2 Εξατομίκευση με Μοντελοποίηση Χαρακτηριστικών Προσωπικότητας

Εξ' όσων γνωρίζουμε, στη διερευνητική αναζήτηση και την αναζήτηση πληροφορίας γενικότερα υπάρχει πολύ περιορισμένη έρευνα αναφορικά με τη μελέτη της επίδρασης της προσωπικότητας. Στην αναζήτηση Ιστού, οι Nie et al. (2014)

εξερεύνησαν τα μοτίβα συσχέτισης μεταξύ χαρακτηριστικών αναζήτησης (π.χ. τάση περιήγησης σε πολλαπλές κατηγορίες, full-text search) και των χαρακτηριστικών προσωπικότητας, ενώ προσπαθούσαν να εκμεταλλευτούν συμπεριφορές αναζήτησης για να προσδιορίσουν χαρακτηριστικά προσωπικότητας.

2.3.4.3 Εξατομίκευση με Επίγνωση Πλαισίου

Η επίγνωση πλαισίου προσαρμόζεται ή προσαρμόζει στην κατάσταση αναζήτησης. Η επίγνωση πλαισίου της πληροφοριακής ανάγκης μπορεί να βοηθήσει στην συστημική αναπαράσταση της πρόθεσης της αναζήτησης πέρα από τις ενδείξεις που συγκεντρώνονται από ερωτήματα και κλικς. Οι μέθοδοι για την υποστήριξη της επίγνωσης πλαισίου αναπτύσσονται συνήθως μέσω πρόσθετων «σημάτων» τα οποία είναι διαθέσιμα στο σύστημα τη στιγμή της αναζήτησης. Αυτά είναι η γεωγραφική τοποθεσία, ο χρόνος και πρόσφατη συμπεριφορά αλληλεπίδρασης. Οι αλληλεπιδράσεις μπορεί να αφορούν είτε α) τον πυρήνα της αναζήτησης όπου παίζουν ρόλο π.χ. τα ερωτήματα, οι επιλογές αποτελεσμάτων και άλλες λειτουργικές αλληλεπιδράσεις, όπως η μετάβαση σε άλλη σελίδα ή προηγούμενο ερώτημα είτε β) την περιήγηση πριν και μετά την εκτέλεση ερωτήματος. Παρόλο που οι περισσότερες μελέτες εστιάζουν στα ερωτήματα και τα clicks, οι αλληλεπιδράσεις επεκτείνονται ακόμα και σε φυσικές ενέργειες όπως στη μετακίνηση και σύρσιμο επάνω σε συσκευές με δυνατότητα αφής, στον προφορικό λόγο ή σε χειρονομίες. Επιλέγεται πολλές φορές να λέγεται ότι οι πρόσφατες ενέργειες είναι σήμα πλαισίου παρά προσωποποίησης, καθώς ο οποιοσδήποτε με τις ίδιες προηγούμενες ενέργειες θα είχε την ίδια προσαρμοσμένη εμπειρία ενώ τα μακροπρόθεσμα προφίλ που συμπεριλαμβάνουν προτιμήσεις είναι διαφορετικά μεταξύ χρηστών. Παρακάτω ενδεικτικά αναφέρονται επιπλέον επεξηγήσεις για τρεις βασικούς τύπους πλαισίου αν και δεν είναι οι μοναδικοί: α) πλαίσιο αλληλεπίδρασης, β) φυσικό πλαίσιο και γ) χρονικό πλαίσιο.

2.3.4.3.1 Πλαίσιο Αλληλεπίδρασης

Το πλαίσιο αλληλεπίδρασης αφορά τις απευθείας αλληλεπιδράσεις με το σύστημα αναζήτησης, όπως ερωτήματα, επιλεγμένα αποτελέσματα και άλλες λειτουργικές αλληλεπιδράσεις όπως πλοήγηση στις σελίδες, γρήγορο γύρισμα σε προηγούμενες σελίδες.

Υπάρχει πλούσια έρευνα που εστιάζει στο τρέχον πλαίσιο αλληλεπίδρασης προκειμένου να βελτιώσει την εμπειρία (Xiang et al. 2010; Bennett et al. 2012). Οι πρώτοι, ανέπτυξαν ευριστικούς αλγορίθμους για να προωθήσουν αποτελέσματα με την ίδια θεματική κατηγορία σε ανώτερες θέσεις της λίστας, εάν διαδοχικά ερωτήματα στη διάρκεια της συνεδρίας συσχετιζόνταν με βάση κάποια μετρική ομοιότητας και δεν ήταν εξειδικεύσεις, γενικεύσεις ή μετασχηματισμοί.

2.3.4.3.2 Φυσικό Πλαίσιο

Για την υποστήριξη της αναζήτησης σε συνθήκες όπου η τοποθεσία μετράει τα συστήματα μπορούν να αξιοποιήσουν την τοποθεσία για αποσαφήνιση ερωτήματος και επανιεράρχηση αποτελεσμάτων. Η τοποθεσία των αναζητητών μιας Διαδικτυακής Υπηρεσίας μπορεί να προσδιοριστεί είτε χρησιμοποιώντας τις συντεταγμένες από τα GPS των κινητών συσκευών είτε με τις αντίστροφες IP. Συνεπώς, άλλη βαρύτητα αποκτούν τα αποτελέσματα ενός ερωτήματος που σχετίζεται με ταινίες στον κινηματογράφο, καθώς συνήθως ενδιαφέρουν περισσότερο τα αποτελέσματα που σχετίζονται με τους κοντινότερους κινηματογράφους σύμφωνα με το στίγμα του αναζητητή.

Οι West et al. 2013, χρησιμοποίησαν καταγεγραμμένα δεδομένα ερωτημάτων που αφορούσαν γεωγραφικές τοποθεσίες, για να δείξουν ότι ένα σύστημα αναζήτησης μπορεί να προβλέψει με μεγάλη ακρίβεια ότι ο αναζητητής πρόκειται να επισκεφθεί μία συγκεκριμένη τοποθεσία. Με αυτόν τον τρόπο το σύστημα μπορεί προδραστικά να ετοιμάσει μία λίστα από δραστηριότητες που μπορεί να πραγματοποιήσει ο αναζητητής εκεί κοντά.

2.3.4.3.3 Χρονικό Πλαίσιο

Η πρόσβαση στην πληροφορία με επίγνωση χρόνου είναι μία περιοχή έρευνας που κερδίζει έδαφος. Η ανάλυση της τάσης στις Διαδικτυακές αναζητήσεις δείχνουν τη συνολική δημοτικότητα των πηγών πληροφορίας του Ιστού (Sellen et al., 2002). Οι μελέτες των τάσεων είναι χρήσιμες ώστε να προβλεφθούν με υψηλότερη ακρίβεια οι δραστηριότητες των χρηστών. Η νεότητα των αποτελεσμάτων είναι πολύ σημαντική ειδικά όσο αφορά αναζητήσεις που είναι για παράδειγμα συνδεδεμένες με εποχιακά γεγονότα, πχ. απονομή των Όσκαρ, επιστημονικά συνέδρια ή αθλητικά γεγονότα. Οι χρήστες αιτούνται να λάβουν πληροφορία σχετική με τα παραπάνω θέματα και περιμένουν να είναι η πιο πρόσφατη. Οι Dong et al. (2010) πρότειναν έναν ταξινομητή για να εντοπίζει τέτοια αιτήματα. Έρευνα που σχετίζεται με εννοιολογική αλλαγή ψάχνει να εντοπίσει αλλαγές στη συμπεριφορά αναζήτησης παρακολουθώντας τη στο χρόνο (Kiseleva et al. 2014). Υπάρχουν μελέτες που έχουν υλοποιήσει ταξινομητές για τον εντοπισμό έκτακτων και επειγόντων ερωτημάτων, όπως αυτά που αφορούν φυσικές καταστροφές (Mishra et al., 2014). Το αποτέλεσμα τέτοιων μελετών έχει εφαρμογή σε αυτόματη συμπλήρωση ερωτημάτων, την αποσαφήνιση ερωτημάτων και επανιεράρχηση αποτελεσμάτων με βάση την «ηλικία» ή την κρισιμότητα.

2.3.5 Επισκόπηση Συστημάτων Διερευνητικής Αναζήτησης

Πολλές μέθοδοι έχουν προταθεί, οι οποίες υιοθετούν το μοντέλο της Διερευνητικής Αναζήτησης και στόχο έχουν να διευκολύνουν τη διαδικασία διερεύνησης. Παρακάτω παρατίθεται επισκόπηση αρκετών συστημάτων Διερευνητικής Αναζήτησης που έχουν προταθεί στη βιβλιογραφία, τα οποία προσφέρουν υποστήριξη για το σχηματισμό ερωτημάτων, για αποσαφήνιση ερωτημάτων, για τη βελτίωση της συνάφειας των αποτελεσμάτων σε σχέση με το στόχο του χρήστη. Παρουσιάζονται επίσης συστήματα που παρέχουν διεπαφές οι οποίες προσφέρουν πολύπλευρη πλοήγηση και οπτικοποιήσεις που συνεισφέρουν στη λήψη αποφάσεων ή στην κατανόηση ενός συγκεκριμένου θέματος, καθώς και συστήματα που στοχεύουν στην ενίσχυση της συνεργασίας μεταξύ αναζητητών που μπορεί να εργάζονται πάνω σε μία κοινή ανάθεση. Τέλος, θα δούμε μία ποικιλία προσεγγίσεων, οι οποίες άλλοτε μπορεί να λειτουργούν επικουρικά σε μία κοινή

Αναζήτηση Ιστού, άλλοτε εισάγουν στους αλγορίθμους τους στοιχεία από τον Κοινωνικό Ιστό ή από τα ανοιχτά Διασυνδεδεμένα Δεδομένα.

2.3.5.1 Συστήματα με χρήση Κοινωνικού Πλαισίου

Από τις πρώτες εργασίες εισαγωγής πληροφορίας από το κοινωνικό πλαίσιο του αναζητητή είναι αυτή των συγγραφέων Donaldson et al. (Donaldson et al., 2008), των οποίων σκοπός ήταν να αξιολογήσουν τη δυνατότητα αξιοποίησης κοινωνικών συνδέσμων στην οπτικοποίηση αποτελεσμάτων αναζήτησης. Για το σκοπό αυτό προτείνουν τον υπολογισμό ομοιότητας μεταξύ διαδικτυακών πηγών, με την εξόρυξη και άθροιση των επισημάνσεων των χρηστών ενός κοινωνικού ιστοτόπου καταγραφής σελιδοδεικτών. Η αξιολόγηση της μεθόδου έδειξε ότι οι χρήστες προτιμούν το υβριδικό σύστημα, όπου η παραδοσιακή ταξινομημένη διάταξη κατάταξης αποτελεσμάτων συμπληρώνεται από την οπτικοποίηση αποτελεσμάτων με σχέσεις κοινωνικής ομοιότητας. Το Mr Taggy είναι ένα σύστημα (Kammerer et al., 2009) στο οποίο υλοποιήθηκε μέθοδος διερευνητικής αναζήτησης, η οποία βασίστηκε στη συλλογική συμπεριφορά χρηστών που επισήμαναν περιεχόμενο. Η μέθοδος συνδυάζει τεχνολογίες αναζήτησης και προτάσεων, οι οποίες παράγονται και αναταξινομούνται αξιοποιώντας την ανατροφοδότηση του χρήστη σχετικά με τη συνάφεια των αποτελεσμάτων και την ομοιότητα των υποψήφιων προτεινόμενων αποτελεσμάτων με το θέμα με βάση την επισήμανσή τους σε κοινωνικούς ιστοτόπους καταγραφής σελιδοδεικτών. Η αξιολόγηση της προτεινόμενης μεθόδου έδειξε ότι α) οι χρήστες εκμεταλλεύτηκαν την επιπρόσθετη λειτουργικότητα του συστήματος και ενεπλάκησαν περισσότερο στη διαδικασία αναζήτησης ενώ β) διαπιστώθηκε ότι η μέθοδος αυτή συνέβαλλε στη μάθηση αναφορικά με την εργασία που τους είχε ανατεθεί. Στην εργασία των Zilincik et al. (Zilincik et al., 2013) οι ερευνητές πρότειναν ένα εργαλείο διερευνητικής αναζήτησης για το Twitter που προσπαθεί να μάθει τις προτιμήσεις του χρήστη και να εκτιμήσει τον αρχικό τους σκοπό εκπαιδύοντας μοντέλο προτιμήσεων βάσει έμμεσης ανατροφοδότησης, π.χ. χρόνος ανάγνωσης ενός tweet και ανατροφοδότησης βάσει άμεσης ανατροφοδότησης, π.χ. βαθμολόγηση ενός tweet σαν ενδιαφέρον ή όχι με τη χρήση του αλγορίθμου SVM. Ο

σκοπός του είναι να επιτρέψει στο χρήστη να αποκτήσει μια αίσθηση του είδους του περιεχομένου που δημοσιεύεται στο Twitter σε σχέση με ένα θέμα και να παρουσιάσει προτάσεις σχετικές με το ενδιαφέρον του χρήστη τη στιγμή της διερεύνησης. Προκειμένου να διευκολύνουν τη διερεύνηση προτείνουν τη δημιουργία θεμάτων εξάγοντας τις πιο συχνά αναφερόμενες φράσεις στα tweets χρησιμοποιώντας ακόμα και το περιεχόμενο των συνδέσμων που αναφέρονται σε αυτά για τον εμπλουτισμό της πληροφορίας που περιγράφει ένα tweet. Η μέθοδος αξιολογήθηκε για την ποιότητα των προτάσεων και φάνηκε να αποτελεί καλή βάση για περαιτέρω έρευνα.

Το TweetMotif (O'Connor, Krieger & Ahn, 2010) εξάγει ένα σύνολο θεμάτων από το Twitter για να προσφέρει μια πολύπλευρη διεπαφή αναζήτησης. Στόχος είναι να προσφερθεί μία συνοπτική περίληψη των Tweets σχετικών με το θέμα της αναζήτησης προσφέροντας θέματα ή κατηγορίες για εύκολη περιήγηση. Παίρνει οποιαδήποτε λέξη ή φράση, βρίσκει τα tweets όπου οι άνθρωποι μιλάνε γι' αυτό, στη συνέχεια τα ομαδοποιεί με στατιστικά απίθανες φράσεις που συνυπάρχουν – και με αυτή τη μέθοδο προκύπτουν τα θέματα της συζήτησης. Το σύστημα ανάλυσης κειμένων βρίσκει τις συστάδες με τα σχεδόν διπλότυπα, προκειμένου να τα αντικαταστήσει με ένα αντιπροσωπευτικό, συγκρίνοντας την ομοιότητα μεταξύ των φράσεων 3 λέξεων (trigrams) που έχουν βρεθεί στα μηνύματα. Μία υβριδική μέθοδος κοινωνικής Διερευνητικής Αναζήτησης ελεγχόμενης από το χρήστη προτάθηκε από τους Sciascio et al (Sciascio et al. 2018), η οποία μπορεί να συνδυάσει 3 μοντέλα: α) το μοντέλο που βασίζεται αυστηρά στο περιεχόμενο, δηλαδή στα αποτελέσματα που θα επιστρέφονταν εάν εφαρμοστούν οι παραδοσιακές μέθοδοι της Ανάκτησης Πληροφορίας, β) το μοντέλο που βασίζεται στους χρήστες, όπου τα αποτελέσματα θα επιστρέφονταν εφαρμόζοντας αλγόριθμους Συνεργατικής Διήθησης, και γ) το μοντέλο που βασίζεται σε ομοιότητα αποτελεσμάτων βάσει των επισημάνσεων χρηστών στο σύστημα όπου έγινε εφαρμογή της μεθόδου. Η αξιολόγηση της μεθόδου έδειξε ότι προάγει την ικανοποίηση του χρήστη.

2.3.5.2 Συστήματα με χρήση Ανοιχτών Διασυνδεδεμένων Δεδομένων

Υπάρχουν μελέτες που έχουν επικεντρωθεί στην διερεύνηση εικόνων και άλλων πόρων πολυμέσων, βλέπε π.χ. (Waitelonis & Sack, 2012, Iijima, Kimura & Yamaguchi, 2010). Στην εργασία των Waitelonis και Sack (2012), για παράδειγμα, οι συγγραφείς προτείνουν ένα διερευνητικό σύστημα αναζήτησης για βίντεο που εκμεταλλεύεται Linked Open Data (Ανοιχτά Διασυνδεδεμένα Δεδομένα) και συγκεκριμένα το ¹DBpedia για να σχηματίσουν συσχετισμούς που βασίζονται σε περιεχόμενο και να καταστήσουν εφικτή την ανακάλυψη απρόσμενων αλλά ευχάριστων πόρων.

Οι Marie, Gandon και Ribière (2013) υλοποίησαν το Discovery Hub, ένα σύστημα διερευνητικής αναζήτησης το οποίο επιλέγει ενδιαφέροντες πόρους και υποστηρίζει πολύπλευρη περιήγηση στα αποτελέσματα κάνοντας χρήση των Ανοιχτών Διασυνδεδεμένων Δεδομένων (LOD). Παρέχει διερευνητική αναζήτηση σε απομακρυσμένα δεδομένα χρησιμοποιώντας τα SPARQL endpoints για να εξασφαλίσει τα πιο πρόσφατα δεδομένα. Επιτρέπει την επεξεργασία των συνδεδεμένων δεδομένων επί τόπου, δηλαδή χωρίς προεπεξεργασία. Για να επιτευχθεί γρήγορη απόκριση του συνόλου των αποτελεσμάτων, πρότειναν έναν αλγόριθμο ενεργοποίησης σημασιολογικής εξάπλωσης και μια τεχνική δειγματοληψίας που το καθιστά εφαρμόσιμο σε απομακρυσμένους γράφους Ανοιχτών Διασυνδεδεμένων Δεδομένων (LOD).

2.3.5.3 Συστήματα Συνεργατικής Διερευνητικής Αναζήτησης

Το Cerchiamo (Golovchinsky et al., 2008) είναι μία διεπαφή συνεργατικής διερευνητικής αναζήτησης. Προτείνει μία μέθοδο που συνδυάζει α) την αλγοριθμική λειτουργικότητα, όπου εκτελεί βασικές αναζητήσεις και παράγει ακατέργαστα αλλά και μετασχηματισμένα αποτελέσματα αναζήτησης, βάσει εισροών από πολλούς

¹ <http://wiki.dbpedia.org/Lookup>

χρήστες και όρους αναζήτησης που χαρακτηρίζουν την τρέχουσα κατάσταση της συνεργασίας, β) τη λειτουργικότητα ανατροφοδότησης συνάφειας και γ) της οντότητας που διερευνά συνεχώς υποσχόμενα μονοπάτια, επομένως ενημερώνει συνεχώς τον κατάλογο των προτάσεων όρων ερωτήματος που βασίζονται σε πρόσφατα εντοπισμένα σχετικά έγγραφα. Οι ερευνητές Yue et al. (Yue et al., 2014) προτείνουν ένα σύστημα συνεργατικής διερευνητικής αναζήτησης Ιστού. Τα ερωτήματα των χρηστών συνδέονται με τη μηχανή αναζήτησης Google και υπάρχει πλαίσιο ιστορικών αναζητήσεων για τους χρήστες και τους συνεργάτες. Υπάρχει δυνατότητα αποθήκευσης αποτελεσμάτων στο χώρο εργασίας της ομάδας. Με χρήση του συστήματος και ανάλυση του πλαισίου αλληλεπίδρασης χρησιμοποιώντας Κρυφά Μαρκοβιανά Μοντέλα, οι ερευνητές βρίσκουν ομοιότητες στα μοτίβα αναζητήσεων με τα στάδια του μοντέλου Information Search Process (Marchionini, 1995). Οι Herceg et al. (2018) πρότειναν μία στρατηγική που ενσωματώνει αρχές συνεργατικής Διερευνητικής Αναζήτησης, στοχεύοντας στη βελτιστοποίηση επιχειρηματικών διαδικασιών για την κλιμάκωση μεγάλων φορτίων πληροφόρησης. Στην εργασία τους ανέλυσαν την απόδοση του συνδυασμού σύγχρονης συνεργατικής διερευνητικής αναζήτησης από αναζητητές που συνέλεξαν αποτελέσματα αναζήτησης για κάθε πληροφοριακή ανάγκη και κατευθυνόμενης περιήγησης από εμπειρογνώμονα, ο οποίος τα ανέλυε, τα ιεραρχούσε και τα απομόνωνε για εστιασμένη ανάγνωση. Σε αντίθεση με την αναζήτηση στο Web, η υβριδική στρατηγική έδειξε ότι ταιριάζει σε οργανισμούς υψηλής απόδοσης με εκθετικούς ρυθμούς αύξησης των δεδομένων καθώς παρέχει σημαντική μείωση των δεδομένων προς επεξεργασία. Η μείωση αυτή διευκολύνεται από τα συγκεντρωτικά αποτελέσματα αναζήτησης, τα οποία παρέχουν δείκτες σε αρχεία που περιέχουν σχετικές πληροφορίες από που έχουν προέλθει από το στάδιο της ιεράρχησης από τον εμπειρογνώμονα. Αυτό έχει θετικό αντίκτυπο καθώς σε περιορισμένο χρονικό ορίζοντα, ο εμπειρογνώμονας μπορεί τώρα να επικεντρωθεί στις πιο σχετικές πληροφορίες. Οι Andolina et al. (2018) πρότειναν μία προσέγγιση που βασίστηκε σε οντότητες για τη διευκόλυνση της αυθόρμητης διερεύνησης μεγάλων χώρων πληροφόρησης που επιτρέπουν τη συνεργασία στον ίδιο τόπο. Υλοποίησαν το

QueryTogether με μία διεπαφή που βασίστηκε σε οντότητες όπως άτομα, τα έγγραφα και οι λέξεις-κλειδιά για την υποστήριξη αναζήτησης σε ετερογενείς συσκευές. Οι οντότητες χρησίμευαν για τη σύνταξη ερωτήσεων που μπορούν να μοιραστούν σε μια δημόσια οθόνη ή συγκεκριμένους χρήστες με αλληλεπίδραση με εύκολη επαφή. Οι συμμετέχοντες βρήκαν τα περισσότερα από τα σχετικά αποτελέσματα νωρίτερα από ότι με ένα κοινό σύστημα συνεργατικής αναζήτησης και για περισσότερα από τα μισά ερωτήματα αποφεύγονταν η εισαγωγή κειμένου με το χειρισμό συνιστώμενων οντοτήτων. Συνολικά η συγκεκριμένη προσέγγιση οδήγησε σε πιο ισορροπημένη συνεργασία και δέσμευση των χρηστών στην αναζήτηση.

2.3.5.4 Έξυπνες Διεπαφές Διερευνητικής Αναζήτησης

Οι ερευνητές (Kangasrääsiö et al. 2015) πρότειναν μία τροποποίηση σε σύστημα που υποστηρίζει τεχνική Ενίσχυσης Μάθησης προκειμένου να βελτιώσουν ζητήματα ελεγχιμότητας από το χρήστη και προβλεψιμότητας του συστήματος. Η προσέγγιση βασίζεται στην εισαγωγή ενός στρώματος που μεταφράζει τις εισροές του χρήστη σε απαιτήσεις για την κατάσταση του συστήματος και καθιστά το σύστημα πιο προβλέψιμο, δείχνοντας στον χρήστη τις εκτιμήσεις των αποτελεσμάτων που θα έχουν οι ενέργειές του στο σύστημα. Θεωρούν την ανατροφοδότηση από το χρήστη σαν στόχο για ένα πρόβλημα βελτιστοποίησης του μοντέλου χρήστη, όπου η βελτιστοποίηση αντιμετωπίζεται σε μια γραμμική άπληστη αυξητική αναζήτηση. Από την άλλη η κατά προσέγγιση πρόβλεψη της κατάστασης του συστήματος αναλόγως της ενέργειας του χρήστη υπολογίστηκε ως γραμμική παρεμβολή ανά λέξη-κλειδί. Ιδέα του αλγορίθμου ήταν να υπολογιστούν τα δυνατά μοντέλα που χρησιμοποιούν τις μέγιστες τιμές συνάφειας για το κλειδί της λέξης και να χρησιμοποιηθεί μια γραμμική παρεμβολή ως πρόβλεψη μοντέλου χρήστη για οποιαδήποτε τιμή συνάφειας μεταξύ των άκρων. Η αξιολόγηση του προτεινόμενου συστήματος έδειξε ότι υπάρχουν ενδείξεις στην απόδοση της εργασίας, στην ευχρηστία, την αντιληπτή χρησιμότητα και αποδοχή του συστήματος.

Το DBExplorer (Singh et al., 2016) είναι ένα διερευνητικό σύστημα αναζήτησης για σχεσιακές βάσεις δεδομένων. Η λύση βασίζεται σε μια νέα τεχνική περίληψης δεδομένων, η οποία εξαρτάται από το πλαίσιο αναζήτησης του χρήστη. Συνδυάζει την περίληψη δεδομένων με πολύπλευρη πλοήγηση, με αποτέλεσμα να υποστηρίζονται ταυτόχρονα γρήγορη αλλά και διερευνητική αναζήτηση. Βασίζονται σε υπολογισμό ομοιότητας πληροφορίας και παρουσιάζουν το σετ αποτελεσμάτων με σύγκριση αυτών προκειμένου οι χρήστες να κατανοήσουν καλύτερα τις τιμές των διαφόρων χαρακτηριστικών κάθε οντότητας.

Οι ερευνητές (Ma & Zhang, 2018) πρότειναν μία μέθοδο προτάσεων ερωτημάτων για την υποστήριξη της διερευνητικής διαδικασίας. Όρισαν το «στόχο αναζήτησης» και τη «μετατόπιση του στόχου αναζήτησης» και βάσει αυτών των ορισμών, αφού ανέλυσαν τη συμπεριφορά των χρηστών και εντόπισαν τις «μετατοπίσεις» στις συνεδρίες με ένα αλγόριθμο μηχανικής μάθησης, κατασκεύασαν ένα «γράφο μετατοπίσεων». Για τις προτάσεις ερωτημάτων χρησιμοποιήθηκε το εκπαιδευμένο μοντέλο για να κριθεί αν οι συμπεριφορές αναζήτησης των χρηστών ανήκουν στην "μετατόπιση στόχου αναζήτησης", και με χρήση του αλγορίθμου Random Walk βρίσκονταν τα top-k πιο συναφή ερωτήματα αναζήτησης.

Με το SearchLens (Chang et al., 2019) προτάθηκε μία έξυπνη διεπαφή στον τομέα της διερεύνησης εστιατορίων, όπου αναγνωρίζει άμεσα ή έμμεσα τις προτιμήσεις του χρήστη χρησιμοποιώντας συλλογές σταθμισμένων λέξεων-κλειδιά, αντιστοιχίζοντάς τις σε σημασιολογικές κατηγορίες αποκαλύπτοντας με αυτόν τον τρόπο τα ενδιαφέροντα του χρήστη. Οι σημασιολογικές κατηγορίες υπάγονται σε ένα εκπαιδευμένο μοντέλο το οποίο χρησιμοποιείται για προτάσεις όρων, ενώ δίδονται οπτικές εξηγήσεις (ακόμα και με στατιστική μορφή για τα επιστρεφόμενα αποτελέσματα), που έχουν να κάνουν με τη δημοτικότητα των αποτελεσμάτων. Για την κατάταξη των αποτελεσμάτων χρησιμοποιήθηκε ο αλγόριθμος Okapi BM25 σταθμισμένος με τον παράγοντα που λαμβάνει υπόψιν τις προτιμήσεις των χρηστών. Η προσέγγιση θέτει το χρήστη στο επίκεντρο της διερεύνησης και αξιολογήθηκε για τη συνολική της χρησιμότητα.

Πίνακας 2 Σύνοψη Συστημάτων Διερευνητικής Αναζήτησης

Authors	Category	Method	Evaluation
Donaldson et al. 2008;	Διερευνητική Αναζήτηση σε κοινωνικό ιστότοπο καταγραφής σελιδοδεικτών.	Οπτικοποίηση κοινωνικών συνδέσεων με χρήση μεθόδου ομοιότητας βασισμένη σε σχολιασμούς σελιδοδεικτών	Αξιολόγηση της διαδικασίας διερεύνησης, της χρησιμότητας της διεπαφής, της ποιότητας της πληροφορίας
Kammerer et al. 2009	Διερευνητική Αναζήτηση σε κοινωνικό ιστότοπο καταγραφής σελιδοδεικτών.	Εμπλουτισμός αποτελεσμάτων με κοινωνικούς δεσμούς - Υπολογισμός πιθανότητας ομοιότητας βάσει κοινωνικών επισημάνσεων Σύστημα MapReduce με βάση την Bayesian εξαγωγή συμπερασμάτων και ενεργοποίηση εξάπλωσης	Αξιολόγηση συμπεριφοράς αλληλεπίδρασης, γνωστικού φορτίου, ποιότητας εργασιών και εκμάθησης
Zilincik et al. 2013	Κοινωνική Διερευνητική Αναζήτηση στο Twitter	Οπτικοποίηση θεμάτων και εμπλουτισμός των αποτελεσμάτων με προτάσεις που βασίζονται σε μοντέλο προτίμησης εκπαιδευμένο με έμμεση ανατροφοδότηση	Ποιότητα προτάσεων
O'Connor, Krieger & Ahn, 2010	Κοινωνική Διερευνητική Αναζήτηση στο Twitter	Περίληψη των θεμάτων με τη μοντελοποίηση της στατιστικής γλώσσας	Χωρίς αξιολόγηση
Sciascio et al. 2018	Κοινωνική Διερευνητική Αναζήτηση	Αξιοποίηση της «κοινωνικής σοφίας» με βάση το ταίριασμα των επισημάνσεων και την ομοιότητα των χρηστών - παραδοσιακές μέθοδοι	Αξιολόγηση Ακρίβειας Αποτελεσμάτων & Ικανοποίησης Χρήστη

		ανάκτησης πληροφοριών και συνεργατική διήθηση.	
Waitelonis & Sack (2012)	Οπτική Διερευνητική Αναζήτηση	Ευριστικοί αλγόριθμοι βασισμένοι στην αξιοποίηση των διασυνδεδεμένων ανοικτών δεδομένων	Αξιολόγηση επιτυχίας εργασιών, ποσοστό ικανοποίησης
Marie, Gandon & Ribière (2013)	Σηματολογική Διερευνητική Αναζήτηση	Αξιοποίηση Διασυνδεδεμένων Ανοικτών Δεδομένων με σηματολογική ενεργοποίηση εξάπλωσης επιτρέποντας επεξεργασία των Διασυνδεδεμένων δεδομένων σε πραγματικό χρόνο	Αξιολόγηση συνάφειας των αποτελεσμάτων και βαθμού ανακάλυψης σε σύγκριση με σύστημα αναφοράς
Golonchinsky et al. 2008	Συνεργατική Διερευνητική Αναζήτηση	Αναδιάταξη των αποτελεσμάτων αναζήτησης και των προτεινόμενων όρων ερωτήματος με βάση τις ενέργειες των ομάδων και άμεση ανατροφοδότηση συνάφειας	Χωρίς αξιολόγηση
Yue et al., 2014	Συνεργατική Διερευνητική Αναζήτηση	Ανάλυση μοτίβων διερεύνησης με βάση τα Κρυφά Μαρκοβιανά Μοντέλα	Χωρίς αξιολόγηση αναφορικά με τη διαδικασία ή το αποτέλεσμα αναζήτησης
Herceg et al. (2018)	Συνεργατική Διερευνητική Αναζήτηση	Αθροιστικά αποτελέσματα αναζήτησης ως διαχωριστικό στρώμα που ασύγχρονα και σειριακά συνδέουν συνεργατική διερευνητική αναζήτηση	Αξιολόγηση της βελτιστοποίησης επιχειρηματικών διαδικασιών και προσαρμογής σε μεγάλα φορτία πληροφορίας

		και κατευθυνόμενη περιήγηση	
Andolina et al. (2018)	Συνεργατική Διερευνητική Αναζήτηση	Διεπαφή αναζήτησης που βασίζεται σε οντολογίες	Αξιολόγηση διευκόλυνσης συνεργασίας και δέσμευσης χρήστη
Kangasrääsiö et al. 2015	Συνεργατική Διερευνητική Αναζήτηση	Ενίσχυση μάθησης δείχνοντας εκτίμηση των αποτελεσμάτων των ενεργειών του χρήστη επιλύοντας πρόβλημα βελτιστοποίησης μοντέλου του χρήστη	Αξιολόγηση απόδοσης της εργασίας, ευχρηστίας, αντιληπτής χρησιμότητας
Singh et al., 2016	Διερευνητική Αναζήτηση σε σχεσιακές βάσεις δεδομένων	Περίληψη δεδομένων αξιοποιώντας το πλαίσιο και μετρικές ομοιότητας	Αξιολόγηση απόδοσης σε ανατεθειμένη εργασία αναφορικά με ποιότητα και ταχύτητα.
Ma & Zhang, 2018	Διερευνητική Αναζήτηση	Προτάσεις ερωτημάτων βάσει ενός Γράφου προσομοίωσης στόχου αναζήτησης	Μέτρηση ακρίβειας των αποτελεσμάτων αναζήτησης, ακρίβειας και καινοτομίας των προτάσεων. Αξιολόγηση εμπειρίας χρηστών (μέτρηση ποσοστού χρησιμοποίησης & χρόνος αναζήτησης)
Chang et al., 2019	Διερευνητική Αναζήτηση για εστιατόρια	Παραδοσιακοί αλγόριθμοι ανάκτησης πληροφοριών για την κατάταξη των αποτελεσμάτων αναζήτησης και την πρόταση όρων που προέρχονται από ένα σημασιολογικό μοντέλο	Συνολική χρησιμότητα

2.3.6 Μετρικές Αξιολόγησης

Παραδοσιακά, η μονάδα αξιολόγησης της ανάκτησης είναι το ερώτημα αναζήτησης. Καθώς όμως καλούμαστε να αξιολογήσουμε συστήματα αναζήτησης επόμενης γενιάς, τα οποία δε στοχεύουν απλώς στη μεγιστοποίηση της συνάφειας ή της ανάκλησης των αποτελεσμάτων, αλλά παρέχουν πιο εκλεπτυσμένη λειτουργικότητα δίνοντας έμφαση στην ολοκλήρωση εργασιών αναζήτησης από την αρχή ως το τέλος - υποστηρίζοντας έτσι μεγαλύτερο εύρος εργασιών -, χρειάζονται νέες μετρικές. Ο Engelbart (1962,ρ.1) ισχυρίστηκε ότι η αυξημένη δυνατότητα των συστημάτων να προσφέρουν στην ανθρώπινη διάνοια λογικά θα οδηγήσει σε αυξημένη κατανόηση θεμάτων που προηγουμένως φάνταζαν σύνθετα, καθώς και σε καλύτερες ποιοτικά λύσεις.

Σε αυτό το κεφάλαιο θα περιγράψουμε μετρικές που σχετίζονται τόσο α) με τη διαδικασία της αναζήτησης, όσο και β) με την επίτευξη αποτελεσμάτων μετά την αναζήτηση. Οι παρακάτω είναι μετρικές που αξιοποιούνται στη Διερευνητική Αναζήτηση, γι' αυτό και εντάσσονται σε υποενότητα του συγκεκριμένου κεφαλαίου.

2.3.6.1 Διαδικασία

2.3.6.1.1 Μάθηση

Η μάθηση περιλαμβάνει την απόκτηση νέας ή μετατροπή υπάρχουσας γνώσης, συμπεριφοράς, ικανοτήτων ή και προτιμήσεων. Οι Pirolli et al. (1996) όρισαν τη μάθηση με βάση τη δομή του θέματος και την ικανότητα σχηματισμού ερωτημάτων από το χρήστη. Άλλοι ερευνητές (Kim et al., 2009; Kammerer et al, 2009) όρισαν τη μάθηση με βάση την πληρότητα της οπτικής του αναζητητή μετά τη διερεύνηση ενός συγκεκριμένου θέματος, συγκρίνοντάς τη με την οπτική πριν την αναζήτηση, με τη χρήση συνεντεύξεων ή εξειδικευμένων αναθέσεων.

Κάποιοι άλλοι όρισαν τη μάθηση σα συνάρτηση του χρόνου ή της προσπάθειας που αναλώθηκε. Οι Pirolli και Rao (1996) έδειξαν ότι οι συμμετέχοντες κατανοήσαν

τις μεταβλητές και τις σχέσεις μεταξύ τους γρηγορότερα με τη χρήση ενός εργαλείου οπτικοποίησης πληροφορίας από ό,τι χωρίς αυτό. Συνεπώς προτείνουν τη μέτρηση της απόδοσης και της μάθησης που αποκτήθηκε μετρώντας χωριστά την απόδοση στην εκπλήρωση της εργασίας και το χρόνο που δαπανήθηκε. Η μέτρηση της μάθησης βάσει του χρόνου καθιστά ευκολότερη τη σύγκριση συστημάτων. Ωστόσο χρειάζεται προσοχή, καθώς η προσκόλληση σε αυτή τη μετρική περισσότερο αξιολογεί την αποδοτικότητα, καθώς οι μεγαλύτεροι χρόνοι που ξοδεύονται για μία εργασία μπορεί κάλλιστα να υποδεικνύουν υψηλότερο βαθμό εμπλοκής και δέσμευσης στην εργασία και στο σύστημα, ένδειξη καλής απόδοσης του συστήματος (White, 2016 p. 311).

2.3.6.1.2 Αποδοτικότητα (Efficiency)

Η αποδοτικότητα μετρά το ποσό της χρήσιμης προσπάθειας στη διάρκεια της διαδικασίας αναζήτησης. Πολλές έρευνες χρησιμοποιούν τη συγκεκριμένη μετρική για να αξιολογήσουν συστήματα αναζήτησης (Zamir and Etzioni, 1999; Goldberg et al. 2002; Gwizdka, 2008). Η κύρια μετρική της αποδοτικότητας είναι ο χρόνος που δαπανήθηκε μέχρι κάποιος να φτάσει την ολοκλήρωση της εργασίας του. Ωστόσο, ίσως η συγκεκριμένη μετρική να είναι πιο κατάλληλη για εργασίες όπου ο στόχος είναι συγκεκριμένος και γνωστός, παρά για εργασίες με διερευνητικό χαρακτήρα.

2.3.6.1.3 Καινοτομία και Ποικιλομορφία

Υπάρχουν πτυχές πέρα από τη συνάφεια, οι οποίες πρέπει να λαμβάνονται υπόψιν όταν αξιολογούνται τα συστήματα ανάκτησης πληροφορίας. Δύο από αυτές είναι η καινοτομία ή νεωτερισμός και η ποικιλομορφία. Οι δύο έννοιες μπορούν να αξιοποιηθούν επίσης ως μετρικές της επίδρασης ή επιτυχίας των συστημάτων ανάκτησης.

Περαιτέρω, στο άρθρο των Clarke et al. (Clarke et al. 2008) η καινοτομία προσεγγίζεται ως ένας τρόπος επίλυσης της απόλυσης. Ιδιαίτερα, στο πεδίο της αναζήτησης ιστού, η καινοτομία αναφέρεται σε αποτελέσματα που ο χρήστης δεν

έχει δει στο παρελθόν. Παρόλα αυτά, η απλοποίηση αυτή κινδυνεύει να υποθέσει ότι η αναζήτηση καινοτομίας των χρηστών είναι η αναζήτηση διαφορετικότητας (Xu and Yin 2007). Τα νέα έγγραφα είναι αυτά που εξειδικεύουν ορισμένο υποτομέα και έτσι ένα σύστημα ανάκτησης πληροφοριών θα πρέπει να παρέχει νέα έγγραφα που να ανταποκρίνονται στην ανάγκη καινοτομίας του χρήστη.

Εάν οι μετρικές αυτές ιδωθούν από την πλευρά της μηχανής αναζήτησης τότε θα μπορούσε κανείς να ισχυριστεί ότι η ποικιλομορφία και η καινοτομία μπορεί να είναι επικαλυπτόμενες έννοιες. Οι (Zhang et al. 2002), (Clarke et al. 2008) και (Xu and Yin 2007) προσεγγίζουν την *καινοτομία* σα μία μέθοδο να αποφύγει κανείς τον πλεονασμό, μια έννοια παρόμοια με την ποικιλομορφία. Από την άλλη, η *ποικιλομορφία* στα συστήματα ανάκτησης πληροφοριών και συστήματα προτάσεων, αναφέρεται στη διαφοροποίηση των αποτελεσμάτων αναζήτησης και των προτάσεων αντίστοιχα. Υπάρχουν δύο γενικές προσεγγίσεις διαφοροποίησης (Minack 2009; Denecke 2012). Στην πρώτη προσέγγιση, η διαφοροποίηση στοχεύει στο *μη-πλεονασμό* των αποτελεσμάτων (Carbonell and Goldstein 1998; Clarke et al. 2008). Για παράδειγμα, οι (Carbonell and Goldstein 1998), προτείνουν τη μέθοδο μέγιστης οριακής συνάφειας (MMR), η οποία επιλέγει ένα έγγραφο που έχει τον υψηλότερο συνδυασμό ομοιότητας σε σχέση με ένα ερώτημα και ένα βαθμό ανομοιότητας σε σχέση με τα έγγραφα που επιλέχθηκαν σε προηγούμενες βαθμίδες. Στη δεύτερη προσέγγιση, η διαφοροποίηση στοχεύει στην *επίτευξη κάλυψης* (*coverage*) και επίλυσης του ζητήματος της ασάφειας των ερωτημάτων (Agrawal et al. 2009; Gollapudi et al. 2010; Santos et al 2010; Skoutas et al. 2010; Capannini et al. 2011; van Leuken et al. 2009).

2.3.6.1.4 Εύνοια Τυχαίων Ανακαλύψεων

Το να μετρήσει κανείς την εύνοια τυχαίων ανακαλύψεων είναι πολύ σημαντικό στα συστήματα που υποστηρίζουν διερεύνηση και ανακάλυψη. Και *ως μετρική σχετίζεται με τη διαδικασία αναζήτησης και όχι τόσο με το αποτέλεσμα αυτής*. Υπάρχουν δύο πτυχές: α) η ευχαρίστηση και η τυχαία φύση της ανακάλυψης και β)

η χρήση της ανακάλυψης για παραγωγικό σκοπό. Στο πλαίσιο της αναζήτησης, η εστίαση παραδοσιακά βρίσκεται στη φύση της πληροφορίας που ευρίσκεται (σχετική ή μη-σχετική) και στον τύπο της πληροφοριακής δραστηριότητας, για παράδειγμα καθοδηγούμενη περιήγηση ή μη (Andre et al. 2009β). Στο πλαίσιο της αναζήτησης η εύνοια των τυχαίων ανακαλύψεων σχετίζεται με αντικείμενο πληροφορίας που είναι απρόσμενα σχετικό, ενδιαφέρον ή χρήσιμο ενώ δεν έψαχνες για αυτό. Αυτό μπορεί να συμβεί στην αναζήτηση ή ακόμα πιο πιθανό όταν προτείνεται περιεχόμενο ή κινήσεις στους αναζητητές καθώς εξετάζουν άλλο περιεχόμενο. Η εύνοια των τυχαίων ανακαλύψεων είναι παρόμοια με το νεωτερισμό, αλλά δίνει έμφαση στην προσωπική ευχαρίστηση και την έκπληξη μαζί με την εύρεση της πληροφορίας. Στην έρευνά του ο André (2009β) ανέδειξε κάποια χαρακτηριστικά που εντοπίζονται σε ερωτήματα που θα μπορούσαν να προκαλέσουν την εύνοια τυχαίων ανακαλύψεων. Μελέτησε ερωτήματα πληροφοριακά, περιηγητικά, σχετιζόμενα με εργασία ή όχι, αν περιείχαν κάποιο όνομα ή όχι. Η εντροπία των clicks, εντοπίζοντας διαφοροποιήσεις στα αποτελέσματα, βρέθηκε να είναι θετικά συσχετισμένη με τον αριθμό των πιθανώς ευχάριστων και αναπάντεχων αποτελεσμάτων.

Από τον τομέα των συστημάτων προτάσεων οι οι Ge, Delgado-Battenfeld και Jannach (2010) πρότειναν για την αξιολόγησή της εύνοιας τυχαίων ανακαλύψεων σε συστήματα προτάσεων τη μέση αναλογία των μη αναμενόμενων (*UNEXP*) και χρήσιμων (*USEFUL*) προτεινόμενων αντικειμένων στη λίστα προτάσεων:

$$SRDP = \sum_u \frac{UNEXP \cap USEFUL}{N}$$

,όπου u ο αριθμός των χρηστών που αξιολογούν τις προτάσεις και N το μέγεθος της λίστας προτάσεων.

2.3.6.1.5 Δέσμευση

Η δέσμευση ή εμπλοκή του αναζητητή αναφέρεται στις συναισθηματικές, γνωστικές και συμπεριφορικές συνδέσεις που δημιουργούνται από την

αλληλεπίδραση με το σύστημα. Ο βαθμός στον οποίο οι χρήστες εμπλέκονται και βιώνουν θετικά συναισθήματα μπορεί να είναι ένας ισχυρός δείκτης της απόδοσης του συστήματος. Η ποσότητα αλληλεπίδρασης που απαιτείται κατά τη διερεύνηση, ο βαθμός στον οποίο ο χρήστης επικεντρώνεται στην εργασία και το περιεχόμενο με την απάντηση του συστήματος μπορεί να δείξει εάν το σύστημα εκπληρώνει το ρόλο του. Ο αριθμός των γεγονότων που συνδέονται με ενέργεια (φόρμες με κλικ, σημειώσεις, ανατροφοδότηση ή εκδηλώσεις προώθησης κ.λπ.) μπορούν να χρησιμοποιηθούν ως μετρική για τον προσεγγιστικό προσδιορισμό του επιπέδου εμπλοκής και δέσμευσης. Για τη μέτρηση της δέσμευσης, έχουν προταθεί ερωτηματολόγια, συνεντεύξεις ή think-aloud μεθοδολογίες από αρκετούς ερευνητές όπως Sauro & Dumais (2009), O'Brien & Toms (2008; 2013). Συγκεκριμένα οι τελευταίοι O'Brien & Toms (2013) πρότειναν ένα προτυποποιημένο ερωτηματολόγιο, την Κλίμακα Δέσμευσης Χρήστη - User Engagement Scale (UES), το οποίο επιτρέπει τη σύγκριση μεταξύ συστημάτων.

Περισσότερο ποσοτικές και συγκεκριμένα Web analytics μέθοδοι προτάθηκαν για παράδειγμα από τους (Yom-Tov et al. 2012), όπως click-through rate, συνολικός αριθμός ερωτημάτων ανά αναζητητή, φορές που επέστρεψε στο σύστημα ο χρήστης.

2.3.6.2 Αποτέλεσμα

2.3.6.2.1 Μετρικές Συνάφειας

Η συνάφεια είναι κεντρική έννοια στην ανάκτηση πληροφορίας. Τα συστήματα διερευνητικής αναζήτησης, από την άλλη, στοχεύουν να υποστηρίξουν τους χρήστες στο να συναντήσουν επαρκή ποσότητα πληροφοριών και λεπτομερειών, που θα τους βοηθήσουν να επιτύχουν το στόχο τους, αντί να φτάσουν απλά σε ένα συγκεκριμένο σχετικό έγγραφο που θα αποτελεί το καλύτερο ταίριασμα. Λόγω των παραπάνω, οι παραδοσιακές μετρικές αξιολόγησης που εφαρμόζονται στην Ανάκτηση Πληροφορίας (IR), εφαρμόζονται πολύ περιορισμένα στην αξιολόγηση συστημάτων Διερευνητικής Αναζήτησης. Από την άλλη, λόγω του γεγονότος ότι η ανάκτηση πληροφορίας ενυπάρχει και στα συστήματα διερευνητικής αναζήτησης θα

αναφερθούμε σύντομα στις πιο γνωστές μετρικές συνάφειας, που είναι η ακρίβεια (precision) και η ανάκληση (recall) αν και υπάρχουν και άλλες όπως: η Μέση Ακρίβεια (Mean Average Precision), η Μέση Βαθμίδα Αμοιβαιότητας (Mean Reciprocal Rank), το Κανονικοποιημένο Μειωμένο Σωρευτικό Κέρδος (Normalized Discounted Cumulative Gain).

Ακρίβεια: Είναι το τμήμα των ανακτηθέντων εγγράφων που σχετίζεται με την αναζήτηση. Η ακρίβεια λαμβάνει υπόψη όλα τα ανακτηθέντα έγγραφα, αλλά μπορεί επίσης να αξιολογηθεί σε μια δεδομένη βαθμίδα αποκοπής λαμβάνοντας υπόψη μόνο τα κορυφαία αποτελέσματα που επέστρεψε το σύστημα. Για παράδειγμα, σε μία αναζήτηση κειμένου σε συλλογή εγγράφων, η ακρίβεια είναι ο αριθμός των σωστών αποτελεσμάτων διαιρούμενος με το συνολικό αριθμό αποτελεσμάτων μέχρι μία καθορισμένη βαθμίδα.

Ανάκληση: Είναι το τμήμα των εγγράφων που είναι σχετικά με το ερώτημα και έχουν επιτυχώς ανακτηθεί. Στο προηγούμενο παράδειγμα, η ανάκληση ορίζεται ως ο αριθμός των σωστών αποτελεσμάτων που ανακτήθηκαν προς τον αριθμό των αποτελεσμάτων που θα έπρεπε να έχουν ανακτηθεί. Στη δυαδική ταξινόμηση, η ανάκληση αποκαλείται και ευαισθησία, και ορίζεται από την πιθανότητα ένα σχετικό έγγραφο να ανακτηθεί από το αίτημα. Η ανάκτηση είναι σημαντική σε εφαρμογές όπου η απουσία μέρους της πληροφορίας δεν είναι ανεκτή, όπως για παράδειγμα σε ιατρικές ή νομικές εφαρμογές.

2.3.6.2.2 Επιτυχία

Η επιτυχία της αναζήτησης μετρά σε τι βαθμό ένα άτομο ολοκληρώνει την εργασία αναζήτησης. Χωρίζεται στην υποκειμενική και αντικειμενική επιτυχία. Η υποκειμενική επιτυχία προσδιορίζεται από τον ίδιο τον αναζητητή όταν θεωρεί ότι βρήκε τις απαντήσεις ή τη λύση στο πρόβλημά του ακόμα και αν αντικειμενικά η απάντηση είναι λάθος. Ο αντικειμενικός ορισμός προσδιορίζεται από τρίτα μέρη για παράδειγμα ειδήμονες επάνω σε ένα θέμα. Οι Hassan et al. έδειξαν ότι η αλληλουχία των ενεργειών, όπως και ο χρόνος μεταξύ των ενεργειών μπορούν να προβλέψουν

την επιτυχία της εργασίας. Ο προσδιορισμός της επιτυχίας από τη δυνατότητα των συμμετεχόντων να εκτιμήσουν το συνολικό χρόνο που αφιέρωσαν στην αναζήτηση και ολοκλήρωση της εργασίας αναφέρεται σε αρκετές εργασίες (Hassan et al. 2010; Czerwinski et al. 2001, Van Bergen 1968; Weybrew 1984).

2.3.6.2.3 Ικανοποίηση

Η ικανοποίηση αξιολογείται στο τέλος της διαδικασίας αναζήτησης και δε σχετίζεται με την επιτυχία (White 2016, p.329). Ένας παράγοντας που μπορεί να προβλέψει την ικανοποίηση είναι οι επιμηκυμένοι χρόνοι παραμονής σε μία σελίδα μετά το κλικ (Fox et al., 2005). Οι Hassan et al. (2011) συνέλεξαν βαθμολογίες ικανοποίησης και εκπαίδευσαν μοντέλο που μπορούσε να εντοπίσει στιγμιότυπα ικανοποίησης.

Πιο πρόσφατες έρευνες λαμβάνουν υπόψιν τη συμπεριφορά του αναζητητή πριν και μετά τα κλικ, για παράδειγμα οι Kim et al. (2014). Για συστήματα αναζήτησης που υποστηρίζουν ανάκτηση γεγονότων, η ικανοποίηση σχετίζεται στενά με την ιεράρχηση των επιστρεφόμενων αποτελεσμάτων και κατά πόσο σχετίζονται με την αρχική πρόθεση του χρήστη.

2.3.6.2.4 Χρόνος εργασίας

Ο χρόνος που αφιερώνεται σε διάφορες δραστηριότητες, όπως ο συνολικός χρόνος που αφιερώνεται, ο χρόνος που αφιερώνεται στην εξέταση άσχετων εγγράφων και το ποσοστό του χρόνου που ξοδεύεται σε διερεύνηση, παρέχουν μια ένδειξη της εμπλοκής του χρήστη καθώς και της αποτελεσματικότητας της διαδικασίας διερεύνησης.

2.3.6.2.5 Δημιουργικότητα

Η Δημιουργικότητα είναι συχνά ζητούμενο αποτέλεσμα της διαδικασίας αναζήτησης, ειδικά όταν οι αναζητήσεις είναι φύσει πιο διερευνητικές.

Εργαστηριακά, η μέτρηση της Δημιουργικότητας δεν είναι εύκολη, κυρίως όταν κάποιος αποπειραθεί να μετρήσει το πόσο δημιουργικός είναι κάποιος.

Σε συνθήκες αναζήτησης, αυτό που ενδιαφέρει είναι να αξιολογηθεί το παραγόμενο που προκύπτει έπειτα από τη συνεδρία αναζήτησης σε σχέση με αυτό που παράγεται εάν ο χρήστης δεν βοηθηθεί με κάποιο σύστημα-εργαλείο.

Μία πολύ σημαντική πτυχή της δημιουργικής διαδικασίας είναι η παραγωγή νέων ιδεών σε σχέση με προηγούμενη γνώση που να είναι διατυπωμένες με τρόπο κατανοητό (Boden, 1996). Οι *μετρικές δημιουργικότητας* χρησιμοποιούνται ευρέως σε μελέτες που συνδέονται με τη δημιουργία ιδεών και τα εργαλεία υποστήριξης της δημιουργικότητας (Dean, Hender, Rodgers και Santanen, 2006). Για παράδειγμα, οι Plucker και Beghetto (2004) πρότειναν ότι οι δημιουργικές ιδέες έχουν δύο διαστάσεις: την καινοτομία και τη χρησιμότητα. Οι MacCrimmon και Wagner (1994) εξέτασαν την πρωτοτυπία των ιδεών που εκφράστηκαν μέσω της καινοτομίας και του πόσο μη -προφανείς ήταν - μια ιδέα που μέχρι τότε ήταν άγνωστη ακόμη και στους ειδικούς στον τομέα. Οι Dormann και Lindgaard (2004) πρότειναν μια τετραπλή κατηγοριοποίηση των κριτηρίων δημιουργικότητας: α) ευχέρεια (ή ποσότητα), β) το επίπεδο της ποικιλίας στη σύνοδο παραγωγής ιδεών, γ) καινοτομία και δ) η χρησιμότητα των παραγόμενων ιδεών. Ομοίως, οι Shah, Smith και Vargas-Hernandez (2003) πρότειναν τις διαστάσεις: *καινοτομία* (novelty) (πόσο ασυνήθιστη, απροσδόκητη ή πρωτότυπη είναι μια ιδέα σε σύγκριση με άλλες ιδέες), *ποικιλία* (μέτρηση του διερευνηθέντος χώρου λύσεων κατά τη διαδικασία δημιουργίας ιδεών), *ποιότητα* (καθορίζει τη σκοπιμότητα μιας ιδέας, τη συγγένειά της με τις προδιαγραφές και πόσο καλά διατυπώνεται) και *ποσότητα* (αριθμός ιδεών) ως κριτήρια δημιουργικότητας.

Άλλοτε έχει ενδιαφέρον να μετρήσει κανείς αν το εργαλείο αναζήτησης βοήθησε τη Δημιουργική Διαδικασία. Για την αξιολόγηση ενός συστήματος ως προς την υποστήριξη της Δημιουργικής Διαδικασίας οι (Cherry and Latulipe, 2014) πρότειναν ένα γενικευμένο ερωτηματολόγιο, το Δείκτη Υποστήριξης Δημιουργικότητας (CSI), το

οποίο μπορεί κάλλιστα να προσαρμοστεί στην αξιολόγηση συστημάτων αναζήτησης που αξιολογεί την ποσοτικοποιημένη επίδραση στους ακόλουθους παράγοντες δημιουργικής διαδικασίας: α) τα *αποτελέσματα άξιζαν την προσπάθεια* - η προσπάθεια που απαιτείται για τη χρήση του εργαλείου δικαιολογεί τα δημιουργικά αποτελέσματα που δημιουργήθηκαν, β) *εκφραστικότητα* (ο βαθμός εκφραστικότητας και δημιουργικότητας που επιτυγχάνεται με το εργαλείο), γ) *διερεύνηση* (η ευκολία με την οποία μπορεί κάποιος να διερευνήσει πολλαπλές επιλογές, ιδέες, σχεδιασμό και αποτελέσματα χωρίς κουραστική, επαναλαμβανόμενη αλληλεπίδραση), δ) *δέσμευση* (το επίπεδο εμπλοκής του χρήστη και της ροής (flow) που επιτυγχάνεται κατά τη χρήση του εργαλείου), ε) *διαφάνεια εργαλείου* (ο βαθμός στον οποίο η διεπαφή ήταν σχεδόν διαφανής κατά τη χρήση του εργαλείου), στ) *συνεργασία* (η ευκολία με την οποία ο χρήστης μπορούσε να εργάζεται με άλλους χρησιμοποιώντας το εργαλείο).

2.4 Δημιουργικότητα

Η ζήτηση και η αναζήτηση πληροφορίας μπορεί να είναι δημιουργικές διαδικασίες. Η δημιουργικότητα μπορεί να ισχυριστεί κάποιος ότι είναι η εφαρμογή της πληροφορίας που έχει βρεθεί. Τα αποτελέσματα έπειτα από τη ζήτηση πληροφορίας και τη συλλογή της τροφοδοτούν άμεσα τη δημιουργική διαδικασία.

Η δημιουργικότητα έχει συζητηθεί σε μοντέλα ζήτησης πληροφορίας, ειδικά κατά τη διερεύνηση και το σχηματισμό θεμάτων. Η επιθυμία να είναι κανείς δημιουργικός οδηγεί σε αναζήτηση πληροφορίας, αλλά η ενέργεια του να είναι τελικά δημιουργικός είναι παράδειγμα χρήσης της πληροφορίας.

Σε αυτήν την ενότητα δίνουμε τον ορισμό της Δημιουργικότητας και της Δημιουργικής Διαδικασίας, ενώ δίνεται έμφαση στη σχέση της αναζήτησης πληροφορίας με τη Δημιουργικότητα. Τέλος παρατίθενται κάποιες από τις γνωστότερες εργασίες συστημάτων που εμπεριέχουν την αναζήτηση για την υποστήριξη της Δημιουργικότητας.

2.4.1 Ορισμός Δημιουργικότητας

Η «δημιουργικότητα είναι η διαδικασία δημιουργίας νέων ιδεών και εννοιών ή η δημιουργία συνδέσεων μεταξύ ιδεών που παράγουν νέες, οι οποίες προηγουμένως δεν υπήρχαν» (Amabile, 1996). Μεταξύ των σημαντικότερων πτυχών της δημιουργικότητας είναι αναμφισβήτητα η γνωστική, η οποία αναφέρεται σε μηχανισμούς και τεχνικές για δημιουργική σκέψη. Η γνωστική πτυχή της δημιουργικότητας απαιτεί ατομικά χαρακτηριστικά, όπως γενική ή συγκεκριμένη γνώση του πεδίου, αντίληψη, πρωτοτυπία, έλξη για την πολυπλοκότητα (π.χ. συνδυασμός, ανάλυση και εφαρμογή διαφορετικών, ανόμοιων ιδεών ή εννοιών), πνεύμα και ευαισθητοποίηση για τη δημιουργικότητα. Επιπλέον, ένα περιβάλλον που προσφέρει ευελιξία και προσβασιμότητα στους πόρους μπορεί να στηρίξει τη δημιουργική σκέψη. Η δημιουργικότητα έχει μια ενδιαφέρουσα ιστορία στις γνωστικές επιστήμες ξεκινώντας από το 1947 με τη θεωρία του Γκίλφορντ που υποστήριζε την αποκλίνουσα σκέψη κατά της σύγκλισης.

Παρόλο που έχει εξεταστεί από πολλές πλευρές, τελικά η ομοφωνία είναι ότι δεν περιλαμβάνει οπωσδήποτε την παραγωγή καινοτόμων χρήσιμων προϊόντων αλλά μπορεί να εστιάσει στη διαδικασία του να παράγει κανείς κάτι.

Η δημιουργικότητα διερευνήθηκε σε λεπτομέρεια στην *έρευνα της ψυχολογίας* από τους Sternberg, Mayer και Csikszentmihalyi. Ανάμεσα στους ερευνητές που μελέτησαν τη δημιουργικότητα, η Boden (1990) ισχυρίστηκε ότι η δημιουργικότητα δεν είναι προνόμιο μόνο λίγων ατόμων, αντιθέτως ο καθένας μας είναι εν δυνάμει δημιουργικός. Με την εφαρμογή αποδεδειγμένων τεχνικών δημιουργικότητας για την ενίσχυση και την τόνωση των δημιουργικών ικανοτήτων, μπορούν να δημιουργηθούν δημιουργικές ιδέες. Η Boden (1990) κάνει διάκριση ανάμεσα στην «*H - δημιουργικότητα*» (ιστορική) και την «*P - δημιουργικότητα*» (ψυχολογική ή προσωπική). Η «*H - δημιουργικότητα*» περιγράφει μια έννοια που δεν έχει δημιουργηθεί ποτέ πριν, δηλαδή η έννοια είναι θεμελιωδώς πρωτότυπη σε σχέση με το σύνολο της ανθρώπινης ιστορίας. Η «*P - δημιουργικότητα*» περιγράφει μια έννοια

που ποτέ πριν δεν δημιουργήθηκε από ένα συγκεκριμένο άτομο, δηλαδή η έννοια είναι θεμελιωδώς πρωτότυπη σε σχέση με το νου που είχε την ιδέα.

Η Boden περιέγραψε τρεις τεχνικές δημιουργικότητας:

Διερευνητική δημιουργικότητα: η δημιουργία νέων ιδεών με τη διερεύνηση δομημένων εννοιολογικών χώρων (μετά από μια οργανωμένη διαδρομή), με αποτέλεσμα νέες και απροσδόκητες ιδέες. Η Boden παρομοιάζει τη διερευνητική δημιουργικότητα με τη χρήση ενός χάρτη για να εξερευνήσει και να περιηγηθεί σε ένα εννοιολογικό χώρο.

Μεταμορφωτική δημιουργικότητα: δημιουργία που προκύπτει μεταμορφώνοντας μία ή περισσότερες διαστάσεις του χώρου για να καταστεί δυνατή η δημιουργία των προηγουμένως αδύνατων ιδεών. Η Boden υποστηρίζει ότι η μετασχηματιστική δημιουργικότητα είναι η διαδικασία δημιουργίας του ίδιου του χάρτη και μπορεί να περιλαμβάνει την αλλαγή του χάρτη (π.χ. Κατάργηση περιορισμών).

Συνδυαστική δημιουργικότητα: δημιουργία που προκύπτει συνδυάζοντας άγνωστες ιδέες για την παραγωγή νέων συνδυασμών γνώριμων ιδεών, π.χ. αναλογική συλλογιστική.

Το έργο της Boden καθοδήγησε την υπολογιστική δημιουργικότητα σε πολύ γενικό επίπεδο. Η Boden (1990) περιγράφει τη δημιουργική πράξη ως την διερεύνηση ενός χώρου μερικών και πλήρων δυνατοτήτων. Η ύπαρξη ενός εννοιολογικού χώρου, που αποτελεί μια αφηρημένη τοποθεσία των «οντοτήτων» που παράγονται από δημιουργικές πράξεις (Ricci, 2006), υποδηλώνει ότι υπάρχουν κανόνες που μπορούν να χρησιμοποιηθούν για τον καθορισμό του χώρου. Η αλλαγή αυτών των κανόνων επιτρέπει τη μεταμόρφωση της δημιουργικότητας. Ο Wiggins (2006) υποστηρίζει ότι με την παροχή διακριτών κανόνων που περιορίζουν τον χώρο αναζήτησης και τους κανόνες με τους οποίους θα γίνει η διέλευση, παράγεται ένα θεωρητικό πλαίσιο με το οποίο πρέπει να μοντελοποιηθεί η διερευνητική και μεταμορφωτική

δημιουργικότητα όπως παρουσιάζεται από την Boden (1990). Αυτό το πλαίσιο δημιουργικών συστημάτων δίνει έμφαση στην έννοια της αναζήτησης ως κεντρικού μηχανισμού προσομοίωσης της δημιουργικότητας. Πιο πρόσφατα, ο Ricci (2012) έχτισε πάνω στο έργο του Wiggin's και τελειοποίησε τους τυπικούς μηχανισμούς για υπολογιστικές δημιουργικές διεργασίες.

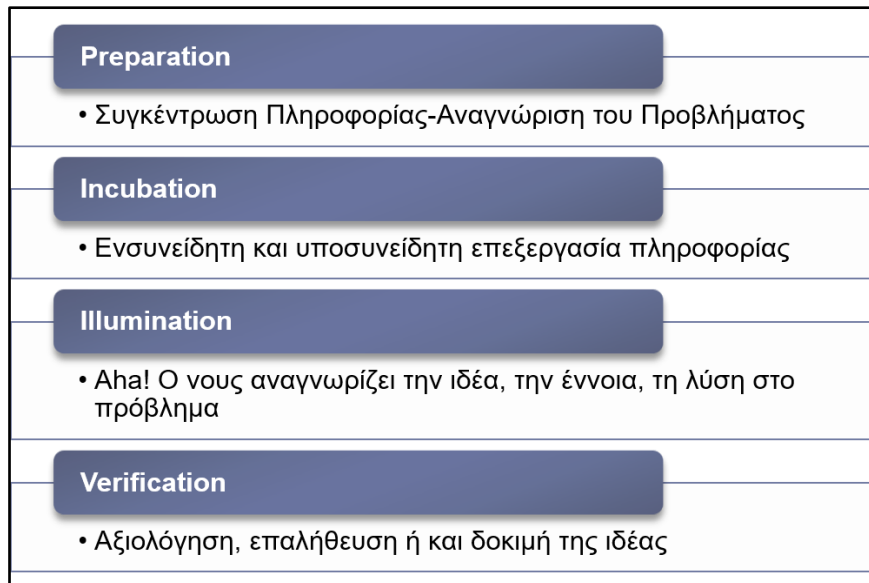
Στη μέτρηση της δημιουργικότητας ενδιαφέρουσες διαστάσεις είναι η συχνότητα με την οποία συμβαίνει, το πόσο σχετική είναι με τον τομέα ενδιαφέροντος, εάν είναι αντικειμενικά μετρήσιμη ή ποιοτικά μετρήσιμη.

2.4.2 Δημιουργική Διαδικασία

Πολλά από τα πρώτα μοντέλα δημιουργικότητας επικεντρώθηκαν στην ανακάλυψη και περιγραφή της φύσης των δημιουργικών ανθρώπων και ασχολήθηκαν με τη δημιουργική διαδικασία από την πλευρά του ατόμου. Στη βιβλιογραφία διακρίνονται τρεις κατηγορίες μοντέλων:

Εμπνευστές

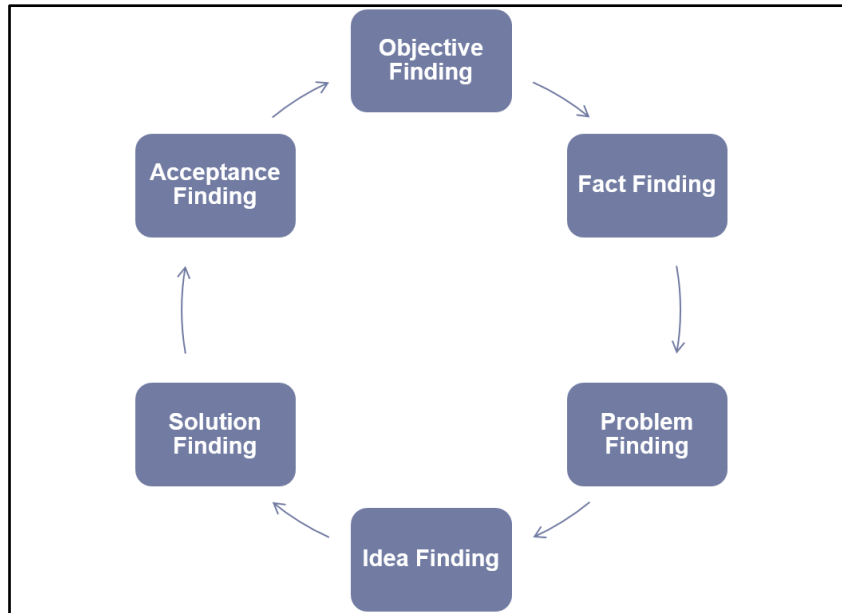
Με τον όρο εμπνευστές ορίζονται τα άτομα τα οποία συλλαμβάνουν ιδέες μέσα από διαλείμματα στιγμών έμπνευσης (Warr and O'Neill, 2005). Το μοντέλο του Wallas (1926), το οποίο θεωρείται ως το κυρίαρχο μοντέλο δημιουργικής διαδικασίας των εμπνευστών, αποτελείται από τέσσερα στάδια (προετοιμασία, επώαση, έκλαμψη και επαλήθευση) και έθεσε τα θεμέλια για άλλα μοντέλα (Aragon and Williams, 2011). Μια κριτική αυτής της μεθόδου είναι ότι δίνει έμφαση στον ρόλο του υποσυνείδητου και επικεντρώνεται στο άτομο, με τον κοινωνικό παράγοντα στην περιφέρεια της δημιουργικότητας (Glaveanu, 2010). Το μοντέλο αυτό προάγει τις τεχνικές δημιουργικότητας για καταιγισμούς ιδεών, ελεύθερη συσχέτιση, πλευρική σκέψη και απόκλιση (Shneiderman, 2000).



Εικόνα 9 Το Μοντέλο Δημιουργικής Διαδικασίας του Wallas

Διαρθρωτικοί

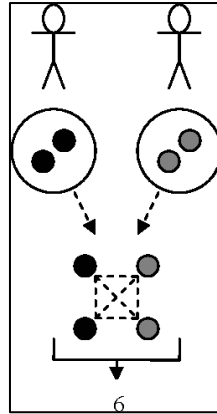
Οι διαρθρωτικοί απομακρύνονται από την υποσυνείδητη διαδικασία που προτείνουν οι εμπνευστές και εφαρμόζουν συστηματικότερες προσεγγίσεις στην διερεύνηση και τη μετατροπή εννοιολογικών χώρων (Warr and O'Neill, 2005). Η διαρθρωτική προσέγγιση της επίλυσης προβλημάτων περιλαμβάνει την "ορθολογική, συστηματική και δομημένη αναζήτηση πληροφοριών" καθώς και την αξιολόγηση και επιλογή εναλλακτικών λύσεων. Ουσιαστικά, ο πυρήνας των διαρθρωτικών δημιουργικών διεργασιών έγκειται στη σκόπιμη δημιουργία και αξιολόγηση ιδεών (Nguyen and Shanks, 2009). Για το λόγο αυτό, υπάρχει μια διαρθρωμένη διαδικασία αποκλίνουσας και συγκλίνουσας σκέψης σε διάφορα μοντέλα διαρθρωτικής δημιουργικής διαδικασίας, π.χ. δημιουργική επίλυση προβλημάτων (CPS) (Parnes, 1967; Osborn 1992). Μια κριτική του μοντέλου διαρθρωτικής δημιουργικής διαδικασίας είναι ότι είναι γραμμική, αλλά στην πραγματικότητα η δημιουργική διαδικασία είναι λιγότερο προβλέψιμη (Glaveanu, 2010).



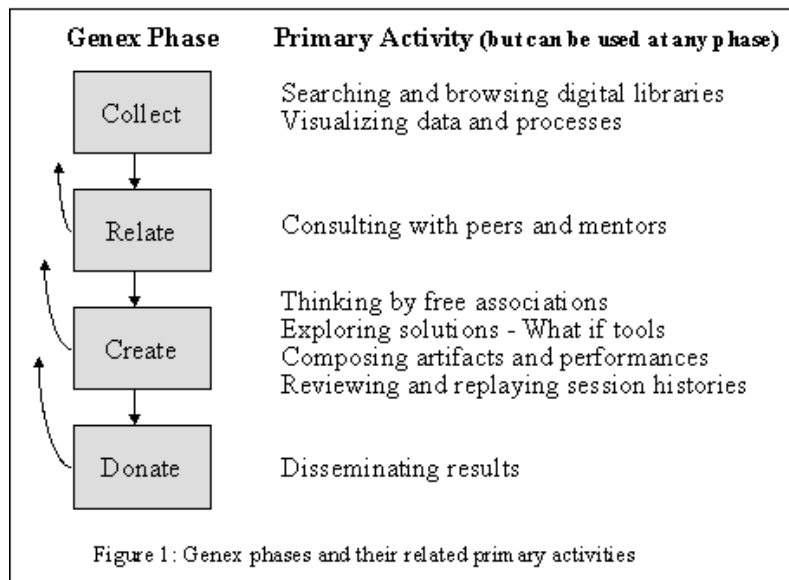
Εικόνα 10 Creative Problem Solving Model (CPS)

Περιπτωσιολόγοι

Οι περιπτωσιολόγοι (situationalists), από την άλλη πλευρά, θεωρούν τη δημιουργικότητα περισσότερο μια κοινωνική διαδικασία, υπογραμμίζοντας τη σημασία της αλληλεπίδρασης και της συνεργασίας με άλλα άτομα (Ariola et al., 2012). Όπως και με τη θεωρητική προοπτική της κατανεμημένης γνώσης (Hollan et al., 2000), οι περιπτωσιολόγοι απομακρύνονται από την άποψη ότι οι δημιουργικές ιδέες δημιουργούνται στο μυαλό ενός ατόμου και θεωρούν ότι η δημιουργική διαδικασία περιλαμβάνει χαρακτηριστικά από το κοινωνικό περιβάλλον που υποστηρίζουν τη γνώση (Warr and O'Neill, 2005). Τοποθετώντας την επικοινωνία και τη συνεργασία στον πυρήνα της δημιουργικής διαδικασίας, οι περιπτωσιολόγοι υπογραμμίζουν το κοινωνικό πλαίσιο ως βασικό συστατικό (Nguyen and Shanks, 2009). Από τα γνωστότερα μοντέλα είναι το GENEX (Shneiderman, 1999) (Εικόνα 12).



Εικόνα 11 Δημιουργικές ιδέες που παράγονται από μία ομάδα (Warr & O'Neill, 2005)



Εικόνα 12 Απεικόνιση των φάσεων του μοντέλου Genex (Shneiderman, 1999)

2.5 Ο ρόλος της Ζήτησης Πληροφορίας στη Δημιουργικότητα

Όταν μελετάμε τη δημιουργικότητα, μπορούμε να μελετήσουμε πώς εμφανίζεται η δημιουργικότητα εστιάζοντας στις διεργασίες και τα περιβάλλοντα που ενεργοποιούν δημιουργικά επινοήματα και διευκολύνουν την ανάπτυξή τους και τον τρόπο με τον οποίο αξιολογούμε ένα επινοήμα για να προσδιορίσουμε αν είναι δημιουργικό» (ML Maher).

Η αναζήτηση και η ζήτηση πληροφορίας είναι μέρος της δημιουργικής διαδικασίας. Ένας αρχιτέκτονας που αναζητά ιδέες για ένα νέο έργο μπορεί να αναζητήσει μια βάση δεδομένων αρχιτεκτονικής. Οι συγγραφείς, οι δημοσιογράφοι και οι καλλιτέχνες μπορούν επίσης να αναζητήσουν νέες ιδέες στο διαδίκτυο. Ένας ιστορικός θα διερευνήσει ένα αρχαιολογικό υλικό για ένα ερευνητικό έργο. Ακόμα και οι ερευνητές ή οι σπουδαστές μπορούν να χρησιμοποιήσουν την αναζήτηση καθώς βελτιώνουν και περιορίζουν το ερευνητικό τους θέμα. Οι διαφημιστές αναζητούν εικόνες ως μέρος της δημιουργικής τους διαδικασίας (Garber & Grunes, 1992). Οι μηχανικοί και οι προγραμματιστές λογισμικού αναζητούν δημιουργικές λύσεις και σε τεχνικά προβλήματα.

Οι προγραμματιστές των εργαλείων αναζήτησης έχουν παραδοσιακά επικεντρωθεί σε αναζητήσεις στις οποίες ο στόχος είναι σαφώς προσδιορισμένος, όπως οι αναζητήσεις γνωστών στοιχείων ή γεγονότων. Οι τυπικές μηχανές αναζήτησης Ιστού και οι βάσεις δεδομένων είναι τώρα πολύ αποτελεσματικές στην ικανοποίηση αυτών των αναζητήσεων με μια απλή λίστα κατάταξης των αποτελεσμάτων όπως αναφέρθηκε στην προηγούμενη ενότητα. Ωστόσο, στο πλαίσιο μιας δημιουργικής εργασίας, οι πληροφορίες που απαιτούνται μπορεί να είναι μόνο μερικώς καθορισμένες ή ασαφείς και ο αναζητητής μπορεί να μην είναι εξοικειωμένος με την ορολογία του τομέα ή της συλλογής που αναζητείται.

Στην επιστήμη της πληροφορίας ο Bawden (1986) περιέγραψε τέσσερις τύπους πληροφορίας που μπορεί να είναι χρήσιμη στην υποστήριξη δημιουργικής σκέψης: α) διεπιστημονική, β) περιφερική, γ) υποθετική, δ) εξαιρέσεις και ασυνέπειες.

Ο Kules (2005) έχτισε επάνω στη θεωρία του Bawden και πρότεινε ότι οι δημιουργικές αναζητήσεις μπορούν να ενσωματώσουν τουλάχιστον τέσσερα χαρακτηριστικά: α) το να έχει κανείς γενικούς σκοπούς, όπως το να μάθει για ένα θέμα ή να παραγάγει νέες ιδέες (τα ερωτήματα μπορεί να είναι διφορούμενα ή μερικώς καθορισμένα), β) το να διευρύνονται οι αναζητήσεις μεταξύ τομέων ή συλλογών (διασταυρούμενα πλαίσια), γ) το να είναι κανείς επαναληπτικός και διερευνητικός. Οι ερευνητές διερευνούν διαδραστικά τα αποτελέσματα αναζήτησης με περιήγηση, φιλτράρισμα ή άλλες τεχνικές. Εφαρμόζουν πολλαπλά, διαδοχικά ερωτήματα καθώς εξελίσσονται οι πληροφορίες τους, πιθανώς σε πολλαπλές περιόδους σύνδεσης. Εμφανίζουν τη συμπεριφορά "berry-picking" ή "information foraging" καθώς συλλέγουν χρήσιμα κομμάτια πληροφοριών, ακολουθούν πολλά υποσχόμενα θέματα και εντοπίζουν νέες πηγές πληροφοριών δ) **να ενθαρρύνεται η εύνοια των τυχαίων ανακαλύψεων και η μη γραμμικότητα** (άρα να υπάρχουν διακλαδώσεις στα ίχνη αναζητήσεων).

Η σχετική έννοια της *εύνοιας τυχαίων ανακαλύψεων* έχει συζητηθεί στη σχετική ερευνητική βιβλιογραφία, με έμφαση στην εκμάθηση πληροφορίας για την οποία κάποιος δεν έψαχνε ενεργητικά (Williamson 1998) και στις τυχαίες ανακαλύψεις πληροφορίας που ενεργοποιούν δημιουργικές διαδικασίες.

Η έννοια της εύνοιας των τυχαίων ανακαλύψεων αναφέρθηκε στην ενότητα 2.3.3 καθώς επίσης αναλύεται στο Κεφάλαιο 4 της παρούσας Διατριβής.

2.5.1 Αρχές Σχεδίασης συστημάτων Αναζήτησης για την υποστήριξη της Δημιουργικότητας

Ορισμοί και ταξονομίες της δημιουργικότητας είναι χρήσιμα για τη σχεδίαση συστημάτων αναζήτησης. Η έρευνα στην ανάπτυξη εργαλείων υποστήριξης

δημιουργικότητας επιζητά την προώθηση της καινοτομίας μέσω της πρόσβασης στην πληροφορία. Το αποτέλεσμα αυτών των συστημάτων ποικίλλει από τεχνουργήματα πληροφορίας μέχρι γενικευμένες υποθέσεις και νέες ανακαλύψεις και συμπεράσματα για τα δεδομένα. Μερικά αποτελέσματα μπορεί να είναι εφήμερα και δύσκολο να μετρηθούν. Τα εργαλεία υποστήριξης δημιουργικότητας χρειάζεται να υποστηρίζουν την εύκολη διερεύνηση εναλλακτικών, όπως επίσης και μηχανισμούς για αποθήκευση κατάστασης, ιστορικότητα βημάτων επιτρέποντας να επιστρέψει ο αναζητητής σε προηγούμενο βήμα στο μονοπάτι αναζήτησης.

Μέσα στα περιβάλλοντα αναζήτησης, υπάρχει ένας αριθμός συστημάτων που υιοθετούν νέα παραδείγματα αλληλεπίδρασης, ώστε να βοηθήσουν τους αναζητητές να γίνουν περισσότερο δημιουργικοί. Τα συστήματα αναζήτησης μπορούν να σχεδιαστούν έτσι ώστε να παρέχουν, για παράδειγμα, επισκοπήσεις των αποτελεσμάτων, οπτικοποιήσεις και υποστήριξη για αποθήκευση και διαχείριση πληροφορίας.

Ο Bill Kules και οι συνεργάτες του το 2005 παρέθεσαν και πιο συγκεκριμένες σχεδιαστικές κατευθύνσεις οι οποίες συνοψίζονται παρακάτω:

- *Ουσιαστικές και σταθερές κατηγορίες:* Οι διαδραστικές επισκοπήσεις των κατηγοριοποιημένων αποτελεσμάτων αναζήτησης Ιστού με χρήση ουσιαστικών και σταθερών διαβαθμίσεων μπορούν να υποστηρίξουν την Διερεύνηση χρηστών, την κατανόηση των μεγάλων συνόλων αποτελεσμάτων και την ανακάλυψη. Οι κατηγορίες μπορούν να εξαχθούν από μεγάλα thesauri, γλωσσάρια ή οντολογίες. Εναλλακτικά, μπορούν να βασίζονται σε απλά συστήματα κατηγοριοποίησης, όπως ο τύπος παραστατικού, οι κωδικοί χώρας ή οι περιοχές μεγέθους του εγγράφου. Υποστηρίζουν τη διαδικασία "Αναζήτηση και περιήγηση" που είναι χαρακτηριστική των ερευνητικών αναζητήσεων, παρέχοντας μια συνεπή οργανωτική δομή για αναζητήσεις λέξεων-κλειδιών.

- *Μεταβλητές κατηγορίες, συμπεριλαμβανομένων συμπλεγμάτων/ομαδοποιημένων επισκοπήσεων:* Οι μεταβλητές κατηγορίες μπορούν να χρησιμοποιηθούν όπως οι σταθερές. Παράγονται όμως δυναμικά από τεχνικές συσταδοποίησης. Μερικές φορές ωστόσο οι χρήστες δυσκολεύονται να κατανοήσουν τις συστάδες ή τις περιγραφικές ετικέτες τους. Χρησιμοποιούνται σε επισκόπηση και αναλυτικές διεπαφές ή οπτικούς χάρτες.

- *Βιβλιογραφική απεικόνιση:* Τα εργαλεία απεικόνισης βιβλιογραφίας παρέχουν επισκοπήσεις ενός τομέα γνώσης ή ενός πεδίου έρευνας, οπτικοποιώντας βιβλιογραφικές ιδιότητες, όπως αναφορές μεταξύ άρθρων ή κοινών θεμάτων. Τα δίκτυα συνεργασίας εντοπίζουν αναφορές μεταξύ εγγράφων από σημαντικούς συγγραφείς σε ένα πεδίο. Μπορούν να απεικονίσουν γραφικά σημαντικά θέματα και υπο-πεδία (Chen, 1999). Αυτοί οι χάρτες μπορεί να βοηθήσουν τους ερευνητές να γεφυρώσουν πολλαπλά πεδία ή να εντοπίσουν τάσεις.

- *Σύνδεση βιβλιογραφίας:* Η βιβλιογραφική σύνδεση είναι μια εξειδικευμένη μορφή δημιουργικών πληροφοριών που επιδιώκουν να ανακαλύψουν νέες συνδέσεις μεταξύ δύο βιβλιογραφιών. Έχει χρησιμοποιηθεί για τον εντοπισμό κρυφών συνδέσεων στην ιατρική βιβλιογραφία, για παράδειγμα μεταξύ ημικρανιών και μαγνησίου με ανάλυση παραπομπών και χειροκίνητη ανασκόπηση των όρων που είναι κοινοί και στις δύο βιβλιογραφίες (Swanson, 1988). Το σύστημα LitLinker ((Pratt & Yetisgen-Yildiz, 2003) είναι ένα παράδειγμα αυτής της τεχνικής. Παρέχει ένα περιβάλλον εργασίας χρήστη που επιτρέπει στους ερευνητές να επιλέγουν την έναρξη και τον στόχο των όρων μέσα σε μια σειρά από βιβλιογραφίες και να εξερευνήσουν διαδραστικά πιθανούς συνδέσμους

- *Μηχανισμοί Τήρησης Ιστορικότητας:* Ο χρήστης μπορεί ανά πάσα στιγμή να επιστρέψει σε προηγούμενο στάδιο αναζήτησης, εάν το μονοπάτι πλοήγησης τον οδηγήσει σε εκφάνσεις του θέματος αναζήτησης που δεν τον ενδιαφέρουν

2.5.2 Επισκόπηση Συστημάτων Αναζήτησης για την Υποστήριξη της Δημιουργικότητας

Στη βιβλιογραφία, αλλά και στη βιομηχανία, συναντάμε μία εξειδικευμένη κατηγορία πληροφοριακών συστημάτων, των οποίων ο ρόλος είναι η υποστήριξη των διαδικασιών καινοτομίας. Τα συστήματα αυτά ονομάζονται Συστήματα Υποστήριξης Δημιουργικότητας. Το εύρος αυτών των συστημάτων είναι εκτεταμένο καθώς τα στάδια της δημιουργικής διαδικασίας είναι περισσότερα του ενός όπως είδαμε και στην προηγούμενη ενότητα (Amabile 1996; Wallas 1926). Αν δούμε την τομή των προτεινόμενων μοντέλων της δημιουργικής διαδικασίας θα οδηγηθούμε σε τέσσερις κατηγορίες συστημάτων: συστήματα για την αναγνώριση του προβλήματος, για την εύρεση πληροφορίας, για την εύρεση ιδεών και για την εύρεση λύσεων με την έννοια των δημιουργικών τεχνουργημάτων. Τα συστήματα που μας απασχολούν στην ενότητα αυτή είναι αυτά τα οποία υποστηρίζουν την εύρεση πληροφορίας μέσω της αναζήτησης.

Οι Kules et al. (2005) κάνοντας μία επισκόπηση εργαλείων που πρότειναν διάφορες οπτικές αναπαραστάσεις των αποτελεσμάτων αναζήτησης διέκριναν τις κατηγορίες που θα μπορούσαν να ενισχύσουν τη δημιουργικότητα. Οι Kerne et al. (2014) πρότειναν μία νέα αναπαράσταση συλλογής εγγράφων συνδυάζοντας κείμενο και στατιστική πληροφορία για να υποστηρίξουν την ανακάλυψη πληροφορίας και τελικά την παραγωγή δημιουργικών ιδεών. Η συγκεκριμένη έρευνα πρότεινε μία ποσοτική μετρική η οποία μετρούσε την ανακάλυψη νέων ιδεών σε συνδυασμό με την εύρεση πληροφορίας. Οι Zachos et al. (2018) πρότειναν ένα plugin που συνοδεύει έγγραφα Google για να παρέχει υποστήριξη σε δημοσιογράφους, προκειμένου να βελτιώσει τη δημιουργική τους ικανότητα. Το εργαλείο εφαρμόζει αποσαφήνιση όρων και επέκτασή ερωτημάτων με βάση τη σημασιολογία, καθώς επίσης παρέχει τα τελευταία νέα από διάφορους ιστότοπους κοινωνικών μέσων. Οι Hugill et al. 2013a πρότειναν ένα εργαλείο το οποίο στοχεύει στη δημιουργία αναπάντεχων και νέων αποτελεσμάτων αναζήτησης παρά σχετικών με τη χρήση Παταφυσικής.

Οι Jenkin,Chan,Skillicorn,Rogers (2013) ανέπτυξαν ένα σύστημα για τον προσδιορισμό νέας πληροφορίας σχετικής με ένα θέμα. Για τον προσδιορισμό της πληροφορίας χρησιμοποίησαν επαναληπτικά τους αρχικούς όρους αναζήτησης μαζί και με νέους που εξήχθησαν από προηγούμενα αποτελέσματα αναζήτησης, προκειμένου να εξαχθεί νέα πληροφορία έμμεσα σχετιζόμενη με το αρχικό θέμα. Μετρήθηκε η καινοτομία των αποτελεσμάτων του εργαλείου καθώς και των αποτελεσμάτων μίας εργασίας θέτοντας βάσεις για τη μέτρηση της εύνοιας τυχαίων ανακαλύψεων ή του μη προφανούς.

Εκτός από μηχανές αναζήτησης ως ερευνητικά πρωτότυπα, οι οποίες έχουν σκοπό την ενίσχυση της δημιουργικότητας, υπάρχουν αρκετές εμπορικές ή ερευνητικές που διατίθενται ανοιχτά στον Ιστό προς χρήση. Κατόπιν εκτεταμένης σχετικής έρευνας συνοψίζονται οι πιο δημοφιλείς στο Παράρτημα 4, παρόλο που αρκετές από αυτές πλέον δε συντηρούνται.

3 Η Πρόταση της Διατριβής

Στο παρόν κεφάλαιο, διατυπώνονται τα ερευνητικά ερωτήματα τα οποία στοχεύει να απαντήσει η συγκεκριμένη Διατριβή. Στη συνέχεια παρουσιάζεται η προτεινόμενη προσέγγιση καθώς και η συνεισφορά της.

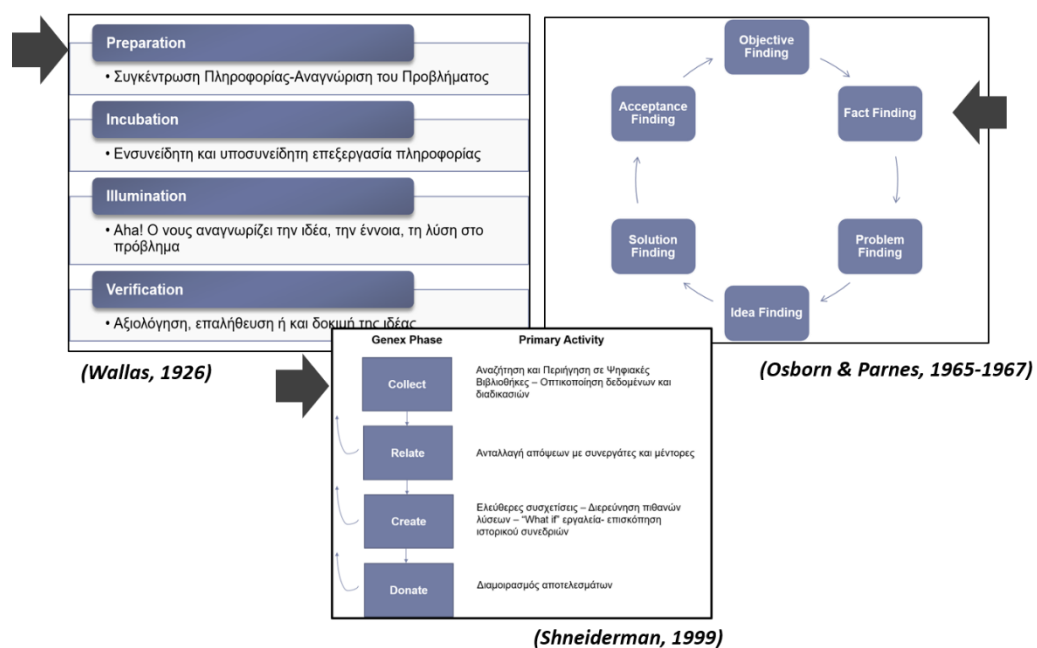
3.1 Διαμόρφωση του Προβλήματος και Ερευνητικές Προκλήσεις

Η αφθονία της πληροφορίας η οποία υπάρχει σε διαδικτυακά πληροφοριακά συστήματα έχει προκαλέσει την ανάγκη να βρεθούν αποτελεσματικοί τρόποι για τη διευκόλυνση της ανακάλυψης χρήσιμης πληροφορίας από τους χρήστες. Η ζήτηση πληροφορίας τυπικά υποθέτει ένα μοντέλο της διαδικασίας αναζήτησης που περιλαμβάνει την αρχική αναγνώριση και τον προσδιορισμό της πληροφοριακής ανάγκης, ακολουθούμενων από την εξέταση των αποτελεσμάτων αναζήτησης και την επανάληψη του κύκλου μέχρι να εντοπιστεί ένα ικανοποιητικό σύνολο αποτελεσμάτων (Schneiderman et al., 1998; Marchionini and White, 2007).

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν τα γνωστότερα μοντέλα ζήτησης πληροφορίας, δίνοντας έμφαση σε εκείνα τα οποία έχουν στο επίκεντρο το χρήστη αναμένοντας ανατροφοδότησή του. Εστίασαμε στο μοντέλο της Διερευνητικής Αναζήτησης, καθώς απαντά στο μεγαλύτερο αριθμό πληροφοριακών αναγκών και ειδικά εκείνων που έχουν ως απώτερο στόχο τη χρήση της πληροφορίας και εν τέλει τη σύνθεση ενός προϊόντος γνώσης.

Η διερευνητική αναζήτηση είναι στενά συνδεδεμένη με εργασίες που απαιτούν έρευνα, μάθηση, δημιουργικότητα και καινοτομία, έχοντας το ρόλο του «μέσου» για την επιτυχή έκβαση της εργασίας. Τέτοιες εργασίες είναι η αναζήτηση ή σύνθεση εμπνευσμένων ιδεών για νέα προϊόντα και υπηρεσίες(εξετάζοντας ερευνητικές εργασίες σε ένα τομέα) ή η εξοικείωση με ένα προηγουμένως άγνωστο τομέα.

Στην παρούσα διατριβή ισχυριζόμαστε ότι η διερευνητική αναζήτηση μπορεί να υποστηρίξει τις δημιουργικές εργασίες. Στόχος μας είναι να προτείνουμε σχετική μέθοδο, την οποία τελικά αξιολογούμε ως προς το αν δύναται να ενισχύσει τη δημιουργικότητα. Συγκεκριμένα, εστιάζοντας στο παράδειγμα της Διερευνητικής Αναζήτησης, προτείνουμε μία μέθοδο διερεύνησης, η οποία έχει ως πρωταρχικό σκοπό την υποστήριξη της εύρεσης και συγκέντρωσης πληροφορίας, κοινό στάδιο της δημιουργικής διαδικασίας σύμφωνα με πολλά από τα γνωστότερα μοντέλα που περιγράφηκαν και στην ενότητα 2.4.2 (Εικόνα 13). Ειδικότερα, υποστηρίζουμε τον αναζητητή στη συγκέντρωση πληροφορίας, προτείνοντας μία διαδικασία προσανατολισμένη στη διερεύνηση και ανακάλυψη, η οποία στόχο έχει να διευκολύνει: α) την πρόσβαση σε ποικίλλου τύπου πληροφορία, η οποία θα είναι τόσο σχετική με το πρόβλημα, όσο και απρόσμενη, ώστε δυνητικά να μπορεί να εμπνεύσει, β) την αλληλεπίδραση με την πληροφορία με σκοπό τη διερεύνηση της συσχέτισης εννοιών, ώστε τελικά με αυτόν τον τρόπο να ο αναζητητής να παραγάγει νέες και περισσότερο δημιουργικές ιδέες.



Εικόνα 13 Στόχος της Διατριβής είναι η υποστήριξη της διερεύνησης, ανακάλυψης και συγκέντρωσης πληροφορίας με την προτεινόμενη προσέγγιση διερεύνησης πληροφορίας για τη σύνθεση δημιουργικών ιδεών.

Όπως προαναφέραμε, το μέσο για την επίτευξη του αποτελέσματος είναι η διερευνητική αναζήτηση, αναζητούνται όμως και οι «ενεργοποιητές», οι παράγοντες εκείνοι οι οποίοι αν εισαχθούν στο σύστημα θα ευοδωθεί ο στόχος.

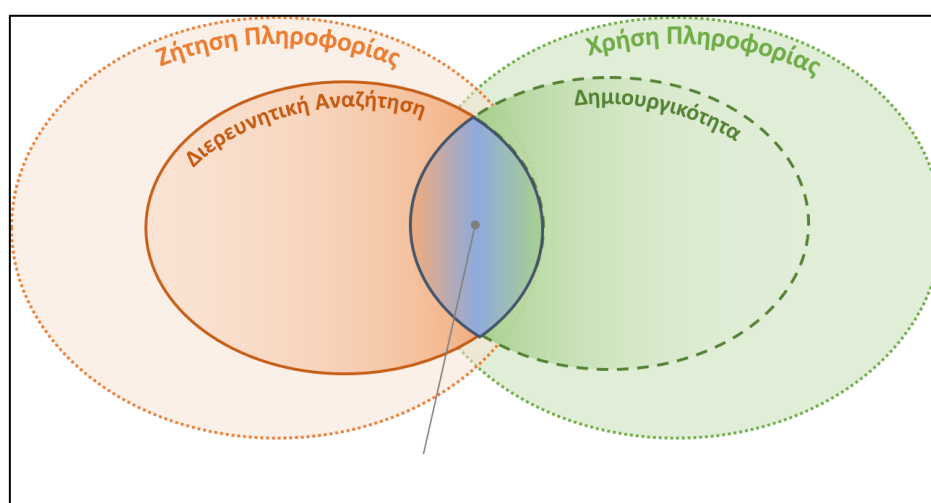
Η επισκόπηση της βιβλιογραφίας ανέδειξε τον παράγοντα της *εύνοιας τυχαίων ανακαλύψεων* ως έναν πρόσθετο παράγοντα για την υποστήριξη της αναζήτησης πληροφορίας (Foster & Ford 2003) και της διερευνητικής αναζήτησης (White et al., 2006) (βλ. ενότητα 2.3.3). Ταυτόχρονα η εύνοια των τυχαίων ανακαλύψεων επιτρέπει την ανάπτυξη δημιουργικών συνδέσεων, καθώς μπορεί να έχει ρόλο στην αποκάλυψη κρυφών συνδέσεων και αναλογιών (Cory, 1999). Ο παράγοντας της εύνοιας τυχαίων ανακαλύψεων αναδύεται ως ένας από τους πυλώνες της ερευνητικής μας μελέτης.

Επιπλέον, στην ενότητα 2.3.2 στις γενικές κατευθύνσεις σχεδίασης αποτελεσματικών συστημάτων διερευνητικής αναζήτησης, αναφέρεται ο παράγοντας της *επίγνωσης πλαισίου*. Τα συστήματα πρέπει να αξιοποιούν τις διαθέσιμες πληροφορίες σχετικά με το χρήστη, την κατάστασή τους και την τρέχουσα εργασία ερευνητικής αναζήτησης (White & Roth, 2009). Εκτός από την ανάπτυξη μοντέλων ενδιαφερόντων των χρηστών, μέσω της εκμετάλλευσης του πλαισίου αλληλεπίδρασης, είναι πολύτιμο για τα συστήματα να γνωρίζουν πότε οι χρήστες αντιμετωπίζουν δυσκολία κατά τις αναζητήσεις τους (Horvitz et al., 1998). Αυτό είναι αναμφισβήτητα μια ιδανική στιγμή για παρέμβαση του συστήματος με προτάσεις σχετικά με ερωτήματα προς εκτέλεση, σελίδες προς επίσκεψη ή με προτάσεις για ενέργειες που βοηθούν τον χρήστη στην επίδιωξη του στόχου τους (White & Roth, 2009). Ο παράγοντας του πλαισίου αλληλεπίδρασης αναδύεται ως δεύτερος πυλώνας της ερευνητικής μας μελέτης.

Η ανθρώπινη συμπεριφορά αναφέρεται στο εύρος των δραστηριοτήτων που επιδεικνύουν οι άνθρωποι και οι οποίες επηρεάζονται, ανάμεσα σε άλλους παράγοντες, από την *προσωπικότητά* τους. Οι μελέτες έχουν συσχετίσει την

προσωπικότητα με προτιμήσεις στη διασκέδαση (ταινίες, μουσική) (Rentfrow, Goldberg & Zilca, 2011), ακαδημαϊκά επιτεύγματα (Komarraju, Karau, Schmeck & Avdic, 2011), πρότυπα λήψης αποφάσεων και επαγγελματικά επιτεύγματα (Kanuka & Nocente, 2003) καθώς και με τη χρήση ιστοτόπων κοινωνικής δικτύωσης (Wu, Chen & He, 2013) και τη συμπεριφορά ζήτησης πληροφορίας (Heinström 2003; Nie, Guan & Zhu, 2014; Marchionini, 2009). Η σχέση μεταξύ προσωπικότητας και γνωσιακών στρατηγικών επεξεργασίας έχει μελετηθεί και τα χαρακτηριστικά προσωπικότητας έχουν χρησιμοποιηθεί ως μέσο πρόβλεψης της τάσης ενός ατόμου να επεξεργάζεται πληροφορίες (Cantor & Kihlstorm, 1981). Γι'αυτό, η προσωπικότητα είναι ένα χαρακτηριστικό του χρήστη που χρειάζεται να λαμβάνεται υπόψιν ως ένας παράγοντας ο οποίος επηρεάζει τη συμπεριφορά αναζήτησης και τις τακτικές διερεύνησης πληροφορίας καθώς μπορεί να αντικατοπτρίζεται στις προτιμήσεις του αναζητητή (Marchionini, 2006). Ο παράγοντας της προσωπικότητας αναδύεται ως τρίτος πυλώνας της ερευνητικής μας μελέτης.

Με βάση τα παραπάνω, η προτεινόμενη μεθοδολογία ενσωματώνει παράγοντες που υποστηρίζουν τις τυχαίες ανακαλύψεις, αξιοποιεί το πλαίσιο αλληλεπίδρασης του χρήστη με το σύστημα Διερευνητικής Αναζήτησης, καθώς και την προσωπικότητά του. Εξετάζουμε τους προαναφερθέντες παράγοντες ως προς την επίδρασή τους στη διαδικασία Διερευνητικής Αναζήτησης και στη Δημιουργικότητα.



Εικόνα 14 Τομείς που στοχεύει να υποστηρίξει η Πρόταση της Διατριβής

3.2 Ερευνητικά Ερωτήματα

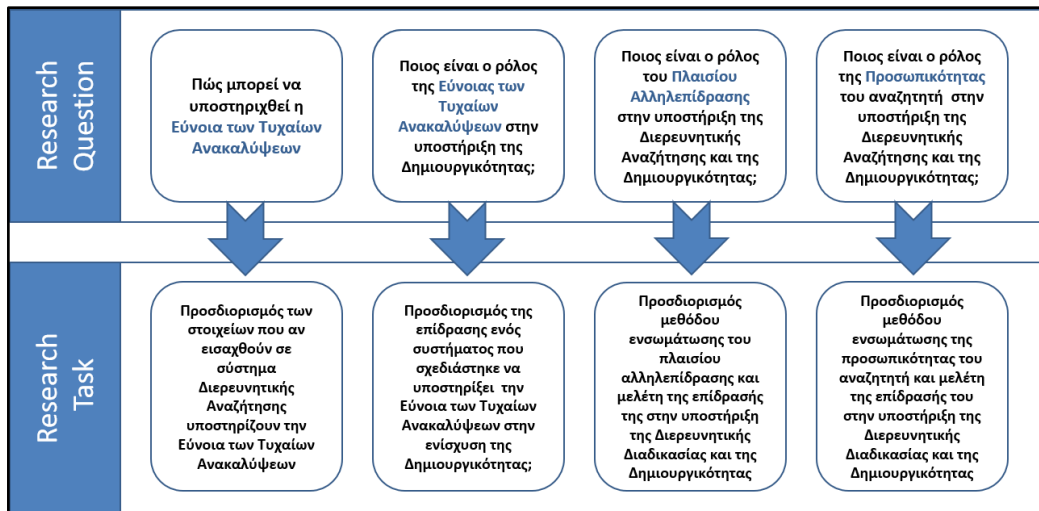
Η παρούσα διατριβή στοχεύει στην υποστήριξη των χρηστών στη διαδικασία Διερεύνησης, η οποία αποτελεί ένα από τα στάδια της Δημιουργικής Διαδικασίας, με απώτερο σκοπό την υποστήριξη της ίδιας της Δημιουργικής Διαδικασίας αλλά και της ενίσχυσης του Δημιουργικού Αποτελέσματος. Για το σκοπό αυτό, η μεθοδολογία που προτείνεται απευθύνεται τόσο στο σχεδιασμό μίας διαδραστικής διεπαφής αναζήτησης που συνδυάζει τεχνολογίες Αναζήτησης (Ανάκτηση Πληροφορίας) και Συστημάτων Προτάσεων (Διήθηση Πληροφορίας), όσο και στην εισαγωγή αλγορίθμων που επηρεάζουν τα επιστρεφόμενα αποτελέσματα των δύο τεχνολογιών και την απεικόνισή τους, έχοντας ως οδηγούς: α) το πώς μπορεί να υποστηριχθεί η ανακάλυψη χρήσιμου και αναπάντεχου περιεχομένου, β) πώς μπορεί να αξιοποιηθεί το πλαίσιο αλληλεπίδρασης του αναζητητή με το σύστημα για τη δημιουργία προτάσεων στη διάρκεια της διερευνητικής διαδικασίας και γ) πώς μπορούν να αξιοποιηθούν τα χαρακτηριστικά της προσωπικότητας του αναζητητή για τη δημιουργία προτάσεων στη διάρκεια της διερευνητικής διαδικασίας.

- **Άξονας 1 - Δημιουργικότητα:** Πώς μπορεί να υποστηριχθεί η δημιουργική διαδικασία και να ενισχυθεί το δημιουργικό αποτέλεσμα εργασιών που απαιτούν διερευνητική αναζήτηση;
- **Άξονας 2 – Διαδικασία Διερεύνησης (Διερευνητική Αναζήτηση):** Πώς μπορούν να υποστηριχθούν οι χρήστες στο να χρησιμοποιήσουν κατάλληλα και αποτελεσματικά ένα σύστημα Διερευνητικής Αναζήτησης και να ενισχυθεί παράλληλα η αίσθηση συμμετοχής και εμπλοκής τους στη διερευνητική διαδικασία.

Ο πίνακας 3 συνοψίζει τα κεντρικά αλλά και επιμέρους ερευνητικά ερωτήματα.

Πίνακας 3 Ερευνητικά Ερωτήματα

<p>Πώς μπορεί να υποστηριχθεί η εύνοια τυχαίων ανακαλύψεων σε σύστημα διερευνητικής αναζήτησης;</p>	<ul style="list-style-type: none"> • Ποιοι είναι οι παράγοντες που αν εισαχθούν σε σύστημα Διερευνητικής Αναζήτησης υποστηρίζονται οι τυχαίες ανακαλύψεις;
<p>Ποιος είναι ο ρόλος της εύνοιας τυχαίων ανακαλύψεων στην υποστήριξη της Δημιουργικότητας;</p>	<ul style="list-style-type: none"> • Μπορεί ένα σύστημα που σχεδιάστηκε να υποστηρίζει τις τυχαίες ανακαλύψεις να ενισχύσει τη δημιουργικότητα;
<p>Ποιος είναι ο ρόλος της υιοθέτησης του πλαισίου αλληλεπίδρασης στην υποστήριξη της Διερευνητικής Αναζήτησης και της Δημιουργικότητας;</p>	<ul style="list-style-type: none"> • Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης να ενσωματώσει το πλαίσιο αλληλεπίδρασης; • Μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει το πλαίσιο αλληλεπίδρασης να ενισχύσει την αποτελεσματικότητα της διερεύνησης και τη συμμετοχή του χρήστη; • Μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει το πλαίσιο αλληλεπίδρασης να υποστηρίξει τη δημιουργικότητα;
<p>Ποιος είναι ο ρόλος της αξιοποίησης της προσωπικότητας του αναζητητή στην υποστήριξη της Διερευνητικής Αναζήτησης και της Δημιουργικότητας;</p>	<ul style="list-style-type: none"> • Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης να ενσωματώσει τον παράγοντα της προσωπικότητας; • Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει την προσωπικότητα να ενισχύσει την αποτελεσματικότητα (ή απόδοση) της διερεύνησης και τη συμμετοχή του χρήστη; • Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει την προσωπικότητα να υποστηρίξει τη δημιουργικότητα;



Εικόνα 15 Τα ερευνητικά ερωτήματα και οι αντίστοιχες ερευνητικές εργασίες που πραγματοποιούνται στη Διατριβή

3.2.1 1^ο Ερευνητικό Ερώτημα: Πώς μπορεί να υποστηριχθεί η εύνοια τυχαίων ανακαλύψεων σε σύστημα διερευνητικής αναζήτησης;

Στην ενότητα 2 αναλύσαμε τη σημασία της αναζήτησης πληροφορίας σε εργασίες έντασης γνώσης (knowledge intensive tasks) και το κατά πόσο αποτελεί αναπόσπαστο συστατικό της δημιουργικής διαδικασίας. Έγινε αναφορά στην εύνοια των τυχαίων ανακαλύψεων και στο συσχετισμό της με τη Δημιουργικότητα καθώς και με τη Διερευνητική Αναζήτηση. Στη βιβλιογραφία έχουμε συναντήσει αρκετές εμπειρικές μελέτες, οι οποίες θέτουν ένα σχεδιαστικό πλαίσιο για την υποστήριξη της εύνοιας τυχαίων ανακαλύψεων και έχουν προταθεί μέθοδοι με κύριες εφαρμογές σε προγράμματα-πράκτορες που εγκαθίστανται σε περιηγητές ή σε συστήματα που δίνουν τη δυνατότητα περιήγησης (π.χ. Liebermann et al. 1999; Khalili et al. 2018). Ωστόσο, εκλείπει η αξιολόγηση της αποτελεσματικότητας των μεθοδολογιών αυτών σε εργασίες διερευνητικής αναζήτησης είτε σε ελεγχόμενες είτε σε πραγματικού χρόνου πιλοτικές μελέτες. Ταυτόχρονα οι περισσότερες υλοποιήσεις στερούνται ενοποίησης μεθοδολογιών αναζήτησης (ανάκτηση πληροφορίας), περιήγησης και αλγορίθμων προτάσεων (διήθηση πληροφορίας).

Ο στόχος της παρούσας Διατριβής είναι, μετά την επισκόπηση της βιβλιογραφίας σε ψηφιακά Πληροφοριακά Συστήματα, να προκύψει ένα σχεδιαστικό πλαίσιο με στόχο την υποστήριξη της Εύνοιας των Τυχαίων Ανακαλύψεων. Στη συνέχεια, αφού εφαρμοστεί το πλαίσιο σε πληροφοριακό σύστημα, το οποίο υποστηρίζει διερευνητικές αναζητήσεις, επιδιώκουμε να επαληθεύσουμε ότι η εφαρμογή των συγκεκριμένων αρχών και μεθοδολογιών πράγματι υποστηρίζει την εύνοια των τυχαίων ανακαλύψεων

3.2.2 2^ο Ερευνητικό Ερώτημα: Ποιος είναι ο ρόλος της εύνοιας τυχαίων ανακαλύψεων στην υποστήριξη της δημιουργικότητας στο πλαίσιο της διερευνητικής αναζήτησης;

Στο πλαίσιο της Διερευνητικής Αναζήτησης, εξ' όσων γνωρίζουμε, δεν έχει αξιολογηθεί η επίδραση μίας μεθοδολογίας που υποστηρίζει τις τυχαίες ανακαλύψεις στη Δημιουργικότητα των αναζητητών.

Ο στόχος της Διατριβής είναι να αξιολογήσουμε κατά πόσο ένα σύστημα που έχει σχεδιαστεί με στόχο την υποστήριξη τυχαίων ανακαλύψεων μπορεί να υποστηρίξει τη Δημιουργικότητα.

3.2.3 3^ο Ερευνητικό Ερώτημα: Ποιος είναι ο ρόλος της υιοθέτησης του πλαισίου αλληλεπίδρασης στην υποστήριξη της Διερευνητικής Αναζήτησης και της Δημιουργικότητας;

Η επισκόπηση της βιβλιογραφίας αναδεικνύει τον παράγοντα της μοντελοποίησης του πλαισίου (context modelling) ως παράγοντα υψηλής σημασίας για την αποτελεσματικότητα ενός πληροφοριακού συστήματος αναζήτησης πληροφορίας. Σύμφωνα με τα όσα γνωρίζουμε, ενώ έχει δοθεί ιδιαίτερη έμφαση στην αξιοποίηση του χρονικού και φυσικού πλαισίου, καθώς και του πλαισίου αλληλεπίδρασης όσον αφορά το περιεχόμενο -κυρίως για την επαύξηση ερωτημάτων-, δεν υπάρχει προσέγγιση που να αξιολογεί την επίδραση του να ληφθούν υπόψιν οι αλληλουχίες ενεργειών του χρήστη (ή τακτικές αναζήτησης) για

τη βελτίωση της αποτελεσματικότητας της διερευνητικής αναζήτησης και της δημιουργικότητας του προϊόντος που προκύπτει έπειτα από μία συνεδρία αναζήτησης.

Ο στόχος της Διατριβής είναι να προτείνει μία μέθοδο μοντελοποίησης των αλληλουχιών ενεργειών των χρηστών στο πλαίσιο ενός συστήματος διερευνητικής αναζήτησης, με σκοπό την πρόβλεψη των πιθανότερων επόμενων ενεργειών του χρήστη, οι οποίες θα προτείνονται για την αποτελεσματικότερη χρήση του συστήματος, αλλά και την καλύτερη απόδοση στις δημιουργικές εργασίες, λόγω των οποίων εκκινεί η αναζήτηση. Επιδιώκουμε να αξιολογήσουμε την επίδραση των προτάσεων -που παράγονται από την αξιοποίηση του πλαισίου αλληλεπίδρασης- στην αποτελεσματικότητα στην αναζήτηση και στην υποστήριξη της Δημιουργικότητας, συγκρίνοντας ένα σύστημα που υλοποιεί την προτεινόμενη μέθοδο παραγωγής προτάσεων με το ίδιο το σύστημα χωρίς την προτεινόμενη μέθοδο.

3.2.4 4ο Ερευνητικό Ερώτημα: *Ποιος είναι ο ρόλος της αξιοποίησης της προσωπικότητας του αναζητητή στην υποστήριξη της Διερευνητικής Αναζήτησης και της Δημιουργικότητας;*

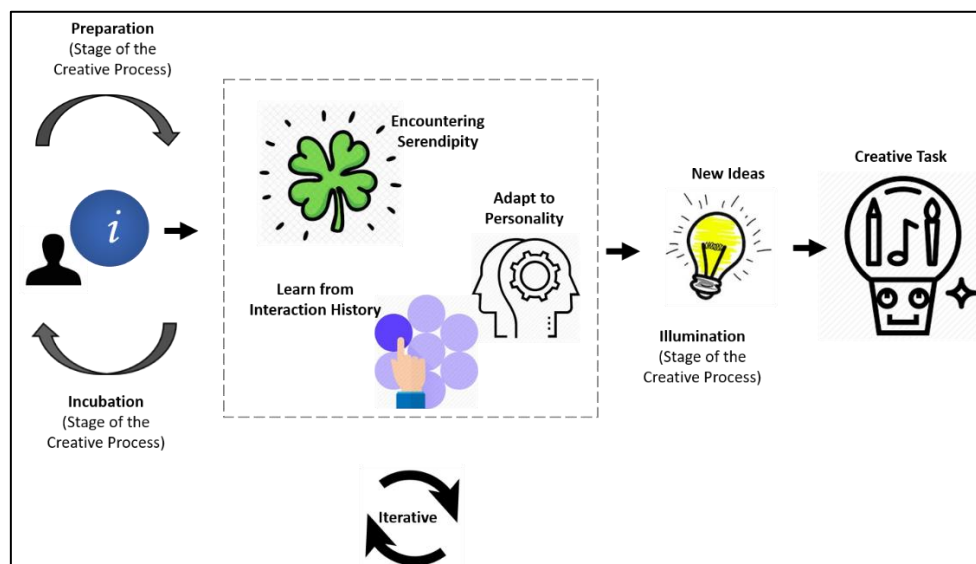
Η επισκόπηση της βιβλιογραφίας ανέδειξε τον παράγοντα της μοντελοποίησης των προτιμήσεων και ενδιαφερόντων του αναζητητή ως καθοριστικό για την αποτελεσματικότητα ενός πληροφοριακού συστήματος που υλοποιεί το μοντέλο της Διερευνητικής Αναζήτησης. Σύμφωνα βέβαια με τα όσα γνωρίζουμε, ενώ υπάρχει πληθώρα προσεγγίσεων που εστιάζουν στη βελτίωση της διαδικασίας αναζήτησης μέσω της σκιαγράφησης των μακροπρόθεσμων προτιμήσεων του χρήστη, απουσιάζουν προσεγγίσεις ενσωμάτωσης της προσωπικότητας στην αναζήτηση και προσαρμογή των δυνατοτήτων του συστήματος στην προσωπικότητα.

Ο στόχος της παρούσας εργασίας είναι να προτείνει μία μέθοδο πρόβλεψης των επόμενων ενεργειών του χρήστη και κατ' επέκταση δημιουργίας προτάσεων στο πλαίσιο της διερευνητικής αναζήτησης με σκοπό την πιθανή χρήση της πληροφορίας

για εργασίες που απαιτούν δημιουργική σκέψη. Η προσέγγιση θα αξιοποιεί αφενός το πλαίσιο αλληλεπίδρασης της κάθε συνεδρίας, και αφετέρου οι προτάσεις θα προσαρμόζονται με βάση την προσωπικότητα του χρήστη. Στη συνέχεια θα αξιολογήσουμε την προτεινόμενη μέθοδο συγκρίνοντας δύο συστήματα: το πρώτο υλοποιεί τη μεθοδολογία παραγωγής προτάσεων του 3^{ου} ερευνητικού ερωτήματος, όπου αξιοποιείται το πλαίσιο αλληλεπίδρασης, χωρίς να γίνεται προσαρμογή στην προσωπικότητα του χρήστη και το δεύτερο υλοποιεί την παρούσα μεθοδολογία προτάσεων βάσει προσωπικότητας. Με αυτόν τον τρόπο θα διαφανεί εάν η όποια πιθανή βελτίωση -είτε στη Διερευνητική αναζήτηση είτε στη Δημιουργικότητα του αποτελέσματος - οφείλεται στον παράγοντα της προσωπικότητας.

3.3 Πρόταση και Συνεισφορά Διατριβής

Όπως προαναφέρθηκε ο σκοπός της Διατριβής είναι να μελετήσει την επίδραση που έχουν συγκεκριμένοι παράγοντες στη Διερευνητική Αναζήτηση και στη χρήση της πληροφορίας με σκοπό να ενισχυθούν εργασίες στις οποίες ενυπάρχει η Δημιουργική σκέψη. Η εικόνα 16 απεικονίζει την εννοιολογική προσέγγιση στο πρόβλημα που



Εικόνα 16 Οι παράγοντες που μελετά η Διατριβή για την υποστήριξη της Δημιουργικότητας και της Διαδικασίας Διερεύνησης

πραγματεύεται η Διατριβή. Στον πίνακα 4 συνοψίζεται η μεθοδολογία που ακολουθήσαμε ανά ερευνητικό ερώτημα και σε ποια ενότητα της Διατριβής μπορεί να βρεθεί η ανάλυσή της.

3.3.1 Ενίσχυση της Εύνοιας Τυχαίων Ανακαλύψεων

Το συγκεκριμένο ερευνητικό ερώτημα αφορά στην ενίσχυση της εύνοιας των τυχαίων ανακαλύψεων σε ένα σύστημα διερευνητικής αναζήτησης.

Η επισκόπηση της σχετικής βιβλιογραφίας μας βοήθησε να εστιάσουμε στις τρεις βασικές περιοχές σχεδίασης ενός Πληροφοριακού Συστήματος Αναζήτησης α) σχεδίαση διαδραστικής διεπαφής χρήστη, β) αποτελέσματα-περιεχόμενο, γ) παρουσίαση αποτελεσμάτων-περιεχομένου, επιλέγοντας για κάθε μία περιοχή την κατάλληλη μεθοδολογία προκειμένου να υποστηριχθεί η εύνοια των τυχαίων ανακαλύψεων.

Στην παρούσα διατριβή προτείνουμε ένα διαδραστικό μοντέλο διερευνητικής αναζήτησης, όπου σε επίπεδο διεπαφής συνδυάζει λειτουργικότητες αναζήτησης, περιήγησης και υπηρεσία προτάσεων. Για την οπτικοποίηση σημαντικών λέξεων και φράσεων, τα οποία εξάγονται κυρίως από κοινωνικά μέσα δικτύωσης, υιοθετείται ένα μοντέλο φολκσονομίας, που υλοποιείται σε μορφή νέφους λέξεων. Σε επίπεδο κατάταξης και παρουσίασης αποτελεσμάτων αναζήτησης προτείνεται α) λίστα κατάταξης και εφαρμογή αλγορίθμου επανακατάταξης με σκοπό την αποκάλυψη πολλαπλών πτυχών του θέματος αναζήτησης στα αποτελέσματα που είναι τοποθετημένα στις υψηλότερες θέσεις, καθώς και β) η ύπαρξη ποικιλομορφίας σε ό,τι αφορά τους τύπους αποτελεσμάτων (δηλαδή, άρθρα, λέξεις ή πολυμέσα). Σε επίπεδο περιεχομένου προτείνεται η εισαγωγή του απροσδόκητου με την παροχή προτάσεων περιεχομένου από Κοινωνικά μέσα Δικτύωσης. Για την επιλογή των Κοινωνικών μέσων Δικτύωσης και συγκεκριμένα του Twitter, βασιστήκαμε στην παρατήρηση ότι οι κοινωνικές προτάσεις μέσω κοινωνικών μέσων, παρέχουν ήδη ένα καθημερινό κανάλι έμπνευσης στους ανθρώπους. Πολλοί χρήστες χρησιμοποιούν

κοινωνικά μέσα όπως το Twitter για την παροχή συνδέσεων με ιστοσελίδες που τους ενδιαφέρουν (Rao, 2010). Τέτοιοι σύνδεσμοι, αν και όχι απαραίτητα σχετικοί με μια ρητή ανάγκη, έχουν τη δυνατότητα να εμπνεύσουν.

Η καινοτομία της Διατριβής είναι διπλή: α) Σχεδιάστηκε για πρώτη φορά σύστημα Διερευνητικής Αναζήτησης με στόχο την Εύνοια των Τυχαίων Ανακαλύψεων και β) αποδείξαμε ότι η παραπάνω επιλογή αλγορίθμων είχε θετική συνεισφορά ως προς την εύρεση χρήσιμων και παράλληλα απρόσμενων αποτελεσμάτων σε σχέση με τυπικό σύστημα αναζήτησης, το οποίο δεν ήταν διαδραστικό, δεν αξιοποίησε οπτικοποίηση εννοιών και αλληλεπίδραση με τα αποτελέσματα, διαφοροποίηση των αποτελεσμάτων και ποικιλία πηγών πληροφορίας, αλλά ούτε και εισήγαγε με κάποια μεθοδολογία το απροσδόκητο.

Στο κεφάλαιο 4 της παρούσας Διατριβής παρουσιάζουμε αναλυτικά τη μέθοδο που συνδυάζει τα παραπάνω στοιχεία και μπορεί, εάν εισαχθεί σε ένα σύστημα αναζήτησης πληροφορίας, να αυξήσει την πιθανότητα για τυχαίες αλλά χρήσιμες ανακαλύψεις.

3.3.2 Υποστήριξη της Δημιουργικότητας μέσω της Εύνοιας Τυχαίων Ανακαλύψεων στο πλαίσιο της Διερευνητικής Αναζήτησης

Η έννοια της εύνοιας τυχαίων ανακαλύψεων συσχετίζεται με τη δημιουργικότητα, όμως η βιβλιογραφία δεν προτείνει σχετική μέθοδο στο πλαίσιο της διερευνητικής αναζήτησης που να επαληθεύει αυτή την αντίληψη.

Η παρούσα Διατριβή στοχεύει να αξιολογήσει την επίδραση ενός συστήματος, σχεδιασμένου με σκοπό την ενίσχυση της εύνοιας τυχαίων ανακαλύψεων, στη δημιουργική διαδικασία και το δημιουργικό αποτέλεσμα. Για το σκοπό αυτό προτείνουμε κατάλληλο σχεδιασμό πειραμάτων, προκειμένου να αξιολογηθεί το δημιουργικό αποτέλεσμα σε εργασίες που απαιτούν δημιουργική σκέψη και αποτελέσματα. Έγινε προσπάθεια να καλυφθούν δύο ειδών εργασίες: α) εργασίες που στοχεύουν σε απόκλιση, όπου κυρίαρχο ρόλο παίζει η πρωτοτυπία, αλλά και η

ποικιλία και ποσότητα (π.χ. εργασίες παραγωγής ιδεών, συνεδρίες καταιγισμού ιδεών) και β) εργασίες που στοχεύουν πρώτα στην απόκλιση και έπειτα στη σύγκλιση, όπου κυρίαρχο ρόλο παίζει εκτός από την πρωτοτυπία, η ποιότητα του προϊόντος ή η ευκρίνεια στη διατύπωση της ιδέας (π.χ. διατύπωση ερευνητικής ιδέας, συγγραφή ενός δημοσιογραφικού άρθρου).

Η καινοτομία της Διατριβής είναι ότι έδειξε πειραματικά, αλλά και σε πιλοτική μελέτη - με τη χρήση προτεινόμενων στη βιβλιογραφία μετρικών- ότι ένα σύστημα διερευνητικής αναζήτησης, μπορεί αφενός να ενισχύσει την εύνοια τυχαίων ανακαλύψεων, και αφετέρου μέσω αυτής να ενισχύσει το Δημιουργικό Αποτέλεσμα καθώς και τη Δημιουργική Διαδικασία.

Στο κεφάλαιο 4 της παρούσας Διατριβής παρουσιάζονται αναλυτικά τα αποτελέσματα των σχετικών πειραμάτων που επιβεβαιώνουν τα ανωτέρω.

3.3.3 Υποστήριξη Διερευνητικής Αναζήτησης και Δημιουργικότητας μέσω ενσωμάτωσης Πλαισίου Αλληλεπίδρασης

Το συγκεκριμένο ερευνητικό ερώτημα αφορά αφενός στη μεθοδολογία αξιοποίησης του πλαισίου αλληλεπίδρασης σε ένα σύστημα διερευνητικής αναζήτησης και αφετέρου στη μελέτη της επίδρασής του στο Δημιουργικό αποτέλεσμα.

Η παρούσα Διατριβή προτείνει τον εμπλουτισμό του μοντέλου Διερευνητικής Αναζήτησης με υπηρεσία παροχής προτάσεων που αξιοποιεί το Πλαίσιο Αλληλεπίδρασης του χρήστη. Η προτεινόμενη μεθοδολογία βασίζεται στη μοντελοποίηση του πλαισίου αλληλεπίδρασης με τη χρήση Κρυφών Μαρκοβιανών Μοντέλων. Οι ενέργειες των χρηστών αναπαρίστανται ως αλληλουχίες ενεργειών και όχι ως μεμονωμένες αλληλεπιδράσεις με το σύστημα. Ο σκοπός είναι ο εντοπισμός μοτίβων αναζήτησης, τα οποία υποθέτουμε ότι αντιστοιχίζονται σε λανθάνουσες γνωστικές (ή πνευματικές) καταστάσεις του χρήστη. Για τη δημιουργία του μοντέλου προτείνεται η διεξαγωγή ενός πειράματος διερευνητικής αναζήτησης με τη

συμμετοχή μεγάλου αριθμού χρηστών, προκειμένου να καταγραφούν οι ενέργειές τους και να τροφοδοτηθεί ο αλγόριθμος εκμάθησης παραμέτρων του μοντέλου. Ο στόχος της μεθόδου είναι, δοθέντων των τρεχουσών ενεργειών στα πλαίσια μίας συνεδρίας διερευνητικής αναζήτησης και του εκπαιδευμένου μοντέλου, να γίνεται πρόβλεψη της πιθανότερης επόμενης ενέργειας του χρήστη, η οποία θα προτείνεται εντός της συνεδρίας αναζήτησης.

Για την αξιολόγησή της μεθόδου σχεδιάστηκε κατάλληλο πείραμα, όπου σαν σύστημα αναφοράς χρησιμοποιήθηκε το σύστημα που σχεδιάστηκε για το Ερευνητικό Ερώτημα 1 και 2 και σαν αξιολογούμενο σύστημα το ανωτέρω σύστημα εμπλουτισμένο με την υπηρεσία προτάσεων βάσει πλαισίου αλληλεπίδρασης.

Η καινοτομία της Διατριβής έγκειται στο ότι α) προτείνει μια καινοτόμο προσέγγιση παραγωγής προτάσεων, όπου β) οι προτάσεις και προτροπές αφενός ωθούν το χρήστη να εκτελέσει περισσότερες ενέργειες ενθαρρύνοντας έτσι μια πιο διερευνητική συμπεριφορά, αφετέρου αποκαλύπτουν πιθανώς ανεξερεύνητες πτυχές του υπό έρευνα θέματος και τέλος γ) μελετά την επίδραση των προτάσεων βάσει Πλαισίου Αλληλεπίδρασης στο Δημιουργικό Αποτέλεσμα της εργασίας του χρήστη, ενώ δείχνει ότι η μέθοδος που αξιοποιεί το πλαίσιο αλληλεπίδρασης υπερέχει της μεθόδου που δεν το αξιοποιεί με στατιστική σημαντικότητα.

3.3.4 Υποστήριξη Διερευνητικής Αναζήτησης και Δημιουργικότητας μέσω ενσωμάτωσης Χαρακτηριστικών Προσωπικότητας

Το συγκεκριμένο ερευνητικό ερώτημα αφορά στη μελέτη της επίδρασης που έχει η προσαρμογή προτάσεων βάσει της προσωπικότητας του αναζητητή στη διαδικασία διερεύνησης και στο δημιουργικό αποτέλεσμα της εργασίας για την οποία ο αναζητητής ξεκινά μια συνεδρία αναζήτησης.

Η παρούσα Διατριβή προτείνει την προσαρμογή των προτάσεων που σχεδιάστηκαν στο ερευνητικό ερώτημα 3 με βάση την προσωπικότητα του αναζητητή. Για να επιτευχθεί αυτό, η μεθοδολογία χρειάζεται να πάρει διαφορετική

διαδρομή ήδη από το στάδιο της προ-επεξεργασίας των δεδομένων και της εκπαίδευσης των μοντέλων. Για τη δημιουργία των μοντέλων, όπως και με το ερευνητικό ερώτημα 3, προτείνεται η διεξαγωγή πειράματος Διερευνητικής αναζήτησης με τη συμμετοχή μεγάλου αριθμού χρηστών προκειμένου να καταγραφούν οι ενέργειές τους. Μετά τη συλλογή των δεδομένων, απαιτείται η συσχέτιση των χαρακτηριστικών προσωπικότητας με τις ενέργειες που πραγματοποιεί ο αναζητητής στη διάρκεια μίας συνεδρίας, προκειμένου να επιβεβαιωθεί αν υπάρχει συσχέτιση μεταξύ προσωπικότητας και των αλληλεπιδράσεων που το σύστημα δίνει τη δυνατότητα στο χρήστη να κάνει. Με την υπόθεση ότι η προσωπικότητα επηρεάζει και τις τακτικές αναζήτησης του αναζητητή, - όπου γράφοντας τακτικές αναζήτησης αναφερόμαστε σε αλληλουχίες ενεργειών και όχι σε μεμονωμένα στιγμιότυπα ενεργειών- για τη δημιουργία των μοντέλων προτείνουμε ως πρώτο βήμα την αναπαράσταση των ενεργειών των αναζητητών ως αλληλουχίες ενεργειών και ως δεύτερο βήμα τη συσταδοποίηση των αναζητητών με βάση τα χαρακτηριστικά προσωπικότητάς τους. Ως τρίτο βήμα, για κάθε συστάδα αναζητητών τροφοδοτούμε τον αλγόριθμο εκπαίδευσης Κρυφών Μαρκοβιανών μοντέλων με τις αλληλουχίες ενεργειών των αναζητητών της συστάδας. Για κάθε συστάδα προκύπτει ένας αριθμός κρυφών καταστάσεων και για κάθε μία η πιθανότητα να παρατηρηθεί μία συγκεκριμένη ενέργεια.

Στην παρούσα Διατριβή προτείνουμε ως μέθοδο παραγωγής προτάσεων: αφού κατηγοριοποιηθεί ο τρέχων αναζητητής στην «πλησιέστερή» του συστάδα, με βάση την μετρική ομοιότητας που χρησιμοποιείται και στον αλγόριθμο συσταδοποίησης, την αξιοποίηση του Κρυφού Μαρκοβιανού Μοντέλου που αντιστοιχεί στη συστάδα για την πρόβλεψη της πιθανότερης επόμενης ενέργειας, δοθέντων των μέχρι εκείνη τη στιγμή ενεργειών του αναζητητή στα πλαίσια της συνεδρίας διερευνητικής αναζήτησης. Ο αλγόριθμος πρόβλεψης, έχοντας ως δεδομένα το μοντέλο, δηλαδή τις πιθανότητες μεταβάσεων από μια κρυφή κατάσταση σε άλλη, καθώς και τις πιθανότητες να παρατηρηθεί μία ενέργεια σε κάθε κρυφή κατάσταση, λαμβάνει ως είσοδο την αλληλουχία ενεργειών του τρέχοντος χρήστη και υπολογίζει τις

πιθανότητες όλων των δυνατών αλληλουχιών που αποτελούνται από τις μέχρι εκείνη τη στιγμή ενέργειες που έχει πραγματοποιήσει ο χρήστης και κάθε μία από τις ενέργειες που δύναται να πραγματοποιήσει στο σύστημα. Η τελευταία ενέργεια από την αλληλουχία με τη μεγαλύτερη πιθανότητα προτείνεται μέσω της υπηρεσίας προτάσεων.

Προτείνουμε την αξιολόγηση της αποτελεσματικότητας της μεθόδου α) ως προς τη δυνατότητα να διευκολύνει τη Διερευνητική Αναζήτηση και β) ως προς τη δυνατότητα να ενισχύσει το Δημιουργικό αποτέλεσμα βάσει της απόδοσης των αξιολογούμενων σε ανάθεση εργασίας, συγκρίνοντας δύο συστήματα: το πρώτο είναι σύστημα που σχεδιάστηκε για το Ερευνητικό Ερώτημα 1 και 2 εμπλουτισμένο με την υπηρεσία προτάσεων βάσει πλαισίου αλληλεπίδρασης, ενώ το δεύτερο –που είναι και το αξιολογούμενο- για την παραγωγή προτάσεων αξιοποιεί τα μοντέλα κρυφών καταστάσεων με βάση το πλαίσιο αλληλεπίδρασης χρηστών όμοιων ως προς τα χαρακτηριστικά προσωπικότητας, καθώς και την προσωπικότητα του τρέχοντος χρήστη.

Η καινοτομία της Διατριβής έγκειται στο ότι α) προτείνει μία καινοτόμο μέθοδο προτάσεων βάσει πλαισίου αλληλεπίδρασης και προσωπικότητας του αναζητητή, β) μελετά την επίδραση της μεθόδου στην υποστήριξη της *διερευνητικής αναζήτησης* και γ) στο *δημιουργικό αποτέλεσμα* της εργασίας του χρήστη.

Πρόέκυψε ότι η υιοθέτηση των χαρακτηριστικών προσωπικότητας και η παροχή προτάσεων προσαρμοσμένων σε αυτή, είχε αντίκτυπο στο να χρησιμοποιήσουν πιο αποτελεσματικά το σύστημα οι χρήστες, ενώ δεν είχε επίδραση με στατιστική σημαντικότητα στο Δημιουργικό αποτέλεσμα.

Πίνακας 4 Σύνοψη των ερευνητικών ερωτημάτων που στοχεύει να απαντήσει η Διατριβή, των προτεινόμενων λύσεων, του τρόπου αξιολόγησης και αναφορά της ενότητας που αναλύονται

Ερευνητικό Ερώτημα	Λύση	Αξιολόγηση	Ενότητα
Πώς μπορεί να υποστηριχθεί η εύνοια τυχαίων ανακαλύψεων σε σύστημα διερευνητικής αναζήτησης;	Προσδιορισμός στοιχείων σχεδιασμού που την υποστηρίζουν, υλοποίηση συστήματος και αξιολόγησή του	Διαμορφωτική Αθροιστική και Πιλοτική Αξιολόγηση	5
Μπορεί ένα σύστημα που σχεδιάστηκε να υποστηρίξει τις τυχαίες ανακαλύψεις να ενισχύσει τη δημιουργικότητα;	Αξιολόγηση δημιουργικής διαδικασίας και του δημιουργικού αποτελέσματος μέσω σύγκρισης με σύστημα αναφοράς	Διαμορφωτική Αθροιστική και Πιλοτική Αξιολόγηση	5
Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης να ενσωματώσει το πλαίσιο αλληλεπίδρασης;	Εκπαίδευση μοντέλου βάσει πλαισίου αλληλεπίδρασης και Υλοποίηση συστήματος προτάσεων επόμενης ενέργειας	Πληθοποριστικό (crowdsourced) πείραμα	6
Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει το πλαίσιο αλληλεπίδρασης να υποστηρίξει τη Διερευνητική Αναζήτηση και τη Δημιουργικότητα;	Σύγκριση συστήματος αναζήτησης που ενσωματώνει τις προτάσεις βάσει πλαισίου αλληλεπίδρασης σε σχέση με σύστημα αναζήτησης χωρίς αυτές	Ελεγχόμενο πείραμα	6
Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης να ενσωματώσει τον παράγοντα της προσωπικότητας;	Εκπαίδευση διαφορετικών μοντέλων αλληλεπίδρασης ανάλογα των ομάδων προσωπικότητας Υλοποίηση συστήματος προτάσεων	Πληθοποριστικό (crowdsourced) πείραμα	6
Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει την προσωπικότητα να υποστηρίξει τη Διερευνητική Αναζήτηση και τη Δημιουργικότητα;	Σύγκριση συστήματος αναζήτησης που ενσωματώνει τις προτάσεις σε σχέση με σύστημα αναζήτησης χωρίς αυτές	Ελεγχόμενο πείραμα	6

4 Η Επίδραση της Εύνοιας των Τυχαίων Ανακαλύψεων

Στο κεφάλαιο αυτό παρουσιάζεται η προτεινόμενη προσέγγιση διαδραστικής διερευνητικής αναζήτησης.

4.1 Εισαγωγή

Όπως συζητήσαμε και στο Κεφάλαιο 2.4, η δημιουργικότητα είναι η διαδικασία δημιουργίας νέων ιδεών και εννοιών ή η δημιουργία συνδέσεων μεταξύ ιδεών που παράγουν νέες, οι οποίες προηγουμένως δεν υπήρχαν (Amabile, 1996). Μεταξύ των σημαντικότερων παραγόντων δημιουργικότητας είναι η δυνατότητα (ή ικανότητα) της έμπνευσης. Η έμπνευση απαιτεί, μεταξύ άλλων, ένα περιβάλλον που προσφέρει χώρο για τη διερεύνηση νέων και ποικίλων πληροφοριακών πόρων (Woodman, Sawyer & Griffin, 1993), που μπορεί να προσφέρει γνωστικά κίνητρα, να διεγείρει τη μεταφορική σκέψη και να οδηγήσει σε νέες κατευθύνσεις για την αντιμετώπιση ενός προβλήματος ή να αποκαλύψει νέες πτυχές του. Οι «μηχανές έμπνευσης» είναι εργαλεία που είναι ικανά να παρέχουν περιεχόμενο, το οποίο μπορεί να είναι απροσδόκητο στις δραστηριότητες οικοδόμησης της γνώσης (knowledge building activities) (Acosta, 2012) και πολλές από τις οποίες υπάρχουν ήδη με τη μορφή κοινωνικών «μηχανών αναζήτησης». Για παράδειγμα, το YossarianLives είναι μια μηχανή οπτικής και μεταφορικής έμπνευσης, της οποίας οι αλγόριθμοι επιστρέφουν αποτελέσματα που είναι διαφορετικά, αλλά πιθανώς σχετίζονται μεταφορικά με το ερώτημα των χρηστών. Το Sprezify είναι ένα παρόμοιο εργαλείο αναζήτησης με στόχο να προκαλέσει έκπληξη διατηρώντας τη συνάφεια στα αποτελέσματα.

Οι μηχανές έμπνευσης συνήθως βασίζονται στον Παγκόσμιο Ιστό ως κύρια πηγή πληροφοριών. Επιπλέον, όλο και περισσότεροι χρήστες του διαδικτύου βρίσκουν πληροφορίες που δυνητικά μπορούν να εμπνεύσουν σε διαδικτυακές ροές πληροφοριών και σε μέσα κοινωνικής δικτύωσης. Για παράδειγμα, οι χρήστες του

κοινωνικού δικτύου Twitter ανταλλάσσουν συνδέσμους με ιστοσελίδες που τους ενδιαφέρουν (το 25% περίπου των 90 εκατομμυρίων tweets ανά ημέρα περιέχει συνδέσμους (Rao, 2010)). Τέτοιοι σύνδεσμοι, αν και όχι απαραίτητα σχετικοί με μια ρητή ανάγκη, έχουν τη δυνατότητα να εμπνεύσουν. Ομοίως, οι χρήστες συχνά συνιστούν πόρους ο ένας στον άλλο μέσω ιστότοπων κοινωνικής δικτύωσης. Οι Parise, Whelan & Todd (2015) στο πρόσφατο άρθρο τους καταδεικνύουν ότι οι εργαζόμενοι που χρησιμοποιούν το Twitter είχαν καλύτερες ιδέες από όσους δεν το έκαναν και επιπλέον με ένα διαφορετικό δίκτυο Twitter -αυτό που τους εκθέτει σε ανθρώπους και ιδέες που δεν γνωρίζουν ήδη- τείνουν να δημιουργούν ιδέες καλύτερης ποιότητας.

Παρά το τεράστιο ποσό πληροφορίας που υπάρχει στον Παγκόσμιο και Κοινωνικό Ιστό, η οποία είναι δυνητικά ικανή να εμπνεύσει, αλλά και την εμφάνιση εξειδικευμένων μηχανών έμπνευσης, δεν είναι εύκολο για τους χρήστες να ανακαλύψουν πληροφορίες που θα τους εμπνεύσουν. Συνήθως, οι χρήστες αναζητούν την πληροφορία και την έμπνευση στον κατάλογο αποτελεσμάτων των μηχανών που κατά κύριο λόγο ακολουθούν το συμβατικό πρότυπο αναζήτησης. Οι χρήστες γράφουν τα ερωτήματά τους και η μηχανή παρουσιάζει μια λίστα με αποτελέσματα, τότε ο χρήστης μπορεί να διατρέξει τη λίστα και ενδεχομένως να ανακαλύψει έναν πόρο έμπνευσης διαβάζοντας την προεπισκόπηση των αποτελεσμάτων. Επιπλέον, το δυναμικό έμπνευσης των μεμονωμένων πληροφοριακών πόρων μπορεί να αυξηθεί εάν συνδυαστεί με άλλες ενδείξεις και κατ' αυτόν τον τρόπο χρησιμοποιείται για την ενίσχυση μιας αναζήτησης που ξεκίνησε από έναν χρήστη.

Υποθέτουμε ότι, αντίθετα με τους τυπικούς μηχανισμούς που κατευθύνουν τους χρήστες σε συγκεκριμένους πόρους πληροφορίας ως απάντηση στα ερωτήματά τους, το «παράδειγμα» διερευνητικής αναζήτησης ταιριάζει καλύτερα όταν ο στόχος αναζήτησης δεν είναι σαφώς καθορισμένος και αλλάζει καθώς η αναζήτηση εξελίσσεται, όπως συμβαίνει με την αναζήτηση έμπνευσης (Marchionini 2006, White and Roth 2009). Οι παραδοσιακές μηχανές αναζήτησης ιστού χρησιμοποιούν

αλγόριθμους ανάκτησης πληροφορίας που βασίζονται σε κείμενο και περιεχόμενο, για να επιστρέψουν τα πιο σχετικά αποτελέσματα σε ένα συγκεκριμένο ερώτημα. Για την κατάταξη των σελίδων του συνόλου των αποτελεσμάτων, λαμβάνουν υπόψιν παράγοντες, όπως η δημοτικότητα συγκεκριμένων ιστοσελίδων και οι προσωπικές προτιμήσεις του χρήστη. Από την άλλη πλευρά, η διερευνητική αναζήτηση, ως ένας ιδιαίτερος κλάδος της Ανθρωπο-Υπολογιστικής Ανάκτησης Πληροφορίας (HCIR), θεωρεί τον χρήστη και το περιβάλλον του και με την αξιοποίηση των στρατηγικών αναζήτησης και περιήγησης στοχεύει στην επίτευξη υψηλότερου επιπέδου αλληλεπίδρασης χρηστών με το σύστημα αναζήτησης. Η διερευνητική μελέτη ενθαρρύνει τη μάθηση και την έρευνα (White and Roth, 2009) και μπορεί να διευκολύνει τον προσανατολισμό του χρήστη και τη μεγαλύτερη δέσμευση του συστήματος, γεγονός που με τη σειρά του μπορεί να οδηγήσει στην ανακάλυψη λανθάνουσών εναλλακτικών κατευθύνσεων στον χώρο πληροφορίας (Marchionini, 2006, Alvarado, Ackerman & Karger, 2004). Ως εκ τούτου, η διερεύνηση εναλλακτικών κατευθύνσεων αναζήτησης μπορεί να οδηγήσει σε τυχαίες, απροσδόκητες, αλλά καλοδεχούμενες ανακαλύψεις για τις οποίες κάποιος δεν έψαχνε.

Η έννοια Serendipity ή αλλιώς εύνοια των τυχαίων ανακαλύψεων, θεωρείται εγγενής έννοια στη δημιουργική σκέψη και στη δημιουργική επίλυση προβλημάτων (Foster & Ford, 2003, McCay-Peet & Toms, 2011). Η διερευνητική αναζήτηση πληροφοριών μπορεί να οδηγήσει σε ένα τυχαίο εύρημα, που μπορεί να εκπλήξει, το οποίο με τη σειρά του μπορεί να εμπνεύσει τον χρήστη και να οδηγήσει σε νέες ιδέες σχετικές με την προηγούμενη γνώση του χρήστη (Boden, 1996). Ο Björneborn (2008), στην έρευνά του για τους παράγοντες που επηρεάζουν την εύνοια τυχαίων ανακαλύψεων, σε μια φυσική και ψηφιακή βιβλιοθήκη, υποστηρίζει ότι τα αναπάντεχα και ευχάριστα ευρήματα συνδέονται με έναν λανθάνοντα τρόπο με τις διερευνητικές, αλλά και τις αποκλίνουσες από την άμεση συνάφεια, συμπεριφορές αναζήτησης. Ως εκ τούτου, για να αντιμετωπιστεί η αναποτελεσματικότητα των τυπικών μηχανών αναζήτησης σε ό,τι αφορά την έμπνευση, προτείνουμε ένα εργαλείο διερευνητικής αναζήτησης που έχει σχεδιαστεί για να αντιμετωπίσει το 1^ο

και 2^ο ερευνητικό ερώτημα που εξετάζουμε στα πλαίσια της Διατριβής, τα οποία μπορούν να συμπυκνωθούν σε ένα «Πώς μπορεί να ενισχυθεί η δημιουργικότητα με υπολογιστική και αλγοριθμική υποστήριξη για ευχάριστες και αναπάντεχες πληροφορίες στο πλαίσιο του διαδραστικού παραδείγματος διερευνητικής αναζήτησης». Υποστηρίζουμε ότι οι προτεινόμενες μέθοδοι που αφορούν τη διεπαφή αναζήτησης αναφορικά με α) το επίπεδο εμφάνισης των αποτελεσμάτων αναζήτησης του εργαλείου, β) το ίδιο το περιεχόμενο των αποτελεσμάτων αναζήτησης καθώς και γ) την υιοθέτηση του διαδραστικού διερευνητικού ερευνητικού παραδείγματος, είναι σημαντικές από την οπτική της Αλληλεπίδρασης Ανθρώπου-Υπολογιστή και θα μας βοηθήσουν στην αντιμετώπιση του ερευνητικού ερωτήματος.

Στις παρακάτω ενότητες θα δούμε, στην παράγραφο 4.2 τους παράγοντες που υποστηρίζουν την εύνοια τυχαίων ανακαλύψεων σε φυσικά και ψηφιακά συστήματα. Στην 4.3 παρουσιάζεται το ερευνητικό υπόβαθρο που σχετίζεται με τους παράγοντες που καθιστούν εφικτή την εύνοια των τυχαίων ανακαλύψεων. Στην 4.4 παρουσιάζεται η προσέγγιση του διαδραστικού μοντέλου διερευνητικής αναζήτησης, ενώ στην 4.5 παρουσιάζεται η υλοποίηση του προηγούμενου μοντέλου στο σύστημα CRUISE (Creative User Centric Inspirational Search) σε συνδυασμό με περιπτώσεις χρήσης του συγκεκριμένου συστήματος. Στην 4.6 παρουσιάζεται η αξιολόγηση της μεθόδου μας από δύο πειράματα και μία πιλοτική μελέτη. Τέλος παρουσιάζονται τα συμπεράσματα και οι προτάσεις μας.

4.2 Σχετικές Εργασίες

4.2.1 Εύνοια Τυχαίων Ανακαλύψεων

Ο όρος serendipity (εύνοια των τυχαίων ανακαλύψεων) δημιουργήθηκε για πρώτη φορά από τον Άγγλο συγγραφέα Horace Walpole το 1754, εμπνευσμένος από το παραμύθι "The Three Princes of Serendip", στο οποίο οι ήρωες πάντα ανακάλυπταν τυχαία πράγματα που δεν σκόπευαν να αναζητήσουν. Σήμερα, η

εύνοια των τυχαίων ανακαλύψεων αναφέρεται συνήθως σε μια απροσδόκητη, αλλά χρήσιμη και ευχάριστη ανακάλυψη. Η έννοια της Εύνοιας των Τυχαίων Ανακαλύψεων έχει διερευνηθεί ως διεπιστημονικό θέμα σε διάφορους τομείς όπως η οικονομία (Landes, 1994), η ιστορία (Delgadillo και Lynch, 1999), η εθνογραφική έρευνα (Dine and Deegan, 1996), η βιβλιοθήκη και η επιστήμη των πληροφοριών (Williamson 1998; Erdelez 1997, Foster και Ford 2003), στον οργανωτικό τομέα (Cunha, Clegg και Mendonça 2005) καθώς και σε ψηφιακούς χώρους πληροφόρησης (Toms 2000, Foster and Ford 2003), όπου ανήκει και η δική μας μελέτη.

4.2.2 Ανθρώπινοι Παράγοντες που επηρεάζουν την Εύνοια Τυχαίων Ανακαλύψεων

Τα άτομα που ανακαλύπτουν ευχάριστη και απρόσμενη πληροφορία χαρακτηρίζονται από αντίληψη, περιέργεια, ευέλικτη σκέψη - παρόμοια με τη θεωρία εναλλακτικής σκέψης (lateral thinking) του de Bono - και εντατική προετοιμασία (Roberts 1989). Προκειμένου η ανακάλυψη να αποκτήσει κάποιο νόημα, είναι απαραίτητη η προηγούμενη γνώση σχετικά με τον τομέα της ανακάλυψης. Αυτή η προετοιμασία του νου που επιτρέπει την επακόλουθη αναγνώριση της εύνοιας των τυχαίων ανακαλύψεων, ονομάζεται «οξύνοια του νου» (*sagacity*) (Seifert, Meyer, Davidson, Patalano & Yaniv 1994). Ο νους που έχει προετοιμασθεί (ήτοι «οξύς νους») υπονοεί ότι το στοιχείο πρέπει να είναι σχετικό ή εν μέρει σχετικό με τις προηγούμενες γνώσεις του χρήστη (Spink, Greisdorf & Bateman, 1998, André, Teevan & Dumais, 2009) προκειμένου ο αναγνώστης να αναγνωρίσει τη χρησιμότητά του.

Λαμβάνοντας ως παράδειγμα την ανακάλυψη της Πενικιλίνης από τον Alexander Flemming, είναι εμφανή τα δύο κύρια χαρακτηριστικά μιας σοβαρής ανακάλυψης: α) το γεγονός ότι απέτυχε να απολυμάνει τις καλλιέργειες βακτηρίων φεύγοντας για διακοπές, για να τα βρει μετά μολυσμένα με καλούπια *Penicillium*, που σκότωσαν τα βακτηρίδια, ήταν ένα τυχαίο γεγονός, β) το γεγονός ότι είχε προηγουμένως εκπονήσει εκτενείς έρευνες για τις αντιβακτηριακές ουσίες και έτσι αυτή η

ανακάλυψη είχε ιδιαίτερη απήχηση για αυτόν, αποδεικνύει ότι η «οξύνοια του πνεύματος», κατόπιν σχετικής προετοιμασίας, ήταν κρίσιμη για αυτή την ανακάλυψη. Επομένως, η εύνοια των τυχαίων ανακαλύψεων δεν είναι το αποτέλεσμα ενός μεμονωμένου τυχαίου περιστατικού, αλλά ένας συνδυασμός των τυχαίων ανακαλύψεων με την περιέργεια, την προηγούμενη γνώση και την ανοιχτή σκέψη.

4.2.3 Μελέτες και Μοντέλα Εισαγωγής Εύνοιας Τυχαίων Ανακαλύψεων στην Ανάκτηση Πληροφορίας

Τα παραδοσιακά συστήματα αναζήτησης με λέξεις-κλειδιά έχει βρεθεί ότι αποτελούν απειλή για την εύνοια τυχαίων ανακαλύψεων, καθώς συστηματικά περιορίζουν τη συλλογή αποτελεσμάτων σε εκείνα που ταιριάζουν αυστηρά με τις λέξεις-κλειδιά (Foster & Ford 2003). Χρειάζεται συνεπώς να δημιουργηθούν διαδικασίες που να ενισχύουν τα αναπάντεχα ευρήματα. Αυτό θα μπορούσε να υλοποιηθεί με μηχανισμούς όπως αυτούς που πρότεινε η Toms το 2000:

- εισαγωγή τυχειότητας
- εισαγωγή ανωμαλιών και εξαιρέσεων στις ήδη χρησιμοποιούμενες μετρικές ομοιότητας
- εύρεση του προετοιμασμένου μυαλού με δημιουργία προφίλ χρηστών
- εκμετάλλευση αναλογιών

Επίσης, ο Kerne το 2008 πρότεινε παροχή πληροφορίας με οπτικά μέσα. Οι André et al. (2009) μελέτησαν το βαθμό στον οποίο η εξατομικευμένη επιστροφή αποτελεσμάτων θα επηρέαζε αρνητικά την πιθανότητα της εύνοιας τυχαίων ανακαλύψεων. Επίσης, η εύνοια τυχαίων ανακαλύψεων μελετήθηκε και στο πλαίσιο τη διήθησης πληροφορίας και συγκεκριμένα στους αλγόριθμους συνεργατικής διήθησης όπου ενδιαφέρον περιεχόμενο εντοπίζεται από το ταίριασμα χρηστών ανάλογα με τα ενδιαφέροντά τους. Εργασίες σχετικές με την συνεργατική διήθηση έχουν προωθήσει το νεωτερισμό και την ανακάλυψη βοηθώντας τους χρήστες να

ανακαλύψουν λιγότερο δημοφιλή και ποικιλόμορφα αποσπάσματα πληροφορίας. Ευφυείς πράκτορες λογισμικού (intelligent software agents), όπως η Letizia (Liebermann 1995), προωθούν την εύνοια τυχαίων ανακαλύψεων προτείνοντας πιθανώς ενδιαφέρον περιεχόμενο στους αναζητητές που σχετίζεται με τις τρέχουσες δραστηριότητές τους.

Παρ' όλ' αυτά η εύνοια των τυχαίων ανακαλύψεων είναι δύσκολο να μετρηθεί, όπως και να εισαχθεί πειραματικά (Erdelez 2004; André 2009). Οι Spink et al. (1998) έδειξαν ότι τα μερικώς σχετικά αποτελέσματα μπορούν να ενισχύσουν την εύνοια των τυχαίων ανακαλύψεων οδηγώντας σε νέες ιδέες και κατευθύνσεις. Στην έρευνά του ο André (2009) ανέδειξε κάποια χαρακτηριστικά που εντοπίζονται σε ερωτήματα που θα μπορούσαν να προκαλέσουν την εύνοια τυχαίων ανακαλύψεων. Μελέτησε ερωτήματα πληροφοριακά, περιηγητικά, σχετιζόμενα με εργασία ή όχι, αν περιείχαν κάποιο όνομα ή όχι. Η εντροπία των clicks, εντοπίζοντας διαφοροποιήσεις στα αποτελέσματα, βρέθηκε να είναι θετικά συσχετισμένη με τον αριθμό των πιθανώς ευχάριστων και αναπάντεχων αποτελεσμάτων. Φαίνεται ότι η ανάλυση του ιστορικού των χρηστών μπορεί να βοηθήσει στον εντοπισμό μοτίβων που προκαλούν την εύνοια τυχαίων ανακαλύψεων.

Σήμερα, οι περισσότερες μελέτες για την υποστήριξη της εύνοιας των τυχαίων ανακαλύψεων στα ψηφιακά πληροφοριακά συστήματα, εστιάζουν στους παράγοντες της τύχης, του προετοιμασμένου νου και της παρουσίασης πληροφορίας η οποία θα μπορούσε να αποδειχτεί "serendipitous". Η Toms (2000) που επικεντρώνεται σε «ευχάριστη και αναπάντεχη ανάκτηση» (serendipitous retrieval) δίνει έμφαση σε: α) το ρόλο της τύχης, β) την ευκαιρία που ευνοεί ένα προετοιμασμένο μυαλό, γ) τις ανωμαλίες και εξαιρέσεις που αντιστοιχούν σε πληροφορίες εντελώς διαφορετικές μεταξύ τους, πράγμα που μπορεί να οδηγήσει σε απροσδόκητες και ευχάριστες πληροφορίες και δ) τη συλλογιστική κατ' αναλογία (reasoning by analogy). Ισχυρίζεται ότι η απόκτηση πληροφοριών μπορεί να προκύψει από την αναζήτηση, την περιήγηση και την πράξη της σάρωσης και ότι ιδιαίτερα στο πλαίσιο της αναζήτησης ή της περιήγησης σε έναν ψηφιακό χώρο, η

ευχάριστη και αναπάντεχη ανάκτηση μπορεί να συμβεί. Στη μελέτη του για τις φυσικές και τις ψηφιακές βιβλιοθήκες, ο Björneborn (2010) αναγνώρισε δέκα διαστάσεις που θα μπορούσαν να διευκολύνουν την εύνοια των τυχαίων ανακαλύψεων, με κάποιες από αυτές να είναι: πλούσια ποικιλία πόρων, εντυπωσιακές αντιθέσεις, διευρυμένη θεματολογία (divergent topics), δυνατότητα διερεύνησης, απεικόνιση που προκαλεί την περιέργεια και άλλες πέντε. Αναφερόμενοι στις διαστάσεις της εύνοιας τυχαίων ανακαλύψεων κατά Björneborn, οι McCay Peet και Toms (2011) προέβησαν σε ανάλυση επάνω σε αυτές, η οποία κατέληξε στη σύνθεση πέντε πρωτογενών παραγόντων που καθιστούν εφικτή την εύνοια τυχαίων ανακαλύψεων: α) δημιουργία συνδέσεων (enabled connection), β) εισαγωγή του απροσδόκητου (introducing the unexpected), γ) έμφαση στην παρουσίαση ποικιλίας (presented variety), δ) ενεργοποίηση της απόκλισης (triggered divergence) και ε) πρόκληση περιέργειας (induced curiosity), οι οποίες σε αντίθεση με τον Björneborn ήταν πλήρως προσαρμοσμένες σε ψηφιακά πληροφοριακά συστήματα. Οι Burkell, Quan-Haase & Rubin (2012) πρότειναν ένα εννοιολογικό μοντέλο των «πτυχών της εύνοιας τυχαίων ανακαλύψεων» που περιλαμβάνει τέσσερις άξονες: α) το "εύρημα", το οποίο μπορεί να είναι πληροφορία, πρόσωπο ή αντικείμενο και θεωρείται ο πυροδότης για το τυχαίο αποτέλεσμα, β) την "τυχαία φύση" (accidental nature) του ευρήματος, το οποίο μπορεί να αφορά απρόσμενη θέση, χρόνο ή έλλειψη γνώσης από εκείνον που το συναντά, γ) "τον προετοιμασμένο νου" (the prepared mind) που αναφέρεται στη συνάφεια του ευρήματος σε σχέση με ό,τι γνωρίζει ήδη ο αναζητητής, καθώς και δ) την "παρατήρηση" (act of noticing) του ευρήματος από τον χρήστη.

4.2.4 Χαρακτηριστικά Πληροφοριακών Συστημάτων που υποστηρίζουν ευχάριστες και αναπάντεχες ανακαλύψεις

Δεδομένου ότι η διερευνητική συμπεριφορά διευκολύνει την εύνοια τυχαίων ανακαλύψεων, ο σχεδιασμός ψηφιακών πληροφοριακών συστημάτων που διευκολύνουν τη διερευνητική αναζήτηση και ορίζουν τον τρόπο με τον οποίο ένα

άτομο αλληλοεπιδρά και διερευνά το περιβάλλον πληροφορίας είναι ζωτικής σημασίας για την αύξηση των ευκαιριών για αναπάντεχες και ευχάριστες ανακαλύψεις. Όπως φανερώνεται στο εννοιολογικό μοντέλο των Burkell, Quan-Haase & Rubin (2012), δεν είναι μόνο οι παράγοντες που είναι σχετικοί με το χρήστη, όπως το «προετοιμασμένο μυαλό» ή η «οξύνοια του νου», που έχουν μεγάλη σημασία για να βιώσει κανείς συναρπαστικές ανακαλύψεις, αλλά και το «εύρημα», που είναι το αντικείμενο ή η πληροφορία που παρέχεται από το σύστημα, καθώς και ο «τυχαίος χαρακτήρας της παρατήρησης», αλλά και η ίδια η «παρατήρηση» ως ενέργεια.

Η πράξη της «παρατήρησης» ή της αναγνώρισης αναπάντεχων και ευχάριστων πληροφοριών είναι κρίσιμη για ένα προετοιμασμένο μυαλό. Ως εκ τούτου, υποστηρίζουμε ότι ο τρόπος με τον οποίο οι πληροφορίες παρουσιάζονται στον χρήστη είναι ένας παράγοντας που διευκολύνει την εύνοια των τυχαίων ανακαλύψεων. Η παρουσίαση των πληροφοριών μπορεί να υποστηριχθεί μέσω διαφορετικών οπτικοποιήσεων, όπως χρώματα, γραμματοσειρές, αναδιοργάνωση περιεχομένου, συνδυασμός διαφορετικών τύπων πληροφοριών σε ένα πλαίσιο, έλξη προσοχής και τόνωση της αντίληψης κάποιου να αναγνωρίσει ενδιαφέρουσες πληροφορίες που διαφορετικά θα ήταν δύσκολο να παρατηρήσει (Burkell, Quan-Haase & Rubin, 2012; Kules, 2005; Björneborn, 2010). Ο σχεδιασμός διεπαφών διερευνητικής αναζήτησης θα πρέπει να επιτρέπει στους χρήστες να δουν πώς οι διαφορετικές εξερευνήσεις μπορούν να οδηγήσουν στα ίδια ή συναφή αποτελέσματα ή διατυπώνοντάς το διαφορετικά να τους βοηθούν να εντοπίσουν αναλογίες σε διαφορετικούς τομείς. Επιπλέον, οι οπτικοποιήσεις της συμπεριφοράς περιήγησης ενός ατόμου, οι οπτικοί τρόποι παρουσίασης αποτελεσμάτων, καθώς και οι προτάσεις για διερεύνηση, μπορούν να υποστηρίξουν το χρήστη στην άντληση γνώσης μέσα από τεράστιες ποσότητες πληροφοριών.

Επιπλέον, όσον αφορά το στοιχείο του "ευρήματος" στην εύνοια τυχαίων ανακαλύψεων, υποστηρίζουμε ότι έννοιες όπως το απροσδόκητο (unexpectedness), η ποικιλομορφία (diversity) και η καινοτομία (novelty), μπορούν άμεσα ή έμμεσα να διαδραματίσουν κάποιο ρόλο στο αποτέλεσμα της έκπληξης και, ως εκ τούτου, να

ενισχύσουν την εύνοια τυχαίων ανακαλύψεων. Αρκετοί μελετητές υπογραμμίζουν τη σημασία αυτών των παραγόντων για τη στήριξη της εύνοιας τυχαίων ανακαλύψεων.

Η εμπειρία που συνοδεύεται από το απροσδόκητο θεωρείται μια βασική πτυχή της εύνοιας τυχαίων ανακαλύψεων σύμφωνα με τους Makri & Blandford (2012), οι οποίοι στο προτεινόμενο πλαίσιο τονίζουν ότι η απροσδόκητη εμπειρία, η άντληση γνώσης και το πολύτιμο αποτέλεσμα αποτελούν ισχυρούς δείκτες των αναπάντεχων και ευχάριστων ανακαλύψεων (*serendipitous encounters*). Οι André, Teevan και Dumais (2009) διερεύνησαν τον τρόπο με τον οποίο η εύνοια τυχαίων ανακαλύψεων μπορεί να εισαχθεί στα συστήματα προτάσεων και αναζήτησης ιστοσελίδων και υπογράμμισε τη σημασία της εξεύρεσης απροσδόκητων και ενδιαφερουσών πληροφοριών, γεγονός που αν στη συνέχεια συνδυαστεί με δημιουργία συνδέσεων με αυτό που είναι ήδη γνωστό, ίσως σε συγκεκριμένο τομέα, μια δημιουργική ιδέα να μπορεί να ακολουθήσει. Στην ίδια κατεύθυνση, οι McCay-Peet & Toms (2011) υπογράμμισαν τη σημασία της εισαγωγής του απροσδόκητου στις εργασίες αναζήτησης, προκειμένου να διευκολυνθούν οι αναπάντεχες και ευχάριστες ανακαλύψεις. Εκτός αυτού, η Boden (1996) ορίζει ως απροσδόκητη ανακάλυψη, μια νέα ανακάλυψη συνοδευόμενη από χρησιμότητα. Η καινοτομία, που ορίζεται ως η εύρεση κάτι μακρινού από αυτό που ο χρήστης γνωρίζει ήδη, μπορεί να αυξήσει την πιθανότητα έκπληξης, αλλά δεν είναι απαραίτητη προϋπόθεση. Αν εξετάσουμε, για παράδειγμα, τη σχέση ανάμεσα στο απροσδόκητο και την εύνοια τυχαίων ανακαλύψεων: μια ευνοημένη τυχαία ανακάλυψη εξ' ορισμού είναι απροσδόκητη. Από την άλλη πλευρά, η καινοτομία εμφανίζεται όταν το σύστημα προτείνει ένα άγνωστο στοιχείο στο χρήστη (Herlocker, 2004). Μια απροσδόκητη ανακάλυψη μπορεί να είναι καινοτόμος ή όχι, και παρόλο που σχετίζεται με τις προτιμήσεις του χρήστη, να είναι δύσκολο να ανακαλυφθεί σκόπιμα από το χρήστη σε ένα συγκεκριμένο πλαίσιο.

Η ποικιλομορφία, εξεταζόμενη από διάφορες πλευρές (McCay-Peet & Toms, 2011; Lewandowski, 2012), μπορεί να ενεργοποιήσει τυχαίες ανακαλύψεις, καθώς βοηθά στην αποκάλυψη άγνωστων πλευρών για πληροφορίες σχετικές με το χρήστη.

Ιδιαίτερα, η ποικιλομορφία σε θέματα, κατηγορίες θεμάτων, ψηφιακά μέσα και δραστηριότητες θεωρείται ότι έχει εξέχουσα θέση υποστηρίζοντας την αποκλίνουσα διερευνητική συμπεριφορά του αιτούντος (Björneborn, 2010). Περαιτέρω, οι McCay-Reet & Toms (2011) μοχλεύουν την έννοια της καινοτομίας με την ποικιλομορφία του περιεχομένου ως παράγοντες που μπορούν να οδηγήσουν σε μια νέα κατεύθυνση μιας κατάστασης ή να αποκαλύψουν μια νέα πτυχή ενός προβλήματος, με τη δυνατότητα να διεγερθεί η μεταφορική σκέψη και να δημιουργηθούν συνδέσεις. Συγκεκριμένα στον τομέα της αναζήτησης στον Ιστό, η ποικιλομορφία προσεγγίζεται ως ένας τρόπος να βρεθεί η καλύτερη ισορροπία μεταξύ συνάφειας και ανομοιότητας των αποτελεσμάτων και, ως εκ τούτου, μπορεί να αξιοποιήσει την αποκάλυψη πολλαπλών πλευρών ενός θέματος που εκπροσωπείται από το ερώτημα.

Πίνακας 5 Αρχές που Υποστηρίζουν την Εύνοια Τυχαίων Ανακαλύψεων

Συστατικά Εύνοιας τυχαίων Ανακαλύψεων κατά Burkell	Στοιχείο συστήματος Αναζήτησης	Ενεργοποιητής Εύνοιας Τυχαίων Ανακαλύψεων	Περιγραφή	Πηγές
‘Η ενέργεια της παρατήρησης’	Σχεδιασμός Διεπαφής Συστήματος Αναζήτησης	Οπτικές ενδείξεις στο επίπεδο διεπαφής και μοντέλα αλληλεπίδρασης	Τι μοντέλα αλληλεπίδρασης θα υιοθετηθούν για να επιτρέψουν στο χρήστη να εξερευνήσει το περιβάλλον πληροφοριών. Πώς το σύστημα επιλέγει να απεικονίσει ενδείξεις δηλ. μέσω χρωμάτων, γραμματοσειρών ή άλλων τεχνικών οπτικοποίησης που υποστηρίζουν το χρήστη στη διατύπωση ερωτήματος	(Burkell et al., 2012), (Björneborn et al., 2008), (Björneborn et al., 2010), (André et al., 2009), (Kules, 2005)

	Αναπαράσταση Αποτελεσμάτων Αναζήτησης	Κατάλληλη Οργάνωση και Ταξινόμηση των Αποτελεσμάτων	Πώς το σύστημα επιλέγει να παρουσιάσει ή να απεικονίσει το τελικό αποτέλεσμα αναζήτησης για να επιστήσει την προσοχή σε διάφορες απόψεις ή απρόσμενες σχέσεις μεταξύ των αποτελεσμάτων αναζήτησης.	
Εύρημα και Τυχαία Φύση του Ευρήματος	Αποτελέσματα Αναζήτησης	Αλγόριθμοι και Μέθοδοι που επηρεάζουν τα αποτελέσματα αναζήτησης	Κατάλληλοι αλγόριθμοι και μέθοδοι θα υιοθετηθούν για την παροχή σχετικών, αλλά και απροσδόκητων και νέων αποτελεσμάτων, λαμβάνοντας υπόψη την απόσταση από το αναμενόμενο ή προωθώντας εξαιρέσεις και ανωμαλίες. Αξιοποίηση πληροφοριών από διαφορετικές πηγές ή είδη, που επηρεάζουν το τελικό σύνολο αποτελεσμάτων αναζήτησης	(McCay-Peet & Toms, 2011), (Lewandowski, 2012), (Björneborn et al., 2008, 2010), (Burkell et al., 2012), (Toms, 2000)

Ο πίνακας 5 συνοψίζει την επισκόπηση της βιβλιογραφίας εστιάζοντας στους παράγοντες που θα πρέπει να εισάγονται στο σχεδιασμό ενός συστήματος αναζήτησης στοχεύοντας στην εύνοια τυχαίων ανακαλύψεων. Οι παράγοντες κατηγοριοποιήθηκαν σε τρεις ευρείες περιοχές που αφορούν το σχεδιασμό ενός

συστήματος αναζήτησης: α) Σχεδιασμός Διεπαφής Αναζήτησης, β) Αναπαράσταση Αποτελεσμάτων Αναζήτησης, γ) Περιεχόμενο Αποτελεσμάτων Αναζήτησης

4.2.5 Ενεργοποιητές Εύνοιας Τυχαίων Ανακαλύψεων στο Σχεδιασμό Διεπαφής Αναζήτησης

Οι ερευνητές έχουν προτείνει ποικίλες διαδραστικές τεχνικές διερευνητικής αναζήτησης που περιλαμβάνουν πλούσιες διεπαφές για να βοηθήσουν τους χρήστες να κατανοήσουν τα αποτελέσματα (Baldonado & Winograd, 1997), οπτικοποιώντας και συνοψίζοντας τα αποτελέσματα ώστε να καταστεί δυνατή η ταχύτερη εκτίμηση της ποιότητας των πληροφοριών που επιστρέφει η μηχανή αναζήτησης (Havre, Hetzler, Perrine, Jurrus & Miller, 2001). Οι ερευνητές Glowacka, Ruotsalo, Kopuyshkova, Kaskia και Jacucci (2013) ανέπτυξαν ένα διαδραστικό σύστημα ανάκτησης πληροφοριών που συνδυάζει τις τεχνικές ενίσχυσης μάθησης (*reinforcement learning*) μαζί με ένα σχεδιασμό διεπαφής χρήστη που επιτρέπει την ενεργό συμμετοχή των χρηστών στο να κατευθύνουν την αναζήτηση.

Οι Devendorf, O'Donovan & Hollerer (2012) πρότειναν μια καινοτόμο διαδραστική διεπαφή για καθοδηγούμενη διερεύνηση μέσω οπτικοποίησης θεμάτων σε μια μεγάλη συλλογή εγγράφων. Οι Millen και Yang (2007) εφαρμόζουν έρευνα σχετικά με τη χρήση ενός κοινωνικού ιστοτόπου καταγραφής σελιδοδεικτών, ενσωματωμένη σε παραδοσιακή μηχανή αναζήτησης, μέσα σε ένα εταιρικό περιβάλλον. Οι συγγραφείς εκμεταλλεύονται τη δραστηριότητα της κοινωνικής επισήμανσης και πειραματίζονται με πολλαπλές διεπαφές αναζήτησης. Το TweetMotif (O'Connor, Krieger & Ahn, 2010) εξάγει ένα σύνολο θεμάτων από το Twitter για να προσφέρει μια πολύπλευρη διεπαφή αναζήτησης. Τέλος, πολλές μελέτες έχουν επικεντρωθεί στη διερεύνηση εικόνων και άλλων πόρων πολυμέσων, βλέπε π.χ. (Waitelonis & Sack, 2012, Iijima, Kimura & Yamaguchi, 2010). Στην εργασία των Waitelonis και Sack (2012), για παράδειγμα, οι συγγραφείς προτείνουν ένα διερευνητικό σύστημα αναζήτησης για βίντεο που εκμεταλλεύεται Ανοιχτά

Διασυνδεδεμένα Δεδομένα (Linked Open Data) και συγκεκριμένα το ²DBpedia για να σχηματίσουν συσχετισμούς που βασίζονται σε περιεχόμενο και να ευνοήσουν την ανακάλυψη απρόσμενων αλλά ευχάριστων πόρων.

Άλλες αρχές στο σχεδιασμό διεπαφών που υποστηρίζουν τη διερευνητική αναζήτηση μπορεί να είναι η οργάνωση αποτελεσμάτων με βάση την κατηγορία τους, προκειμένου να γίνει πιο διαδραστική η αναζήτηση. Οι κατηγορίες μπορούν να είναι σταθερές, είτε να προέρχονται από θησαυρούς, οντολογίες ή άλλα σχήματα κατηγοριοποίησης (για παράδειγμα με βάση τον τύπο του εγγράφου) είτε δυναμικές με τη συσταδοποίηση αποτελεσμάτων σε δυναμικά δημιουργούμενες κατηγορίες όπως για παράδειγμα στο www.clusty.com (Kules, 2005).

Οι Zamir & Etzioni (1999) πρότειναν μια διεπαφή με βάση τη συσταδοποίηση σε μια υπηρεσία μετα-αναζήτησης ακολουθώντας έναν αλγόριθμο συσταδοποίησης εγγράφων. Οι Beale, McNab, & Witten (1999) προτείνουν το «HyperSpace», ένα εργαλείο απεικόνισης που παρέχει στους χρήστες μια γενική εικόνα της στρατηγικής τους στην αναζήτηση πληροφορίας, που βοηθά τους χρήστες να εντοπίζουν τις σχέσεις και δυναμικά όχι τόσο προφανείς συνδέσεις μεταξύ των ερωτημάτων. Οι Devendorf, O'Donovan και Hollerer (2012) χρησιμοποιούν τη μοντελοποίηση θεμάτων (topic modelling) και ένα υβριδικό μοντέλο οπτικοποίησης για να υλοποιήσουν ένα διαδραστικό εργαλείο για τη διερεύνηση και τη δημιουργία προτάσεων σε θέματα, έγγραφα και χρήστες μέσα σε μεγάλο σώμα περιεχομένου. Οι Melo και Carvalhais (2013) προσπάθησαν να εκμεταλλευτούν τη σκοπιμότητα, την αμεσότητα, την ποικιλομορφία, την περιέργεια, καθώς και την τυχαιότητα για να υποστηρίξουν την αναπάντεχη και ευχάριστη περιήγηση στο σύστημά τους. Πρότειναν μία ρευστή, τροχιακή διάταξη, η οποία αντιπροσώπευε τρεις

² <http://wiki.dbpedia.org/Lookup>

διαφορετικούς τύπους περιεχομένου, δηλαδή περιεχόμενο το οποίο ο χρήστης έχει επισημάνει ως σχετικό και τυχαίες πληροφορίες για τον ιστό.

Ο Kerne, ο Smith, ο Koh, ο Choi και ο Graeber (2008) παρείχαν ένα εργαλείο ανακάλυψης πληροφοριών που υποστήριζε δημιουργικές εργασίες μέσω ενός διαδραστικού σχεδιασμού διεπαφής που επιτρέπει στους χρήστες να αναζητούν και να διερευνούν αποτελέσματα από πολλαπλά ερωτήματα, συγκρίνοντας πληροφορία που βρέθηκε μέσω εικόνας και αναπαραστάσεων κειμένου σχηματίζοντας εννοιολογικές συσχετίσεις. Σύγκριναν ιδιαίτερα την εικόνα και τις αναπαραστάσεις κειμένου που δημιουργήθηκαν με γραμμικές αναπαραστάσεις κειμένου. Η αξιολόγηση έδειξε ότι η έκθεση σε αναπαραστάση σύνθεσης εικόνας-κειμένου αυξάνει τον αριθμό των συνδυασμών ιδεών που αντλούνται από τους πόρους ψηφιακών πληροφοριών.

Με σκοπό τη διευκόλυνση της επίγνωση των χρηστών μέσω της περιήγησης στο Web, ο Beale (2007) υλοποιεί το «Mitsikeru», ένα σύστημα βασισμένο σε πράκτορες λογισμικού (software agents) που μοντελοποιεί τη συμπεριφορά των χρηστών και προσδιορίζει το πλαίσιο της αλληλεπίδρασής τους με το σύστημα. Εμφανίζει σελίδες με μεγάλη συνάφεια ως έντονα χρωματιστές και αυτές με χαμηλή συνάφεια με τον κανονικό χρωματισμό. Οι Han, Shih & Carroll (2014) υποστηρίζουν τους χρήστες που καταναλώνουν τις τοπικές ειδήσεις και έχουν πρόσβαση σε διαφορετικές απόψεις, παρέχοντάς τους ένα εργαλείο που απεικονίζει τις πληροφορίες της κοινότητας χρησιμοποιώντας ένα νέφος επισημάνσεων.

4.2.6 Ενεργοποιητές Εύνοιας Τυχαίων Ανακαλύψεων στην Παρουσίαση Αποτελεσμάτων Αναζήτησης

Αναφορικά με την απεικόνιση αποτελεσμάτων, οι περισσότερες εμπορικές, δημοφιλείς μηχανές αναζήτησης γενικής χρήσης, χρησιμοποιούν τη λίστα γραμμικής κατάταξης. Από την άλλη, μπορούν να βρεθούν εμπορικές ή πειραματικές μηχανές αναζήτησης/διερεύνησης που υποστηρίζουν αναπαραστάσεις αποτελεσμάτων 2D

αναζήτησης όπως προβολές πλέγματος (<http://www.spezify.com>) ή 3D αναπαραστάσεις αποτελεσμάτων αναζήτησης (Benford, Taylor, Brailsford, Koleva, Craven, Fraser, .. & Greenhalgh, 1999).

Για παράδειγμα, στο σύστημα διερεύνησης που προσανατολίζεται στην υποστήριξη της εύνοιας τυχαίων ανακαλύψεων, οι Laere, Bordino, Mejona και Lalmas (2014), όπου το περιεχόμενο αντιπροσωπεύεται ως δίκτυο οντοτήτων, οι συντάκτες πρότειναν έναν αλγόριθμο που οργανώνει τα αποτελέσματα αναζήτησης σε τοπικά συνεκτικές δέσμες, με βάση την κατηγορία του ερωτήματος. Σε αυτή την ενότητα θα επικεντρωθούμε στο θέμα της κατάταξης αποτελεσμάτων αναζήτησης.

Ένας αριθμός πρόσφατων εργασιών επικεντρώνεται στην εισαγωγή της ποικιλομορφίας στα ανακτημένα αποτελέσματα για να α) αντιμετωπιστεί η αβεβαιότητα της αμφισημίας του ερωτήματος και να αποκαλύψει πολλαπλές πτυχές του ερωτήματος και β) να επιλύσει το μη πλεονασμό των αποτελεσμάτων προκειμένου να μεγιστοποιηθεί η ικανοποίηση όλων των χρηστών ενός συστήματος αναζήτησης.

Πιο συγκεκριμένα, αυτές οι μελέτες αντιμετωπίζουν το πρόβλημα της «διαφοροποίησης των αποτελεσμάτων» (diversification) ή την πολυδιάστατη αναζήτηση (diversity-aware search) προσπαθώντας να βρουν το σημείο ισοροπίας μεταξύ συνάφειας και διαφοροποίησης ή ανομοιομορφίας των αποτελεσμάτων. Για παράδειγμα, οι Carbonell και Goldstein (1998) προτείνουν τη μέθοδο της Μέγιστης Οριακής Συνάφειας, κατά την οποία ένα έγγραφο-αποτέλεσμα επιλέγεται βάσει του υψηλότερου σκορ συνδυασμού της βαθμολογίας ομοιότητας με το ερώτημα και ανομοιότητας με τα έγγραφα-αποτελέσματα, που επιλέχθηκαν σε προηγούμενες επαναλήψεις. Οι ερευνητές (Agrawal, Gollapudi, Halverson, & Jeong, 2009) θεωρούν τα έγγραφα-αποτελέσματα ανεξάρτητα το ένα από το άλλο και συγκρίνουν τη συνάφειά τους με κάθε πτυχή του ερωτήματος, παρέχοντας αποτελέσματα αναλογικά με την πιθανότητα να ανήκουν σε κάθε μία από τις πτυχές του ερωτήματος.

Οι Vallet & Castells (2012) συνδύασαν την προσωποποίηση με την ποικιλία στα αποτελέσματα, όπου η αναπαράσταση κάθε χρήστη προέρχεται από επισημάνσεις που δημιουργούνται από χρήστες ενός κοινωνικού ιστότοπου αποθήκευσης σελιδοδεικτών. Για να αντιμετωπίσουν το θέμα της αμφισημίας ενός ερωτήματος και να επιτύχουν τη μέγιστη κάλυψη, οι Glowacka, Ruotsalo, Kopuyshkova, Kaskia και Jacucci (2013) δειγματοληπτούν τα αποτελέσματα, τα οποία ανακτώνται από τη μηχανή αναζήτησης χρησιμοποιώντας τον αλγόριθμο Δειγματοληψίας (Dirichlet Sampling Algorithm). Τα μεταδεδομένα περιεχομένου έχουν χρησιμοποιηθεί για τη διαφοροποίηση με βάση την κατηγορηματική απόσταση χρησιμοποιώντας την ταξινόμια του Open Directory Project (Agrawal, Gollapudi, Halverson, & Jeong, 2009). Το xQuAD (Macdonald, & Ounis, 2010) μοντελοποιεί ένα αμφίσημο ερώτημα σαν ένα σύνολο υπο-ερωτημάτων και υπολογίζει τη συνάφεια των αποτελεσμάτων αναζήτησης συγκρίνοντάς τα όχι μεταξύ τους αλλά με κάθε υπο-ερώτημα. Ο αλγόριθμος OptSelect (Carannini, Nardini, Perego, & Silvestri, 2011) αναγνωρίζει τους διαφορετικούς σκοπούς που εμφανίζονται στις βελτιώσεις του ερωτήματος στις περισσότερες συνεδρίες ερωτημάτων του χρήστη και υπολογίζει τις κατανομές πιθανότητας για να διασφαλίσει ότι οι διαφορετικοί σκοποί καλύπτονται αναλογικά με τις σχετικές πιθανότητες. Οι Gollapudi & Sharma (2011) πρότειναν: α) τον αλγόριθμο Max-Min, που μεγιστοποιεί την ελάχιστη συνάφεια (για το θέμα) και την ανομοιότητα (μεταξύ δύο αποτελεσμάτων) του επιλεγμένου συνόλου και β) τον αλγόριθμο Max-Sum, που μεγιστοποιεί το άθροισμα της συνάφειας και της ανομοιότητας του επιλεγμένου συνόλου.

4.2.7 Ενεργοποιητές Εύνοιας Τυχαίων Ανακαλύψεων στο περιεχόμενο Αποτελεσμάτων Αναζήτησης

Πολλά συστήματα ανάκτησης πληροφορίας έχουν προταθεί τα οποία προσπαθούν να εισάγουν εκτός από τη συνάφεια με το ερώτημα του χρήστη, έννοιες όπως η ποικιλία, η καινοτομία, το απροσδόκητο, οι οποίες ευνοούν τις τυχαίες αλλά

χρήσιμες ανακαλύψεις. Οι Marie, Gandon και Ribière (2013) υλοποίησαν το Discovery Hub, ένα σύστημα διερευνητικής αναζήτησης το οποίο επιλέγει ενδιαφέροντες πόρους και υποστηρίζει πολύπλευρη περιήγηση στα αποτελέσματα. Παρέχει διερευνητική αναζήτηση σε απομακρυσμένα δεδομένα χρησιμοποιώντας τα SPARQL endpoints για να εξασφαλίσει τα πιο πρόσφατα δεδομένα και να επιτρέψει στους χρήστες να επηρεάσουν τα αποτελέσματα καθορίζοντας τα ερωτήματά τους. Για να επιτευχθεί γρήγορη απόκριση του συνόλου των αποτελεσμάτων, πρότειναν έναν αλγόριθμο ενεργοποίησης σημασιολογικής εξάπλωσης και μια τεχνική δειγματοληψίας που το καθιστά εφαρμόσιμο σε απομακρυσμένους γράφους Ανοιχτών Διασυνδεδεμένων Δεδομένων (LOD) που εισάγουν μη προβλεψιμότητα για να προσεγγίσουν την εύνοια τυχαίων ανακαλύψεων. Οι Bordino, Mejona και Lalmas (2013) προσπάθησαν να προσεγγίσουν την εύνοια τυχαίων ανακαλύψεων στη διερευνητική αναζήτηση αναπαριστώντας το περιεχόμενο των πηγών δεδομένων των πληροφοριών που δημιουργούν οι χρήστες σαν ένα δίκτυο οντοτήτων, το οποίο εμπλουτίστηκε περαιτέρω με μετα-δεδομένα που αφορούσαν το συναίσθημα, την ποιότητα της γραφής και την θεματική κατηγορία. Χρησιμοποίησαν επίσης αλγόριθμο Random Walk με επανεκκίνηση για προτάσεις ερωτημάτων.

Οι Αδαμόπουλος και Tuzhilin (2014) προτείνουν μια μέθοδο παραγωγής απροσδόκητων προτάσεων ορίζοντας μια συγκεκριμένη απόσταση από το αναμενόμενο σετ που αποτελείται από όλα τα παρόμοια στοιχεία που ο χρήστης έχει ήδη βαθμολογήσει και όλα τα στοιχεία που δημιουργούνται χρησιμοποιώντας τεχνικές συνεργατικής διήθησης. Οι Akiyama, Obara και Tanizaki (2010) προσδιορίζουν το απροσδόκητο ως ένα απίθανο συνδυασμό χαρακτηριστικών. Υποθέτουν ότι η χαμηλή συνύπαρξη όρων σε ένα προτεινόμενο αντικείμενο οδηγεί σε υψηλού βαθμού απροσδόκητο και πρότειναν μία γενική μετρική η οποία προσαρμόζει τις προτιμήσεις των χρηστών με τη γενική απροσδόκητη εμφάνιση του στοιχείου. Ένας ορισμός για την αναπάντεχη πληροφορία είναι αυτός που δόθηκε από τους συγγραφείς Sun, Zhang και Mei (2013) στο πλαίσιο ενός κοινωνικού δικτύου, σύμφωνα με τον οποίο η πληροφορία η οποία διαφέρει από το προφίλ της

πηγής, αλλά είναι συναφής με την ανάγκη του παραλήπτη, θεωρείται τότε από τον παραλήπτη αναπάντεχη. Μία διαφορετική προσέγγιση για να προταθούν απροσδόκητα αλλά χρήσιμα αντικείμενα προτάθηκε από τους de Gemmis, Lops, Semerago και Musto (2015). Οι συγγραφείς προτείνουν μία διεργασία έντασης γνώσης η οποία εισάγεται σε έναν ιδιαίτερου τύπου αλγόριθμο Random Walk. Ειδικότερα, εκμεταλλεύονται τις γλωσσικές γνώσεις που υπάρχουν στο Wordnet και τη Βικιπαίδεια και τις μετατρέπουν σε Γνωσιακές Μονάδες μέσω μιας φάσης εναρμόνισης που χρησιμοποιεί μηχανισμούς Επεξεργασίας Φυσικής Γλώσσας. Στη φάση συλλογιστικής, ένα μοντέλο ενεργοποίησης διάδοσης (spreading activation model) είναι υπεύθυνο για την ανεύρεση πληροφοριών δυνητικά χρήσιμων για την τροφοδότηση του αλγορίθμου προτάσεων.

Ο παράγοντας της καινοτομίας έχει προσεγγιστεί στο πεδίο των συστημάτων προτάσεων από τους Iaquinta, De Gemmis, Lops, Semerago, Filannino και Molino (2008), οι οποίοι αξιολόγησαν την καινοτομία βάσει της απόστασης μεταξύ αυτών που ο χρήστης γνωρίζει και προτιμά και αυτών που του προτείνονται. Είναι προφανές ότι, σε κάποιες περιπτώσεις, η καινοτομία προσεγγίζεται με παρόμοιο τρόπο με την κατάσταση απροσδόκητου (unexpectedness). Εναλλακτικά στη μελέτη των Clarke, Kolla, Cormack, Vechtomova, Ashkan, Büttcher και MacKinnon (2008), η καινοτομία προσεγγίζεται ως ένας τρόπος να αντιμετωπιστεί ο πλεονασμός των αποτελεσμάτων, μία έννοια συγγενική της ποικιλομορφίας.

Σε μια άλλη μελέτη, στον τομέα της μουσικής, οι συγγραφείς (Taramigkou, Bothos, Christidis, Apostolou & Mentzas, 2013) στοχεύουν στην ενίσχυση της διερευνητικής αναζήτησης σε διαφορετικούς χώρους μουσικών προτιμήσεων. Καθιστούν εφικτή την καθοδηγούμενη μετάβαση ανάμεσα σε επιθυμητά μουσικά θέματα, αλλά και την ανακάλυψη τυχαίων, απρόβλεπτων και νέων καλλιτεχνών ή τραγουδιών.

Στο ίδιο πεδίο οι Zhang, Séaghdha, Quercia και Jambor (2012), πρότειναν ένα πλαίσιο-σύμβαση για δημιουργία προτάσεων αξιοποιώντας τις έννοιες ακρίβεια,

καινοτομία, ποικιλία και εύνοια των τυχαίων ανακαλύψεων. Παρουσίασαν δύο καινούριες τεχνικές οι οποίες ενισχύουν την εύνοια τυχαίων ανακαλύψεων βασισμένες στη Λανθάνουσα Κατανομή Dirichlet, η οποία θα μπορούσε να συνδυαστεί με υπάρχουσες μεθόδους ταξινόμησης μέσω υβριδοποίησης. Έδειξαν ότι οι αλγόριθμοί τους παράγουν σημαντικά περισσότερες τυχαίες, αναπάντεχες, αλλά ευχάριστες προτάσεις και προκαλούν μεγαλύτερη ικανοποίηση στο χρήστη «θυσιάζοντας» κάποια ακρίβεια συγκρινόμενοι με βασικούς αλγόριθμους.

Εξ' όσων γνωρίζουμε, και όπως τελικά προκύπτει από την ανασκόπηση της βιβλιογραφίας, η δική μας μελέτη είναι η πρώτη προσπάθεια να σχεδιαστεί και να αναπτυχθεί ένα σύστημα διερευνητικής αναζήτησης που στοχεύει στη μόχλευση των αναγνωρισμένων μηχανισμών στο επίπεδο της διεπαφής αναζήτησης, τα αποτελέσματα αναζήτησης και την παρουσίαση των αποτελεσμάτων αναζήτησης εις τρόπον που ενισχύουν την εύνοια των τυχαίων ανακαλύψεων, με απώτερο στόχο την αξιολόγηση της σχέσης των αναπάντεχων ανακαλύψεων με τη δημιουργική διαδικασία και τα δημιουργικά αποτελέσματα κατά τη διάρκεια της διαδικασίας αναζήτησης.

4.3 Ερευνητική Προσέγγιση

Λαμβάνοντας υπόψιν τους παράγοντες εκείνους που καθιστούν εφικτή την εύνοια τυχαίων ανακαλύψεων, οι οποίοι περιγράφηκαν στην προηγούμενη ενότητα, σχεδιάσαμε και υλοποιήσαμε ένα διαδραστικό/αλληλεπιδραστικό σύστημα διερευνητικής αναζήτησης, το οποίο επικεντρώνεται στο να εξάγει στοιχεία από την *κοινωνική συζήτηση* (social chatter) που μπορούν να προσφέρουν έμπνευση και τα χρησιμοποιεί για να βοηθήσει τους χρήστες να ανακαλύψουν πληροφορία που δυνητικά μπορεί να εμπνεύσει. Η συζήτηση σε κοινωνικά μέσα μπορεί να περιέχει χρήσιμη πληροφορία που μπορεί να εμπνεύσει, παρόλο που είναι δύσκολο για ένα χρήστη να επεξεργαστεί μεγάλο όγκο πληροφορίας χειροκίνητα. Επιπλέον η δυναμική έμπνευσης κάθε πόρου πληροφορίας μπορεί να ενισχυθεί εάν συνδυαστεί

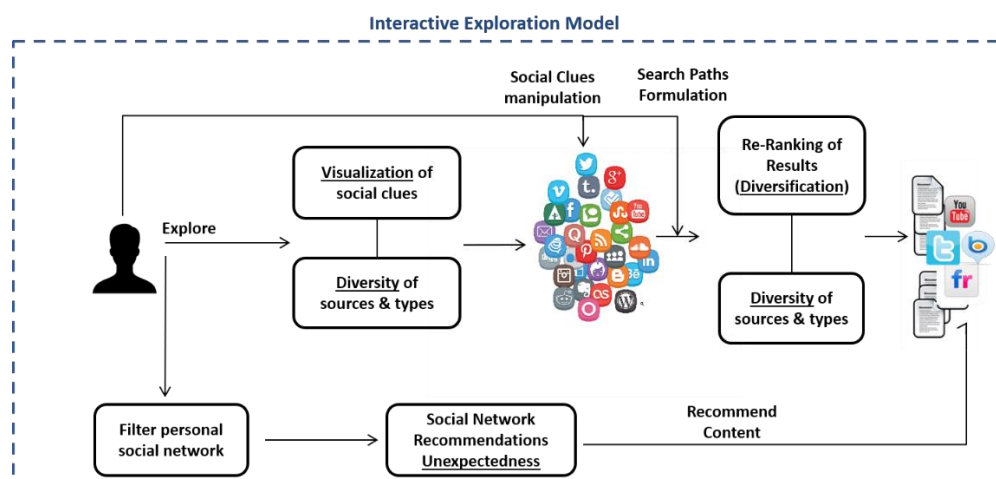
με άλλα στοιχεία που προέρχονται από το κοινωνικό δίκτυο του χρήστη. Ακολουθούμε την προσέγγιση της διοχέτευσης της κοινωνικής συνομιλίας σε μια υπηρεσία αναζήτησης πληροφοριών που επιτρέπει στους χρήστες να καταναλώνουν έμμεσα ροές πληροφορίας, όχι με τη ρητή ανάγνωσή, αλλά με τη χρήση των πληροφοριών που είναι ενσωματωμένες σε αυτές για να επεκτείνουν τα ερωτήματα αναζήτησης.

Σε αυτή την ενότητα περιγράφουμε το μοντέλο και τις μεθόδους, εστιάζοντας στις πληροφορίες που διακινούνται στα κοινωνικά δίκτυα και διαμορφώνονται σε ιστότοπους μικρο-blogging, όπως το twitter, στοχεύοντας να απαντήσουμε στο 1^ο ερευνητικό μας ερώτημα. Επιπλέον, συζητάμε με ποιο τρόπο το CRUISE, η διερευνητική μηχανή αναζήτησης που εφαρμόζει το μοντέλο και τις μεθόδους μας, αντιμετωπίζει το 2^ο ερευνητικό μας ερώτημα, το οποίο είναι να επιβεβαιώσει εάν οι κατευθυντήριες γραμμές και οι παράγοντες που αναφέρθηκαν για την ενίσχυση της εύνοιας των τυχαίων ανακαλύψεων (Πίνακας 5) μπορούν επίσης να ενισχύσουν τη δημιουργική διαδικασία και το δημιουργικό αποτέλεσμα. Υποστηρίζουμε ότι οι προτεινόμενες παρεμβάσεις στη διεπαφή αναζήτησης, τα αποτελέσματα αναζήτησης και το επίπεδο παρουσίασης των αποτελεσμάτων αναζήτησης, (λαμβάνοντας υπόψιν τους παράγοντες που ενισχύουν την εύνοια τυχαίων ανακαλύψεων που συζητήθηκαν στην ενότητα 4.2), καθώς και η υιοθέτηση της προσέγγισης διαδραστικής διερευνητικής αναζήτησης είναι σημαντικές από την οπτική γωνία της Αλληλεπίδρασης Ανθρώπου-Υπολογιστή και μπορούν να υποστηρίξουν το ερευνητικό μας ερώτημα.

4.3.1 Διαδραστικό Μοντέλο Διερεύνησης

Το μοντέλο διερεύνησης που προτείνουμε χρησιμοποιεί πληροφορίες από κοινωνικά μέσα όπως το Twitter, καθώς και άλλο περιεχόμενο του Παγκόσμιου Ιστού για να υποστηρίξει τους χρήστες να διερευνήσουν τους λεγόμενους «χώρους αναζήτησης». Οι «χώροι αναζήτησης» χρησιμοποιούν τα API των δημόσιων μηχανών αναζήτησης όπως το Bing, το Flickr, το Youtube και το μελετητή της Google, καθώς

και προσαρμοσμένες μηχανές αναζήτησης για την αναζήτηση ιδιωτικών πηγών πληροφοριών σε εταιρικές πύλες ή εταιρικά ιδιωτικά δίκτυα. Το μοντέλο μας (Εικόνα 17) λαμβάνει υπόψιν τους ακόλουθους ενεργοποιητές της εύνοιας τυχαίων ανακαλύψεων: α) την **οπτικοποίηση** των δημόσιων στοιχείων των κοινωνικών μέσων στο επίπεδο του σχεδιασμού της διεπαφής, β) την **ποικιλομορφία** των πηγών πληροφορίας στα τελικά αποτελέσματα και τον αλγόριθμο δημιουργίας **απροσδόκητης** πληροφορίας εφαρμοσμένων σε εξατομικευμένα δεδομένα που επηρεάζουν τα αποτελέσματα αναζήτησης, γ) τον αλγόριθμο **διαφοροποίησης**, που χρησιμοποιείται για την κατάταξη του τελικού συνόλου αποτελεσμάτων στο επίπεδο της παρουσίασης αποτελεσμάτων αναζήτησης, και τέλος δ) τη συμμετοχή του χρήστη στην προσαρμογή των αποτελεσμάτων (**διαδραστικότητα**).



Εικόνα 17 Διαδραστικό Μοντέλο Διερεύνησης Πληροφορίας

Η διερεύνηση αρχίζει με το χρήστη να εισάγει ένα σύνολο όρων ως αρχικό σημείο εισόδου στην διερεύνηση. Το εργαλείο χρησιμοποιεί αυτούς τους όρους και «ερωτά» δημοφιλή κοινωνικά μέσα και μηχανές αναζήτησης Ιστού. Για παράδειγμα, στην περίπτωση του Twitter, ερωτά το API αναζήτησης του Twitter για τα πιο πρόσφατα δημοφιλή σχετικά tweets. Στη συνέχεια, κατασκευάζει ένα νέφος λέξεων με συναφείς με το ερώτημα όρους-λέξεις, που βρίσκονται με υψηλή συχνότητα στα πρόσφατα δημοφιλή tweets. Επιπλέον, το CRUISE προσφέρει την δυνατότητα εισαγωγής όρων

στο νέφος λέξεων, οι οποίοι προέρχονται από tweets συγκεκριμένων χρηστών ή ροών hashtags, που σχετίζονται με ένα συγκεκριμένο πλαίσιο ή σενάριο.

Οι χρήστες μπορούν να κάνουν πολλαπλές επιλογές όρων που εμφανίζονται στο νέφος λέξεων. Αυτή η αλληλεπίδραση έχει ως αποτέλεσμα νέοι όροι να εμφανίζονται στο νέφος καθώς και ένα νέο σύνολο πόρων να παρουσιάζονται στο χρήστη. Χρησιμοποιώντας ένα κουμπί δημιουργικού «ανακατέματος», ο χρήστης μπορεί να τροποποιήσει το βάθος του χώρου αναζήτησης, το οποίο σχετίζεται άμεσα με τη δημοτικότητα των όρων. Μπορεί να κάνει drill down στο νέφος για να αποκαλύψει όρους, οι οποίοι προηγουμένως κρύβονταν λόγω χαμηλής συχνότητας σε σχέση με άλλους όρους. Περνώντας το δρομέα πάνω από τους όρους εμφανίζεται προεπισκόπηση του αρχικού πόρου στον οποίο εμπεριέχεται ο όρος, για παράδειγμα ένα tweet που περιλαμβάνει τον όρο. Αυτό βοηθά τους χρήστες να δουν πώς ο όρος σχετίζεται με το αρχικό ερώτημα.

Βοηθούμε τους χρήστες να αλληλεπιδράσουν με λέξεις όχι μόνο με το να περιηγούνται στις υποβολές των άλλων χρηστών, αλλά με το να χρησιμοποιούν τις λέξεις για να σχηματίσουν **μονοπάτια αναζήτησης**. Με δυνατότητα επιλογής ή απαλοιφής όρων που έχουν προηγουμένως επιλεγεί, δίνεται η δυνατότητα στους χρήστες να σχηματίσουν ή να τροποποιήσουν μονοπάτια αναζήτησης, τα οποία στη συνέχεια χρησιμοποιούνται σαν ερωτήματα προς τους διαθέσιμους χώρους αναζήτησης. Με αυτόν τον τρόπο, παρέχεται στους χρήστες η δυνατότητα να περιορίσουν τα μονοπάτια αναζήτησης με την επιλογή περισσότερων όρων από το «νέφος» ή να τα «χαλαρώσουν» αφαιρώντας όρους. Το εργαλείο παρέχει άμεση ανατροφοδότηση στο χρήστη ανανεώνοντας την οθόνη με νέα αποτελέσματα από τους διαθέσιμους χώρους αναζήτησης και επιτρέπει για περαιτέρω αλληλεπίδραση τροποποιώντας τα μονοπάτια αναζήτησης. Με αυτήν τη λειτουργικότητα, στοχεύουμε να καταστήσουμε εφικτό για τους χρήστες να αναγνωρίσουν και να δημιουργήσουν συσχετίσεις μεταξύ στοιχείων που δεν είχαν σκεφτεί νωρίτερα και να βρουν πληροφορία που μπορεί να οδηγήσει σε απρόσμενες και χρήσιμες ανακαλύψεις.

4.3.2 Δημιουργία και Απεικόνιση Ενδείξεων (Clues)

Χρησιμοποιώντας το μοντέλο της φολκσονομίας, για να δημιουργήσουμε το νέφος λέξεων, εντοπίζουμε τις πιο συχνές λέξεις καθώς και τις λέξεις που εμφανίζονται συχνά μαζί. Συγκεκριμένα, χρησιμοποιώντας τα επιλεγμένα API αναζήτησης ιστού και κοινωνικών μέσων μαζικής ενημέρωσης, αρχικά ανακτούμε τις 100 συχνότερες λέξεις που περιέχονται στα επιστρεφόμενα αποτελέσματα κάθε ερωτήματος. Για να αποδώσουμε μια στάθμιση για τις λέξεις, προσαρμόζουμε τον τυπικό τύπο του TF-IDF (term frequency-inverse document frequency) που ορίζεται ως εξής:

$$\text{tfidf}(t, c, N) = \text{tf}(t, c) * \text{idf}(t, N)$$

Θεωρώντας $\text{tf}(t, c)$ πόσες φορές εμφανίζεται στα αποτελέσματα η λέξη c . Στη δική μας περίπτωση, η παραλλαγή σε σχέση με τις τυπικές μηχανές αναζήτησης, είναι ότι ο αριθμός των εμφανίσεων αφορά το σύνολο των tweets και όχι κάθε αποτέλεσμα ξεχωριστά, καθώς στο Twitter οι λέξεις συνήθως εμφανίζονται μόνο μία ή δύο φορές στα αποσπάσματα. Η συχνότητα των όρων πολλαπλασιάζεται με το $\text{idf}(t, N) = \log N / \text{dft}$, όπου N ο συνολικός αριθμός των αποτελεσμάτων και dft ο αριθμός των αποτελεσμάτων που εμφανίζεται η λέξη. Θα μπορούσε εναλλακτικά να είχε χρησιμοποιηθεί ως συντελεστής στάθμισης ο αριθμός των εμφανίσεων μιας λέξης στα αποτελέσματα, ωστόσο με την εφαρμογή του IDF επιτυγχάνουμε ισορροπία μεταξύ των συχνών λέξεων που είναι λίγες και που συνήθως κυριαρχούν στο σύνολο των αποτελεσμάτων και των λιγότερο συχνών λέξεων, οι οποίες ωστόσο θα πρέπει να είναι ορατές στο νέφος καθώς μπορεί να αποκαλύψουν ενδιαφέροντα στοιχεία.

Στη συνέχεια δημιουργούμε 2-γράμματα (bigrams), τα οποία τυπικά είναι n -γράμματα (n -grams) για το $n = 2$ (για παράδειγμα, μια φράση όπως "Αφήστε με τώρα" θα οδηγούσε σε διακεκομμένες φράσεις όπως "Αφήστε με", "με τώρα"). Η βασική ιδέα πίσω από αυτό είναι να συλλάβουμε λέξεις που συνυπάρχουν περισσότερο από μία φορά στη συλλογή εγγράφων και να αποφευχθεί η διάσπαση και η διασπορά τους γύρω από το νέφος. Με τον τρόπο αυτό, δίνουμε τη δυνατότητα στους χρήστες

να ενθαρρύνουν τις συσχετίσεις μεταξύ των λέξεων, οι οποίες μπορούν με τη σειρά τους να προκαλέσουν περαιτέρω διερευνήσεις.

N-grams generation pseudocode

```

T=set of results (t) in the index

n=minGram

N=maxGram

NGrams{} → ∅ // hash to store ngrams and their numberofoccurrences

for ti in T do

words() ← Split(ti)

    for i=n to N do

        for j = 0 to length(words) – i do

ngram = “

            for k = j to i do

ngram = Concat(ngram, words(k))

            end for

            if ngram in NGrams

NGrams(ngram) += 1

            Else

NGrams(ngram) = 1

            end for

        end for

    end for

RemoveSingleOccurences(NGrams) //removes ngarms that occur only once

NgramFilter(NGrams) //filters ngrams according to regular expressions

ReturnNGrams

```

Εικόνα 18 Υλοποίηση Αλγορίθμου n-Grams

Ο ψευδοκώδικας που παρατίθεται στην Εικόνα 18 περιγράφει τον αλγόριθμο για την εξαγωγή των n-grams. Ο αλγόριθμος επαναλαμβάνεται για κάθε αποτέλεσμα που ανήκει στο αρχικό σύνολο αποτελεσμάτων, διαιρεί τις λέξεις που περιέχονται στο απόσπασμα αποτελεσμάτων και στη συνέχεια δημιουργεί n-grams με τη σύζευξη

διαδοχικών λέξεων. Στην περίπτωση μας, το minGram και το maxGram έχουν οριστεί στην τιμή 2, καθώς ψάχνουμε μόνο bi-grams. Στην περίπτωση, π.χ., που τροφοδοτούμε τη μέθοδο με minGram = 2 και maxGram = 3, τότε θα πάρουμε bi-grams και 3-grams. Η μέθοδος RemoveSingleOccurrences συνεχίζει με τον έλεγχο της συχνότητας των n-grams, εξαλείφοντας εκείνα που εμφανίζονται μόνο μία φορά στα αποτελέσματα. Τέλος, καλείται η μέθοδος NgramFilter για να φιλτράρει και να καθαρίσει τα n-grams όπως συνέβη και στην περίπτωση μεμονωμένων όρων. Αν ένας όρος του n-gram είναι μη αποδεκτός λόγω των κανόνων βάσει των οποίων απομακρύνουμε το «θόρυβο» από το περιεχόμενο των tweets τότε και το n-gram το ίδιο εξαλείφεται επίσης από τον κατάλογο.

4.3.3 Χώρος Αναζήτησης και Διαφοροποίηση των Αποτελεσμάτων

Η ποικιλομορφία (Denepcke, 2012) μπορεί να υποστηρίξει τις απρόσμενες και ευχάριστες ανακαλύψεις. Αντιμετωπίζουμε την ποικιλομορφία ή διαφορετικότητα με δύο τρόπους: α) διαφοροποιώντας τον χώρο αναζήτησης και β) διαφοροποιώντας τα επιστρεφόμενα τελικά αποτελέσματα αναζήτησης. Διευρύνουμε τον χώρο αναζήτησης, προκειμένου να επιτρέψουμε στο χρήστη να διερευνήσει ένα θέμα από διαφορετικές οπτικές γωνίες υιοθετώντας την τεχνική της επέκτασης ερωτημάτων. Η επέκταση των ερωτημάτων χρησιμοποιήθηκε ευρέως προκειμένου να παραφράσουμε μία ανάγκη πληροφόρησης στην ανάκτηση πληροφορίας. Οι διευρυμένοι όροι μπορεί να προέρχονται από έγγραφα για τα οποία υπάρχει ανατροφοδότηση από το χρήστη, συλλογή εγγράφων-στόχων ή από εξωτερικούς πόρους γνώσης (Baeza-Yates & Ribeiro-Neto, 1999). Όταν ο χρήστης υποβάλει ένα ερώτημα για να διερευνήσει ένα θέμα, ιδιαίτερα ένα θέμα στο οποίο δεν έχει εκτεταμένη προηγούμενη γνώση, πρέπει να αποκτήσει όσο το δυνατόν περισσότερες γνώσεις, αλλά χρειάζεται επίσης να αποκτήσει πληροφορίες σχετικά με τα υποθέματα, στα οποία το αρχικό θέμα μπορεί ενδεχομένως να επεκταθεί. Για το σκοπό αυτό, εφαρμόζουμε επέκταση ερωτήματος, υιοθετώντας την υπηρεσία αναζήτησης

του DBpedia (DBpedia Lookup Service³). Η υπηρεσία DBpedia Lookup Service μπορεί να χρησιμοποιηθεί για την αναζήτηση DBpedia URI με σχετικές λέξεις-κλειδιά. Σχετικές σημαίνει ότι είτε η ετικέτα ενός πόρου είτε το κείμενο που χρησιμοποιήθηκε στο Wikipedia για να αναφερθεί σε έναν συγκεκριμένο πόρο, αντιστοιχούν στο ερώτημα. Στην πράξη, μόλις ο χρήστης υποβάλει ένα ερώτημα, το σύστημα ερωτά την υπηρεσία DBpedia προκειμένου να βρει όλες τις οντότητες που σχετίζονται με το θέμα του ερωτήματος σύμφωνα με τη Wikipedia. Στη συνέχεια, το σύστημα παράγει το νέφος λέξεων όχι μόνο από το αρχικό ερώτημα του χρήστη, αλλά και από όλες τις σχετικές οντότητες που επιστρέφει η DBpedia, αν υπάρχουν. Έτσι, το νέφος λέξεων εμπλουτίζεται περαιτέρω με διαφορετικές απόψεις του ίδιου θέματος.

Λαμβάνοντας επίσης υπόψη ότι η δυνατότητα των χρηστών να αποκλίνουν ενισχύεται κυρίως από την αντίληψη ότι το περιβάλλον και τα στοιχεία είναι διαφορετικά, καθώς και ότι ο τυπικός χρήστης ελέγχει τα πρώτα δέκα έως δεκαπέντε αποτελέσματα προτού προχωρήσει στην αναδιατύπωση του ερωτήματός του, διευρύνουμε τα αποτελέσματα τα οποία προκύπτουν από τα APIs των μηχανών αναζήτησης χρησιμοποιώντας έναν αλγόριθμο που ανακατατάσσει τα αποτελέσματα με στόχο να αποκαλύψει στα πρώτα δέκα αποτελέσματα περιεχόμενο που σχετίζεται με όλες τις πιθανές πτυχές του ερωτήματος. Παράλληλα, η επιθυμία μας δεν είναι να θυσιάσουμε τη συνάφεια για την ποικιλομορφία, δεδομένου ότι η παρουσίαση αξιόπιστων αποτελεσμάτων αυξάνει την εμπλοκή του χρήστη και την εμπιστοσύνη προς το σύστημα αναζήτησης (Kammerer & Gerjets, 2014). Για το λόγο αυτό, επιλέξαμε να χρησιμοποιήσουμε την κανονική έκδοση του πλαισίου μέγιστης οριακής συνάφειας (MMR-Maximal Marginal Relevance framework):

$$\max_{D_i \in \mathcal{A}_S} (\lambda \text{Sim}(D_i, Q) - (1 - \lambda) \max_{D_j \in S} \text{Sim}(D_i, D_j))$$

³ <http://wiki.dbpedia.org/Lookup>

όπου $Sim(D_i, Q)$ είναι η ομοιότητα του εγγράφου σε σχέση με το ερώτημα Q , $Sim(D_i, D_j)$ είναι η ομοιότητα μεταξύ του τρέχοντος εγγράφου και ενός εγγράφου που επιλέχθηκε σε προηγούμενες επαναλήψεις και λ είναι μια παράμετρος που βελτιστοποιεί έναν γραμμικό συνδυασμό των κριτηρίων συνάφειας και ποικιλομορφίας. Όταν η τιμή του λ ισούται με 1 τότε παράγεται η τυπική λίστα κατάταξης με βάση τη συνάφεια ενώ όταν η τιμή του λ ισούται με 0 παράγεται μια λίστα εγγράφων μέγιστης ποικιλομορφίας. Η προτεινόμενη τιμή του λ από τους Carbonell & Goldstein ήταν το 0.3, στην περίπτωση μας καθώς τα αποτελέσματα είναι ήδη ταξινομημένα με βάση τη συνάφεια από την εκάστοτε μηχανή αναζήτησης ή site δημιουργίας περιεχομένου προκειμένου να πετύχουμε μεγαλύτερο βαθμό διεύρυνσης-διαφοροποίησης, η τιμή που προτείνεται στην παρούσα Διατριβή είναι από 0.1 έως 0.2. Θεωρούμε ως έγγραφο την περίληψη που εμφανίζεται σε κάθε αποτέλεσμα του Bing, τον τίτλο στην περίπτωση του Flickr και το περιεχόμενο της σελίδας στην περίπτωση των αποτελεσμάτων που προέρχονται από την εταιρική Πύλη (portal).

Καθώς βασιζόμαστε σε «δημόσια και ανοιχτά» APIs αναζήτησης και γι' αυτό δεν μπορούμε να γνωρίζουμε το πραγματικό σκορ ομοιότητας των αποτελεσμάτων αναφορικά με το ερώτημα, το υπολογίζουμε νοητά με βάση τη θέση του εγγράφου στη λίστα που επιστρέφει το API, ως ακολούθως:

$$Sim(D_i, Q) = \frac{N - Pos(D_i)}{N}$$

όπου $Pos(D_i)$ είναι η θέση του εγγράφου D στη λίστα αποτελεσμάτων του ερωτήματος που επιστρέφει η μηχανή αναζήτησης και N είναι το μέγεθος της λίστας. Για το έγγραφο που κατατάσσεται πρώτο στη λίστα η παράσταση παίρνει την τιμή $Sim(D_i, Q) = 1$, ενώ για το τελευταίο την τιμή $Sim(D_i, Q) = 1/N$. Ο παραπάνω μαθηματικός τύπος έχει προκύψει με το δεδομένο ότι το 1^ο στοιχείο κατέχει τη θέση $Pos(D_1) = 0$ και το τελευταίο την $Pos(D_N) = (N-1)$.

Για να υπολογίσουμε την ομοιότητα μεταξύ των εγγράφων στο σετ αποτελεσμάτων R χρησιμοποιούμε ομοιότητα συνημίτονου:

$$\text{Sim}(D_i, D_j) = \cos(\theta) = \frac{D_i D_j}{\|D_i\| \|D_j\|}$$

όπου θ είναι η γωνία μεταξύ των διανυσμάτων των εγγράφων. Χρησιμοποιούμε το Μοντέλο Διανυσματικού Χώρου (Vector Space Model-VSM) για να αναπαραστήσουμε κάθε έγγραφο σαν ένα διάνυσμα, οι συνιστώσες του οποίου αντιπροσωπεύουν τη σημασία ενός όρου χρησιμοποιώντας τις μετρικές TF-IDF, δεδομένου ενός συνόλου λέξεων που προκύπτουν από τα έγγραφα στο σετ αποτελεσμάτων R . Για τη λεξικογραφική ανάλυση των εγγράφων στο R , δηλαδή την ανάλυση τους σε λεκτικές μονάδες, την αφαίρεση κοινών λέξεων και την εύρεση της ρίζας των λέξεων (tokenization, stop words removal, stemming), χρησιμοποιούμε τον Apache Lucene Standard analyser⁴. Η Εικόνα 19 δείχνει λεπτομερέστερα την εφαρμογή του MMR στο CRUISE. Ο αλγόριθμος ξεκινά με την αναπαράσταση κάθε αποτελέσματος ως διάνυσμα με συνιστώσες όλες τις λέξεις που περιλαμβάνονται στο ευρετήριο και βάρη τον την ομοιότητας κάθε εγγράφου με το ερώτημα. Στη συνέχεια προχωράει με την εφαρμογή του MMR που ξεκινά με την τοποθέτηση του πρώτου εγγράφου στην τελική λίστα των εγγράφων, η οποία θα είναι η διαφοροποιημένη λίστα εγγράφων. Η διαδικασία επαναλαμβάνεται για τα υπόλοιπα έγγραφα και για κάθε ένα υπολογίζει το αποτέλεσμα MMR που αθροίζει τις ομοιότητές του εγγράφου με κάθε ένα από τα έγγραφα που βρίσκονται ήδη στην τελική λίστα. Το έγγραφο με τη μέγιστη περιθωριακή συνάφεια θα καταλάβει τη θέση του επόμενου στοιχείου στην τελική λίστα των εγγράφων. Η διαδικασία συνεχίζεται μέχρι να τοποθετηθούν όλα τα έγγραφα στη λίστα (βλ. υλοποίηση αλγορίθμου στην Εικόνα 19).

⁴ <https://lucene.apache.org/>

MMR Implementation

Input: S = set of documents (D) returned by the search engine, Q =query

Output: Final Ranked list = re-ranked documents

$FinalRank\{\} \rightarrow \emptyset$

$Temp\{\} \rightarrow \emptyset$ // List that holds the candidate documents

for the next rank

For D_i in S **do**

Calculate TF*IDF vector

Calculate Sim(D_i , Q)

end for

while $|FinalRank| < |S|$ **do**

For d_i in S **do**

for d_j in $FinalRank$ **do**

Score _{i} ← Similarity(D_i , D_j)

end for

Score _{i} ← $0,1 * Sim(D_i, Q) - 0,9 * Score_i$

Temp ← { d_i , Score _{i} }

end for

maxScoreDoc ← Call FindMaxScore(Temp)

FinalRank{ } ← {maxScoreDoc}

end while

Return FinalRank

Εικόνα 19 Υλοποίηση Αλγόριθμου Διαφοροποίησης

4.3.4 Παραδίδοντας Απροσδόκητα Αποτελέσματα

Η προσέγγισή μας στοχεύει στην ανακάλυψη απροσδόκητου περιεχομένου από το Twitter υιοθετώντας τον ορισμό από τους Sun, Zhang and Mei (2013), σύμφωνα με τον οποίο οι πληροφορίες που αποκλίνουν από το προφίλ της πηγής ενώ ταυτόχρονα είναι σχετικές και χρήσιμες για τον δέκτη, θεωρούνται απροσδόκητες από τον δέκτη. Τα tweets που προέρχονται από τους χρήστες που ακολουθούμε, τους λεγόμενους ακολουθούμενους (Chen, Nairn, Nelson, Bernstein &

Chi, 2010), βρίσκονται κοντά στα ενδιαφέροντά μας, καθώς απορρέουν από τους λογαριασμούς που επιλέγουμε να ακολουθήσουμε.

Τις περισσότερες φορές, όμως, ο μεγάλος αριθμός εκείνων που ακολουθούμε, παράλληλα με το γεγονός ότι πολλοί από αυτούς συχνά κάνουν tweets καθημερινά, οδηγούν σε πολύ πυκνά χρονολόγια που δεν επιτρέπουν πάντα στους χρήστες να εντοπίζουν χρήσιμες και ενδιαφέρουσες πληροφορίες. Επιπλέον αρκετοί λογαριασμοί κάνουν tweets σχετικά με ένα συγκεκριμένο θέμα αρκετά συχνά, με αποτέλεσμα ο χρήστης να σχηματίζει τελικά μια συγκεκριμένη αντίληψη για τα ενδιαφέροντα του εκάστοτε λογαριασμού (Sun, Zhang & Mei, 2013).

Πρακτικά, για να αξιοποιήσουμε την παραπάνω διατύπωση ακολουθούμε τα παρακάτω βήματα: α) ανακτούμε τους ακολουθούμενους από ένα χρήστη και β) ανακτούμε το χρονολόγιο του καθενός. Τα tweets που ανήκουν στο χρονολόγιο του ακολουθούμενου θα είναι υποψήφια για να προταθούν στο χρήστη όταν εκείνος αναζητά πληροφορία για ένα συγκεκριμένο θέμα.

Στο παράδειγμα του Twitter (paradigm), δέκτης θεωρείται ο χρήστης ο οποίος παρατηρεί ένα tweet ή retweet, ενώ πηγή θεωρείται ο χρήστης που ανήκει στο άμεσο κοινωνικό δίκτυο του δέκτη από τον οποίο ξεκίνησε ένα tweet ή re-tweet.

Προκειμένου να εκτιμηθούν ποιες πληροφορίες θα θεωρηθούν απροσδόκητες από τον χρήστη, πρέπει να διαμορφώσουμε τα προφίλ των ακολουθούμενων και να επιλέξουμε τα tweets που είναι υποψήφια προς πρόταση, εξαιρώντας ή κατατάσσοντας χαμηλότερα στη λίστα προτάσεων εκείνα που προέρχονται από τους λογαριασμούς που κάνουν tweets περισσότερο συχνά για το θέμα ενδιαφέροντος. Τα προφίλ με τη λιγότερη ομοιότητα με το θέμα θεωρούνται οι πιο απροσδόκητες πηγές.

Αξιολογούμε το απροσδόκητο περιεχόμενο λαμβάνοντας υπόψιν δύο παράγοντες: τη συνάφεια και την απόσταση. Στην περίπτωση μας διασφαλίζουμε τη συνάφεια ψάχνοντας tweets σχετικά με το θέμα ενδιαφέροντος. Έπειτα αξιολογούμε

την ομοιότητα του προφίλ του χρήστη από όπου προέρχεται ένα συγκεκριμένο tweet σε σχέση με το θέμα ενδιαφέροντος, το οποίο μπορεί να είναι είτε ένα ερώτημα αναζήτησης, ένα κομμάτι πληροφορίας που ο χρήστης β έχει αλληλοεπιδράσει ή το ίδιο το προφίλ του χρήστη β. Στην περίπτωση μας ο σχηματισμός των προφίλ των ακολουθούμενων είναι προαπαιτούμενο πριν τον υπολογισμό ομοιότητας. Για το σχηματισμό των προφίλ ακολουθούμε το μοντέλο του *συνόλου λέξεων* (bag of words) (Zhang, Jin & Zhou, 2010) κατασκευάζοντας ένα λεξικό σχετικών συχνοτήτων ανά χρήστη.

Για τη λεξικογραφική ανάλυση (tokenization, stop words removal and stemming) των tweets, χρησιμοποιήσαμε τον αναλυτή Apache Lucene Standard. Επιπλέον, εφαρμόσαμε ένα προσαρμοσμένο φίλτρο στα αποτελέσματα για να απομακρύνουμε τις περιττές λέξεις. Επίσης, εφαρμόσαμε κανόνες για τον καθαρισμό δεδομένων, όπως τον αποκλεισμό των ονομάτων χρηστών, των όρων με επαναλαμβανόμενους χαρακτήρες (π.χ. 'booooh', 'looo!' κ.λπ.), συνδέσμους καθώς και τον καθαρισμό λέξεων που περιλαμβάνουν ειδικούς χαρακτήρες.

Τέλος, κάθε σύνολο περιέχει τις λέξεις που περιλαμβάνονται στα tweets του χρήστη-πηγή, μαζί με το πόσες φορές έχουν χρησιμοποιηθεί. Κάθε προφίλ τώρα αντιπροσωπεύεται από ένα n-entry διάνυσμα όπου n ο αριθμός των διακριτών λέξεων στο σύνολο και με τιμές τις αντίστοιχες συχνότητες όρων. Συμπληρωματικά με τα προφίλ, το θέμα ενδιαφέροντος θα εκπροσωπείται επίσης ως διάνυσμα όρων με τον ίδιο τρόπο όπως τα προφίλ. Για να μετρήσουμε την ομοιότητα, επιλέξαμε να εφαρμόσουμε ομοιότητα συνημίτονου στα δύο διανύσματα για να εκτιμήσουμε την απόκλιση ή την εγγύτητα:

$$Sim(U_i, T_j) = \cos(\theta) = \frac{U_i T_j}{\|U_i\| \|T_j\|}$$

όπου U_i διάνυσμα του προφίλ χρήστη, T_j διάνυσμα του θέματος ενδιαφέροντος, θ η γωνία μεταξύ των δύο διανυσμάτων. Έχουμε πειραματιστεί με πολλές τιμές ομοιότητας ως ένα κατώφλι για την ανάδειξη απροσδόκητου και συνάμα σχετικού περιεχομένου. Οι δοκιμές μας έδειξαν ότι μια τιμή ομοιότητας μικρότερη από 0.15

παρήγαγε αρκετά απροσδόκητα, αλλά σχετικά αποτελέσματα. Εφαρμόσαμε φίλτρο στο σύνολο των σχετικών tweets με βάση αυτή την τιμή για να προχωρήσουμε περαιτέρω στην κατάταξη των αποτελεσμάτων. Το όριο αυτό, ωστόσο, εξαρτάται από την αντίληψη κάθε χρήστη και επομένως η αξιολόγηση με μεγάλο αριθμό χρηστών είναι απαραίτητη, ώστε η τιμή αυτή να αποτελέσει και πρόταση με γενικευμένη αξία.

Για την κατάταξη των αποτελεσμάτων εκμεταλλευτήκαμε δύο διαφορετικές αλλά σχετικές πτυχές της κοινωνικής δύναμης. Ο πρώτος ορισμός που ακολουθούμε είναι ότι όσο υψηλότερος ο αριθμός των ακόλουθων, τόσο πιο κοινωνικά ισχυρός είναι ο χρήστης άρα και με μεγαλύτερη επιρροή τα tweets του αν προταθούν. Για κάθε χρήστη στον οποίο ανήκει κάθε tweet από το σύνολο των σχετικών και απροσδόκητων tweets, χρησιμοποιούμε το Twitter API για να ζητήσουμε τον αριθμό των ακολούθων του, ώστε να σχηματίσουμε μία λίστα με το ζεύγος πληροφορίας: χρήστη και αριθμός των ακολούθων του. Στη συνέχεια ταξινομούμε τα tweets σύμφωνα με τον αριθμό των ακολούθων του χρήστη από τον οποίο προέρχονται, προωθώντας στις υψηλότερες θέσεις αυτά που προέρχονται από τους χρήστες με τον υψηλότερο αριθμό ακολούθων.

4.4 Υλοποίηση

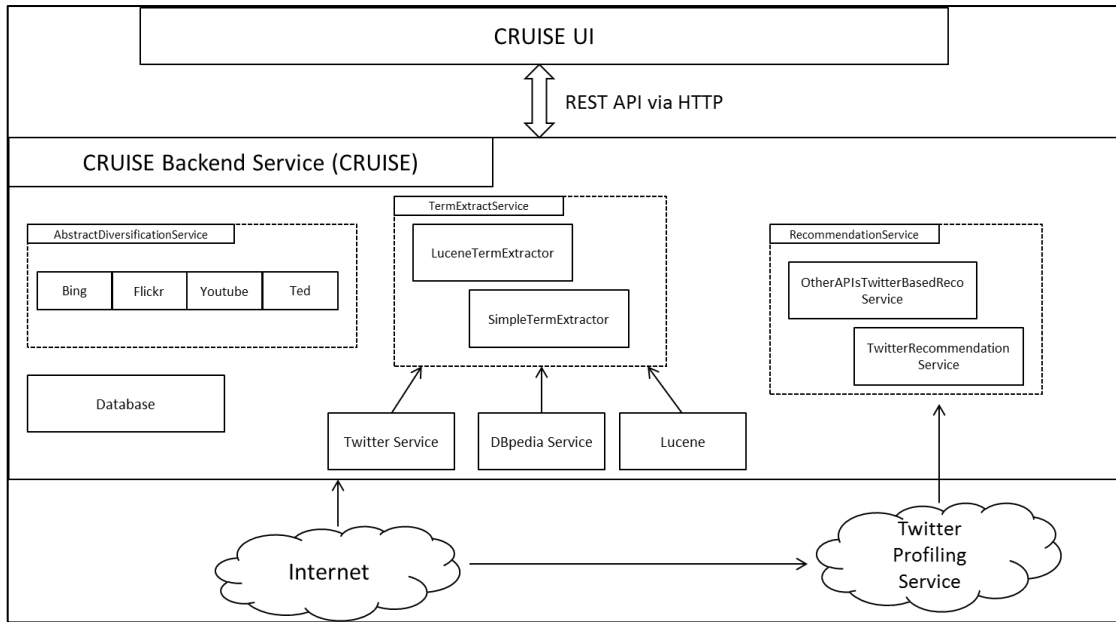
4.4.1 Αρχιτεκτονική Σχεδίαση

Το CRUISE έχει αναπτυχθεί σύμφωνα με την αρχιτεκτονική που παρουσιάζεται στην Εικόνα 20. Ο πυρήνας του συστήματος είναι το συστατικό «Υπηρεσίες», από το οποίο απορρέουν τρεις κύριες λειτουργικότητες. Η πρώτη παρέχει ένα ταξινομημένο σύνολο όρων που εξάγονται από τις ροές των κοινωνικών δικτύων και ταξινομούνται σύμφωνα με τις συχνότητες των όρων τους. Η δεύτερη παρέχει διαφοροποιημένα αποτελέσματα που προκύπτουν από ερωτήματα προς ποικίλες πηγές πληροφοριών, όπως δημόσιες μηχανές αναζήτησης, υπηρεσίες εικόνων ή αποθήκες πληροφορίας ιδιωτικών δικτύων (intranets). Οι πληροφορίες από τις ροές κοινωνικών δικτύων αντιμετωπίζονται με υλοποιήσεις της διεπαφής TermExtractService. Η υπηρεσία

TermExtractService θα πρέπει να ικανοποιεί ένα σύνολο κλήσεων API και να επιστρέφει μια σταθμισμένη λίστα όρων βάσει των συχνοτήτων τους. Η τρέχουσα έκδοση του συστήματος επικεντρώνεται σε πληροφορίες που συλλέγονται από το κοινωνικό δίκτυο Twitter, το οποίο είναι προσβάσιμο μέσω του API αναζήτησης Twitter (Twitter Search API). Επιπλέον, υπάρχουν δύο διαθέσιμες υλοποιήσεις του TermExtractService: το LuceneTermExtractor και το SimpleTermExtractor. Η τρίτη λειτουργικότητα παρέχει διαφορετικούς τύπους προτάσεων προς τους χρήστες, όπως tweets, έγγραφα ιστού, εικόνες και βίντεο.

Το LuceneTermExtractor υπολογίζει τη στάθμιση των όρων στα tweets που παρέχονται από το API αναζήτησης του Twitter, όπως περιγράφεται στην προηγούμενη ενότητα και τα αποθηκεύει σε ένα Ευρετήριο Lucene. Είναι επίσης υπεύθυνο για τον συνδυασμό των όρων σε N-Grams. Ο SimpleTermExtractor χρησιμοποιείται όταν υπάρχουν μόνο λίγα αποτελέσματα και το TF / IDF δεν είναι η κατάλληλη προσέγγιση για τον προσδιορισμό της σημασίας ενός συγκεκριμένου όρου. Για παράδειγμα, αν τα αποτελέσματα που επιστρέφονται από το API του Twitter είναι λιγότερα από 100, τότε επιλέγεται η χρήση του SimpleTermExtractor, λόγω του μικρού κειμένου του εγγράφου και του κειμένου που βρέθηκε μέσα.

Ο SimpleTermExtractor χρησιμοποιεί τη μετρική που παρέχεται από το Lucene προκειμένου να αποδίδει βάρη στους όρους χωρίς να συμπεριλαμβάνονται υπολογισμοί TF/IDF. Η διαφοροποίηση των αποτελεσμάτων από τις πηγές πληροφοριών παρέχεται από τις εφαρμογές της διεπαφής AbstractDiversificationService. Τη στιγμή της παρούσας συγγραφής, υποστηρίζουμε πέντε πηγές πληροφοριών, όπως το Bing, το Flickr, το YouTube, το Twitter και το TED. Κάθε εφαρμογή είναι υπεύθυνη για την ανάκτηση των αποτελεσμάτων των ερωτημάτων από τις επιλεγμένες πηγές πληροφοριών (αποτελέσματα αναζήτησης, εικόνες κ.λπ.), είτε χρησιμοποιώντας διαθέσιμες διασυνδέσεις προγραμματισμού εφαρμογών (API) είτε μέσω αυτόματης ανίχνευσης των διαθέσιμων πληροφοριών (π.χ. screen scraping).



Εικόνα 20 Αρχιτεκτονική CRUISE

Η υπηρεσία `TwitterRecommendationService` ερωτά πρώτα την υπηρεσία δημιουργίας προφίλ Twitter (βλ. Εικόνα 21) για να επιλέξει μόνο τα σχετικά tweets που προέρχονται από το προσωπικό δίκτυο του χρήστη, ενώ είναι υπεύθυνη για την εφαρμογή του αλγόριθμου απροσδόκητων αποτελεσμάτων, προκειμένου να ενισχύσει τα πιο απροσδόκητα tweets από το σετ. Το `OtherAPITwitterBasedRecoService` είναι η υπηρεσία που βασίζεται στο κείμενο των απροσδόκητων tweets που βρέθηκαν από την προηγούμενη υπηρεσία ζητά από άλλα APIS να ανακτήσουν εικόνες ή βίντεο απροσδόκητων εγγράφων. Το CRUISE, καθώς και το CRUISE UI, έχει υλοποιηθεί χρησιμοποιώντας αποκλειστικά βιβλιοθήκες ανοιχτού κώδικα, δεδομένου ότι η ίδια η υλοποίηση είναι ανοιχτή και προέρχεται από την LGPL. Λόγω της φύσης του έργου επιλέξαμε τη Γλώσσα Προγραμματισμού Java και τις προδιαγραφές J2EE ως βάση για την υλοποίηση και επέκταση.

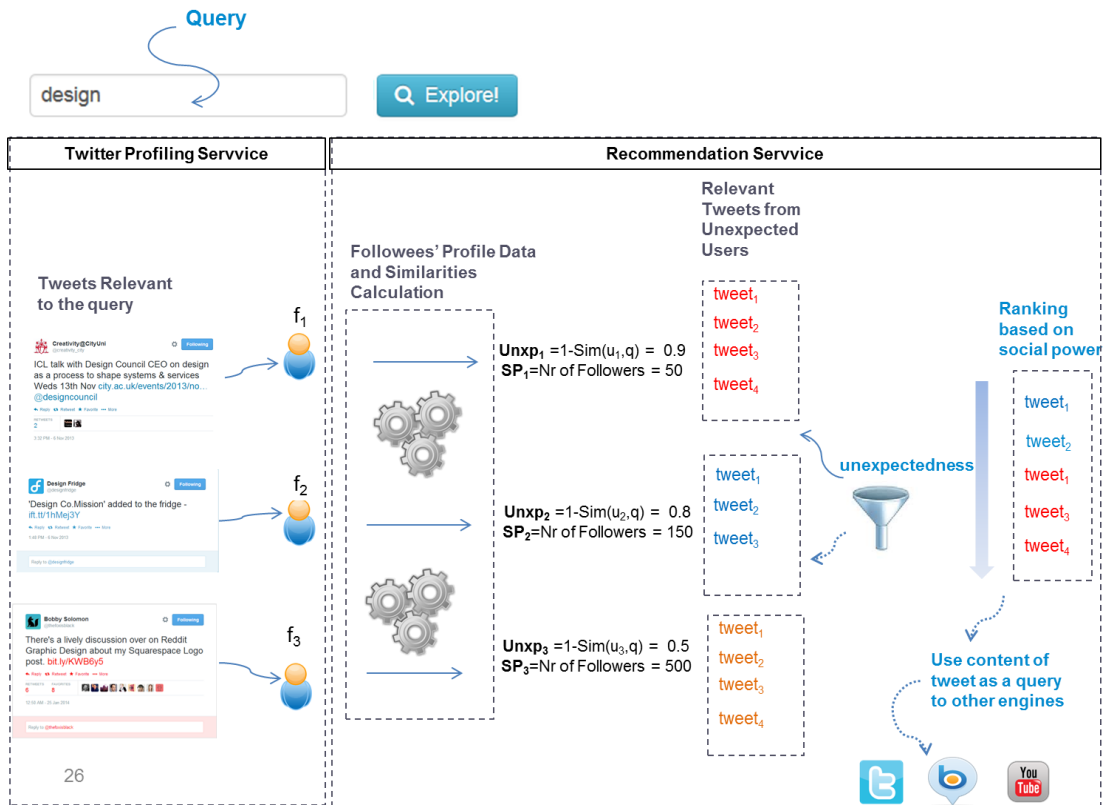
Οι πληροφορίες που προέρχονται από ροές κοινωνικών πληροφοριών, δηλαδή το Twitter στην περίπτωσή μας, συνήθως περιλαμβάνουν «θόρυβο». Προκειμένου να αποφευχθεί η εμφάνιση ακατάλληλων λέξεων στο νέφος λέξεων, εφαρμόσαμε ένα προσαρμοσμένο φίλτρο στα αποτελέσματα χρησιμοποιώντας το `lucene.apache.org`

API για να απομακρύνουμε τις περιττές λέξεις που υπάρχουν στα tweets. Εφαρμόσαμε ακόμα επιπλέον κανόνες για τον καθαρισμό της πληροφορίας, όπως τον αποκλεισμό των ονομάτων χρηστών, των όρων με επαναλαμβανόμενους χαρακτήρες (π'χ. 'b'oo', 'l'ool' κ.λπ.), την αφαίρεση συνδέσμων καθώς και τον «καθαρισμό» λέξεων που περιλαμβάνουν ειδικούς χαρακτήρες.

4.4.2 Σχεδίαση Υπηρεσίας Προτάσεων

Το CRUISE παρέχει τη δυνατότητα στον τελικό χρήστη να συνδεθεί, επιτρέποντάς του να λαμβάνει εξατομικευμένες προτάσεις. Για το σκοπό αυτό, ο χρήστης θα συνδεθεί χρησιμοποιώντας τα προσωπικά του διαπιστευτήρια και στη συνέχεια θα συνδέσει το λογαριασμό του στο Twitter για να επιτρέψει στην εφαρμογή να χρησιμοποιήσει τις πληροφορίες που έχουν ήδη συγκεντρωθεί από την υπηρεσία προφίλ Twitter.

Η υπηρεσία δημιουργίας προφίλ Twitter ανακτά τις σχετικές πληροφορίες από το προφίλ του τρέχοντος χρήστη και το άμεσο προσωπικό του δίκτυο (τους ακόλουθους και τα tweets τους, τους ακολουθούμενους και τα tweets τους). Με αυτό τον τρόπο, προετοιμάζεται το σύνολο δεδομένων για την εφαρμογή του αλγορίθμου για την ανίχνευση απροσδόκητου περιεχομένου. Σε πρώτη φάση, όταν ο χρήστης αναζητά περιεχόμενο σχετικό με ένα θέμα όπως π.χ. «design», το CRUISE ερωτά την υπηρεσία δημιουργίας προφίλ Twitter για να εντοπίσει τα tweets που περιέχουν τη λέξη «design» καθώς και τον ιδιοκτήτη τους. Μόλις βρεθούν όλα τα σχετικά tweets που προέρχονται από τους λογαριασμούς που ακολουθεί ο χρήστης, η Υπηρεσία Προτάσεων είναι υπεύθυνη για την κατηγοριοποίηση του περιεχομένου ως απροσδόκητο ή όχι, μετρώντας πόσο παρόμοιο είναι το προφίλ του κάθε ακολουθούμενου, μεταξύ αυτών που βρέθηκαν, για το ερώτημα 'design'. Η ομοιότητα αντανακλά το πόσο συχνά οι ακολουθούμενοι δημοσιεύουν tweets για το προαναφερθέν θέμα. Σύμφωνα με την εικόνα 21, αφού η ομοιότητα του followee1 με το ερώτημα είναι $Sim(f1,q) = 0.1$ ο βαθμός απροσδόκητου είναι $1-Sim(f1,q) = 0.9$.



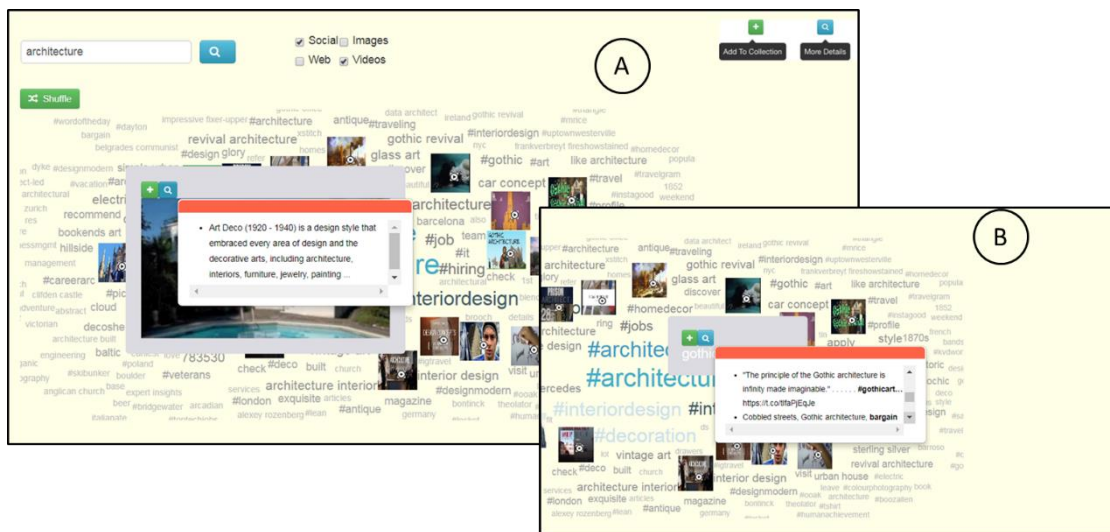
Εικόνα 21 Επισκόπηση Υπηρεσίας Προτάσεων

Μόλις εντοπιστεί το σύνολο των απροσδόκητων χρηστών, τότε γίνεται η κατάταξη λαμβάνοντας υπόψιν και την κοινωνική ισχύ των ακολουθούμενων όπως ορίστηκε προηγουμένως. Όσο μεγαλύτερος ο αριθμός των οπαδών του ακολουθούμενου, τόσο υψηλότερη είναι η θέση του tweet στη λίστα των tweets. Συμπληρωματικά με τα tweets που προσφέρει η Υπηρεσία Προτάσεων, το CRUISE προτείνει εκτός από tweets, ένα σύνολο ποικίλων -όσον αφορά τον τύπο- πόρων που σχετίζονται με το περιεχόμενο των tweets. Αναλυτικότερα, το περιεχόμενο κάθε tweet χρησιμοποιείται ως ερώτημα για τα Bing, Youtube, Flickr APIs για να ανακτηθούν δυνητικά ευχάριστα και απρόσμενα, άρθρα, βίντεο και εικόνες από το διαδίκτυο.

4.4.3 Παρουσίαση Εργαλείου & Σενάρια Χρήσης

Παρακάτω παρουσιάζουμε το CRUISE, το οποίο υποστηρίζει την αναζήτηση σε δημοφιλή κοινωνικά μέσα και μηχανές αναζήτησης Ιστού. Η διερευνητική

προσέγγιση του CRUISE βασίζεται στη «μεταφορά του νέφους» (cloud metaphor), η οποία κατασκευάζεται σε πραγματικό χρόνο και απεικονίζει τα αποτελέσματα αναζήτησης (λέξεις, όροι, εικόνες, βίντεο και άρθρα διαδικτύου), όπου το μέγεθος κάθε στοιχείου αποτελεσμάτων εξαρτάται από τη δημοτικότητά του. Με το CRUISE, οι χρήστες μπορούν να κάνουν κλικ σε οποιοδήποτε στοιχείο αποτελέσματος για να αποκτήσουν περισσότερες πληροφορίες διαβάζοντας ή παρακολουθώντας σχετικά tweets, άρθρα και βίντεο) (Εικόνα 22).



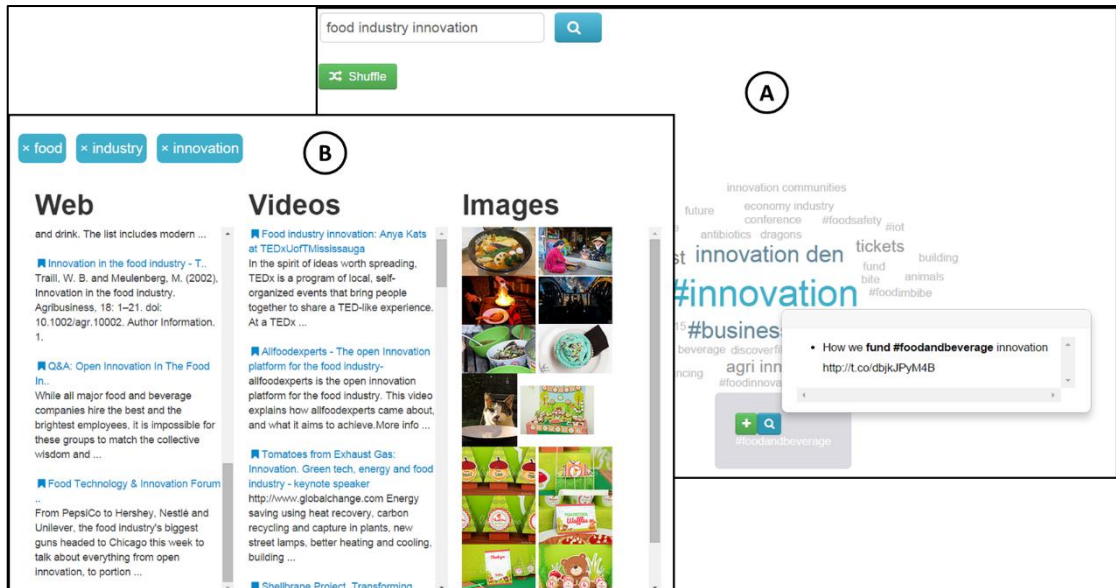
Εικόνα 22 Το Cruise με αναδυόμενα παράθυρα που εμφανίζουν επιπλέον πληροφορία όταν ο χρήστης κάνει κλικ είτε σε όρους είτε σε video.

Επιπλέον, οι χρήστες μπορούν να κάνουν πολλαπλές επιλογές αντικειμένων που εμφανίζονται στο νέφος. Αυτή η αλληλεπίδραση έχει ως αποτέλεσμα τη διαμόρφωση μιας συλλογής αγαπημένων αντικειμένων, η οποία επιτρέπει στους χρήστες να διερευνήσουν περαιτέρω και να ανακτήσουν νέα αποτελέσματα. Οι χρήστες είναι επίσης σε θέση να πατήσουν το κουμπί Shuffle, που τους δίνει τη δυνατότητα να αποκαλύψουν όρους που προηγουμένως ήταν κρυμμένοι λόγω της χαμηλής προβολής τους σε σχέση με άλλους όρους στο σύνολο αποτελεσμάτων.

Σενάριο Χρήσης A

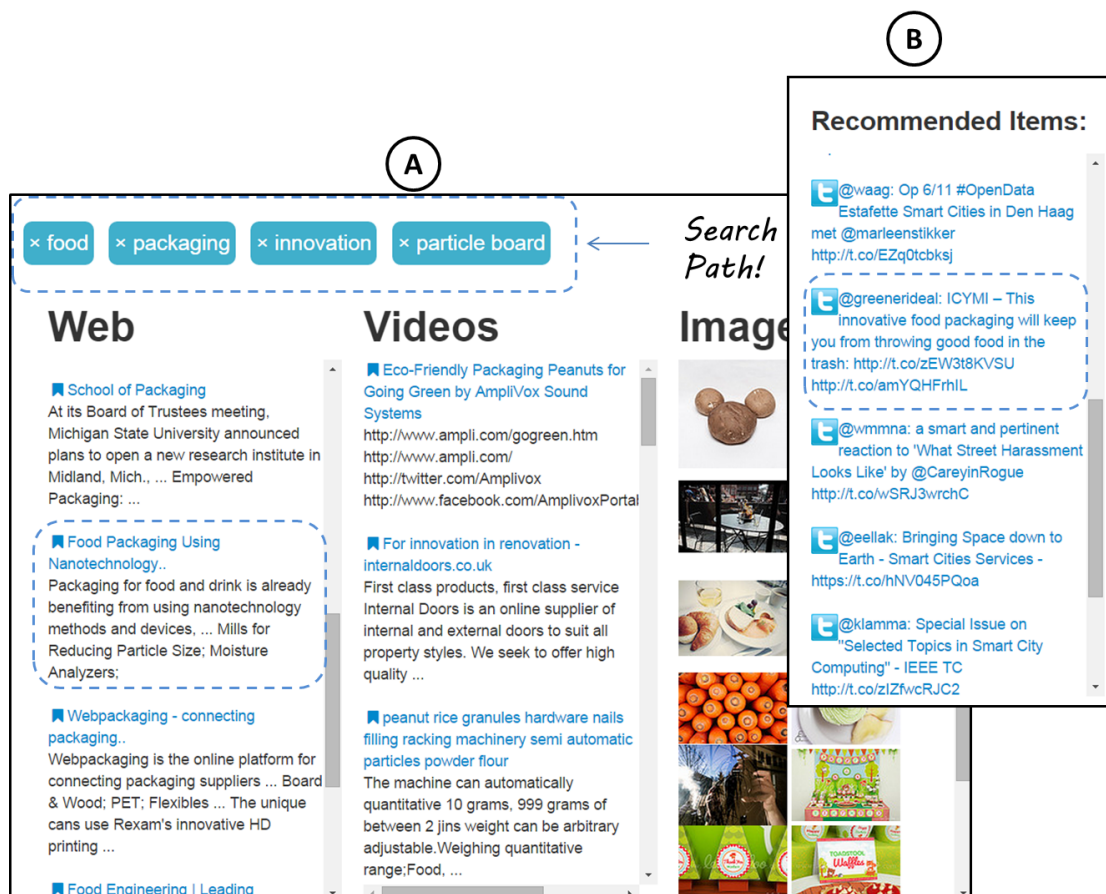
Παρουσιάζουμε ένα απλό παράδειγμα που εξηγεί τη λειτουργικότητα του εργαλείου και το σχεδιασμό αλληλεπίδρασης που συνοδεύει μια εργασία διερευνητικής αναζήτησης. Ας πάρουμε για παράδειγμα ένα *σχεδιαστή εννοιών* (concept designer) που εργάζεται σε ένα έργο, το οποίο σχετίζεται με τις μελλοντικές προκλήσεις και την ανοιχτή καινοτομία στη βιομηχανία τροφίμων. Ο κύριος στόχος του είναι η διερεύνηση του πεδίου, ώστε να παρουσιάσει νέες ιδέες. Ο σχεδιαστής ξεκινά μια διερεύνηση χρησιμοποιώντας το ερώτημα «καινοτομία στη βιομηχανία τροφίμων».

Δημιουργείται νέφος λέξεων με όρους που εξάγονται από τα πιο δημοφιλή tweets που σχετίζονται με το ερώτημα. Όροι όπως "#foodandbeverage" και "#foodsafety" εμφανίζονται στο νέφος λέξεων. Ο σχεδιαστής εννοιών περνάει το ποντίκι του πάνω από αυτούς τους όρους (hovering), χωρίς να κλικάρει, για να εξετάσει πώς σχετίζονται με το ερώτημά του (Εικόνα 23A). Βρίσκει tweets σχετικά με την «έξυπνη συσκευασία τροφίμων και ποτών», tweets σχετικά με την «έξυπνη συσκευασία για την ασφάλεια των τροφίμων», καθώς και tweets σχετικά με «το πώς η ρομποτική, η εκτύπωση 3D και η νανοτεχνολογία κάνουν τη βιομηχανία πιο έξυπνη». Συνεχίζει εξετάζοντας τα αποτελέσματα στο κάτω μέρος της οθόνης, ώσπου βρίσκει ένα ενδιαφέρον άρθρο σχετικά με την «ανοιχτή καινοτομία» στη βιομηχανία τροφίμων που περιγράφει την ανάγκη των εταιρειών τροφίμων να βελτιώσουν τη γεύση, όπως και τη διάρκεια ζωής των προϊόντων τους, καθώς και την ευκολία μέσω έξυπνης συσκευασίας (Εικόνα 23B).



Εικόνα 23 Διεπαφή Χρήστη CRUISE

Αναφέρει ότι η ανάγκη για έξυπνη συσκευασία τροφίμων αποτελεί μείζονα πρόκληση για τη βιομηχανία τροφίμων και συνεχίζει πραγματοποιώντας ένα δεύτερο ερώτημα, αντικαθιστώντας τη λέξη 'industry' με τη λέξη 'packaging', καθιστώντας το ερώτημα πιο συγκεκριμένο μετασηματίζοντάς το σε «καινοτομία συσκευασίας τροφίμων». Νέες λέξεις βρίσκονται διάσπαρτες στο νέφος λέξεων και η φράση «μοριοσανίδα» προσελκύει την προσοχή του. Το tweet αφορά σε «μια βιώσιμη πηγή τροφίμων που μπορεί να δέσει τούβλα, μοριοσανίδες & συσκευασίες». Θέλει να μάθει περισσότερα για 'αυτό και συνεχίζει κάνοντας κλικ στο 'particle board' για να το προσθέσει στο «μονοπάτι αναζήτησης» (Εικόνα 24A). Η ενέργεια αυτή ενεργοποιεί αυτόματα μια νέα αναζήτηση στους διαθέσιμους χώρους αναζήτησης, με ερώτημα το συνδυασμό {food packaging innovation particle board}. Βρίσκει συναφή άρθρα, όπως "Μανιτάρια αντικαθιστούν φορμαλδεΐδες στην παραγωγή μοριοσανίδων" ή "τεχνικές συσκευασίας τροφίμων με τη χρήση νανοτεχνολογίας". Ταυτόχρονα, παρατηρεί ένα προτεινόμενο tweet από το προσωπικό του δίκτυο (Εικόνα 24B) σχετικά με τις έξυπνες συσκευασίες που εμποδίζουν τη ρίψη καλής τροφής συνειδητοποιώντας μια ακόμη πτυχή του τρόπου με τον οποίο η καινοτομία συσκευασίας τροφίμων επηρεάζει όχι μόνο τις εταιρείες αλλά και τους καταναλωτές.



Εικόνα 24 Ο χρήστης διατυπώνει ένα νέο μονοπάτι αναζήτησης και εξετάζει τα σχετικά αποτελέσματα. Δημιουργείται ένα νέο σύνολο προτάσεων.

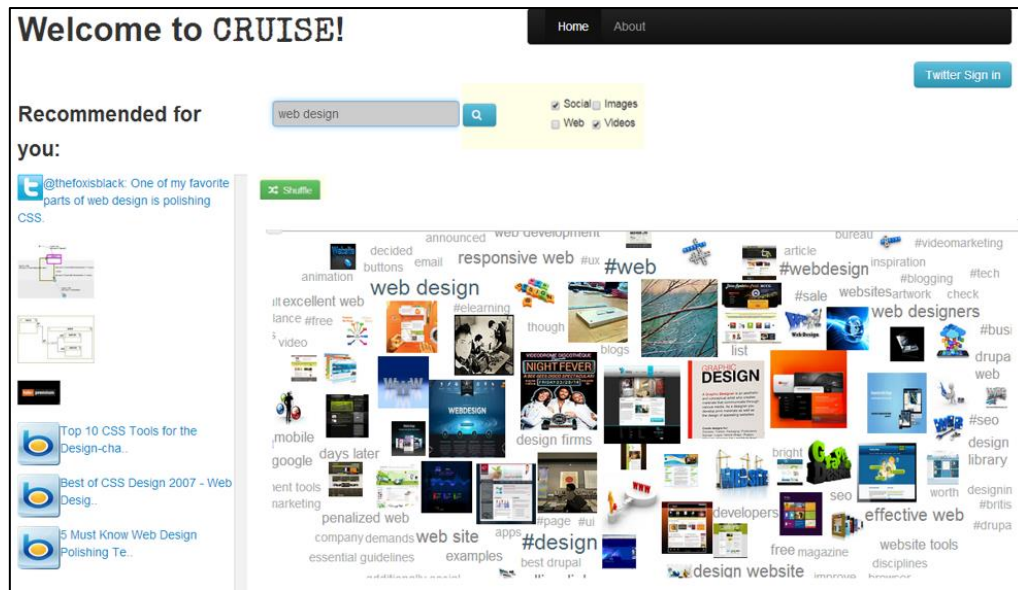
Σενάριο B

Το συγκεκριμένο σενάριο παρουσιάζει ένα παρόμοιο σενάριο χρήσης, όπου όμως απεικονίζεται πώς μπορούν να εμπλουτιστούν οι προτάσεις με εικόνες και άρθρα και άλλα μέσα εκτός από tweets.

Ένας σχεδιαστής εννοιών που εργάζεται στις καθημερινές του δημιουργικές αναθέσεις χρησιμοποιεί το CRUISE για να διερευνήσει τις τάσεις στο σχεδιασμό ιστοσελίδων. Ξεκινά μια διερευνητική συνεδρία ψάχνοντας με τον όρο 'web design'. Λαμβάνει το τυπικό νέφος λέξεων με αποτελέσματα από τη γενική ροή του twitter εμπλουτισμένο επίσης και με εικόνες. Έχει επίσης τη δυνατότητα να λάβει πιο εξατομικευμένα αποτελέσματα που προέρχονται από το προσωπικό του δίκτυο στο twitter. Με την είσοδο στο twitter -μέσω του CRUISE- ο χρήστης μπορεί να πάρει πολύ

πρόσφατα αποτελέσματα ακόμη και από το χρονολόγιο του twitter, τα οποία εμφανίζονται με τη μορφή όρων στο νέφος. Επιπλέον, μπορεί να λάβει προτάσεις που σχετίζονται με τον «σχεδιασμό ιστοσελίδων» που προέρχονται από το προσωπικό του δίκτυο. Ο σχεδιαστής εννοιών ακολουθεί ήδη συγκεκριμένους λογαριασμούς πλατφορμών σχεδιασμού όπως @ropurcity, @FastCoDesign, @designmilk. Όταν χρησιμοποιεί το twitter ξεχωριστά από το CRUISE, παρατηρεί συνήθως τα πιο πρόσφατα tweets από τους λογαριασμούς πλατφορμών σχεδιασμού που ακολουθεί λόγω της συχνότητας του tweeting σχετικά με το συγκεκριμένο θέμα. Η υπηρεσία προτάσεων από το Twitter ανακτά όλα τα tweets που σχετίζονται με το «σχεδιασμό ιστοσελίδων» και υπάρχουν στο χρονολόγιο όσων ακολουθεί. Ανακτώνται tweets που σχετίζονται με το 'web design' από το @ropurcity, @FastCoDesign, @designmilk και άλλους λογαριασμούς. Δεδομένου ότι αυτοί οι λογαριασμοί είναι πολύ παρόμοιοι με τον «σχεδιασμό ιστοσελίδων», δεν ψηφίζονται ως πρόταση. Το σύστημα ευνοεί τελικά ένα tweet που είναι σχετικό και δυνητικά χρήσιμο, το οποίο προέρχεται από έναν λογαριασμό @thefoxisblack με χαμηλή ομοιότητα με το ερώτημα, αλλά υψηλή κοινωνική δύναμη. Μόλις ο χρήστης λάβει τα tweets σχετικά με το σχεδιασμό που προέρχονται από έναν λογαριασμό διαφορετικό από τους συνήθεις λογαριασμούς πλατφόρμας σχεδιασμού, εξάπτεται το ενδιαφέρον του (αριστερό παράθυρο της εικονας 25). Παρατηρεί επίσης τις υπόλοιπες προτάσεις που βασίζονται στο περιεχόμενο του tweet, αλλά είναι διαφορετικού τύπου.

Εκτός από την έκπληξη που αποδίδεται στην προέλευση του tweet, ο σχεδιαστής εννοιών βρίσκει μερικά από τα συνιστώμενα αντικείμενα ενδιαφέροντα που συμπληρώνουν το σύνολο των αντικειμένων που εμφανίζονται στο νέφος. Για παράδειγμα, προτείνεται ένα ενδιαφέρον εργαλείο τεχνολογίας σχετικό με το σχεδιασμό ιστοσελίδων. Ο σχεδιαστής εννοιών μπορεί να κάνει κλικ στην πρόταση για να πάρει περισσότερες πληροφορίες σχετικά με το εργαλείο.



Εικόνα 25 Ο χρήστης λαμβάνει εικόνες και λέξεις στο νέφος αντικειμένων σχετικών με το ερώτημά του αλλά και προτάσεις από tweets, εικόνες και άρθρα Ιστού

4.5 Αξιολόγηση

Ο κύριος στόχος της πειραματικής αξιολόγησης ήταν να απαντήσει στο 1^ο και 2^ο ερευνητικό ερώτημα και πιο συγκεκριμένα: να επιβεβαιώσει εάν το CRUISE είναι υποκειμενικά καλύτερο για τη μόχλευση της εύνοιας τυχαίων ανακαλύψεων και της έμπνευσης στη διαδικασία αναζήτησης, καθώς και της δημιουργικότητας στο αποτέλεσμα σε σχέση με τις τυπικές διεπαφές που βασίζονται σε ερωτήματα. Επιπλέον, στόχος ήταν να συλλέξουμε ανατροφοδότηση για τη βελτίωση της χρηστικότητας της εφαρμογής CRUISE. Με την πιλοτική μελέτη, επιδιώξαμε να προσδιορίσουμε την ικανότητα του CRUISE να βοηθήσει επαγγελματίες «σχεδιαστές εννοιών» που ασχολούνται με δημιουργική εργασία και να αξιολογήσουμε την ικανότητα της προτεινόμενης προσέγγισης να υποστηρίξει τη δημιουργική διαδικασία (μέρος του 2^{ου} ερευνητικού ερωτήματος).

4.5.1 Μετρικές Δημιουργικότητας και Εύνοιας Τυχαίων Ανακαλύψεων

Δεν ζητήσαμε από τους συμμετέχοντες να αξιολογήσουν άμεσα την εύνοια τυχαίων ανακαλύψεων, καθώς πολύ λίγοι είναι εκείνοι οι οποίοι γνωρίζουν τον όρο ή έχουν κατανοήσει την έννοια. Αξιολογήσαμε την εύνοια τυχαίων ανακαλύψεων στη διαδικασία αναζήτησης, χρησιμοποιώντας τη μετρική που πρότειναν οι Ge, Delgado-Battenfeld και Jannach (2010) για την αξιολόγησή της σε συστήματα προτάσεων, που βασίζεται στη μέση αναλογία μη αναμενόμενων και χρήσιμων προτεινόμενων αντικειμένων στη λίστα προτάσεων:

$$SRDP = \sum_u \frac{UNEXP \cap USEFUL}{N}$$

όπου u ο αριθμός των χρηστών που αξιολογούν τις προτάσεις και N το μέγεθος της λίστας προτάσεων. Στη δική μας περίπτωση το u αντιπροσωπεύει τους χρήστες που αξιολογούν τα αποτελέσματα αναζήτησης και το N αντιπροσωπεύει τη λίστα αποτελεσμάτων αναζήτησης.

Ως μετρική δημιουργικού αποτελέσματος επιλέξαμε τα κριτήρια που προτάθηκαν από τους Shah, Smith και Vargas-Hernandez (2003). Η συγκεκριμένη μετρική και άλλες όμοιές της αναφέρθηκαν λεπτομερώς στην ενότητα 2.3.6. Επιλέξαμε να διατηρήσουμε τον όρο πρωτοτυπία στη θέση της καινοτομίας προκειμένου να μη συγχέεται με την έννοια της καινοτομίας που συζητήθηκε στην ενότητα 4.2 και έτσι τα κριτήρια δημιουργικότητας που τελικά υιοθετήσαμε είναι: α) η **πρωτοτυπία**, β) η **ποιότητα**, γ) η **ποσότητα** και δ) η **ποικιλία**.

Στην αθροιστική αξιολόγηση, επιδιώξαμε επίσης να αξιολογήσουμε το CRUISE ανεξάρτητα από το δημιουργικό αποτέλεσμα εστιάζοντας στο κατά πόσο η προσέγγισή μας μπορεί να υποστηρίξει τη δημιουργική διαδικασία. Για αυτό το σκοπό χρησιμοποιήσαμε το Δείκτη Υποστήριξης Δημιουργικότητας (CSI) (Cherry and Latulipe, 2014) που αξιολογεί την ποσοτικοποιημένη επίδραση στους ακόλουθους παράγοντες δημιουργικής διαδικασίας: (α) **τα αποτελέσματα άξιζαν την προσπάθεια** – δηλαδή τα δημιουργικά αποτελέσματα που δημιουργήθηκαν δικαιολογούν την προσπάθεια που απαιτείται για τη χρήση του εργαλείου, (β)

εκφραστικότητα - ο βαθμός εκφραστικότητας και δημιουργικότητας που επιτυγχάνεται με το εργαλείο, (γ) **διερεύνηση** - η ευκολία με την οποία μπορείτε να εξερευνήσετε πολλαπλές επιλογές, ιδέες, σχεδιασμό και αποτελέσματα χωρίς κουραστική, επαναλαμβανόμενη αλληλεπίδραση, (δ) **δέσμευση** - το επίπεδο εμπλοκής του χρήστη και ροής (flow) που επιτυγχάνεται κατά τη χρήση του εργαλείου. (ε) **διαφάνεια εργαλείου** - ο βαθμός στον οποίο η διεπαφή ήταν σχεδόν διαφανής κατά τη χρήση του εργαλείου (στ) **συνεργασία** - η ευκολία με την οποία ο χρήστης μπορούσε να εργάζεται με άλλους χρησιμοποιώντας το εργαλείο.

4.5.2 Εργαστηριακές Μελέτες

Σχεδιάσαμε δύο ελαφρώς διαφορετικά πειράματα, λαμβάνοντας υπόψη τις διάφορες ήδη δημοσιευμένες πειραματικές διαδικασίες για την αξιολόγηση της δημιουργικότητας, όπως ζητώντας από τους συμμετέχοντες να βρουν πρωτότυπες ιδέες για ανοιχτά προβλήματα, εργασίες με αναπαραστάσεις προβλημάτων που χρειάζονται να αναδομηθούν, παραγωγή δημιουργικών ιστοριών, μελωδιών, πινάκων (Arden, Chavez, Grazioplene & Jung, 2010, Dietrich & Kanso, 2010). Και τα δύο πειράματα αφορούσαν τη χρήση του συστήματος CRUISE και ενός βασικού συστήματος. Το βασικό σύστημα ήταν ένα τυπικό σύστημα αναζήτησης και ανάκτησης πληροφορίας με βάση τα ερωτήματα, στο οποίο οι χρήστες μπορούσαν να εκφράσουν την ανάγκη τους για πληροφόρηση μόνο με την πληκτρολόγηση ερωτημάτων και τα αποτελέσματα παρουσιάζονταν ως λίστα ιστολογίων, βίντεο και εικόνων ιστού. Το βασικό σύστημα δεν χρησιμοποίησε ούτε την εξαγωγή όρων από κοινωνικές συζητήσεις και την απεικόνιση τους, ούτε την επέκταση ερωτημάτων, τη διερεύνηση της πορείας αναζήτησης, τη διαφοροποίηση των πόρων και τις κοινωνικές προτάσεις. Τα APIs μηχανών αναζήτησης ήταν τα ίδια και στα δύο συστήματα. Στο πρώτο πείραμα ακολουθήθηκε ο σχεδιασμός μεταξύ των υποκειμένων και το CRUISE αξιολογήθηκε ως ένα σύστημα για την παροχή έμπνευσης και διευκόλυνσης της ανακάλυψης απρόσμενων και χρήσιμων αποτελεσμάτων σε μια ερευνητική εργασία. Στο δεύτερο ακολουθήσαμε το σχεδιασμό εντός των

αντικειμένων και το CRUISE κρίθηκε ως μέσο για την παροχή έμπνευσης και πρόκλησης δημιουργικής παραγωγής ιδεών σε καθημερινά θέματα ευρύτερου ενδιαφέροντος.

Στο πρώτο πείραμα, συμμετείχαν 20 διδακτορικοί φοιτητές (12 άνδρες και 8 γυναίκες) από το πανεπιστήμιο μας. Η εργασία που τους ανατέθηκε ήταν να διατυπώσουν μία πρόταση για διδακτορική διατριβή στον τομέα του "cloud computing" (Πίνακας 6).

Πίνακας 6 Περιγραφή πρώτου Πειράματος

Experiment Description
<p>Suppose you have to suggest an idea that will be the subject of a new doctorate in the area of cloud computing * Take just 15 minutes to search for articles that will give you inspiration to think the idea with the following tool.</p> <p>*Cloud computing is a model that allows an easy way to access a farm computing resources (networks, servers, storage spaces, applications) with fast way and little effort or supervision. Main characteristics of cloud computing are:</p> <ul style="list-style-type: none">Organized with storage standardsKeeping backupsAbroad application for easy data exchange between industriesSaving computing resources and energy

Οι συμμετέχοντες ήταν τόσο άνδρες όσο και γυναίκες, περίπου στο ίδιο επίπεδο γνώσης -καθώς ήταν όλοι στο ίδιο έτος της διδακτορικής τους διατριβής στην Επιστήμη Υπολογιστών- και κανένας από αυτούς δεν εκπονούσε διατριβή στον τομέα του cloud computing, ώστε να διασφαλιστεί η διερευνητική συμπεριφορά με τη

χρήση των παρεχόμενων εργαλείων. Οι μισοί συμμετέχοντες πραγματοποίησαν την εργασία χρησιμοποιώντας την εφαρμογή CRUISE και οι άλλοι μισοί χρησιμοποιώντας το σύστημα αναφοράς. Αναθέσαμε τυχαία τους συμμετέχοντες στην ομάδα εστίασης και ελέγχου, προκειμένου να μειώσουμε το ρίσκο από μεταβλητές συγχύσεως των αποτελεσμάτων, όπως τα χαρακτηριστικά της προσωπικότητας (για παράδειγμα η "περιέργεια", το "πόσο ανοιχτός είναι κάποιος στην διερεύνηση" κ.λπ.). Προκειμένου να μειώσουμε την επίδραση της προσδοκίας των συμμετεχόντων, οι συμμετέχοντες και των δύο ομάδων δεν γνώριζαν ποιο σύστημα ήταν το αναφοράς και ποιο το CRUISE. Περιορίσαμε το διαθέσιμο χρόνο ολοκλήρωσης της εργασίας σε 15 λεπτά για να διασφαλίσουμε ότι οι συμμετέχοντες αναζητούσαν ενεργά κατά τη διάρκεια του πειράματος, καθώς επίσης ότι θα είχαν ίσο χρόνο για να ολοκληρώσουν την εργασία. Σκοπός μας ήταν να δείξουμε ότι η έκθεση σε χαρακτηριστικά όπως (α) οι *οπτικοποιημένες ενδείξεις* από την κοινωνική συζήτηση, (β) η συμμετοχή σε μια *διαδραστική διαδικασία*, (γ) η *ανακατάταξη των αποτελεσμάτων αναζήτησης*, η (δ) *πολυμορφία* των πηγών και ε) το *απροσδόκητο* ενθαρρύνει την ανακάλυψη και έμπνευση σε σύγκριση με τις διαδικασίες άμεσης αναζήτησης. Τα δύο συστήματα παρουσιάστηκαν στους συμμετέχοντες και των δύο ομάδων λίγο πριν την έναρξη του πειράματος. Οι συμμετέχοντες κλήθηκαν να αξιολογήσουν τα αποτελέσματα αναζήτησης της έρευνάς τους (περιορισμένη στα 10 κορυφαία αποτελέσματα) για την καινοτομία (αν ένας δεδομένος πόρος παρουσιάζει μια νέα πτυχή), για τη συνάφεια (τα αποτελέσματα ήταν σχετικά βοηθώντας τον χρήστη να διευκρινίσει την ιδέα του σχετικά με το θέμα), τη χρησιμότητα (τα αποτελέσματα παρείχαν τουλάχιστον μία πιθανή απάντηση στο θέμα) και το βαθμό απροσδόκητου (αν αυτός ο πόρος ήταν σχετικός με την ανάθεση, αλλά δεν σχετιζόταν άμεσα και δεν αναμενόταν να βρεθεί). Δεν περιοριστήκαμε μόνο στα κριτήρια που σχετίζονται με τη χρησιμότητα και την απροσδόκητη εμφάνιση των αποτελεσμάτων, με την υπόθεση ότι δίνοντας περισσότερες επιλογές στους χρήστες θα μας επέτρεπε να αποκτήσουμε μια πιο ακριβή εικόνα της αντίληψής τους, αποφεύγοντας κάθε μεροληψία εκ μέρους μας που θα μπορούσε να οδηγήσει σε συμπεράσματα υπέρ της εύνοιας των τυχαίων ανακαλύψεων

Μετά την ολοκλήρωση του πειράματος, ζητήσαμε από δυο καθηγητές να αξιολογήσουν την πρωτοτυπία των προτάσεων για διατριβή και την ποιότητά τους με βάση τη συνάφειά τους στον τομέα του cloud computing, την εφικτότητά τους και τη λεπτομέρειά τους στη διατύπωση. Άλλες μετρικές που σχετίζονται με τα δημιουργικά αποτελέσματα, όπως η ποσότητα και η ποικιλία ιδεών, δεν εφαρμόστηκαν σε αυτό το πείραμα, καθώς οι συμμετέχοντες κλήθηκαν να διατυπώσουν μόνο μία πρόταση για έρευνα. Υποθέσαμε ότι οι συμμετέχοντες χρήστες που είχαν χρήσιμες και απρόσμενες ανακαλύψεις θα εμπνέονταν και θα μπορούσαν να διαμορφώσουν δημιουργικές ιδέες. Προκειμένου να ουδετεροποιήσουμε τυχόν προκατάληψη των καθηγητών για το σύστημα, δεν τους παρείχαμε πληροφορίες σχετικά με το σύστημα από το οποίο προέρχονται οι απαντήσεις. Για την αξιολόγηση της αξιοπιστίας της συμφωνίας μεταξύ των καθηγητών, χρησιμοποιήσαμε το στατιστικό *karrra* του Cohen (Cohen, 1960). Η μέθοδος μπορεί να εφαρμοστεί σε δεδομένα που δεν ακολουθούν κανονική κατανομή, ακόμα και δυαδικά (όχι / ναι), αλλά ταιριάζει καλύτερα σε κανονική κλίμακα, όπως η κλίμακα βαθμολόγησης 1 έως 5. Επειδή οι αξιολογήσεις του πειράματος ακολουθούν την κλίμακα 1 έως 5, χρησιμοποιήσαμε το σταθμισμένο *karrra*, καθώς έπρεπε να ληφθεί υπόψη η απόσταση από τη συμφωνία (Cohen, 1968). Τέλος, ζητήσαμε από τους συμμετέχοντες χρήστες να απαντήσουν σε ένα ερωτηματολόγιο που βασίστηκε στο μοντέλο ResQue, ένα πλαίσιο αξιολόγησης που βασίζεται στο χρήστη, το οποίο προτάθηκε από τους Chen και Pu (2010), προκειμένου να μας βοηθήσει να αποκαλυφθούν τυχόν προβλήματα χρηστικότητας και να εντοπιστούν πιθανές βελτιώσεις. Το ResQue επιλέχθηκε καθώς μετράει συγκεκριμένα τις ικανότητες αλληλεπίδρασης και την έκφραση προτιμήσεων που προσφέρει ένα σύστημα.

Στο δεύτερο πείραμα συμμετείχαν 94 μεταπτυχιακοί φοιτητές (72 άνδρες και 22 γυναίκες). Σε όλους δόθηκε ένας κατάλογος έξι ερωτήσεων από τις οποίες έπρεπε να επιλέξουν μία και στη συνέχεια να ολοκληρώσουν μια αποστολή τριών βημάτων προσπαθώντας να προσεγγίσουν την ανάθεση δημιουργικά. Τους αφήσαμε να επιλέξουν την ερώτησή αναλόγως των προσωπικών τους ενδιαφερόντων και των

βασικών γνώσεών τους για να αποφύγουν να έχουν οι συμμετέχοντες μια ερώτηση που βρίσκεται έξω από τον τομέα ενδιαφέροντός τους. Στο πρώτο βήμα της αποστολής, κλήθηκαν να σημειώσουν οποιεσδήποτε ιδέες θα μπορούσαν να σκεφτούν χωρίς τη χρήση ψηφιακών εργαλείων. Στο δεύτερο βήμα, κλήθηκαν να χρησιμοποιήσουν το σύστημα αναφοράς για να τους βοηθήσει να βρουν νέες ιδέες ή απαντήσεις στην ερώτηση. Παρομοίως, στο τρίτο βήμα, τους ζητήθηκε να παράγουν όσο το δυνατόν περισσότερες ιδέες χρησιμοποιώντας το CRUISE. Ομοίως όπως και στο πρώτο πείραμα, το CRUISE παρουσιάστηκε στους συμμετέχοντες λίγο πριν την έναρξη της μελέτης. Τους εξηγήσαμε ότι οι ιδέες τους θα μπορούσαν να προέρχονται από τα ευρήματά τους ή από τη χρήση των ευρημάτων τους ως ερεθίσματα. Η συνολική διάρκεια της ανάθεσης καθορίστηκε περίπου στα 30-40 λεπτά, προτείνοντάς τους να δαπανήσουν 5-10 λεπτά για την πρώτη εργασία, 10 λεπτά για κάθε μία από τις επόμενες δύο εργασίες και 5-10 για το ερωτηματολόγιο. Με αυτόν τον τρόπο, διασφαλίσουμε ότι η κατανομή του χρόνου τους έγινε ανάλογα με τις απαιτήσεις του συγκεκριμένου έργου και ότι συμμετείχαν παρομοίως σε όλα τα στάδια της διαδικασίας. Οι συμμετέχοντες κλήθηκαν να αξιολογήσουν τη συνολική τους εμπειρία με το τελευταίο εργαλείο που χρησιμοποιήθηκε (δηλ. το CRUISE) κατά τη διάρκεια των περιόδων διερεύνησης τους σχετικά με παραμέτρους όπως καινοτομία, συνάφεια, χρησιμότητα και απροσδόκητο των αποτελεσμάτων. Τους ζητήσαμε επίσης να αξιολογήσουν τη συμβολή του «νέφους λέξεων» στην παραγωγή νέων ιδεών καθώς και τη συμβολή των πηγών (ιστολόγια, βίντεο και εικόνες) τα αποτελέσματα των οποίων ήταν απόρροια της διαμόρφωσης του «μονοπατιού αναζήτησης» κατά την αλληλεπίδρασή τους με το νέφος λέξεων. Από εκείνους τους συμμετέχοντες που είχαν συνδεθεί στην υπηρεσία προτάσεων άρα και εκμεταλλεύτηκαν τις προτάσεις, ζητήθηκε επιπλέον να αξιολογήσουν πόσο βοηθήθηκαν από τα προτεινόμενα αντικείμενα στην παραγωγή των ιδεών τους. Επιπλέον ζητήσαμε, από όλους τους συμμετέχοντες να απαντήσουν (α) σε ποιο βαθμό οι ιδέες τους προήλθαν κατευθείαν από τα αποτελέσματα αναζήτησης ή τα προτεινόμενα στοιχεία και (β) σε ποιο βαθμό, οι ιδέες τους ήταν ένας συνδυασμός της φαντασίας τους ή των βασικών γνώσεων που είχαν πριν χρησιμοποιήσουν το

σύστημα. Οι βαθμολογίες τους θα μας επέτρεπαν να εκτιμήσουμε σε ποιο βαθμό το CRUISE θα μπορούσε να συμβάλει στην γνωστική τους ανάπτυξη επάνω στο επιλεγμένο θέμα ή αν οι απαντήσεις τους στο τρίτο βήμα της ανάθεσης προήλθαν κυρίως από την επίδραση του «φαινομένου της εκμάθησης» που συμβαίνει λόγω έρευνας στο θέμα σε προηγούμενα βήματα της ανάθεσης. Προσπαθήσαμε να εκτιμήσουμε εάν το CRUISE παρείχε αυξημένες ευκαιρίες για ευχάριστα και αναπάντεχα ευρήματα και αν τέτοιου είδους ανακαλύψεις-ευρήματα ενισχύουν την παραγωγή ιδεών με αυξημένο δημιουργικό περιεχόμενο, αξιολογώντας την **πρωτοτυπία**, την **ποιότητα** την **ποσότητα** και την **ποικιλία των ιδεών**. Σε αυτό το πλαίσιο, ζητήσαμε από δύο εμπειρογνώμονες να αξιολογήσουν τις ιδέες που δημιούργησαν οι ερωτηθέντες όσον αφορά τις τέσσερις προαναφερθείσες μετρικές. Οι εμπειρογνώμονες εξέτασαν διεξοδικά τις απαντήσεις και απέκλεισαν από το σύνολο δεδομένων τις απαντήσεις 12 συμμετεχόντων που τις έκριναν ως ανεπαρκείς ή εντελώς άσχετες. Τρεις εξ αυτών εξαιρέθηκαν από τη διαδικασία αξιολόγησης, διότι στο πρώτο και δεύτερο βήμα δεν προέκυψε καμία ή μόνο μία ιδέα.

4.6 Αποτελέσματα

4.6.1 Πρώτη Εργαστηριακή Μελέτη

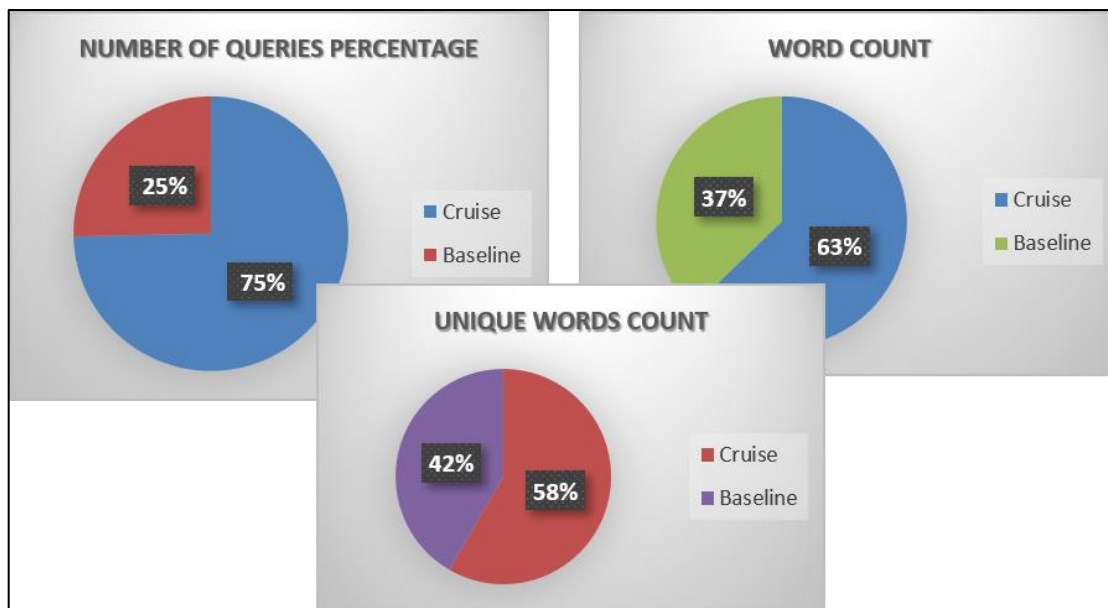
Τα ποσοτικά αποτελέσματα από την πρώτη εργαστηριακή μελέτη (βλέπε επίσης Πίνακα 7) ευνόησαν το CRUISE από την άποψη της επάρκειάς να υποστηρίξει τους χρήστες να εκφράσουν τις ανάγκες τους ($m = 5$, $std\ dev = 0.7$) και να εξοικειωθούν με τη χρήση του ($m = 4.7$, $std\ dev = 0.4$). Όσον αφορά την ποιότητα του περιεχομένου, οι χρήστες έτειναν να συμφωνήσουν ότι το CRUISE προσφέρει πολύτιμο, νέο και απροσδόκητο περιεχόμενο ($m = 4.2$, $std\ dev = 0.8$) με ισχυρή εμπιστοσύνη στη δυναμική του να τους εμπνεύσει ($m = 4.7$, $std\ dev = 0.4$).

Τα δεδομένα που καταγράφηκαν από τις αλληλοεπιδράσεις με το σύστημα CRUISE και το σύστημα αναφοράς, συμπεριλαμβανομένων των clicks για την εμφάνιση λεπτομερειών των αποτελεσμάτων που εμφανίστηκαν, των όρων που

Πίνακας 7 Μέσος Όρος, Τυπική Απόκλιση, και Πηλίκιο των απαντήσεων του ResQue βάσει κλίμακας από το 1 (Διαφωνώ Απόλυτα) έως το 5 (Συμφωνώ Απόλυτα).

Evaluation Aspect	Mean	Std	Mode
Quality of Results (Novel, Unexpected, Ser/pitous, Diverse)	4.2	0.8	4
Interaction Adequacy (to express, revise needs)	4	0.7	4
Interface Adequacy (attractive, adequate layout)	4	0.7	4
Perceived Ease of Use (familiarity with the tool)	4.8	0.4	5
Perceived Usefulness (supported in exploring)	4.7	0.5	5
Control/Transparency (control to express preferences)	4.2	0.6	4
Attitudes (confidence of getting inspired)	4.8	0.4	5
Behavioral Intention (use again, tell my friends)	4.8	0.4	5

εμφανίστηκαν στο σύστημα, των clicks επάνω σε όρους, των ερωτημάτων που εκτελέστηκαν στη «μηχανή» και των αξιολογήσεων των αποτελεσμάτων από τους χρήστες, αποκάλυψαν ότι οι χρήστες του CRUISE πραγματοποίησαν περισσότερα ερωτήματα ($m = 1.8$) σε σχέση με τους χρήστες του συστήματος αναφοράς ($m = 1.14$). Η διαφορά ήταν στατιστικά σημαντική ($p < 0.05$) (Εικόνες 26-27-28).



Εικόνα 26 Η απόδοση των αναζητητών κατά τη διάρκεια της Διερεύνησης

cloud computing optimisation entrepreneurship social governance
 brokeragerevalue mobile telecom aspects enterprise interoperability creationcloudcomputing
 creationeconomic challenges workshop business trends services analytics creation lte mobility value recommender
 systems sla living optimisationbrokeragerevalue infographic cybersecurity interaction design computingtop10 brokerage
 applications institutional institution 2013 computingsocial drive service marketing manual advancements future equitisation platform
 economic

Εικόνα 27 Νέφος Λέξεων που απεικονίζει τις λέξεις και τη συχνότητα στο σύνολο των ερωτημάτων στο CRUISE

cloud computing video processing research on grid bioinformatics with federated agenda
 security ontology mapreduce topics the annotate analysis 2013 workshop subgraph safety in distributed densest
 mapreducevideo enterprise interaction business trends design semantic cideo processingnull annotation

Εικόνα 28 Νέφος Λέξεων που απεικονίζει τις λέξεις και τη συχνότητα στο σύνολο των ερωτημάτων στο σύστημα αναφοράς.

Τα δεδομένα που οι χρήστες παρείχαν χρησιμοποιώντας τα δύο συστήματα εντοπίστηκαν σε 64 συνεδρίες. Ένας αριθμός 34 συνεδριών αντιστοιχήθηκε στους χρήστες που χρησιμοποίησαν το CRUISE, ενώ οι υπόλοιπες αντιστοιχήθηκαν στο σύστημα αναφοράς. Απορρίψαμε τυχαία τέσσερις συνεδρίες με το CRUISE, για να

εξισορροπηθούν τα αποτελέσματα από πλευράς αριθμού στα δύο συστήματα. Όπως φαίνεται στον Πίνακα 8, ο υπολογισμένος μέσος όρος ανά συνεδρία έδειξε ότι οι

Πίνακας 8 Επισκόπηση των μέσων των βαθμολογιών των επιστρεφόμενων αποτελεσμάτων για το CRUISE και το σύστημα αναφοράς

Criterion	Mean per session
Novel	CRUISE: 1.58 Baseline: 1.23
Useful	CRUISE: 2.35 Baseline: 2.00
Unexpected	CRUISE: 2.18 Baseline: 0.78
Unexpected and Useful	CRUISE: 0.09 Baseline: 0.03

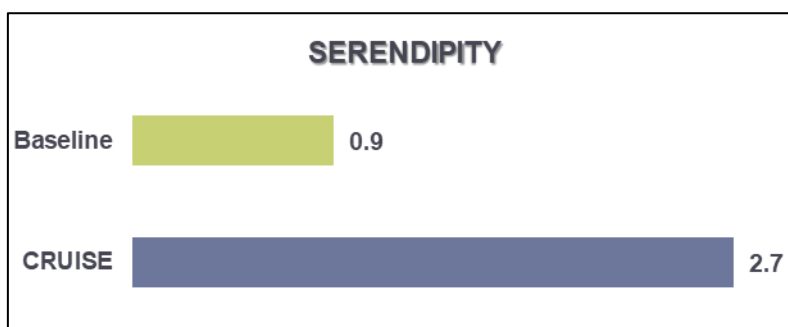
χρήστες του CRUISE εντόπισαν περισσότερο απροσδόκητα ($m_{\text{CRUISE}} = 2.18$; $m_{\text{Baseline}} = 0.78$) και χρήσιμα ($m_{\text{CRUISE}} = 2.35$; $m_{\text{Baseline}} = 2.00$) αποτελέσματα σε σχέση με τους χρήστες του συστήματος αναφοράς. Καταγράψαμε επίσης τα στοιχεία που αξιολογήθηκαν τόσο χρήσιμα όσο και απροσδόκητα και στα δύο συστήματα για να υπολογίσουμε μία προσέγγιση της «εύνοιας τυχαίων ανακαλύψεων» σύμφωνα με τον ορισμό των Ge et al. (2010). Τα ευρήματα αποκάλυψαν ότι το CRUISE επικρατεί αναφορικά με την εύρεση απροσδόκητων και χρήσιμων αντικειμένων ανά συνεδρία σε σύγκριση με το σύστημα αναφοράς ($m_{\text{CRUISE}} = 0.09$, $m_{\text{Baseline}} = 0.03$). Στον Πίνακα 9, μια μέτρηση της «εύνοιας τυχαίων ανακαλύψεων» απεικονίζεται για το σύστημα αναφοράς και το CRUISE όπως υπολογίστηκε λαμβάνοντας υπόψη όλες τις συνεδρίες για κάθε σύστημα. Για να μετρήσουμε την «εύνοια τυχαίων ανακαλύψεων», προσαρμόσαμε την εξίσωση που γνωρίσαμε νωρίτερα

$$SRDP = \sum_u \frac{UNEXP \cap USEFUL}{N}$$

αντιστοιχίζοντας το μ στον αριθμό των συνεδριών και το N στον αριθμό των αποτελεσμάτων αναζήτησης που ανακτήθηκε και παρουσιάστηκε στο χρήστη ανά συνεδρία.

Πίνακας 9 Ποσοτική εκτίμηση της Εύνοιας Τυχαίων Ανακαλύψεων στα επιστρεφόμενα αποτελέσματα για το CRUISE και το Σύστημα Αναφοράς βάσει της εξίσωσης που παρατέθηκε στην ενότητα 4.5.1

System	Serendipity
CRUISE	2.7
Baseline	0.9



Εικόνα 29 Συγκριτική Απεικόνιση Εύνοιας Τυχαίων Ανακαλύψεων στα δύο υπό αξιολόγηση συστήματα

Συνολικά, το πείραμα έδειξε ότι με το CRUISE, οι χρήστες βρήκαν πιο χρήσιμα, καινοφανή και απροσδόκητα αποτελέσματα και κατ'επέκταση πιο ευχάριστες και αναπάντεχες ανακαλύψεις από ότι με το σύστημα αναφοράς. Αυτά τα ευρήματα δικαιολογούνται από την παρατήρηση ότι οι συμμετέχοντες που χρησιμοποίησαν το CRUISE ασχολήθηκαν με πιο επιμηκυμένες συνεδρίες χρησιμοποιώντας τους όρους που οπτικοποιούνται μέσω του νέφους λέξεων - όπως αποκάλυψαν τα καταγεγραμμένα στοιχεία - γεγονός που οδήγησε στην εύρεση πιο ενδιαφερόντων άρθρων, πράγμα το οποίο συνάδει και με τις αξιολογήσεις τους. Λαμβάνοντας υπόψη την αξιολόγηση των δημιουργικών αποτελεσμάτων από τους καθηγητές, διαπιστώσαμε ότι η προσέγγισή μας είχε κάποια επίδραση σε σχέση με την ποιότητα

(p-value: 0.05) και την πρωτοτυπία (p-value: 0.01) (Πίνακας 10), που δείχνει ότι οι χρήστες που είχαν πρόσβαση σε πόρους μέσω του CRUISE ήταν σε θέση να διατυπώσουν τις ιδέες τους καλύτερα από εκείνους που δεν το χρησιμοποίησαν. Μια πιθανή εξήγηση για αυτό το εύρημα θα μπορούσε να είναι ότι με το CRUISE οι χρήστες είχαν πρόσβαση σε διάφορους πόρους και κατάφεραν να εντοπίσουν περιεχόμενο καλύτερης ποιότητας από ό, τι με το σύστημα αναφοράς. Επιπλέον, το CRUISE πιθανότατα καθοδήγησε τους χρήστες προς πιο στοχευμένες αναζητήσεις με αποτέλεσμα την εξεύρεση καινοτόμων ιδεών.

Πίνακας 10 Ανάλυση των προτάσεων διδακτορικών διατριβών από δύο καθηγητές ως προς: α) την ποιότητα και β) την πρωτοτυπία

Metric	CRUISE	Baseline
Quality mean	3.8	2.7
Quality STD	1.4	0.9
Originality mean	3.8	2.4
Originality STD	0.6	1.2

Αναφορικά με την αξιοπιστία των βαθμολογιών των δύο καθηγητών, ο πίνακας 11 αναφέρει τον συνολικό αριθμό των βαθμολογιών όπου υπήρχε συμφωνία, την αντίστοιχη στατιστική μέτρηση του *kappa*, την τυπική τιμή σφάλματος καθώς και το εύρος του διαστήματος εμπιστοσύνης. Ο Cohen (1960) πρότεινε το αποτέλεσμα *kappa* να ερμηνευτεί ως εξής: οι τιμές ≤ 0 δεν δείχνουν συμφωνία οι 0.01-0.20 από καθόλου ως ελάχιστη συμφωνία, 0.21-0.40 ως αποδεκτές, 0.41-0.60 ως μέτριες, 0.61-0.80 ουσιαστικές και 0.81-1.00 ως σχεδόν τέλεια συμφωνία. Παρατηρούμε ότι με 40 επιλεγμένα θέματα και σύμφωνα με τα προτεινόμενα όρια από την Cohen, η τιμή *kappa* ($k = 68.18\%$) θεωρείται σημαντική, αν και το εύρος του διαστήματος εμπιστοσύνης είναι ευρύ για να θεωρηθεί η τιμή του *Kappa* ιδιαίτερα ακριβής και αυτό επηρεάζεται από δύο παράγοντες : (α) το μικρό αριθμό των συμμετεχόντων β)

τον μη δυαδικό χαρακτήρα των αξιολογήσεων του καθηγητή. Παρόλα αυτά, δεδομένης της φύσης της μελέτης και του γεγονότος ότι οι βαθμολογητές ακολούθησαν μια κλίμακα βαθμολόγησης 1 έως 5, ένα ποσοστό διαφωνίας 31,82% είναι αποδεκτό, και η τιμή αυτή του *kappa* μεταφράζεται ως ότι είτε το 31,82% των δεδομένων δεν είναι αξιόπιστα είτε ότι είναι λανθασμένα.

Πίνακας 11 Συμφωνία μεταξύ των καθηγητών στην αξιολόγηση της ποιότητας και της πρωτοτυπίας.

Number of Raters	Number of Agreements	kappa	Standard Error	CI 95%
2	32	68.18%	12.16%	44.35% to 92%

4.6.2 Δεύτερη Εργαστηριακή Μελέτη

Στη δεύτερη μελέτη, οι συμμετέχοντες ήταν σε θέση να επιλέξουν το προτιμώμενό τους θέμα τους ανάμεσα σε τέσσερις ερωτήσεις. Ο Πίνακας 12 δείχνει τον αριθμό των συμμετεχόντων που επέλεξαν κάθε ερώτηση. Υπάρχει σαφής προτίμηση για τις ερωτήσεις σχετικά με τη «δημιουργική εκπαίδευση», τον «καθαρισμό του νερού» και την «3d εκτύπωση».

Πίνακας 12 Δημοτικότητα Ερωτήσεων Μεταξύ Συμμετεχόντων

Question	No of participants
How to use 3d printing in everyday life?	20
How to use energy in sustainable ways?	10
How to make education more creative?	27
How to purify water?	25

Σε αυτό το πείραμα επιχειρήσαμε να δημιουργήσουμε μια τιμή αναφοράς για την εύνοια τυχαίων ανακαλύψεων στα πλαίσια του CRUISE με την προσαρμογή του

ορισμού των Ge, Delgado-Battenfeld και Jannach (2010). Θεωρήσαμε ότι η αντιληπτή εύνοια τυχαίων ανακαλύψεων για το CRUISE ορίζεται από το άθροισμα των εκτιμήσεων του χρήστη που σχετίζονται με το απροσδόκητο και τη χρησιμότητα. Ο Πίνακας 13 παρουσιάζει τον μέσο όρο, την τυπική απόκλιση και το mode των απαντήσεων των συμμετεχόντων σχετικά με το πόσο καινοτόμα, χρήσιμα, απροσδόκητα και ευχάριστα και αναπάντεχα (serendipitous) ήταν τα αποτελέσματα που βρέθηκαν με τη χρήση του CRUISE. Αξίζει να σημειωθεί ότι οι συμμετέχοντες

Πίνακας 13 Ο μέσος όρος, η τυπική απόκλιση και το mode αποκρίσεων για το CRUISE με βάση μια κλίμακα Likert πέντε βαθμών από το 1 (διαφωνώ απόλυτα) μέχρι το 5 (συμφωνώ απόλυτα). Η αντιληπτή δυνατότητα για εύνοια τυχαίων ανακαλύψεων κανονικοποιήθηκε από το 1 έως το 5

Criterion	Mean	StDev	Mode
Novelty	4.2	0.8	4
Usefulness	4.28	0.94	5
Unexpectedness	3.95	0.89	4
Perceived Serendipitousness (Unexpected and Useful)	4.3	0.48	4.6

βρήκαν τα αποτελέσματα που το CRUISE επέστρεψε νέα και ιδιαίτερα χρήσιμα για την εργασία τους. Επιπλέον, τα αποτελέσματα δείχνουν ότι το CRUISE βοήθησε τους περισσότερους συμμετέχοντες να κάνουν απροσδόκητες συσχετίσεις μεταξύ ιδεών που δεν είχαν μελετηθεί πριν και τελικά η έμμεση αξιολόγηση της εύνοιας τυχαίων ανακαλύψεων διατηρείται επίσης σε υψηλά επίπεδα ($m=4.3$, $stdev=0.48$). Όσον αφορά το πρώτο κριτήριο αξιολόγησης ενός αποτελέσματος ως δημιουργικού, το οποίο αφορά την παραγωγικότητα των ιδεών σε κάθε βήμα της ανάθεσης, ο Πίνακας 14 απεικονίζει τη μέση ποσότητα ιδεών που παράγονται με «κανένα εργαλείο», με το «σύστημα αναφοράς» και με το «CRUISE». Παρατηρούμε ότι τόσο το σύστημα αναφοράς όσο και το CRUISE είναι σε θέση να υποστηρίξουν τους

χρήστες στη δημιουργία επιπλέον ιδεών συγκριτικά με εκείνες που θα παρήγαγαν χωρίς ψηφιακή υποστήριξη.

Αναμφισβήτητα, η συμβολή διαφόρων πηγών, διαδικτυακών εγγράφων, βίντεο και εικόνων διαδραματίζει καθοριστικό ρόλο στην ενίσχυση της παραγωγικότητας ιδεών, από τη μη ύπαρξη ψηφιακής βοήθειας. Σημειώστε ότι η μέση ποσότητα ιδεών που παράγονται χρησιμοποιώντας το CRUISE ($m = 3$) σε σύγκριση με τη μέση ποσότητα ιδεών που παράγονται χρησιμοποιώντας το σύστημα αναφοράς ($m = 4$) μειώνεται, αλλά αυτό συμβαίνει πιθανώς επειδή οι συμμετέχοντες που ολοκλήρωσαν και τις τρεις αναθέσεις («κανένα εργαλείο», «σύστημα αναφοράς» και «CRUISE») είναι οι ίδιοι, εκτελώντας τα διαδοχικά. Φαίνεται λογικό ότι έκρηξη στην παραγωγικότητα ιδεών θα συνέβαινε στο δεύτερο βήμα της ανάθεσης, στην οποία οι συμμετέχοντες χρησιμοποίησαν για πρώτη φορά ψηφιακό μέσο, το οποίο ήταν το σύστημα αναφοράς, ενώ το CRUISE χρησιμοποιήθηκε τελευταίο και αφού τους είχε ζητηθεί η παραγωγή ιδεών πρώτα χωρίς ψηφιακή βοήθεια και έπειτα με την υποστήριξη του βασικού συστήματος. Ωστόσο, οι συμμετέχοντες ήταν σε θέση να παράγουν επιπλέον ιδέες χρησιμοποιώντας το CRUISE, αν και με μειωμένο ρυθμό. Η πλειοψηφία των συμμετεχόντων ανέφερε ότι «με το CRUISE μπορούσαν να βρουν ακόμα τα περισσότερα από τα ευρήματα του δεύτερου βήματος, αλλά και ότι ήταν το CRUISE που τους ενέπνευσε να βρουν νέα ερωτήματα και να δημιουργήσουν νέους σύνθετους συνδυασμούς λέξεων που διαμόρφωσαν νέα «μονοπάτια αναζήτησης», τα οποία συνετέλεσαν στην αποκάλυψη ενδιαφέροντος υλικού, γεγονός που τους βοήθησε να παραγάγουν νέες ιδέες. Επιπλέον, δεν υπάρχουν αξιοσημείωτες διαφορές όσον αφορά τις μέσες τιμές ποικιλίας των ιδεών που παράγονται μεταξύ των τριών βημάτων της ανάθεσης, αν και φαίνεται ότι με το σύστημα αναφοράς ($m = 3.1$) και με το CRUISE ($m = 2.88$) οι συμμετέχοντες ήταν πιο πιθανό να παραγάγουν ιδέες που ανήκουν σε διαφορετικές πτυχές του θέματος από ότι με «κανένα εργαλείο» ($m = 2.23$) (Πίνακας 14).

Πίνακας 14 Ανάλυση των ιδεών αναφορικά με την πρωτοτυπία, ποιότητα, ποσότητα και ποικιλία με βάση τις βαθμολογίες των εμπειρογνομόνων

Metric	CRUISE	Baseline	No System
	Mean Std	Mean Std	Mean Std
Originality	3.62 1.9	3.4 1.82	3.1 1.54
Quality	2.8 1.21	0.9 0.87	1.56 1.2
Quantity	3 2.1	4 2.4	3 2.32
Variety	2.88 1.73	3.1 1.8	2.23 1.4

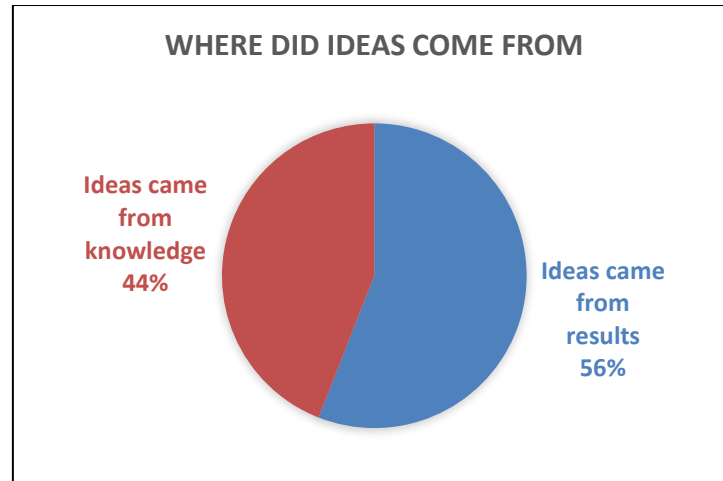
Λαμβάνοντας υπόψη το γεγονός ότι το CRUISE διατηρεί τις ίδιες πηγές πληροφορίας με το σύστημα αναφοράς, μπορούμε να υποστηρίξουμε ότι οι παράγοντες που προωθούν την παραγωγικότητα, αλλά και την ποιότητα των ιδεών είναι οι λέξεις που εμφανίζονται στο νέφος λέξεων, η δημιουργική ανακατανομή των αποτελεσμάτων που κατέστη δυνατή από τον αλγόριθμο διαφοροποίησης, καθώς και το προτεινόμενο περιεχόμενο από την υπηρεσία προτάσεων για όσους χρήστες είχαν τη δυνατότητα να την εκμεταλλευθούν. Ο Πίνακας 15 συνοψίζει τη συμβολή της λειτουργικότητας κάθε εργαλείου έτσι όπως έγινε αντιληπτή στους συμμετέχοντες, δίνοντας ένα μικρό προβάδισμα στη συνιστώσα «πηγές», το οποίο ενδεχομένως δικαιολογείται από το γεγονός ότι ο κύριος όγκος πληροφοριών θα μπορούσε να βρεθεί εκεί, με μεγαλύτερη λεπτομέρεια και σε διάφορες μορφές (όπως π.χ. η μορφή των άρθρων). Η οπτικοποίηση των ενδείξεων με τη χρήση του «νέφους λέξεων» από την άλλη πλευρά θεωρείται επίσης ένας ισχυρός παράγοντας έμπνευσης με ($m = 3.7$). Παρατηρούμε επίσης ότι στην αντίληψη των χρηστών η συνεισφορά του νέφους λέξεων είναι ισάξια με αυτή της υπηρεσίας προτάσεων, πιθανώς λόγω του γεγονότος ότι και τα δύο στοιχεία εκμεταλλεύονται πληροφορίες που προέρχονται από το Twitter.

Πίνακας 15 Η συνεισφορά των λειτουργιών του CRUISE (δηλ. νέφος λέξεων, είδος πηγών, προτάσεις από το twitter) στην έμπνευση, όπως έγινε αντιληπτή από τους χρήστες.

Element	Mean	StDev	Mode
Word Cloud	3.7	0.95	4
Sources	3.93	0.88	4
SocialMedia Recommendations	3.75	0.95	3

Για να αξιολογηθεί κατά πόσο οι ιδέες που παράχθηκαν χρησιμοποιώντας το CRUISE στο τρίτο βήμα της ανάθεσης ήταν πράγματι το αποτέλεσμα του εργαλείου και όχι συνέπεια της διαδικασίας του πειράματος (δηλαδή οι συμμετέχοντες έγιναν πιο έμπειροι κατά τη διάρκεια του πειράματος), ζητήσαμε από τους συμμετέχοντες τα σχόλιά τους για το «πώς οι ιδέες τους προέκυψαν στο τρίτο βήμα της ανάθεσης», «ήταν κατ'ευθείαν από τα αποτελέσματα αναζήτησης» ή «από τη φαντασία/προηγούμενο γνωστικό υπόβαθρο» (βλ. περιγραφή πειράματος) (Εικόνα 30). Ο μέσος αριθμός αξιολογήσεων για το αν «οι ιδέες ήρθαν από τα αποτελέσματα αναζήτησης ($m = 4.6$) σε σχέση με το αν «οι ιδέες προέκυψαν από τα αποτελέσματα αναζήτησης και τη φαντασία / προηγούμενο γνωστικό υπόβαθρο» ($m = 3.62$), δείχνει ότι η «επίδραση της δοκιμής» (testing effect) που προκύπτει από την έρευνα σχετικά με το θέμα πριν από τη χρήση του CRUISE, παρόλο που δεν είναι αμελητέα, δεν επηρεάζει σε μεγάλο βαθμό την εγκυρότητα της υπόθεσής μας.

Εκτός από την ποσότητα και την ποικιλία των παραγόμενων ιδεών, προκειμένου να αξιολογηθεί η δημιουργικότητα του αποτελέσματος κάθε εργασίας, οι εμπειρογνώμονες αξιολόγησαν επίσης την ποιότητα και την πρωτοτυπία. Όπως φαίνεται στον Πίνακα 14, οι ιδέες που δημιουργήθηκαν με το CRUISE ήταν πιο ποιοτικές ($m = 2.8$), σε σύγκριση με την κατάσταση «χωρίς χρήση συστήματος» ($m = 1.56$) και «με χρήση του συστήματος αναφοράς» ($m = 0.9$), το οποίο σημαίνει ότι οι



Εικόνα 30 Οι απαντήσεις των συμμετεχόντων σχετικά με το αν η διαδικασία τους βοήθησε να σχηματίσουν ιδέες κατά το τρίτο βήμα της ανάθεσης

χρήστες που είχαν πρόσβαση σε πηγές μέσω του CRUISE ήταν σε θέση να διατυπώσουν ιδέες που θεωρήθηκαν σχετικές με την εργασία, εφικτές και περιγράφηκαν πιο λεπτομερώς και ευκρινώς από εκείνους που δε χρησιμοποίησαν το CRUISE. Είναι προφανές ότι δεν μπορεί να αγνοηθεί η μαθησιακή επίδραση από τα προηγούμενα βήματα της ανάθεσης. Ωστόσο, η διαφορά στη μέση τιμή των αξιολογήσεων αναφορικά με την ποιότητα χρησιμοποιώντας το CRUISE, υπερβαίνει κατά πολύ την αντίστοιχη τιμή για τον αριθμό των ιδεών που παράχθηκαν χρησιμοποιώντας το σύστημα αναφοράς ή χωρίς σύστημα. Αυτό το αποτέλεσμα δείχνει ότι οι απαντήσεις των χρηστών στο τρίτο βήμα της εργασίας δεν προήλθαν μόνο ως αποτέλεσμα της επίδρασης των προηγούμενων βημάτων (testing effect), καθώς οι γνώσεις τους ήταν πιο ώριμες κατά τη διάρκεια της ανάθεσης, αλλά ήταν και αποτέλεσμα των επιπρόσθετων λειτουργιών CRUISE σε σύγκριση με το σύστημα αναφοράς. Από την άλλη πλευρά, η συνολικά χαμηλή μέση τιμή αποκαλύπτει ότι οι αξιολογήσεις ποιότητας διατηρήθηκαν χαμηλές από την πλευρά των εμπειρογνομώνων, πράγμα αναμενόμενο δεδομένου ότι στην εργασία αυτή δε ζητήθηκε ρητά από τους συμμετέχοντες λεπτομερή ανάλυση των ιδεών που βρέθηκαν, αλλά αφήθηκε στη διακριτική ευχέρεια του συμμετέχοντα. Η σχετικά μικρή χρονική διάρκεια του πειράματος θεωρήθηκε επίσης ως περιορισμός για μια λεπτομερέστερη περιγραφή των ιδεών.

Όσον αφορά την πρωτοτυπία, οι ειδικοί εξέτασαν κάθε μία από τις ιδέες που παρουσιάστηκαν και στα τρία στάδια της μελέτης. Ομαδοποίησαν παρόμοιες ιδέες σε θέματα και ανέθεσαν έναν κωδικό ως αναγνωριστικό σε κάθε θέμα. Στη συνέχεια, έθεσαν τον αντίστοιχο κωδικό σε κάθε ιδέα και υπολόγισαν πόσες ιδέες για κάθε κωδικό δημιουργήθηκαν και στα τρία στάδια της διαδικασίας. Στη συνέχεια, σύμφωνα με την ευριστική τεχνική «όσο πιο δημοφιλής είναι η ιδέα, τόσο λιγότερο πρωτότυπη θεωρείται» βαθμολόγησαν κάθε ιδέα ανάλογα με το επίπεδο δημοτικότητας. Ορίστηκε μια κοινή κλίμακα για κάθε θέμα, όπου όταν ο αριθμός των παρόμοιων ιδεών ήταν πάνω από 12, τότε το επίπεδο πρωτοτυπίας καθορίστηκε στο 1 ενώ όταν το ποσό των ιδεών ήταν από 1 έως, 3 το επίπεδο πρωτοτυπίας ήταν το υψηλότερο και τέθηκε στο 5. Μετά την ολοκλήρωση αυτής τη διαδικασίας, υπολογίστηκε ο μέσος όρος για κάθε θέμα και βήμα του πειράματος (δηλαδή κανένα σύστημα, σύστημα αναφοράς, CRUISE) και, τέλος, ο μέσος όρος για όλα τα θέματα ανά βήμα. Ο πίνακας 14 δείχνει ότι με το CRUISE ($m = 3.62$) οι συμμετέχοντες ήταν σε θέση να καταλήξουν σε πιο πρωτότυπες ιδέες, σε σύγκριση με το «χωρίς εργαλείο» ($m = 3.1$) και το σύστημα βάσης ($m = 3.4$). Αυτό το αποτέλεσμα μπορεί να αποδοθεί στο νέφος λέξεων καθώς και στη συνεισφορά από την κοινωνική συζήτηση, όπως παρουσιάζεται με τη μορφή του νέφους διάσπαρτων λέξεων. Το νέφος έδωσε τη δυνατότητα στους χρήστες να σχηματίσουν νέα μονοπάτια αναζήτησης συνδυάζοντας λέξεις από αυτό, το οποίο με τη σειρά του γίνεται το μέσο για την ανάκτηση νέων αποτελεσμάτων, τα οποία αν και σχετικά είναι ταυτόχρονα και ελαφρώς απομακρυσμένα από το θέμα, γεγονός που θα μπορούσε να υποστηρίξει τις ευχάριστες και αναπάντεχες ανακαλύψεις. (André, Teevan & Dumais, 2009).

Για να διερευνήσουμε τη σχέση μεταξύ της αντιληπτής χρησιμότητας, του απροσδόκητου και της εύνοιας των τυχαίων ανακαλύψεων στα αποτελέσματα αναζήτησης σε σχέση με τα δημιουργικά αποτελέσματα, υπολογίσαμε τους συντελεστές συσχέτισης Pearson όλων των προαναφερθέντων παραγόντων με όλα τα κριτήρια που καθορίζουν ένα δημιουργικό αποτέλεσμα (δηλ. την πρωτοτυπία, την ποιότητα, την ποσότητα και την ποικιλία των ιδεών που δημιουργήθηκαν με το

CRUISE.) Θεωρήσαμε, ωστόσο, μόνο ουδέτερα ή θετικά εγγραφές όπου και εύνοια των τυχαίων ανακαλύψεων και η χρησιμότητα ήταν πάνω από 2 ("Διαφωνώ λίγο").

Πίνακας 16 Ανάλυση συντελεστή συσχέτισης pearson

Criterion	Originality	Quality	Quantity	Variety
Perceived Serendipitousness	0.18	0.35	-0.04	-0.07
Unexpectedness	0.1	0.27	-0.02	-0.07
Usefulness	0.18	0.17	-0.05	-0.05

Όπως φαίνεται στον πίνακα 16, υπάρχει *θετική γραμμική συσχέτιση* (pearson=0.27) ανάμεσα στο απροσδόκητο και την ποιότητα και μια ισχυρότερη μεταξύ της προδιάθεσης για την εύνοια τυχαίων ανακαλύψεων και της ποιότητας (pearson=0.35), ενώ η χρησιμότητα δεν παίζει σημαντικό ρόλο στο δημιουργικό αποτέλεσμα, παρόλο που αναφέρεται στον ορισμό τη εύνοιας τυχαίων ανακαλύψεων. Τα αποτελέσματα υποβάθμισαν τη σχέση μεταξύ απροσδόκητου και εύνοιας τυχαίων ανακαλύψεων με την πρωτοτυπία των ιδεών. Αυτό είναι πιο πιθανό να έχει συμβεί, επειδή στην αξιολόγησή μας η πρωτοτυπία αξιολογείται σε ένα ευρύτερο πλαίσιο μεταξύ όλων των ιδεών που παρέχονται από τους συμμετέχοντες και δεν λαμβάνει υπόψη το υπόβαθρο γνώσης κάθε συμμετέχοντος ή το άλμα γνώσης που τυχόν επιτυγχάνει. Ακόμα πιο αδύναμη είναι η σχέση μεταξύ της εύνοιας τυχαίων ανακαλύψεων και της ποσότητας των ιδεών. Μία πιθανή εξήγηση μπορεί να είναι ότι τα απροσδόκητα και χρήσιμα αποτελέσματα αυξάνουν την εμπλοκή του χρήστη με το σύστημα, οδηγώντας σε μεγαλύτερη κατανόηση του θέματος που υποστηρίζει το σχηματισμό ιδεών, γεγονός που δεν συνδέεται απαραίτητα με την ποσότητα των ιδεών. Ο Πίνακας 17 συνοψίζει τις μετρήσεις αξιοπιστίας σχετικά με τη συμφωνία μεταξύ των δύο καθηγητών στην αξιολόγηση ποιότητας και πρωτοτυπίας. Με 125

συμφωνίες σε 162 ονομαστικά θέματα , η τιμή Kappa ($kappa=0.7$) υποδεικνύει μια ισχυρή συμφωνία μεταξύ των βαθμολογητών. Παρατηρούμε ότι το εύρος διαστήματος εμπιστοσύνης κυμαίνεται μεταξύ 62.44% και 78.,37%. Το εύρος είναι στενότερο από αυτό του πρώτου πειράματος λόγω του μεγαλύτερου αριθμού συμμετεχόντων και κατά συνέπεια αξιολογούμενων θεμάτων.

Πίνακας 17 Συμφωνία μεταξύ καθηγητών για την αξιολόγηση τόσο της ποιότητας όσο και της πρωτοτυπίας

#Raters	# Agreements	kappa	Standard Error	CI 95%
2	125	70.41%	4.07%	62.44% to 78.37%

Τα σχόλια που λάβαμε από τα δύο εργαστηριακά πειράματα κάλυψαν τρεις πτυχές: α) φιλικότητα προς το χρήστη β) χρηστικότητα και γ) ικανότητα απόκρισης. Όσον αφορά στη φιλικότητα προς το χρήστη, οι συμμετέχοντες πρότειναν να παρέχονται περισσότερες οδηγίες στους χρήστες σχετικά με τον τρόπο διεξαγωγής της αναζήτησης, γεγονός που μας οδήγησε σε αλλαγές στη διεπαφή έτσι ώστε να καθοδηγεί τον χρήστη σε κάθε βήμα της διαδικασίας. Όσον αφορά τη χρηστικότητα, υπογράμμισαν τη σημασία του φιλτραρίσματος του περιεχομένου του νέφους λέξεων. Για τη βελτίωση του νέφους λέξεων εφαρμόσαμε εκτεταμένους κανόνες για την εξάλειψη μη χρήσιμων λέξεων παράλληλα με την επεξεργασία των tweets, προκειμένου να εμφανιστούν φράσεις που θα μπορούσαν να δώσουν το ερέθισμα στους χρήστες προς μια συγκεκριμένη κατεύθυνση και να τους βοηθήσουν να προσδιορίσουν συσχετίσεις εννοιών. Τέλος, δήλωσαν ότι αντιλήφθηκαν το χρόνο που διήρκεσε η επεξεργασία, γεγονός που θα μπορούσε να μειώσει τον πραγματικό χρόνο διερεύνησης, ώστε να προχωρήσουμε επίσης σε βελτιώσεις όσον αφορά την υλοποίηση της εφαρμογής για τη μείωση του χρόνου απόκρισης.

4.6.3 Πιλοτική Μελέτη

Αναθέσαμε σε μια ομάδα επτά επαγγελματιών σχεδιαστών εννοιών (*concept designers*) από το Ινστιτούτο για την τέχνη, την επιστήμη και την τεχνολογία της

Waag⁵ να χρησιμοποιήσουν το CRUISE για την ενίσχυση των δημιουργικών αποτελεσμάτων της εργασίας τους. Η ομάδα των σχεδιαστών εννοιών, διενήργησαν ένα εργαστήριο διάρκειας 2 ημερών για τα «υλικά ανοικτού κώδικα». Σε αυτό το εργαστήριο, στόχος ήταν να παραχθούν ιδέες για το σχεδιασμό και την κατασκευή ανοιχτών υλικών για τη μονάδα Fablab. Το πρόγραμμα περιλάμβανε τον ορισμό του «υλικού ανοικτού κώδικα» και τον προβληματισμό σχετικά με το τι μπορεί να γίνει με αυτά, με τον πειραματισμό και την παραγωγή υλικών, καθώς και την ανάπτυξη δύο ολοκληρωμένων εννοιών (concepts). Λόγω του γεγονότος ότι η φύση του έργου τους βασίζεται κυρίως σε οπτικά ερεθίσματα, τους παρείχαμε μια βελτιωμένη έκδοση του συστήματος που περιλαμβάνει όχι μόνο τις βελτιώσεις που έγιναν μετά τις διαμορφωτικές αξιολογήσεις, αλλά και μια παραλλαγή του CRUISE, όπου το νέφος λέξεων αντικαταστάθηκε από ένα νέφος εμπλουτισμένο με ποικίλα μέσα ενημέρωσης όπως λέξεις, φράσεις, εικόνες καθώς και βίντεο και συνδέσμους. Είχαν επίσης τη δυνατότητα να σχηματίσουν συλλογές προτιμώμενων αντικειμένων που επιλέγονταν από το νέφος, παράλληλα με το σχηματισμό μονοπατιών αναζήτησης. Κατά τη διάρκεια του εργαστηρίου οι συμμετέχοντες της Εταιρείας Waag χρησιμοποίησαν το CRUISE για να διερευνήσουν το θέμα του «ανοικτού υλικού». Χρησιμοποίησαν ερωτήματα αναζήτησης όπως «ανοικτό σχέδιο», «απόβλητο» και «υλικό».

Το CRUISE χρησιμοποιήθηκε με επιτυχία, με αποτέλεσμα να δημιουργηθούν σε σύντομες περιόδους περισσότερες ιδέες από εκείνες που δημιουργούνται με οποιοδήποτε άλλο σχετικό εργαλείο χρησιμοποιήθηκε προηγουμένως σε παρόμοιες εργασίες από τους σχεδιαστές εννοιών. Οι συμμετέχοντες ήταν θετικοί σχετικά με τη συνολική χρησιμότητα του εργαλείου για την υποστήριξη της έμπνευσης και της δημιουργικής ικανότητάς τους, καθώς και την εύνοια τυχαίων ανακαλύψεων στα αποτελέσματα αναζήτησης. Οι συμμετέχοντες εκτίμησαν ιδιαίτερα την προσθήκη

⁵ <https://waag.org/en/home>

βίντεο στο νέφος λέξεων καθώς και την ιδέα της δημιουργίας μιας προσωπικής συλλογής προτιμώμενων αντικειμένων, τα οποία θα ήθελαν να έχουν τη δυνατότητα να αποθηκεύουν για μεταγενέστερη χρήση. Οι συμμετέχοντες εκτίμησαν επίσης την ιδέα της οπτικής αναζήτησης κλικάροντας συχνά σε εικόνες και βίντεο για να προστεθούν στην προσωπική συλλογή, αλλά πρότειναν τη βελτίωση της επισκόπησης των στοιχείων που έχουν προστεθεί στη συλλογή. Καθώς το εργαλείο προορίζεται και για την πρόκληση έκπληξης, δεν ήταν σαφές εάν η έκπληξη προκλήθηκε από τα αποτελέσματα του ερωτήματος και τον τρόπο με τον οποίο παρουσιάστηκαν στο χρήστη ή εάν το σύστημα έκανε περαιτέρω επεξεργασία για την πρόκληση έκπληξης. Όσον αφορά την ταχύτητα της υπηρεσίας, τόνισαν ότι "μερικές φορές η επεξεργασία πήρε λίγο χρόνο παραπάνω, πράγμα το οποίο μπορούσε να προκαλέσει δυσανασχέτηση".

Οι συμμετέχοντες κλήθηκαν να δώσουν τη βαθμολογία τους ως προς τη χρησιμότητα του CRUISE χρησιμοποιώντας μια κλίμακα αξιολόγησης πέντε βαθμών από το 1 που δηλώνει «Διαφωνώ απόλυτα» μέχρι το 5 που δηλώνει «Συμφωνώ απόλυτα» (Πίνακας 18). Τα βασικά ζητήματα που εντοπίστηκαν στην πιλοτική μελέτη σχετίζονται με την αντίληψη των χρηστών όσον αφορά τη χρησιμότητα των αποτελεσμάτων αναζήτησης. Παρατίθενται ενδεικτικά οι σχετικές ερωτήσεις και ο μέσος όρος βαθμολογίας: «Τα περισσότερα από τα αποτελέσματα που παρείχε το εργαλείο ήταν χρήσιμα και συναφή με την ανάγκη μου» με $m=3$, «Κάνοντας την εργασία βρήκα αντικείμενα που θα μπορούσα να μοιραστώ με τους φίλους μου» ($m=3$). Υπήρχαν και ερωτήσεις που συνδέονται με την ικανότητα για έμπνευση: «Ένωσα σίγουρος/η ότι κάποια από τα αποτελέσματα θα μου παρείχαν έμπνευση» ($m=3$). Σημειώστε ότι το τελευταίο έρχεται σε αντίθεση με τη θετική ανατροφοδότηση που λάβαμε σχετικά με τη συνολική χρησιμότητα του εργαλείου για την υποστήριξη της έμπνευσης και της δημιουργικής του ικανότητας, καθώς και τα αποτελέσματα της δήλωσης «Υπήρχαν στιγμές που η φαντασία μου προκλήθηκε» με $m=4$. Περαιτέρω έρευνα αποκάλυψε ότι οι σχεδιαστές εννοιών ενοχλήθηκαν με τις σποραδικά εμφανιζόμενες διαφημίσεις του twitter στα αποτελέσματα του νέφους λέξεων.

Προφανώς, παρά τις βελτιώσεις που κάναμε στο φιλτράρισμα του νέφους λέξεων μετά τις προτάσεις που προέκυψαν από τις πρώτες αξιολογήσεις, εξακολουθεί να υπάρχει περιθώριο βελτίωσης, όχι μόνο όσον αφορά τις λέξεις που εμφανίζονται στο νέφος λέξεων, αλλά και από την άποψη των πηγών πληροφόρησης που κρύβονται πίσω από κάθε στοιχείο του νέφους.

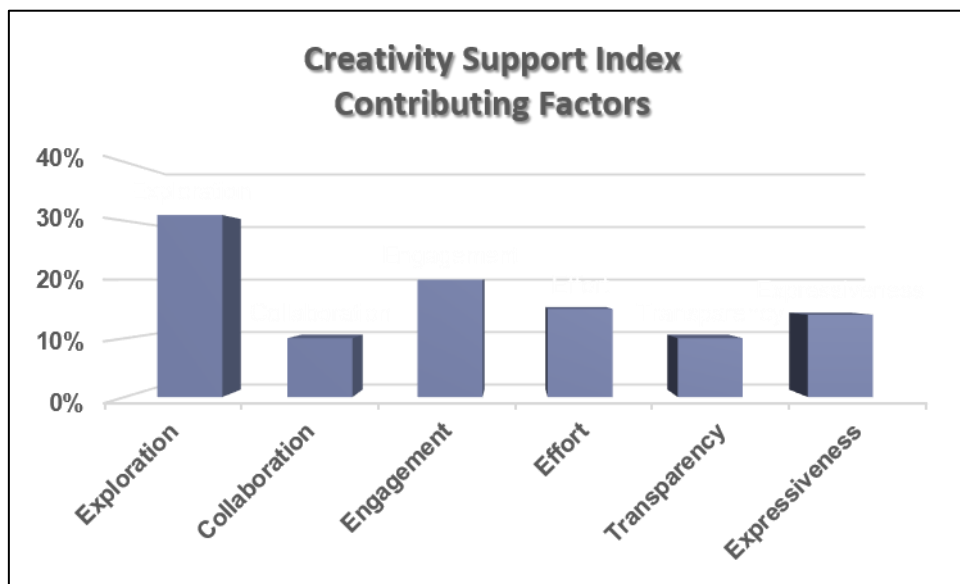
Δεδομένου ότι οι «σχεδιαστές εννοιών» ήταν οι εμπειρογνώμονες στη διεξαγωγή του συγκεκριμένου εργαστηρίου, δεν αξιολογήσαμε το αποτέλεσμα των ιδεών τους, αλλά χρησιμοποιήσαμε τον Δείκτη Υποστήριξης Δημιουργικότητας (CSI) για να

Πίνακας 18 Αντιληπτή Χρησιμότητα του CRUISE

Διατύπωση Ερώτησης	Μέσος Όρος Βαθμολογίας
Most of the results the tool provided were useful and relevant to my need	3
Some of the results were novel to me	3.67
Some of the results were surprising discoveries that I wasn't expecting to find.	3.6
Using the tool increased my knowledge about the subject I was investigating	3.17
There were times when my imagination was aroused	4
Doing the task I found items I could share with my friends.	3
I felt confident I would be inspired with some of the results.	3
I became familiar with the tool quickly.	3.17
I think that I would use the tool quickly.	3.33

αξιολογήσουμε την επίδραση του CRUISE στους ακόλουθους παράγοντες της δημιουργικής εργασίας: *διερεύνηση, συνεργασία, δέσμευση, προσπάθεια, διαφάνεια και εκφραστικότητα*. Λόγω της φύσης της δουλειάς τους και της μεγάλης εμπειρίας στη δημιουργική εργασία, θεωρούμε την επιλογή του ερωτηματολογίου CSI ιδιαίτερα σχετική. Το ερωτηματολόγιο CSI, αποτελείται από δύο μέρη: ένα μέρος βαθμολόγησης, όπου οι συμμετέχοντες αξιολόγησαν το CRUISE σε 12 χαρακτηριστικά δημιουργικότητας και ένα συνδυαστικό τμήμα σύγκρισης, το οποίο χρησιμοποιήθηκε για την ανάθεση βαρών στα χαρακτηριστικά που είχαν προηγουμένως αξιολογηθεί. Στις 12 ερωτήσεις, οι χρήστες αξιολογούν τους έξι προαναφερθέντες παράγοντες.

Τα αποτελέσματα του δείκτη CSI (71.3 από τα 100) καθώς και οι σχετικές συνεισφορές των πτυχών *διερεύνηση, συνεργασία, δέσμευση, προσπάθεια, διαφάνεια και εκφραστικότητα* στο συνολικό μέσο όρο του CSI απεικονίζονται στην εικόνα 31. Διαφαίνεται ότι ο παράγοντας «διερεύνησης» συμβάλλει το περισσότερο ανάμεσα στους έξι παράγοντες που εξετάστηκαν.



Εικόνα 31 Οι παράγοντες που συνεισφέρουν στο Δείκτη Υποστήριξης Δημιουργικότητας

4.7 Συμπεράσματα

Στην ενότητα αυτή παρουσιάσαμε το πρώτο μέρος της προτεινόμενης προσέγγισης της Διατριβής που είναι η διαδραστική διερευνητική αναζήτηση με αξιοποίηση ενδείξεων από την κοινωνική συνομιλία που στόχο έχει να βοηθήσει τους χρήστες να αναζητήσουν πληροφορίες στους διαθέσιμους χώρους πληροφορίας. Δείξαμε ότι οι χρήστες μπορούν να κατευθύνουν την αναζήτησή τους αλληλεπιδρώντας με όρους που εξάγονται από την κοινωνική συζήτηση, ώστε να διατυπώσουν νέες διαδρομές αναζήτησης. Μια μελέτη χρηστών με συγκεκριμένη ανατεθειμένη εργασία έδειξε ότι το σύστημά μας βελτιώνει τις αναπάντεχες και ευχάριστες ανακαλύψεις παρέχοντας παράλληλα αρκετά χρήσιμες, σχετικές και απροσδόκητες πληροφορίες. Τα αποτελέσματα ενισχύουν το συμπέρασμα ότι η προτεινόμενη προσέγγιση διερεύνησης και η χρήση των κοινωνικών μέσων μπορούν να ενισχύσουν τις δυνατότητες για αναπάντεχες και ευχάριστες ανακαλύψεις, που μπορεί να οδηγήσουν σε δημιουργικά αποτελέσματα.

Κατά την ανάπτυξη και αξιολόγηση της προσέγγισής μας και του αντίστοιχου λογισμικού, συναντήσαμε μερικούς περιορισμούς. Πρώτον, ο περιορισμένος αριθμός διαθέσιμων συμμετεχόντων δεν μας επέτρεψε να τους χωρίσουμε σε ξεχωριστές ομάδες (εκείνους που δεν χρησιμοποιούν ψηφιακό εργαλείο, εκείνους που χρησιμοποιούν το βασικό σύστημα και εκείνους που χρησιμοποιούν το σύστημα CRUISE). Εάν ακολουθούσαμε τη συγκεκριμένη προσέγγιση, θα μπορούσε να πραγματοποιηθεί άμεση σύγκριση μεταξύ του συστήματος CRUISE και του βασικού συστήματος. Ωστόσο, αυτό θα απαιτούσε την εξισορρόπηση των δημογραφικών δεδομένων, καθώς και την κατανόηση του γνωστικού υπόβαθρου των συμμετεχόντων, όπως και τη δυνατότητα για διερεύνηση, αλλά και την εξοικείωση με το σύστημα. Επιπλέον, έπρεπε να ασχοληθούμε με θέματα που αφορούν το σύστημα προτάσεων, το οποίο βασίζεται σε λογαριασμούς χρηστών στο Twitter. Οι πιο πολλοί από τους συμμετέχοντες δεν είχαν δραστηριότητα στο Twitter και κατά συνέπεια δεν ήταν εγγεγραμμένοι στην υπηρεσία, ενώ όσοι είχαν λογαριασμό

ανησυχούσαν για θέματα ιδιωτικότητας. Κατά συνέπεια, τα αποτελέσματά που παρουσιάζονται και είναι σχετικά με το εν λόγω σύστημα προτάσεων βασίζονται σε περιορισμένα δεδομένα. Παρ' όλα αυτά, οι συμμετέχοντες επεσήμαναν τη σπουδαιότητα της ύπαρξής του, ενώ οι περισσότεροι σχολίασαν τη σημασία των παραγόντων της κοινωνικής δύναμης, καθώς και το πόσο πρόσφατο είναι το προτεινόμενο περιεχόμενο. Οι περισσότεροι δήλωσαν ότι η κοινωνική δύναμη ενός λογαριασμού από τον οποίο προέρχεται ένα προτεινόμενο tweet παίζει πράγματι ρόλο στη χρησιμότητα όπως την αντιλαμβάνεται ο χρήστης, αφού «αναφορικά με την επιρροή, θεωρούμε χρήστες με μεγάλο αριθμό ακολούθων, γνώστες του θέματος με το οποίο συνδέεται το tweet τους, ακόμα κι αν δεν είναι πάντα η περίπτωση». Από την άλλη πλευρά, οι περισσότεροι συμμετέχοντες δε θεώρησαν το χρόνο που δημοσιεύθηκε πρώτη φορά ένα προτεινόμενο tweet ισχυρό παράγοντα, ενώ κάποιοι από αυτούς υπογράμμισαν ότι «εξαρτάται από το θέμα, για παράδειγμα τα tweets που αφορούν τεχνολογικά θέματα δε θα πρέπει να δίνουν ανεπίκαιρη πληροφορία». Επάνω στις παραπάνω έννοιες απαιτούνται εξειδικευμένες μελέτες, προκειμένου τα αποτελέσματα να έχουν στατιστική σημαντικότητα.

5 Η Επίδραση του Πλαισίου Αλληλεπίδρασης και της Προσωπικότητας

5.1 Εισαγωγή

Μία τυπική συνεδρία διερευνητικής αναζήτησης μπορεί να παραγάγει πλούσιο περιεχόμενο από σύνθετες αλληλεπιδράσεις συστήματος-χρήστη συμπεριλαμβάνοντας πολλαπλά ερωτήματα και κλικς για κάθε ερώτημα. Συλλογικά, όλες οι κινήσεις του χρήστη παρέχουν μία συνολική εικόνα της πρόθεσής του καθώς αλληλοεπιδρά με το εργαλείο αναζήτησης. Τα πολλαπλά ερωτήματα, τα επιλεγμένα αποτελέσματα για ανάγνωση, τα έγγραφα που κρύβονται πίσω από αυτά, είναι όλα χρήσιμα ίχνη που μπορούν να βοηθήσουν στην αποκάλυψη του σκοπού αναζήτησης του χρήστη (Zhang, Chen, Wang & Yang, 2011) και, από μία ευρεία οπτική, την συμπεριφορά διερεύνησης και αναζήτησης του χρήστη (Yue, Han & He, 2014). Σαφώς, η διερευνητική αναζήτηση απαιτεί σύνθετη αλληλεπίδραση χρήστη-συστήματος-αποτελεσμάτων, η οποία δεν μπορεί να είναι προδιαγεγραμμένη ή ίδια για όλους τους χρήστες. Αναμφισβήτητα, η προσωπικότητα των χρηστών διαδραματίζει σημαντικό ρόλο στις κινήσεις τους καθώς και στη συμπεριφορά τους γενικότερα σε μία συνεδρία διερευνητικής αναζήτησης. Μελέτες έχουν δείξει ότι υπάρχει σχέση μεταξύ της προσωπικότητας και της συμπεριφοράς ζήτησης των χρηστών (Nie, Guan & Zhu, 2014; Marchionini, 2009). Γι' αυτό, η προσωπικότητα είναι ένα χαρακτηριστικό που χρειάζεται να λαμβάνεται υπόψιν ως παράγοντας που επηρεάζει τη συμπεριφορά αναζήτησης και τις τακτικές διερεύνησης πληροφορίας, καθώς μπορεί να αντικατοπτριστεί στις προτιμήσεις του αναζητητή (Marchionini, 2006).

Η έρευνά μας στοχεύει να αναλύσει τις τακτικές του αναζητητή συνδέοντάς τις με τα χαρακτηριστικά της προσωπικότητάς του και να χρησιμοποιήσει τις αναλύσεις

για την παραγωγή προτάσεων κατά τη διάρκεια μιας συνεδρίας διερευνητικής αναζήτησης με σκοπό τη βελτίωση των δημιουργικών αποτελεσμάτων έπειτα από εργασίες διερευνητικής αναζήτησης. Προτείνουμε μία προσέγγιση για να προβλέψουμε τις ανάγκες των χρηστών και τις αμέσως επόμενες αλληλεπιδράσεις τους, δεδομένου του κάθε φορά τρέχοντος πλαισίου, τις παρελθοντικές αλληλεπιδράσεις και την προσωπικότητα. Επιδιώκουμε επίσης, περαιτέρω δημιουργία προτάσεων που θα τους βοηθήσουν να ολοκληρώσουν το σκοπό αναζήτησης πληροφορίας. Επιπλέον, η προσέγγισή μας στοχεύει στη βελτίωση της εμπειρίας και της αίσθησης εμπιστοσύνης των χρηστών, διατηρώντας το ενδιαφέρον τους κατά τη διάρκεια συνεδριών διερευνητικής αναζήτησης και αναβαθμίζοντας τα δημιουργικά αποτελέσματα που εν δυνάμει προκύπτουν μετά την αναζήτηση πληροφορίας.

Η προσέγγισή μας εξάγει μοτίβα διερεύνησης βασισμένα στην καταγεγραμμένη συμπεριφορά χρηστών, προκειμένου να παράγουν μοντέλα συμπεριφοράς, τα οποία με τη σειρά τους χρησιμοποιούνται για την πρόβλεψη της επόμενης δράσης στο τρέχον πλαίσιο.

Δύο τύποι μοντέλων παράγονται με το πρώτο να είναι ένα «γενικευμένο μοντέλο» με βάση τις καταγεγραμμένες συνεδρίες του συνολικού πληθυσμού. Προκειμένου να δημιουργηθεί ο δεύτερος τύπος μοντέλων, που ονομάζονται «μοντέλα προσωπικότητας», κάθε χρήστης κατατάσσεται σε μια συγκεκριμένη ομάδα χρηστών που παρουσιάζουν συγκεκριμένα χαρακτηριστικά προσωπικότητας.

Αναλύουμε τα μοντέλα που εξετάζουν κατά πόσο και ποιες ομάδες παρουσιάζουν παρόμοιες τακτικές διερεύνησης (δηλαδή το είδος των ενεργειών ή την ακολουθία των ενεργειών που εκτελούνται). Έχοντας τα μοντέλα, παρουσιάζουμε πώς μπορούμε να προβλέψουμε την επόμενη αλληλεπίδραση του χρήστη, δεδομένης μιας συγκεκριμένης ακολουθίας ενεργειών εντός της τρέχουσας συνεδρίας. Με αυτόν τον τρόπο, δημιουργούμε προτάσεις για να βοηθήσουμε τους χρήστες στη διαδικασία

διερεύνησής τους, παρέχοντας οδηγίες σχετικά με ενδεχομένως ανεξερεύνητες πτυχές του υπό έρευνα θέματος.

5.2 Σχετικές Εργασίες

5.2.1 Μοντελοποίηση Πλαισίου Αλληλεπίδρασης

Η χρήση του πλαισίου για την υποστήριξη της διαδραστικής ανάκτησης πληροφορίας και της διερευνητικής αναζήτησης είναι μία τεχνική που χρησιμοποιείται συχνά για να διευκολύνει την ανακάλυψη πληροφορίας. Υπάρχουν πολλών ειδών πλαίσια που μπορούμε να αξιοποιήσουμε. Με την προϋπόθεση της απαίτησης ότι ένας χρήστης μπορεί να παρέχει ανατροφοδότηση άμεσα, οι προσεγγίσεις ανατροφοδότησης σχετικότητας (Rocchio 1971) έχουν βρει περιορισμένες εφαρμογές στην πράξη και γι'αυτό έχουν ανοίξει το δρόμο για προσεγγίσεις που βασίζονται σε έμμεση ανατροφοδότηση, όπως των Joachims (2002), Kelly and Teevan (2003), Kelly and Belkin (2004), Sugiyama et al. (2004), White et al. (2004). Γενικά, όλο το ιστορικό αλληλεπίδρασης μπορεί να αξιοποιηθεί σαν έμμεση ανατροφοδότηση για τη βελτίωση της απόδοσης ανάκτησης πληροφορίας, όπως παρελθοντικά ερωτήματα, πληροφορία σχετική με το ποια έγγραφα έχει επιλέξει ο χρήστης να δει και ακόμα πώς ο χρήστης έχει διαβάσει ένα έγγραφο(π.χ. σε ποιο μέρος του εγγράφου ο χρήστης έχει δαπανήσει τον περισσότερο χρόνο). Οι Shen et al. (2005) μελέτησαν τέσσερα στατιστικά μοντέλα γλώσσας (FixInt, BayesInt, On-lineUp and BatchUp) για να δείξουν ότι η χρήση έμμεσης ανατροφοδότησης, ειδικά η ιστορικότητα των κλικ, μπορεί να βελτιώσει ουσιαστικά την απόδοση της ανάκτησης χωρίς να απαιτεί επιπλέον προσπάθεια από το χρήστη.

Προηγούμενη έρευνα έχει δείξει ότι η πληροφορία του πλαισίου μπορεί να βελτιώσει την απόδοση των συστημάτων προτάσεων σε πολλές εφαρμογές. Το πλαίσιο γενικά διακρίνεται σε «πλαίσιο αναπαράστασης» και «πλαίσιο αλληλεπίδρασης» (Dourish, 2004; Åman & Liikkanen, 2017). Η πλειοψηφία των συστημάτων προτάσεων με επίγνωση πλαισίου έχουν ερευνήσει τη χρήση του

πλαisiού αναπαράστασης, όπως ο χρόνος, η τοποθεσία, ο καιρός (Baltrunas, 2011). Οι περισσότερες προσεγγίσεις λαμβάνουν υπόψιν την πληροφορία πλαisiού για να βελτιώσουν τις προτάσεις συστημάτων συνεργατικής διήθησης, δηλαδή ενσωματώνουν πληροφορία πλαisiού στην ομοιότητα των χρηστών (Chen, 2005; Hu, Dou and Liu, 2012; Natarajan, Shin & Dhillon, 2013).

Για συστήματα προτάσεων που ενσωματώνουν το πλαίσιο αλληλεπίδρασης υπάρχει περιορισμένη έρευνα. Οι White et al. (2009) μελέτησαν την αλληλεπίδραση του χρήστη μαζί με ακόμα τέσσερις πηγές πληροφόρησης πλαisiού (κοινωνικές, ιστορικές, εργασία και συλλογή) αναφορικά με το πόσο αποτελεσματικά μπορούν να προβλέψουν τα μελλοντικά ενδιαφέροντα του χρήστη στο πλαίσιο προτάσεων Ιστού. Τα ευρήματά τους επιδεικνύουν ότι οι πηγές αποδίδουν διαφορετικά και εξαρτάται από τη διάρκεια του παραθύρου που χρησιμοποιήθηκε για μελλοντική πρόβλεψη και ότι η επικάλυψη με το πλαίσιο ξεπερνά κάθε μεμονωμένη πηγή.

Στον τομέα της μουσικής, οι Hariri, Mobasher and Burke (2012) καθώς και οι Natarajan et al. (2013) πραγματεύονται το θέμα του να προτείνεται το επόμενο μουσικό κομμάτι, δοσμένης μίας αλληλουχίας κομματιών που έπαιξαν πρόσφατα. Η πρώτη εργασία χρησιμοποιεί μοντελοποίηση θεμάτων για να αναπαραστήσει κάθε αλληλουχία θεμάτων σε μία χειρωνακτικά επεξεργασμένη βάση λιστών αναπαραγωγής, σαν αλληλουχία λανθανόντων θεμάτων για την ανακάλυψη συχνών μοτίβων θεμάτων χρησιμοποιώντας έναν αλγόριθμο εξόρυξης μοτίβων. Η δεύτερη ερευνητική εργασία πραγματοποιεί συσταδοποίηση χρηστών βάσει των αλληλεπιδράσεών τους με την βάση μουσικής, οι οποίες αναπαρίστανται σαν Μαρκοβιανές αλυσίδες και χρησιμοποιεί μία συνάρτηση κατάταξης εμπνευσμένης από τον αλγόριθμο PageRank για την πρόβλεψη της συνάφειας του επόμενου κομματιού.

Ένα σχετικό ρεύμα έρευνας επικεντρώνεται στην ενσωμάτωση πληροφορίας πλαisiού στην αναζήτηση. Η Y!Q αναζήτηση με ενσωμάτωση πλαisiού (Kraft, Mahoul & Chang, 2005) παράγει επαυξήσεις ερωτημάτων από το περιεχόμενο Ιστοσελίδων

και έχει μία λειτουργία χωρίς ερωτήματα. Στην εφαρμογή Implicit Query οι ερευνητές (Broder, Fontoura, Josifovski & Riedl, 2007) μελέτησαν ένα παρόμοιο πρόβλημα, αλλά για ένα περιβάλλον επιφάνειας εργασίας χρησιμοποιώντας τα e-mails και άλλη δραστηριότητα του χρήστη για να δημιουργηθούν ερωτήματα και προδραστικά να βρεθούν σχετικά έγγραφα. Οι Henzinger, Chang, Milch and Brin (2005) εστίασαν στην αυτόματη ανάκτηση πληροφορίας κατά τη διάρκεια ζωντανής αναμετάδοσης ειδήσεων εξαγοντας λέξεις κλειδιά. Οι Liebling, Bennett and White (2012) έχουν μελετήσει το πρόβλημα της πραγματοποίησης αναζήτησης χωρίς τη χρήση ερωτημάτων, για την πρόβλεψη αναγκών του χρήστη στη γενική Αναζήτηση Ιστού προβλέποντας τα αποτελέσματα για τα οποία οι χρήστες θα ψάξουν και θα επισκεφθούν σε επόμενο βήμα, δεδομένου μόνο του τρέχοντος συνδέσμου περιήγησης.

5.2.2 Μοντελοποίηση Συμπεριφοράς στην Αναζήτηση

Υπάρχουν πολλά γνωστά μοντέλα διαδικασίας αναζήτησης, όπως της Kuhlthau (1999), του Ellis(1993), του Marchionini (1995) και το μοντέλο συλλογής άγριων καρπών (berry-picking) της Bates (1989). Η Kuhlthau ανέπτυξε ένα μοντέλο το οποίο αναγνωρίζει και τονίζει τη σημασία των μεμονωμένων σταδίων τα οποία περιλαμβάνονται σε εργασίες εκμάθησης και επίλυσης προβλημάτων. Άλλες μελέτες όπως του Ellis ψάχνουν σε μοτίβα αναζήτησης μέσω της ανάλυσης των αλληλοεπιδράσεων χρήστη-συστήματος. Η μελέτη του, με βάση ερευνητές κοινωνικών επιστημών αποκάλυψε έξι γενικά χαρακτηριστικά συμπεριφοράς αναζήτησης πληροφορίας. Περιγράφονται στην ενότητα 2. Από την άλλη πλευρά, ο Marchionini του οποίου η έρευνα περιλάμβανε χρήστες ηλεκτρονικού συστήματος, αποκάλυψε τρία στάδια: α) ορισμός προβλήματος, β) σχηματισμός ερωτήματος και γ) διερεύνηση, εξαγωγή και σκέψη(reflection). Οι Xie and Joo (2010) έφεραν στην επιφάνεια τη σημασία της διερεύνησης μεταβάσεων στις τακτικές αναζήτησης σα μέσω εξέτασης διαδικασιών αναζήτησης με τη χρήση Μαρκοβιανών αλυσίδων. Ο Elswailer (2009) πρότεινε τη διερεύνηση διαδικασιών αναζήτησης με τη χρήση

Μαρκοβιανού Μοντέλου στη διαχείριση προσωπικής πληροφορίας. Ο Fuhr (2008) στηρίχθηκε στην υπόθεση ότι οι χρήστες κινούνται μεταξύ καταστάσεων και μοντελοποίησε τις διαδικασίες αναζήτησης σαν Μαρκοβιανή αλυσίδα. Σε κάθε κατάσταση, οι χρήστες έχουν μία λίστα από επιλογές και κάθε επιλογή (π.χ. μία κίνηση χρήστη) έχει το κόστος της, πιθανότητα αποδοχής και όφελος. Οι He, Qvarfordt, Halvey & Golonchinsky (2016) αξιοποίησαν την καταγραφή των κινήσεων των χρηστών κατά τη διάρκεια της χρήσης δύο συστημάτων αναζήτησης για να εξάγουν τακτικές αναζήτησης ενώ οι Shah, Hendaheva & González-Ibáñez (2015) ανέλυσαν διαδικασίες αναζήτησης σε συνεδρίες διερευνητικής αναζήτησης για την αξιολόγηση της απόδοσης της αναζήτησης προκειμένου να προβλέψουν το πώς οι χρήστες είναι πιθανό να διεξάγουν αναζήτηση και να κατακτήσουν το στόχο αναζήτησης. Η προσέγγισή τους, αν και παρόμοια με τη δική μας, δεν ενσωματώνει πληροφορία σχετική με την προσωπικότητα. Η εστίαση στη δική μας δουλειά είναι διμερής: α) αρχικά η αξιοποίηση του ιστορικού αλληλεπίδρασης των χρηστών σε σύστημα διερευνητικής αναζήτησης με στόχο την εκπαίδευση μοντέλων τακτικών διερεύνησης, β) αφετέρου η εκπαίδευση των μοντέλων διερεύνησης ανά ομάδα χρηστών που αντιπροσωπεύεται από συγκεκριμένες τιμές χαρακτηριστικών προσωπικότητας. Ο στόχος της προσέγγισής μας είναι η πρόβλεψη της επόμενης πιθανότερης αλληλεπίδρασης του χρήστη και παροχή πρότασης στο χρήστη, με σκοπό την υποστήριξή του στη διερευνητική διαδικασία για την επίτευξη πιο δημιουργικών αποτελεσμάτων σε σχέση με ένα βασικό σύστημα αναζήτησης πληροφορίας. Η δημιουργικότητα μπορεί να θεωρηθεί ως μια διαδικασία αναζήτησης ιδεών στη μακροχρόνια μνήμη, στην οποία θεωρείται ότι οι έννοιες αποθηκεύονται σε δικτυακές δομές, δηλαδή στη συνειρμική μνήμη (Nijstad & Stroebe, 2006). Η σκέψη μιας έννοιας μπορεί να ενεργοποιήσει μια άλλη έννοια μέσω κωδικοποιημένων συσχετίσεων. Επομένως, η δημιουργία μιας μεταγενέστερης ιδέας μπορεί να έχει εξάρτηση από προηγούμενη παραγωγή ιδεών (Wang, 2008). Οι Brown, Tumeo, Larey & Paulus (1998) μοντελοποίησαν την παραγωγή ιδεών ως αλυσίδα Markov που αποτελείται από μεταβατικές πιθανότητες για ένα σύνολο κατηγοριών ιδεών. Σε αυτό το μοντέλο, κάθε συγκεκριμένη ιδέα θεωρείται ότι έχει μόνο μία

ετικέτα που υποδηλώνει την κατηγορία της. Για να δημιουργήσετε μια συγκεκριμένη ιδέα απαιτείται η πρόσβαση στην προκαθορισμένη κατηγορία με την οποία συνδέεται η ιδέα. Σε κάθε στιγμιότυπο του χρόνου, μπορεί κανείς να δημιουργήσει μια ιδέα που βρίσκεται κάτω από την ίδια κατηγορία με την προηγούμενη ιδέα ή να μεταβεί σε μια άλλη κατηγορία και να δημιουργήσει μια ιδέα από τη νέα. Οι Nijstad & Stroebe (2006) πρότειναν το μοντέλο SIAM (Search for Ideas in Associative Memory) ως μια προσπάθεια ενσωμάτωσης της θεωρίας της συνειρμικής μνήμης σε ένα μοντέλο το οποίο υποθέτει την ανάκτηση μνήμης που αποτελείται κυρίως από δύο στάδια. Πρώτον, μια εικόνα που περιέχει μια κεντρική ιδέα, σχετικές έννοιες και ιδιότητες που σχετίζονται με τις έννοιες, ανακτάται από τη μακροχρόνια μνήμη στη μνήμη εργασίας και στη συνέχεια δημιουργούνται δεύτερες ιδέες από την ενεργοποιημένη εικόνα. Το μοντέλο βρίσκει τα κοινά σημεία μεταξύ δημιουργίας ιδεών και ελεύθερης ανάκλησης για να αποφευχθεί η υπερβολικά εξειδικευμένη παραγωγή ιδεών και να εξηγηθούν συναφή φαινόμενα βασισμένα σε γενικές γνωστικές θεωρίες. Ο Wang (2008) χρησιμοποιεί Κρυφά Μαρκοβιανά Μοντέλα (HMM) ως υπολογιστική συσκευή μοντελοποίησης που καταγράφει παρατηρήσιμες εξόδους ιδεών και θεωρητική, μη παρατηρήσιμη ενεργοποίηση των υποδομών μνήμης (αποκαλούμενες νοητικές εικόνες). Οι νοητικές εικόνες αντιπροσωπεύονται ως κρυφές καταστάσεις στα Κρυφά Μαρκοβιανά Μοντέλα. Οι ιδέες είναι σύμβολα εξόδου που εκπέμπονται από τις κρυμμένες καταστάσεις. Η παρατηρούμενη σημασιολογική εξάρτηση μεταξύ διαδοχικών ιδεών μπορεί να εξηγηθεί με ιδέες που προέρχονται από τις ίδιες ή διαφορετικές καταστάσεις.

5.2.3 Ο ρόλος της Προσωπικότητας στη Συμπεριφορά Ζήτησης Πληροφορίας

Η προσωπικότητα διαδραματίζει σημαντικό ρόλο στη λήψη αποφάσεων και γι' αυτό οι ερευνητές έχουν αρχίσει να υλοποιούν συστήματα προτάσεων που βασίζονται στην προσωπικότητα. Η προσωπικότητα έχει προταθεί για την επίλυση του προβλήματος της ψυχρής εκκίνησης. Στην ερευνητική εργασία τους (Hu & Pu, 2011; Fernández-Tobías, Braunhofer, Elahi, Ricci & Cantador, 2015) οι συγγραφείς

πρότειναν νέα μέτρα ομοιότητας χρηστών με βάση τα χαρακτηριστικά της προσωπικότητας και δημιούργησαν τρεις διαφορετικές προσεγγίσεις για προτάσεις. Οι συγγραφείς (Wu, Chen & He, 2013) ανέπτυξαν ένα σύστημα προτάσεων εισάγοντας μία ιδιόμορφη στρατηγική προσαρμογής της ποικιλίας βάσει προηγούμενης έρευνας (Chen, Wu & He, 2013) που ερεύνησε τη συσχέτιση μεταξύ προσωπικότητας και της ανάγκης για ποικιλία στις προτάσεις. Στην ίδια κατεύθυνση οι Ferwerda, Graus, Vall, Tkalcic, & Schedl (2016) χρησιμοποιώντας Wilcoxon signed-rank test για τη δοκιμή των αντιληπτών επιπέδων ποικιλομορφίας των λιστών προτάσεων, δημιούργησαν διαφορετικούς καταλόγους προτάσεων με τρία επίπεδα διαφοροποίησης (χαμηλό, μεσαίο, υψηλό). Η έρευνά τους αποκάλυψε ότι οι ευσυνείδητοι άνθρωποι συσχετίζονταν θετικά με τον βαθμό διαφοροποίησης, ενώ οι ευχάριστοι άνθρωποι (McCrae & John, 1992) ήταν πιο ικανοποιημένοι με ένα μεσαίο βαθμό διαφοροποίησης. Οι Potash and Rumshisky (2016) πρότειναν ένα σύστημα προτάσεων που εισήγαγε την προσωπικότητα του χρήστη σε Παραγοντοποίηση Μήτρας για την πρόβλεψη της βαθμολογίας του χρήστη. Η πρωτοτυπία της δουλειάς τους συνίσταται κυρίως στο ότι εξήγαγαν τα προφίλ προσωπικότητας από καταχωρημένες αξιολογήσεις χρηστών βάσει ενός μοντέλου το οποίο εκπαιδεύθηκε μέσω αλγορίθμου με επίβλεψη πάνω στο σύνολο δεδομένων MyPersonality⁶ (Kosinski, Stillwell & Graepel, 2013).

Στην αναζήτηση Ιστού, οι Nie, Guan & Zhu (2014) διερεύνησαν τα μοτίβα συσχέτισης των χαρακτηριστικών αναζήτησης Ιστού (δηλαδή την τάση περιήγησης πολλαπλών κατηγοριών ή την αναζήτηση πλήρους κειμένου κ.λπ.) και τα χαρακτηριστικά της προσωπικότητας, ενώ ταυτόχρονα εκμεταλλεύτηκαν τις συμπεριφορές αναζήτησης για να προσδιορίσουν τα χαρακτηριστικά της προσωπικότητας.

⁶ <http://www.socialmediacentre.eu/resource/mypersonality-dataset/>

Με βάση τη γνώση μας, στον τομέα της διερευνητικής αναζήτησης, υπάρχει περιορισμένη εργασία που εξετάζει τον αντίκτυπο των χαρακτηριστικών προσωπικότητας στη συμπεριφορά αναζήτησης του χρήστη. Οι Al-Samarraie, Eldenfria & Dawoud (2017) προσπάθησαν να καλύψουν αυτό το κενό εξετάζοντας τη συμπεριφορά κίνησης των ματιών σε διαφορετικές εργασίες αναζήτησης πληροφοριών. Διερεύνησαν τις παραμέτρους κίνησης των ματιών που συσχετίζονται με τα χαρακτηριστικά της προσωπικότητας σε συμπεριφορές ζήτησης πληροφορίας με ηλεκτρονικά μέσα, το κύριο αποτέλεσμα των χαρακτηριστικών προσωπικότητας στη συμπεριφορά αναζήτησης χρηστών και κατά πόσο οι παράμετροι κίνησης των ματιών μπορούν να προβλέψουν τα προφίλ προσωπικότητας των χρηστών στην ηλεκτρονική ζήτηση πληροφορίας.

5.3 Ερευνητική Προσέγγιση

Σε αυτή την ενότητα περιγράφουμε μια μέθοδο που χρησιμοποιεί πλαίσιο αλληλεπίδρασης και χαρακτηριστικά προσωπικότητας, προκειμένου να προτρέψει τους χρήστες να αναλάβουν δράση για την επίτευξη του στόχου διερευνητικής αναζήτησης. Η μέθοδος αναλύει τις διαφορές στις τακτικές διερεύνησης μεταξύ ομάδων ερευνητών που χαρακτηρίζονται από συγκεκριμένους συνδυασμούς χαρακτηριστικών προσωπικότητας. Βασιζόμαστε στα μοντέλα Markov που χρησιμοποιούνται συνήθως για τη μοντελοποίηση της συμπεριφοράς των χρηστών του διαδικτύου (Kiseleva, Lam, Pechenizkiy & Calders, 2013) ως εξής:

1. Πρώτα εκπαιδεύουμε γενικευμένα Κρυφά-Μαρκοβιανά-Μοντέλα (Hidden Markov Models-HMMs) με εμπειρικά δεδομένα από χρήστες που διεξάγουν συνεδρίες διερευνητικής αναζήτησης και σε δεύτερη φάση εκπαιδεύουμε Κρυφά-Μαρκοβιανά-Μοντέλα με δεδομένα από χρήστες με παρόμοια γνωρίσματα προσωπικότητας.

2. Χρησιμοποιούμε τα παραγόμενα Κρυφά-Μαρκοβιανά-Μοντέλα για να δημιουργούμε προβλέψεις για τις επόμενες πιθανότερες ενέργειες του χρήστη, ακολουθώντας την εκάστοτε διαθέσιμη μερική ακολουθία των ενεργειών του.

3. Με βάση τις παραγόμενες προβλέψεις, προτείνουμε προτροπές για την καθοδήγηση του χρήστη στην προβλεπόμενη επόμενη ενέργεια στη συνεδρία διερευνητικής αναζήτησης.

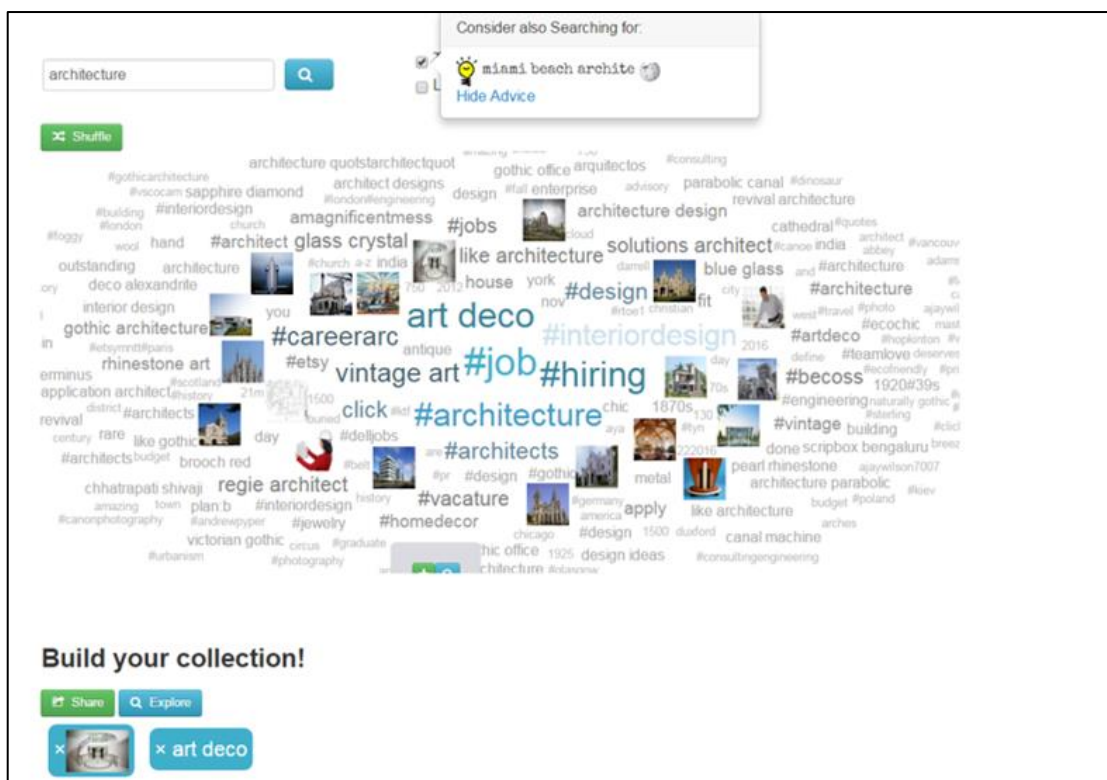
4. Τέλος, πραγματοποιούμε μια συγκριτική ανάλυση, σε μια προσπάθεια να αξιολογήσουμε εάν οι προτάσεις που βασίζονται στα Κρυφά-Μαρκοβιανά-Μοντέλα με βάση και την προσωπικότητα υπερτερούν των γενικευμένων Κρυφών-Μαρκοβιανών-Μοντέλων .

Προκειμένου να απεικονίσουμε τη μέθοδο μας, χρησιμοποιούμε το σύστημα διερευνητικής αναζήτησης CRUISE, το οποίο παρουσιάστηκε στην προηγούμενη ενότητα. Σε αυτή την ενότητα θα περιγράψουμε τον τρόπο με τον οποίο έχουμε επεκτείνει το CRUISE με τη δημιουργία υπηρεσίας παροχής προτάσεων προς τους χρήστες με σκοπό να τους βοηθήσουμε στην διερευνητική αναζήτηση (Εικόνα 32).

Στις επόμενες ενότητες όταν αναφερόμαστε σε Κρυφά-Μαρκοβιανά-Μοντέλα θα αναφερόμαστε είτε ολογράφως στα ελληνικά είτε με την αγγλική συντομογραφία HMMs.

5.3.1 Μοντελοποίηση Πλαισίου Αλληλεπίδρασης

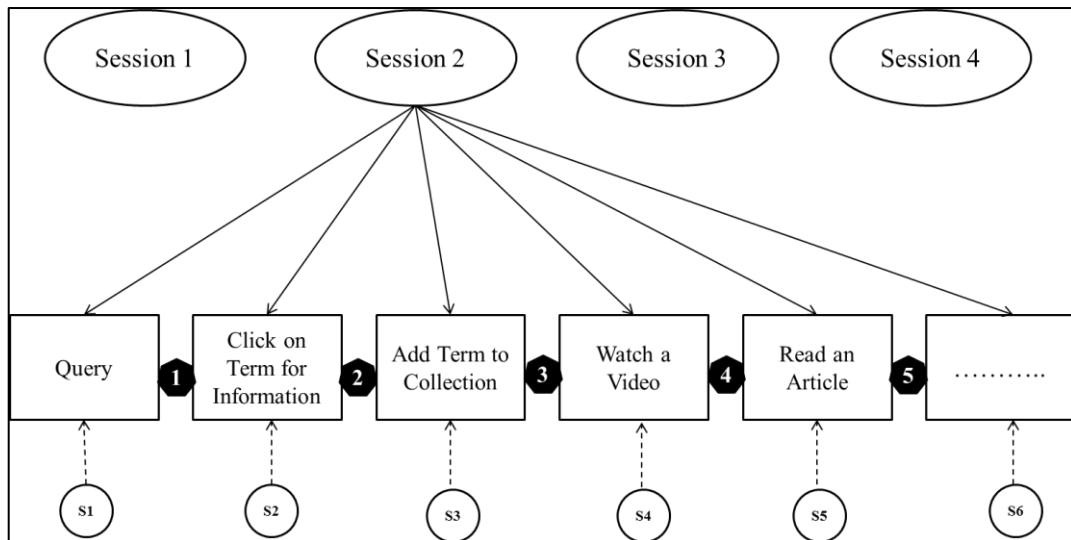
Καταγράψαμε τις αλληλεπιδράσεις του χρήστη με το διερευνητικό σύστημα αναζήτησης σε όλη τη διάρκεια μιας συνεδρίας (π.χ. αναζήτηση, κλικ για περισσότερες πληροφορίες, κλικ για προβολή ή παρακολούθηση κλπ.) ως μεταβάσεις με συγκεκριμένη σειρά και τις μοντελοποιήσαμε ως μια στοχαστική διαδικασία Markov, υποθέτοντας ότι η συμπεριφορά διερεύνησης είναι μια διαδικασία στην οποία οι ενέργειες δράσης συχνά επηρεάζονται από προηγούμενες.



Εικόνα 32 Το σύστημα CRUISE με προτάσεις. Σε αυτό το παράδειγμα η προβλεπόμενη επόμενη ενέργεια είναι ένα «Νέο ερώτημα» και το σημασιολογικό πλαίσιο είναι η «Αρχιτεκτονική» που οδηγεί στο προτεινόμενο ερώτημα 'architecture miami beach', που εμφανίζεται στην κορυφή της οθόνης.

Συνεπώς, υποθέτουμε ότι η συμπεριφορά διερεύνησης, στο πλαίσιο της δημιουργικής εργασίας έντασης γνώσης (knowledge intensive task), μπορεί να κατευθύνεται από τη γνωστική ή νοητική κατάσταση του χρήστη που είναι προφανώς άγνωστη και έτσι κρυμμένη. Αυτό οδηγεί στην επιλογή των Κρυφών Μαρκοβιανών Μοντέλων ως μηχανισμό πρότυπης αλληλεπίδρασης με το σύστημα.

Τα Κρυφά-Μαρκοβιανά-Μοντέλα είναι στατιστικά μοντέλα τα οποία χρησιμοποιούνται για τη μοντελοποίηση σειριακών δεδομένων ή χρονοσειρών, και έχουν επιτυχώς χρησιμοποιηθεί σε τομείς όπως η αναγνώριση ομιλίας, η ανάλυση ακολουθίας πρωτεϊνών ή DNA, στο ρομποτικό έλεγχο και στην εξαγωγή πληροφορίας από δεδομένα κειμένου. Αυτού του τύπου τα δεδομένα μοντελοποιούνται εύκολα με τη χρήση στοχαστικών διαδικασιών. Τα Μαρκοβιανά Μοντέλα παρέχουν μία φθηνή και διαισθητική λύση για τη μοντελοποίηση στοχαστικών διαδικασιών. Η δομή ενός



Εικόνα 33 Οι αλληλεπιδράσεις χρήστη-συστήματος σε μια συνεδρία, αναπαρίστανται ως σειριακές μεταβάσεις από μία ενέργεια σε άλλη. Κάθε ενέργεια είναι μια παρατήρηση που προκύπτει από μια συγκεκριμένη κρυφή κατάσταση.

Κρυφού-Μαρκοβιανού-Μοντέλου περιλαμβάνει καταστάσεις και παρατηρήσεις. Ένα Μαρκοβιανό Μοντέλο μπορεί να απεικονιστεί σαν μία μηχανή πεπερασμένων καταστάσεων με πιθανοτικές κορυφές. Σε ένα κρυφό Μαρκοβιανό Μοντέλο η κατάσταση της διαδικασίας δεν είναι απευθείας παρατηρήσιμη, έτσι δεν μπορούμε ποτέ να είμαστε σίγουροι για την τρέχουσα κατάσταση της διαδικασίας. Αντίθετα, τα Κρυφά-Μαρκοβιανά-Μοντέλα περιλαμβάνουν μία πιθανοτική συνάρτηση που αντιστοιχίζει καταστάσεις με παρατηρήσιμες ενδείξεις—αυτό μας επιτρέπει να εκτιμήσουμε την τρέχουσα κατάσταση βάσει ενός ιστορικού παρατηρήσιμων ενδείξεων (Baum & Petrie, 1966; Baum & Eagon, 1967; Baum & Sell, 1968; Baum & Petrie, 1970). Στην περίπτωση μας ένα Κρυφό-Μαρκοβιανό-Μοντέλο αποτελείται από N αριθμό καταστάσεων οι οποίες αντιστοιχίζονται στις γνωστικές ή νοητικές καταστάσεις στις οποίες μεταβαίνει ο χρήστης στη διάρκεια μιας συνεδρίας, κάθε μία από τις οποίες συσχετίζεται με ένα σύνολο πιθανών παρατηρήσεων. Οι παρατηρήσεις μας αντιστοιχίζονται στις πιθανές ενέργειες του χρήστη όπως ορίζονται στον Πίνακα 19.

Πίνακας 19 Κωδικοποίηση και Περιγραφή Αλληλεπιδράσεων Χρήστη με το CRUISE

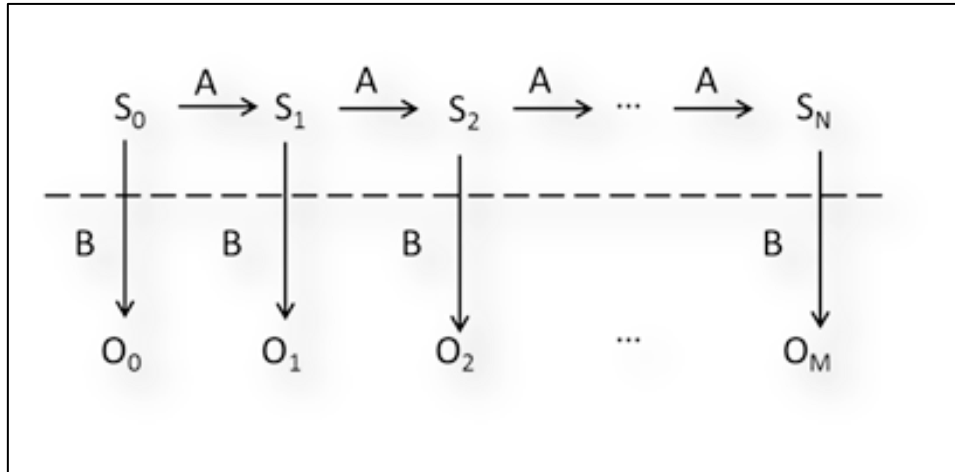
Interaction	Description	Code	Interaction	Description	Code
Query (Q)	A user issues a query	1	Idle (I)	A user stays inactive for 1 min	13
Click on Term or Image or Video or Web Link for info (CT,CI,CV,CA)	A user clicks on a cloud term to view relevant tweets or image/video/link description	2-5	Explore Collection (EC)	A user explores the combination of the items of his/her collection	14
Click to view Image or watch video or read an article (VI,WV,RA)	A user clicks on a cloud image to navigate to the original source	6-8	Shuffle (SH)	A user clicks on 'Shuffle' to reveal more items	15
Add term or image or video or link to collection (TC,IC,VC,AC)	A user selects a term or image or video or link to add to personal collection	9-12			

Ο ορισμός του προβλήματος περιλαμβάνει την αρχική παράμετρο π , έναν πίνακα μεταβάσεων A και έναν πίνακα παρατηρήσεων B , και μπορεί να οριστεί από το σύνολο παραμέτρων του μοντέλου HMM, $\Theta = (A, B, \pi)$.

Υπάρχουν τρεις κύριες περιγραφές προβλημάτων που επιλύονται από τα Κρυφά-Μαρκοβιανά-Μοντέλα:

- Αξιολόγηση: Αξιολόγηση της πιθανότητας μίας αλληλουχίας συμβόλων παρατηρήσεων, δοθέντος ενός συγκεκριμένου HMM.
- Αποκωδικοποίηση: Εύρεση του πιο πιθανού μονοπατιού μετάβασης καταστάσεων δοθείσης μίας αλληλουχίας παρατηρήσεων.
- Εκμάθηση: Η μάθηση των παραμέτρων του μοντέλου $\Theta = (A, B, \pi)$ βάσει του ιστορικού των παρατηρηθέντων αλληλουχιών και δοθισών των διαστάσεων N, M , όπου $N = o$ αριθμός των γνωστικών ή νοητικών καταστάσεων $S_{0...N}$, $M = o$

αριθμός των παρατηρήσεων $O_0 \dots O_M$ ή ενεργειών, π =οι αρχικές καταστάσεις, A =πίνακας πιθανοτήτων μεταβάσεων σε καταστάσεις, B =πίνακας πιθανοτήτων εκπομπών/παρατηρήσεων.



Εικόνα 34 Κρυφό Μαρκοβιανό Μοντέλο

Χρησιμοποιούμε τα Κρυφά-Μαρκοβιανά-Μοντέλα για να λύσουμε το τρίτο πρόβλημα δεδομένων ακολουθιών παρατηρήσεων O . Προκειμένου να εκπαιδύσουμε το μοντέλο μας, δηλαδή να εκτιμήσουμε τις παραμέτρους $\Theta=(A,B,\pi)$, αξιοποιούμε τον αλγόριθμο Baum-Welch ο οποίος είναι αλγόριθμος εκπαίδευσης χωρίς επίβλεψη. Ανήκει στην κατηγορία των Αλγορίθμων Μεγιστοποίησης Προσδοκίας (EM) και ξεκινάει με κάποια τυχαία εικασία των τιμών των παραμέτρων. Σε κάθε επανάληψη εκτελείται το βήμα προσδοκίας (βήμα E) για να βρεθεί μια κατανομή πιθανότητας για κάθε πιθανό μοντέλο, χρησιμοποιώντας τις τρέχουσες παραμέτρους. Στο βήμα μεγιστοποίησης (βήμα M), υπολογίζονται νέες τιμές των παραμέτρων χρησιμοποιώντας τις τρέχουσες πιθανότητες. Αυτές οι νέες εκτιμώμενες παράμετροι από το βήμα M, στη συνέχεια χρησιμοποιούνται για να προσδιοριστεί η κατανομή πιθανότητας των κρυφών μεταβλητών στο βήμα E που ακολουθεί στην επόμενη επανάληψη. Μετά από αρκετές επαναλήψεις, ο αλγόριθμος EM συγκλίνει σε ένα τοπικό ανώτατο όριο ή τοπικό μέγιστο. Η διαδικασία επαναλαμβάνεται έως ότου συγκλίνει η πιθανότητα, δηλαδή το HMM που προκύπτει από μία επανάληψη να μη διαφέρει σημαντικά από της προηγούμενης επανάληψης.

5.3.2 Παραγωγή Προβλέψεων

Δοθέντων των εκπαιδευμένων μοντέλων, είτε των γενικευμένων είτε αυτών που βασίζονται στην προσωπικότητα των χρηστών και την τρέχουσα αλληλουχία ενεργειών χρήστη, προβλέπουμε ποια θα ήταν η πιθανότερη επόμενη ενέργεια. Έστω $y_{1:M}=Y$ οι M παρατηρήσεις μέχρι τώρα και θ οι παράμετροι του HMM. Στη συνέχεια θέλουμε να εκμαιεύσουμε την πιθανότητα της επόμενης παρατήρησης δοθέντων των παρατηρήσεων που έχουν ιδωθεί, που εκφράζεται ως:

$$P(y_{M+1}|y_{1:M} = Y, \theta)$$

Η παραπάνω υποθετική έκφραση είναι ισοδύναμη με:

$$P(y_{M+1}|y_{1:M} = Y, \theta) = \frac{y_{1:M}=Y, y_{M+1}|\theta}{y_{1:M}=Y|\theta}$$

Σημειώστε ότι ο παρονομαστής είναι ανεξάρτητος από την παρατήρηση y_{M+1} , έτσι γίνεται:

$$P(y_{M+1}|y_{1:M} = Y, \theta) \propto P(y_{1:M} = Y, y_{M+1}|\theta)$$

Μία βίαια απλουστευμένη προσέγγιση είναι η ακόλουθη: Για κάθε μία από όλες τις πιθανές παρατηρήσεις, $y_{M+1}='Click'$, $y_{M+1}='Explore'$, κλπ, χρειαζόμαστε να υπολογίσουμε την πιθανότητα των αλληλουχιών $y_{1:M+1}$. Αυτό σημαίνει $P(y_{M+1}='Query', y_{1:M}=Y|\theta)$, $P(y_{M+1}='Explore Collection', y_{1:M}=Y|\theta)$, κλπ για κάθε μία από τις πιθανές αλληλουχίες παρατηρήσεων. Τότε η y_{M+1} , η οποία δίνει τη μέγιστη πιθανότητα, μπορεί να εκτιμηθεί ως η καλύτερη υπόθεση για την επόμενη παρατήρηση. Λαμβάνοντας αυτό υπόψιν, μπορούμε να δηλώσουμε ότι κάθε ένας από αυτούς τους υπολογισμούς πιθανότητας είναι μία ευθεία εφαρμογή του αλγόριθμου προώθησης (forward algorithm), που αντιστοιχεί στο 1^ο από τα τρία προβλήματα των HMMs (αξιολόγηση της πιθανότητας μίας αλληλουχίας συμβόλων παρατηρήσεων δοθέντος ενός συγκεκριμένου HMM), όπως δηλώθηκε στην

προηγούμενη ενότητα. Αργότερα στην ενότητα της υλοποίησης, θα επιδειχθούν σχετικά παραδείγματα της εφαρμογής του αλγορίθμου προώθησης.

5.3.3 Σχεδίαση Προτάσεων

Όπως έχει εξηγηθεί, η ανάλυση του πλαισίου αλληλεπίδρασης μπορεί να παρέχει ένα μέσο να προκύπτουν προβλέψεις για τις επόμενες πιθανές ενέργειες. Για να ληφθεί υπόψιν αυτή η δυνατότητα και να παρέχονται χρήσιμες προτάσεις στους χρήστες, κάποιος χρειάζεται να λαμβάνει υπόψιν του τις προτιμήσεις του χρήστη. Η εργασία μας δίνει έμφαση στις προτιμήσεις που προκύπτουν από το τρέχον πλαίσιο της διερεύνησης και όχι από τις μακροπρόθεσμες προτιμήσεις όπως αυτές αποτυπώνονται για παράδειγμα στα προφίλ χρηστών. Τα κλικς επάνω σε αντικείμενα για τη λήψη επιπλέον πληροφορίας και ιδιαίτερα τα κλικς σε αντικείμενα για τη δημιουργία συλλογών από αντικείμενα είναι χρήσιμες ενδείξεις για να εξαχθούν οι προτιμήσεις μέσα στο συγκεκριμένο πλαίσιο. Συνεπώς, οι προτάσεις θα πρέπει να είναι συναφείς και με τις ενέργειες που προβλέπονται και με τις προτιμήσεις όπως σκιαγραφούνται βάσει πλαισίου. Συνδυάζοντας τις προβλέψεις επόμενων ενεργειών με τις προτιμήσεις καθίστανται εφικτές προτάσεις όπως οι ακόλουθες: Εάν ένας χρήστης, για παράδειγμα, δείχνει ενδιαφέρον για βιομηχανία ηλιακών αυτοκινήτων και η προβλεπόμενη επόμενη κίνηση θα ήταν η παρακολούθηση ενός video, τότε η τελική πρόταση θα ήταν ένα video σχετικό με ηλιακά αυτοκίνητα.

Στην προσπάθειά μας να παράγουμε χρήσιμες προτάσεις, εμπλουτίζουμε τις προτιμήσεις του πλαισίου με συναφείς έννοιες. Αν θεωρήσουμε, για παράδειγμα, την προτίμηση κάποιων χρηστών σχετικά με την "ανανεώσιμη ενέργεια", προκύπτουν συγγενικές έννοιες όπως, "ανανεώσιμες πηγές", "βιώσιμη ενέργεια", "εμπορία εκπομπών", ακόμη και "plug-in hybrid". Οι όροι αναζήτησης ενδέχεται να αποκαλύπτουν διαφορετικές πτυχές των προτιμήσεων των χρηστών. Με αυτόν τον τρόπο μπορεί να αποκαλυφθούν διαφορετικές ή ακόμα και απροσδόκητες πτυχές των προτιμήσεων των χρηστών. Για να εμπλουτίσουμε τις προτάσεις, χρησιμοποιούμε την υπηρεσία DBpedia Lookup, η οποία αναζητά URI στο DBpedia με συναφείς λέξεις-

κλειδιά. Συναφείς σημαίνει ότι είτε η ετικέτα μίας πηγής ταιριάζει, είτε το κείμενο που χρησιμοποιείται συχνά στο Wikipedia για να αναφερθεί σε έναν πόρο ταιριάζει. Ειδικότερα, η υπηρεσία που χρησιμοποιήθηκε ήταν το Keyword Search API για να βρεθούν σχετικοί πόροι από το DBpedia δοθείσης μίας λέξης⁷. Η λέξη μπορεί να αποτελείται από μία ή πολλαπλές λέξεις. Τα διαφορετικά αποτελέσματα που επιστρέφει η υπηρεσία Lookup χρησιμοποιήθηκαν σαν επιπλέον ερωτήματα του αρχικού, προκειμένου να εξαχθούν περισσότερες πληροφορίες από τα διάφορα APIs που πλαισιώνουν τη λειτουργία του Cruise (δηλαδή το Twitter, Bing, Flickr, Youtube). Η μέθοδος υπολογισμού των τρεχουσών προτιμήσεων του χρήστη εφαρμόζεται στους παρακάτω τύπους προτάσεων:

- **Πρόταση ενός νέου ερωτήματος:** Αυτός ο τύπος πρότασης εφαρμόζεται όταν η προβλεπόμενη επόμενη κίνηση είναι η ενέργεια με κωδικό 1 (βλ. Πίνακα 19) συνδυασμένη με τις δυναμικές προτιμήσεις του χρήστη αναφορικά με το περιεχόμενο.
- **Πρόταση Πολυμέσων και Άρθρων για παρακολούθηση ή ανάγνωση:** Αυτός ο τύπος πρότασης εφαρμόζεται όταν η προβλεπόμενη επόμενη ενέργεια είναι μεταξύ κωδικών 6 και 8 (βλ. Πίνακα 19) σε συνδυασμό με τις δυναμικές προτιμήσεις χρήστη όσον αφορά το περιεχόμενο.
- **Προτάσεις όρων ή εικόνων ή βίντεο για προσθήκη στη συλλογή (Καταγραφή):** Αυτός ο τύπος πρότασης εφαρμόζεται όταν η προβλεπόμενη επόμενη ενέργεια είναι μία μεταξύ των κωδικών 9 και 12 (βλ. Πίνακα 19) σε συνδυασμό με δυναμικές προτιμήσεις χρήστη όσον αφορά το περιεχόμενο.

⁷ <http://lookup.dbpedia.org/api/search.aspx/KeywordSearch?QueryClass=&QueryString=solar%20energy>

Η προσέγγισή μας περιγράφει τρεις ακόμα τύπους προτάσεων που καταλληλότερο είναι να ονομάζονται προτροπές, ανεξάρτητα από τις προτιμήσεις του δυναμικού χρήστη, οι οποίες βασίζονται κυρίως στην προβλεπόμενη επόμενη ενέργεια.

- **Προτροπή Σάρωσης Αποτελεσμάτων:** Αυτός ο τύπος προτροπής εφαρμόζεται όταν η επόμενη ενέργεια προβλέπεται να είναι μία μεταξύ των κωδικών 2 και 5, καθώς επίσης εμφανίζεται όταν ο τρέχων χρήστης δεν είναι πραγματικά ενεργός.
- **Προτροπή Χρηστών να πιέσουν το κουμπί Shuffle για ανακατανομή των αποτελεσμάτων:** Αυτό το είδος προτροπής εφαρμόζεται όταν προβλεπόμενη επόμενη ενέργεια είναι ο κωδικός 15 και ο τρέχων χρήστης δεν έχει ανακαλύψει τη λειτουργία.
- **Προτροπή Χρηστών να διερευνήσουν βάσει της Συλλογής:** αυτό το είδος προτροπής εφαρμόζεται όταν η επόμενη ενέργεια που προβλέπεται είναι ο κωδικός 14 και ο τρέχων χρήστης έχει περισσότερα από ένα στοιχεία στη συλλογή του ώστε να διερευνηθεί ο συνδυασμός τους.

5.4 Υλοποίηση

Επιλέξαμε η πηγή προέλευσης των δεδομένων μας να είναι μια ηλεκτρονική εργαστηριακή αγορά, η Amazon Mechanical Turk ⁸, όπου οι εργαζόμενοι προσλαμβάνονται από εργοδότες (οι οποίοι καλούνται αιτούντες) για την εκτέλεση

⁸ <https://www.mturk.com/mturk/welcome>

εργασιών που ονομάζονται HITs⁹. Η απόφασή μας καθοδηγήθηκε από την ετερογένεια των συμμετεχόντων, καθώς και από τον υψηλότερο αριθμό συμμετεχόντων, την οποία μπορούσαμε να επιτύχουμε με αυτόν τον τρόπο, δύο στοιχεία απολύτως επιθυμητά για ανάλυση δεδομένων. Εάν και θα υπήρχαν ανησυχίες σχετικά με την ποιότητα των δεδομένων, δοθέντος ότι οι εργαζόμενοι του Mechanical Turk πληρώνονται πολύ λίγο, ωστόσο δεν υπάρχει ισχυρή απόδειξη ότι οι απαντήσεις από το Mechanical Turk έχουν χειρότερη ποιότητα από αυτών μίας εργαστηριακής ή παραδοσιακής μελέτης στον Ιστό (Paolacci, Chandler, Ipeirotis 2010; Behrend, Sharek, Meade, & Wiebe, 2011). Επιπλέον, το γεγονός ότι οι εργαζόμενοι του Amazon Mechanical Turk μπορούν να εκτελούν πειράματα χωρίς να έχουν επαφή με τους «πειραματιστές», αποφεύγοντας έτσι το διάλογο (Edlund et al., 2009), συνεισφέρει θετικά στην εσωτερική εγκυρότητα. Ζητήσαμε από τριακόσιους χρήστες του Amazon Mechanical Turk να διεξάγουν έρευνα με τη χρήση του CRUISE για ένα θέμα της προτίμησής τους επιλέγοντας από μια λίστα με συγκεκριμένα θέματα (π.χ. διαφήμιση, αρχιτεκτονική, τέχνη, βιοτεχνία, σχεδιασμός, μόδα, μουσική, τέχνες, εκδόσεις, βιβλιογραφία, επιστήμη) (Πίνακας 20). Καταγράψαμε τις αλληλεπιδράσεις των χρηστών με το εργαλείο, κάτι που οι χρήστες γνώριζαν. Για την καταγραφή των αλληλεπιδράσεων του χρήστη με το εργαλείο, εφαρμόσαμε ένα μηχανισμό, ο οποίος κατέγραφε τις ενέργειες του χρήστη, όπως νέες ερωτήσεις, κλικ σε αντικείμενα, πέρασμα ποντικιού πάνω από αντικείμενα, συλλογή στοιχείων, σχηματισμό συλλογών ευρημάτων (όρων, πολυμέσων, συνδέσμων σε άρθρα) και περιόδους αδράνειας ανά συνεδρία. Η έναρξη και το τέλος μιας συνεδρίας επισημάνθηκαν με ένα μοναδικό αναγνωριστικό που διέκρινε κάθε συνεδρία. Δεδομένου ότι το πείραμα δεν διεξήχθη σε εργαστήριο, όπου οι συμμετέχοντες θα μπορούσαν να εκπαιδευτούν

⁹ Οι εργασίες που ανατίθενται μέσω της crowdsourcing Διαδικτυακής Υπηρεσίας Mechanical Turk ονομάζονται HITs (acronym for Human Intelligence Tasks)

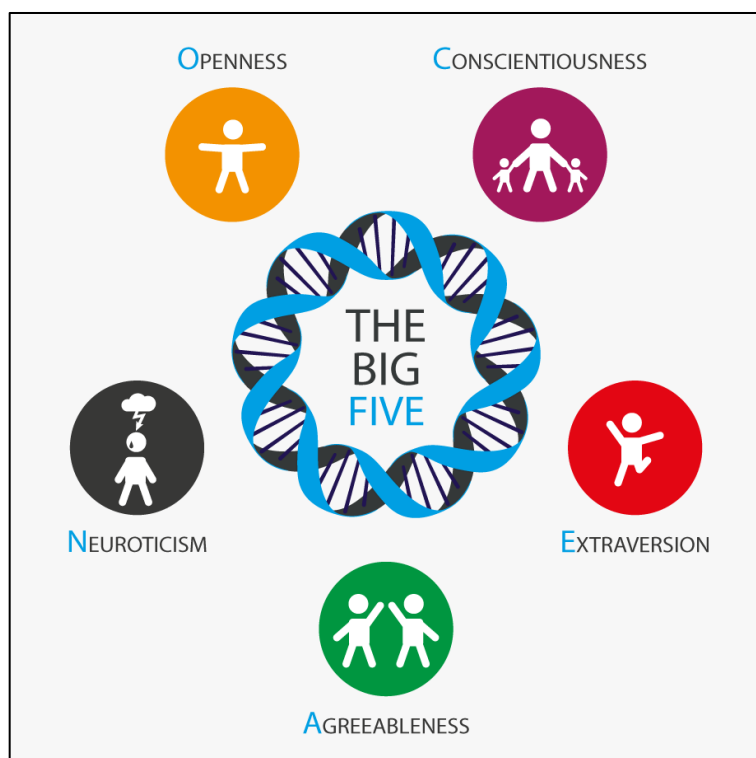
Πίνακας 20 Περιγραφή Amazon Mechanical Turk HIT για τη Συλλογή Δεδομένων Αλληλοεπίδρασης

HIT Instructions
<ul style="list-style-type: none">• Choose one of the following topics: advertising, architecture, art, crafts, design, fashion, music, performing arts, publishing, literature, R&D, software and science.• Keeping the topic in mind, go to the search engine Cruise.• In Cruise, do some research about the chosen topic of interest by performing search/click/browsing activities. The purpose of your exploration is to learn as much as possible about your topic, being able to identify as many aspects as possible (or concepts related to it), which you will discover using the tool. Afterwards you will be asked to note down your findings, make a short summary describing them and how they are related to each other. Watch this short video describing Cruise.• When you feel you have completed your topic research press the 'Finish' button that will appear on the search engine's screen. A pop up with your personal Unique-Id will be displayed. Please, note it down, in order to answer the first question.

κατάλληλα στο εργαλείο, παρείχαμε ένα σύντομο βίντεο που εξηγούσε τις δυνητικές ενέργειες που υποστηρίζονται από το εργαλείο. Δεν παρουσιάσαμε μια περιγραφή που να περιγράφει ένα συγκεκριμένο σενάριο για να αποφευχθεί τυχόν προκατάληψη σχετικά με τη διαδικασία διερεύνησης.

Τους ζητήσαμε να κάνουν μια σύντομη περιγραφή του θέματος για το οποίο διεξήγαγαν έρευνα και αξιολογήσαμε τις απαντήσεις τους σε συνδυασμό με τα καταγεγραμμένα αντικείμενα που είχαν αλληλοεπιδράσει, προκειμένου να εντοπιστούν οι πιο αξιόλογοι συμμετέχοντες και να απορριφθούν οι υπόλοιποι. Επιπλέον, συγκεντρώσαμε πληροφορίες για την προσωπικότητα των χρηστών καθώς και μια αξιολόγηση της χρηστικότητας και της αντίληψής τους για τη χρησιμότητα της εφαρμογής CRUISE.

Προκειμένου να αποκτήσουμε πληροφορίες για την προσωπικότητα, ζητήσαμε από τους συμμετέχοντες να απαντήσουν στις 44 ερωτήσεις του μοντέλου Big Five Factor (McCrae & John, 1992), που εξετάζει τους ακόλουθους πέντε παράγοντες προσωπικότητας: α) εξωστρέφεια (extraversion), β) ευγένεια (agreeableness), γ) ευσυνειδησία (conscientiousness), δ) νευρωτισμό (neuroticism) και ε) ανοιχτό πνεύμα (openness). Η απάντηση στο ερωτηματολόγιο 44 ερωτήσεων δίνει μια βαθμολογία για κάθε ένα από τους παράγοντες του Μοντέλου Πέντε Παραγόντων (Big Five Personality Factors).



Εικόνα 35 Οι πέντε Διστάσεις της Προσωπικότητας σύμφωνα με το μοντέλο Big Five

Οι συμμετέχοντες που δεν πληρούσαν τα ακόλουθα κριτήρια απορρίφθηκαν (παρόλο που η εργασία πειραματισμού ή αλλιώς το HIT ήταν ορατό μόνο σε εργαζόμενους με ποσοστό αποδοχής μεγαλύτερο από 95% για να εξασφαλιστεί η ποιότητα των δεδομένων): α) δεν παρείχαν τον κωδικό αναγνώρισης που παράγεται από το εργαλείο, προκειμένου να συνδέσουν τις απαντήσεις τους στο ερωτηματολόγιο με την καταγεγραμμένη συμπεριφορά διερεύνησης, β) οι

απαντήσεις τους στις ερωτήσεις σχετικά με το επιλεγμένο θέμα ήταν κακής ποιότητας, πιθανώς επειδή δεν ασχολήθηκαν πλήρως με αυτό που τους ζητήθηκε να κάνουν, γ) αφιέρωσαν πολύ λίγο χρόνο και κατά συνέπεια πραγματοποίησαν λίγες ενέργειες. Τελικά, διατηρήθηκαν οι συνεδρίες 224 συμμετεχόντων, οι οποίοι κυμαίνονταν από ηλικία 20 έως 66 ετών ($M = 33$, $SD = 9,54$), βλέπε Πίνακα 21.

Πίνακας 21 Χαρακτηριστικά Συμμετεχόντων

Subject Pool	%Females	Average Age
Mechanical Turk	42%	33

5.4.1 Συλλογή Δεδομένων Πλαισίου Αλληλεπίδρασης και Προσωπικότητας

5.4.1.1 Δεδομένα Αλληλεπίδρασης

Η καταγραφή των αλληλεπιδράσεων των χρηστών με το εργαλείο κατά τις περιόδους διερεύνησης μας έδωσε τη δυνατότητα να αναλύσουμε τα γενικά χαρακτηριστικά της διερευνητικής τους συμπεριφοράς ανά συνεδρία, όπως φαίνεται

Πίνακας 22 Ανάλυση ενδεικτικών χαρακτηριστικών συνεδρίας – Ο χρόνος Δέσμευσης μετρείται σε δευτερόλεπτα

Session Characteristics	Mean	StDev	Mode
#Queries	2.83	2.31	1
#Actions	16.9	16.2	9
Engagement Time	480	320	74

στον Πίνακα 22. Για να κωδικοποιήσουμε τις καταγεγραμμένες ενέργειες υιοθετήσαμε την κωδικοποίηση που περιγράφηκε στον Πίνακα 19. Η καταγραφή των αλληλεπιδράσεων των χρηστών με το εργαλείο έδειξε ότι κάθε χρήστης παρουσίασε ιδιαίτερα μεταβαλλόμενη συμπεριφορά στις διαφορετικές συνεδρίες. Η μεγάλη τυπική απόκλιση του συνολικού χρόνου εμπλοκής ($SD = 320$) που παρουσιάζεται στον

Πίνακα 22 δείχνει ότι οι χρόνοι που συμμετείχαν στη διερευνητική εργασία ποικίλουν πολύ, γεγονός που μπορεί να αποδοθεί μεταξύ άλλων στους παράγοντες της προσωπικότητας, του βαθμού εμπλοκής ή της ταχύτητας που ανέπτυξαν για την επίτευξη του επιθυμητού αποτελέσματος. Δεν αποτελεί έκπληξη το γεγονός ότι η

Πίνακας 23 Κατανομή Ενεργειών Συμμετεχόντων

Action	Code	Distribution
Query (Q)	1	0.176
Add term to collection (TC)	9	0.145
Click to view Image (VI)	6	0.117
Shuffle (SH)	15	0.116
Click to watch video (WV)	7	0.115
Add image to collection (IC)	10	0.074
Add link to collection (AC)	12	0.072
Add video to collection (VC)	11	0.063
Idle (I)	13	0.055
Explore Collection (EC)	14	0.054
Click on Term for info (CT)	2	0.010
Click Web Link for info (CA)	5	0.003
Click on Image for info (CI)	3	0.001
Click on Video for info (CV)	4	< 0.001
Click to Read an Article (RA)	8	< 0.001

ενέργεια «Ερώτημα-Query (Q)» ήταν η πιο συχνή ενέργεια που έκαναν οι χρήστες στην αναζήτηση χρήσιμων πληροφοριών (Πίνακας 23). Από την άλλη πλευρά, διαπιστώνουμε ότι η ενέργεια «Κλικ στην εικόνα για πληροφορίες» ήταν σπάνια, πιθανότατα επειδή οι εικόνες μεταβίβαζαν χρήσιμες ενδείξεις μέσω των thumbnails ή με απλό πέρασμα του ποντικιού πάνω από τα αποτελέσματα, καθώς και επειδή οι σχολιασμοί δεν ήταν πάντα διαθέσιμοι. Οι ενέργειες παρακολούθησης βίντεο (WV) και προβολής εικόνων (VI) ήταν επίσης αρκετά συχνές, γεγονός που, σε συνδυασμό με το ότι η πιθανότητα της ενέργειας «ανάγνωση ενός άρθρου (RA)» είναι μικρότερη από 0.001, αποτελεί πιθανή ένδειξη ότι η έμπνευση προέρχεται περισσότερο από οπτικά ή στοιχεία κειμένου αντί από τη λεπτομερή ανάγνωση πληροφοριών σχετικά με ένα αποτέλεσμα (Πίνακας 23).

5.4.1.2 Δεδομένα Προσωπικότητας

Ο Πίνακας 24 δείχνει τη συσχέτιση μεταξύ χαρακτηριστικών προσωπικότητας των συμμετεχόντων και χαρακτηριστικών μιας συνεδρίας (αριθμός ενεργειών, αριθμός ερωτημάτων, χρόνος δέσμευσης). Συγκεκριμένα, πραγματοποιούμε ανάλυση ζεύγους-συσχέτισης μεταξύ των χαρακτηριστικών της συνεδρίας και των σκορ για κάθε παράγοντα προσωπικότητας του Μοντέλου των Πέντε Παραγόντων (Big Five Factor).

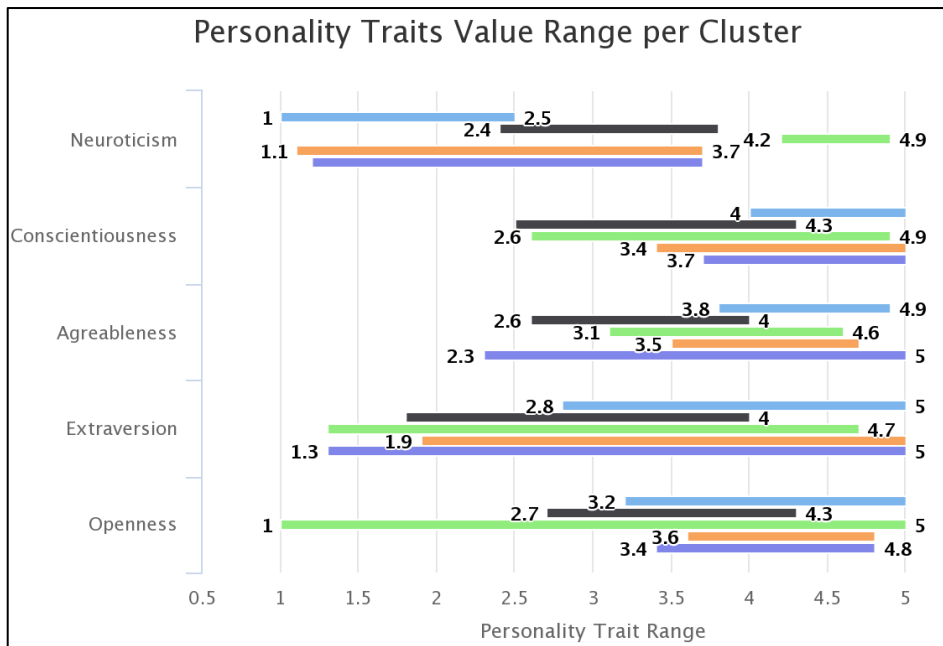
Πίνακας 24 Συσχέτιση χαρακτηριστικών προσωπικότητας των συμμετεχόντων και χαρακτηριστικών συνεδρίας

Personality Trait	#Actions	#Queries	Eng. Time
Neuroticism	0.98	0.97	-0.13
Conscientiousness	0.98	0.95	0.021
Agreeableness	0.93	0.97	-0.087
Extraversion	0.98	0.96	-0.042
Openness	0.98	0.95	-0.007

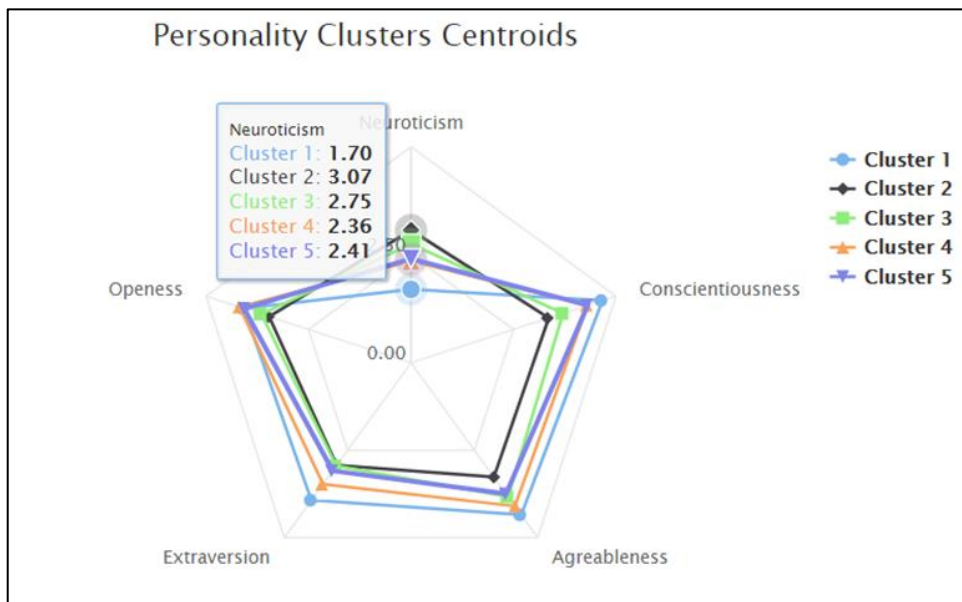
Προκειμένου να υπολογιστούν οι συσχετισμοί, χρησιμοποιούμε τη μη παραμετρική συσχέτιση Spearman λόγω του γεγονότος ότι η φύση της κατανομής των χαρακτηριστικών της συνεδρίας είναι μη κανονική και τα δεδομένα εξαιρετικά επικαλυμένα. Για τον έλεγχο της στατιστικής σημαντικότητας της συσχέτισης εφαρμόσαμε μετασχηματισμό Fisher (Caruso & Cliff, 1997).

Η ευσυνειδησία, η εξωστρέφεια, η ευγένεια, ο νευρωτισμός και το ανοιχτό πνεύμα εμφανίστηκαν θετικά μονοτονικά συσχετισμένα χαρακτηριστικά με τον αριθμό των ερωτημάτων και τον αριθμό των ενεργειών ($p < 0.05$), ενώ δεν συσχετίστηκαν σημαντικά με το χρόνο δέσμευσης. Συνολικά, τα χαρακτηριστικά της προσωπικότητας φαίνεται ότι επηρεάζουν τα χαρακτηριστικά διερευνητικής συμπεριφοράς και ως εκ τούτου προχωράμε σε συσταδοποίηση χρηστών με βάση τα χαρακτηριστικά της προσωπικότητάς τους.

Για να κατατάξουμε τους συμμετέχοντες σε ομάδες παρόμοιων χρηστών με βάση τα χαρακτηριστικά της προσωπικότητάς τους, πραγματοποιήσαμε συσταδοποίηση χωρίς επίβλεψη χρησιμοποιώντας μία παραλλαγή του αλγορίθμου Μεγιστοποίησης Προσδοκίας - Expectation Maximization (Dempster, Laird & Rubin, 1977), τον K-Means, έχοντας ως κριτήριο την ομοιότητα των βαθμολογιών προσωπικότητας σε κάθε παράγοντα του μοντέλου Big Five και χρησιμοποιώντας ως μετρική το τετράγωνο της Ευκλείδειας απόστασης. Αυτή η μέθοδος είχε ως αποτέλεσμα να παραχθούν συνολικά πέντε ομάδες χρηστών. Οι συστάδες που δημιουργήθηκαν απεικονίζονται στην εικόνα 36 και 37. Η εικόνα 36 δείχνει την περιοχή τιμών κάθε χαρακτηριστικού προσωπικότητας για κάθε ομάδα. Για παράδειγμα, οι συστάδες 4 και 5 είναι πανομοιότυπες όσον αφορά το Νευρωτισμό, αλλά διαφέρουν κυρίως όσον αφορά την Ευγένεια και λιγότερο όσον αφορά την Εξωστρέφεια. Το διάγραμμα αράχνης της εικόνας 37 απεικονίζει τις μέσες τιμές των χαρακτηριστικών προσωπικότητας ανά συστάδα (κεντροειδή). Ο νευρωτισμός φαίνεται να είναι το χαρακτηριστικό που διαφέρει περισσότερο μεταξύ των πέντε ομάδων.



Εικόνα 36 Το διάγραμμα ράβδων απεικονίζει το εύρος τιμών κάθε χαρακτηριστικού προσωπικότητας ανά συστάδα. (Ανοιχτό μπλε για τη συστάδα 1, μαύρο για τη συστάδα 2, πράσινο για τη συστάδα 3, πορτοκαλί για τη συστάδα 4 και μοβ για τη συστάδα 5.)



Εικόνα 37 Το διάγραμμα Spiderweb περιγράφει τα κεντροειδή κάθε συστάδας. Τα διανύσματα κεντροειδών αποτελούνται από τις μέσες τιμές κάθε χαρακτηριστικού προσωπικότητας στην αντίστοιχη συστάδα

Οι Πίνακες 25, 26 και 27 συνοψίζουν τα χαρακτηριστικά διερευνητικής συμπεριφοράς (δηλ. το μέσο αριθμό ενεργειών, ερωτημάτων και χρόνου εμπλοκής) κάθε συστάδας.

Αριθμός Ενεργειών

Ο έλεγχος ανεξάρτητων δειγμάτων t-test έδειξε ότι οι χρήστες της συστάδας 1 πραγματοποίησαν περισσότερες ενέργειες στα πλαίσια της συνεδρίας ($M = 19.8$, $SD = 24.7$) σε σύγκριση με τη συστάδα 3 ($M = 13.1$, $SD = 8.5$), $t^{10} (75) = 1.99$, $p < 0.05$.

Αριθμός Ερωτημάτων

Αναφορικά με τον αριθμο ερωτημάτων, η συστάδα 1 ($M = 3.22$, $SD = 2.34$) και η συστάδα 5 απέδωσαν καλύτερα σε σχέση με τη συστάδα 2 ($M = 3.48$, $SD = 2.20$), $t^{11}_{1-2} (104) = 1.98$, $t_{5-2}^{12} (52) = 2.00$, $p < 0.05$, καθώς και σε σύγκριση με τη συστάδα 4 ($M = 2.76$, $SD = 1.65$), $t_{1-4} (101) = 1.98$, $t_{5-4} (52) = 2.00$, $p < 0.05$. Δεν βρέθηκε στατιστική σημαντικότητα συγκρίνοντας τη συστάδα 1 με τη συστάδα 3 ή τη συστάδα 5 με τη συστάδα 3.

Χρόνος Δέσμευσης (Engagement Time)

Αναφορικά με το «Χρόνο Δέσμευσης», η συστάδα 1 ($M = 559$, $SD = 524$) και η συστάδα 5 ($M = 579$, $SD = 523$) απέδωσαν καλύτερα σε σύγκριση και με τη συστάδα 2 ($M = 360$,

¹⁰ t αντιπροσωπεύει t-statistic. Οι βαθμοί ελευθερίας αντιπροσωπεύονται από τον αριθμό μέσα στην παρένθεση.

¹¹ t_{x-y} αντιπροσωπεύει την τιμή t statistic που παράγεται συγκρίνοντας Cluster X με Cluster Y. Χρησιμοποιούμε το συμβολισμό, όταν αναφερόμαστε σε περισσότερες από μία συγκρίσεις μέσα σε μία πρόταση.

SD = 374), $t_{1-2}(104) = 1,98$, $t_{5-2}(49) = 2.00$, $p < 0.05$ και τη συστάδα 3 ($M=353$, $SD=292$), $t_{1-3}(94) = 1.98$, $t_{5-3}(43) = 2.01$, $p < 0.05$.

Παρατηρούμε ότι το χαρακτηριστικό ευσυνειδησίας είναι υψηλότερο στις συστάδες 1, 4 και 5 που είναι και οι συστάδες με τις υψηλότερες βαθμολογίες στα χαρακτηριστικά προσωπικότητας από ότι στις συστάδες 2 και 3.

Πίνακας 25 Ανάλυση χαρακτηριστικών συνεδρίας ανά συστάδα χρηστών.

Cluster	#Actions		#Queries		Eng. Time	
	Mean	SD	Mean	SD	Mean	SD
CI1	19.8	24.7	3.22	2.34	559	524
CI2	14.8	11.7	2.49	1.70	360	374
CI3	13.1	8.5	2.55	3.20	353	292
CI4	16.2	15.3	2.76	1.65	420	326
CI5	17.9	12.6	3.48	2.20	579	523

Πίνακας 26 Μέσες τιμές χαρακτηριστικών προσωπικότητας ανά σύμπλεγμα (κεντροειδή).

Personality Trait (m)	CI1	CI2	CI3	CI4	CI5
Neuroticism	1.70	3.07	2.75	2.36	2.41
Conscientiousness	4.70	3.36	3.65	4.31	4.47
Agreeableness	4.35	3.27	3.83	4.10	3.75
Extraversion	3.93	2.93	2.95	3.47	3.09
Openness	4.08	3.46	3.65	4.18	4.07

Στη συνέχεια, εφαρμόσαμε ανάλυση συσχέτισης για κάθε ομάδα (Πίνακας 27). Συγκεκριμένα, χρησιμοποιήσαμε τη μη παραμετρική συσχέτιση Spearman για τον υπολογισμό των συσχετισμών. Τα αποτελέσματα δείχνουν ότι η ευσυνειδησία συσχετίζεται σε όλες τις ομάδες σε μεγάλο βαθμό με τον «Αριθμό ενεργειών» που εκτελείται από το χρήστη κατά τη διάρκεια της συνεδρίας ($p < .01$). Ιδιαίτερα, η ευσυνειδησία και η ευγένεια συνδέονται σε μεγάλο βαθμό με τον «Αριθμό ερωτημάτων» που υποβλήθηκαν κατά τη διάρκεια της συνόδου στη συστάδα 1 ($r_s(57)^{13} = 0.94, p < 0.01$ and $r_s(57) = 0.95, p < 0.01$), η οποία αποτελείται από χρήστες με υψηλές βαθμολογίες στην ευσυνειδησία, αλλά χαμηλές στο νευρωτισμό. Στις συστάδες 4 και 5, με χρήστες που σκοράρουν ψηλά στο ανοιχτό πνεύμα, μπορούμε να δούμε ότι το ανοιχτό πνεύμα είναι θετικά συσχετισμένο με τον «Αριθμό Ερωτημάτων» που διενεργούνται ($r_s(43) = 0.91, p < .01$ and $r_s(30) = .87, p < 0.01$). Το ίδιο συμβαίνει στη συστάδα 4 με την ευγένεια και τον «Αριθμό Ερωτημάτων» ($r_s(43) = 0.91, p < 0.01$). Φαίνεται ότι η εξωστρέφεια δεν παίζει κυρίαρχο ρόλο παρόλο που είναι θετικά συσχετισμένη με τον «Αριθμό Ενεργειών» και με τον «Αριθμό Ερωτημάτων» σε όλες τις συστάδες, αλλά είναι θετικά συσχετισμένη με το χρόνο δέσμευσης ($r_s(43) = 0.4, p < 0.01$) μόνο στη συστάδα 4 (πιο ευσυνειδητοί χρήστες) σε σύγκριση με άλλες συστάδες. Στη συστάδα 3 με πιο νευρωτικούς χρήστες, ο νευρωτισμός είναι αρνητικά γραμμικά συσχετισμός με τον «Αριθμό Ενεργειών» ($r(43)^{14} = -0.34, p < 0.05$) και τον «Αριθμό Ερωτημάτων» ($r(43) = -0.26, p = 0.574$).

¹³ r_s αντιπροσωπεύει τη συσχέτιση Spearman's. Οι βαθμοί ελευθερίας δηλώνονται από τον αριθμό στην παρένθεση.

¹⁴ r αντιπροσωπεύει το συντελεστή συσχέτισης Pearson. Οι βαθμοί ελευθερίας δηλώνονται από τον αριθμό στην παρένθεση.

Πίνακας 27 Συσχέτιση Spearman's ή Pearson χαρακτηριστικών προσωπικότητας με χαρακτηριστικά συνεδρίας ανά συστάδα

Action	#Actions					#Queries					Engagement Time				
	C11	C12	C13	C14	C15	C11	C12	C13	C14	C15	C11	C12	C13	C14	C15
Neuroticism	0.78**	0.79**	-0.34*	0.79**	0.82**	0.78	0.91**	-0.26	0.78**	0.78**	-0.39	-0.12	-0.02	0.18	-0.09
Conscientiousness	0.82**	0.82**	0.85**	0.83**	0.71**	0.94**	0.83**	0.69**	0.89**	0.81**	-0.27	0.04	0.15	0.17	0.04
Agreeableness	0.83**	0.73**	0.84**	0.80**	0.69**	0.95**	0.94**	0.87**	0.91**	0.78**	-0.3	-0.26	-0.16	0.09	0.09
Extraversion	0.84**	0.82**	0.83**	0.79**	0.57**	0.78**	0.82**	0.64**	0.61**	0.48**	0.02	0.01	0.10	0.40**	0.01
Openness	0.84**	0.79**	0.71**	0.81**	0.69**	0.83**	0.85**	0.49**	0.91**	0.87**	-0.09	-0.02	0.14	0.12	-0.09

Linear Pearson correlation typeset in **bold**

* $p < .05$

** $p < .01$

5.4.2 Εκπαίδευση Μοντέλων

5.4.2.1 Προεπεξεργασία Δεδομένων και Εκμάθηση Παραμέτρων

Για την εκπαίδευση του μοντέλου χρησιμοποιήσαμε το Matlab και συγκεκριμένα μια εφαρμογή του αλγορίθμου Baum Welch που παρέχεται στο HMM Toolbox για το Matlab. Για να απλουστευθεί η αναπαράσταση των δεδομένων, προτού εφαρμόσουμε τον αλγόριθμο Baum-Welch, μετασηματίσαμε τα καταγεγραμμένα δεδομένα της εφαρμογής μας βάσει του Πίνακα 19 που σχηματίζει ακολουθίες αριθμών που αντιστοιχούν σε ενέργειες. Στην περίπτωση μας, έχουμε μια ακολουθία πιθανών ενεργειών χρήστη από το 1 έως το 15, όπως φαίνεται στον Πίνακα 19 του Κεφαλαίου 5.3.1. Το Κρυφό-Μαρκοβιανό-Μοντέλο υποθέτει ότι έχουμε επίσης ένα σύνολο κρυφών καταστάσεων και κάθε ενέργεια παράγεται από μια αντίστοιχη κρυφή κατάσταση, αλλά διαφορετικές ενέργειες μπορούν να δημιουργηθούν από την ίδια κρυφή κατάσταση με διαφορετικές πιθανότητες. Οι καταγεγραμμένες ακολουθίες των ενεργειών των χρηστών κωδικοποιήθηκαν χρησιμοποιώντας τις κωδικοποιήσεις αλληλεπιδράσεων που απεικονίζονται στον προαναφερθέντα πίνακα. Ένα κρίσιμο ζήτημα ήταν ότι στο σύνολο δεδομένων εκπαίδευσης δεν είχαν όλες οι συνεδρίες (δηλαδή αλληλουχίες αλληλεπιδράσεων) το ίδιο μήκος. Υπήρχαν αλληλουχίες που αποτελούνταν από μόνο 3 παρατηρήσεις, ενώ η μακρύτερη αλληλουχία ήταν 180 παρατηρήσεις. Επιπλέον, έπρεπε να υπολογίσουμε τις αρχικές παραμέτρους για το μοντέλο. Ο αλγόριθμος ξεκινά με μια αρχική εκτίμηση μοντέλου και ενημερώνεται επανειλημμένα μέχρι τη σύγκλιση. Δεδομένου ότι ο αλγόριθμος Baum-Welch αναζητά ένα τοπικό βέλτιστο HMM σε σχέση με τη λειτουργία πιθανότητας, η επιλογή ενός αρχικού μοντέλου είναι κρίσιμη. Βλέποντας ότι στην περίπτωση μας δεν υπήρχε λογική εικασία, αρχικοποιούμε $\Theta = (A, B, \pi)$ πρώτα με τιμές τέτοιες ώστε $\pi_i \approx 1 / N$ και $a_{ij} \approx 1 / N$ και $b_j(k) \approx 1 / M$ και τέλος με τυχαίες τιμές. Είναι σημαντικό να αποφευχθεί ο καθορισμός ίσων τιμών για τα A, B και π , αφού οι ακριβώς ομοιόμορφες τιμές θα οδηγήσουν σε ένα τοπικό μέγιστο από το οποίο το μοντέλο δεν μπορεί να αναρριχηθεί (Stamp 2004). Μια εναλλακτική προσέγγιση για την αρχικοποίηση του

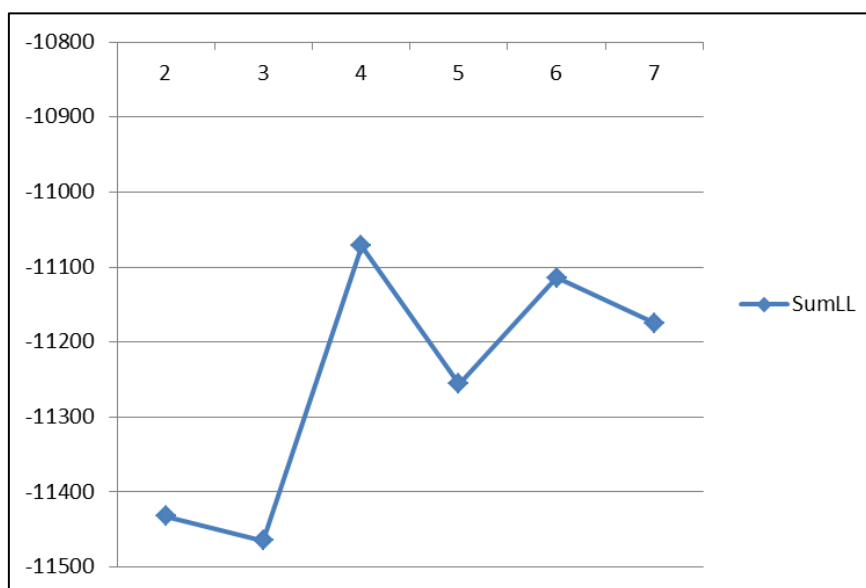
μοντέλου θα ήταν η υιοθέτηση τυχαίων τιμών για τα A και π και η χρήση της κατηγοριοποίησης ομαδοποίησης K -μέσων για μια καλύτερη εκτίμηση του πίνακα πιθανοτήτων εκπομπών (Rabiner, 1989), προσέγγιση που όμως δεν ακολουθήσαμε. Έπειτα, έπρεπε να αποφασίσουμε για τον αριθμό των κρυφών καταστάσεων του HMM μας, δεδομένου ότι με βάση τις ακολουθίες, ο αλγόριθμος Baum Welch χρειάζεται επίσης να έχει ως είσοδο τον αριθμό των καταστάσεων (N) και τον αριθμό των παρατηρήσεων (M). Δεδομένου ότι δεν είναι δυνατό να είμαστε σε θέση να γνωρίζουμε τον αριθμό των καταστάσεων που ταιριάζουν στα δεδομένα μας εκ των προτέρων, δοκιμάσαμε με n καταστάσεις, όπου $2 \leq n \leq 7$. Για να αποφευχθεί το υπερταίριασμα (overfitting) που μπορεί να συμβεί με μικρά σύνολα δεδομένων, χρησιμοποιήσαμε "Leave one out" Διασταυρούμενη Επικύρωση (Cross Validation), η οποία είναι K -fold Cross Validation (CV) όπου το K ισούται με τον αριθμό των ακολουθιών δεδομένων (δηλαδή 224 φορές CV) και επιλέγει το πιο γενικευμένο μοντέλο. Σε κάθε επανάληψη, διατηρήσαμε μία ακολουθία ως δεδομένα δοκιμής και εκπαιδεύσαμε ένα μοντέλο n -state HMM χρησιμοποιώντας τις υπόλοιπες ακολουθίες 223. Για τον υπολογισμό των πιθανοτήτων ελέγχου των δεδομένων, λαμβάνοντας υπόψη το εκπαιδευμένο μοντέλο, εφαρμόσαμε τον αλγόριθμο προώθησης (Rabiner, 1989).

5.4.2.2 Επιλογή Μοντέλου

Για την επιλογή του βέλτιστου αριθμού καταστάσεων, χρησιμοποιήσαμε ένα δείκτη καλής προσαρμογής, που στην περίπτωσή μας ήταν η λογαριθμική πιθανότητα του εκπαιδευμένου μοντέλου. Το άθροισμα των λογαριθμικών πιθανοτήτων που προέκυψαν από κάθε fold μέσω της διασταυρούμενης επικύρωσης χρησιμοποιήθηκε ως το συνολικό μέτρο ή του δείκτη καλής προσαρμογής (Celeux and Durand, 2008, Wang, 2008). Τα μέτρα καλής προσαρμογής τυπικά συνοψίζουν την απόκλιση μεταξύ των τιμών που έχουν παρατηρηθεί και των τιμών που αναμένονται στο υπό εξέταση μοντέλο.

5.4.2.2.1 Γενικευμένο Μοντέλο

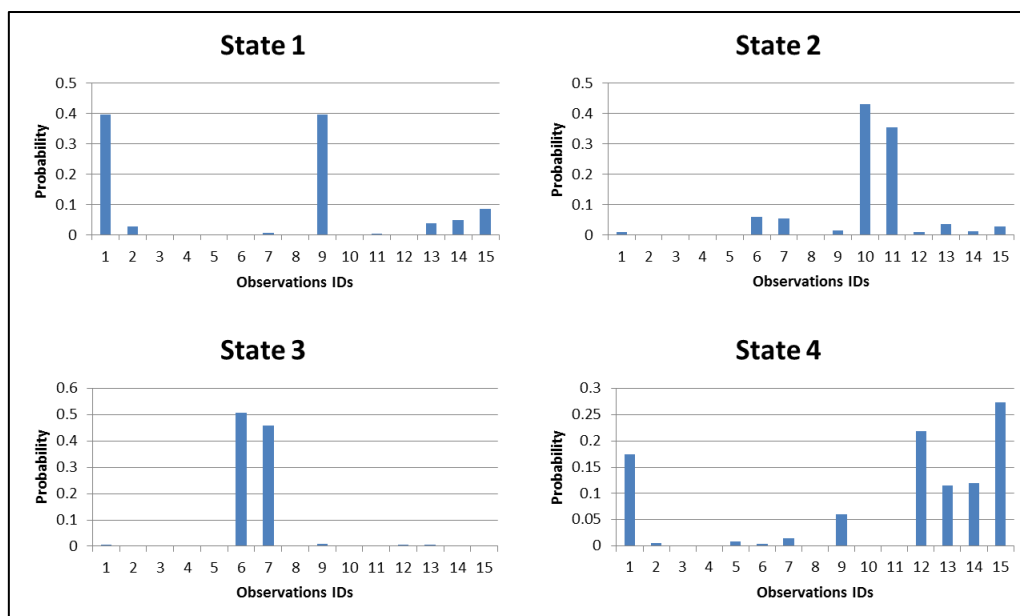
Τα αποτελέσματα που προέκυψαν από τη διασταυρούμενη επικύρωση για $2 \leq n \leq 6$ συγκλίνουν στον ίδιο αριθμό κρυφών καταστάσεων. Το μοντέλο 4-state HMM φαίνεται να επικρατεί με την υψηλότερη αξία του καλού δείκτη προσαρμογής και, κατά συνέπεια, πρόκειται να είναι το εκπαιδευμένο μοντέλο που θα χρησιμοποιηθεί για τα επόμενα πειράματά μας. Η εικόνα 38 παρουσιάζει το άθροισμα των λογαριθμικών πιθανοτήτων σε σχέση με τον αριθμό των κρυφών καταστάσεων.



Εικόνα 38 Άθροισμα της εκτίμησης της Λογαριθμικής Πιθανότητας των παραμέτρων του HMM

Το διάγραμμα στην εικόνα 39 απεικονίζει την κατανομή των πιθανών εκπομπών στο επιλεγμένο μοντέλο 4 καταστάσεων. Οι κατανομές αποκαλύπτουν τον τρόπο συγκέντρωσης των ενεργειών γύρω από τις καταστάσεις στο HMM.

Λαμβάνοντας υπόψη την Εικόνα 39 και τον Πίνακα 28, καταλήγουμε στο συμπέρασμα ότι η πρώτη κατάσταση είναι η κατάσταση κατά την οποία οι χρήστες ερωτούν το σύστημα για πληροφορίες και αυτό μπορεί να χαρακτηριστεί ως κατάσταση «**Διεξαγωγή Ερωτημάτων (Query)**». Σε μια προσπάθεια εμπειρικής ταξινόμησης όλων των πιθανών ενεργειών σε σημαντικές κατηγορίες ή καταστάσεις, καταλήγουμε στο συμπέρασμα ότι οι ενέργειες με κωδικό από 2 έως 5 και η ενέργεια



Εικόνα 39 Διαγράμματα Πιθανοτήτων Εκπομπών για HMM 4 καταστάσεων με προσεγγιστικές $1/N$, $1/M$ αρχικές παραμέτρους.

Πίνακας 28 Απεικόνιση των υψηλότερων τιμών Πιθανοτήτων Εκπομπών για κάθε κατάσταση.

States	Q	IC	VI	SH
State 1 (Query)	0.40			
State 2 (Capturing)		0.43		
State 3 (Examination)			0.51	
State 4 (Scanning)				0.28

15 σχετίζονται κυρίως με τις ενέργειες σάρωσης, οι οποίες μπορούν να χαρτογραφηθούν περαιτέρω σε μια κατάσταση «**Αναζήτηση ενδείξεων (Search for clues)**». Οι ενέργειες από 6 έως 8 που οδηγούν στην προβολή της αρχικής πηγής κάθε

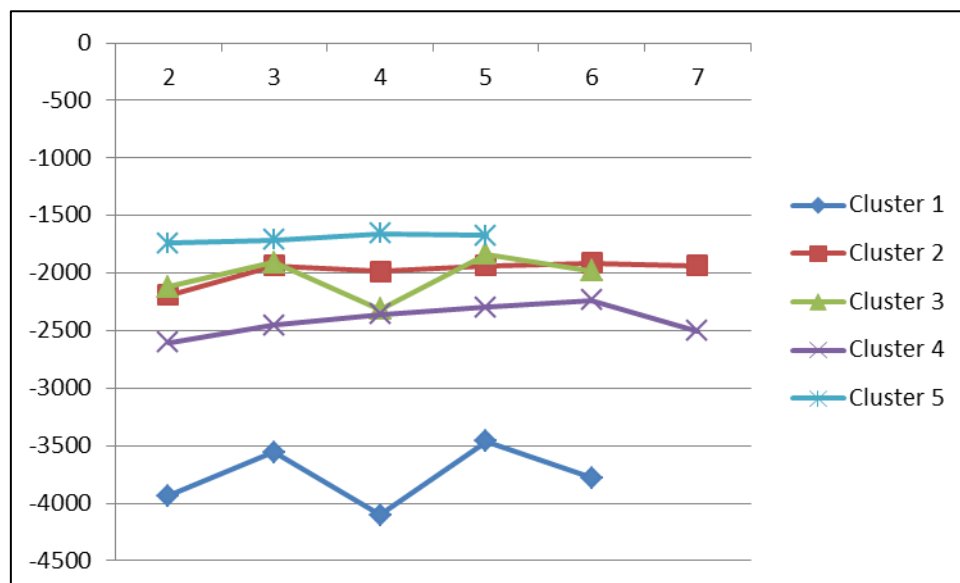
στοιχείου πάνω στο οποίο έχουν κάνει κλικ έχουν ως επί το πλείστον σχέση με την κατάσταση «**Εξέταση των Ενδείξεων (Clues Examination)**». Οι ενέργειες 9 έως 12, οι οποίες προσθέτουν ένα στοιχείο στη συλλογή του χρήστη, μπορούν να σχετίζονται με την κατάσταση «**Καταγραφή Ενδείξεων (Clues Capturing)**» ενώ η ενέργεια 14 αντιστοιχεί στην κατάσταση «**Συνδυασμός Ενδείξεων (Clues combination)**». Στην **κατάσταση 2**, η πιθανότητα για τον χρήστη να προσθέσει εικόνες ή βίντεο στη συλλογή είναι υψηλότερη και ως εκ τούτου μπορούμε να την ονομάσουμε κατάσταση «**Καταγραφή Ενδείξεων (Clues Capturing)**». Η **κατάσταση 3** είναι η κατάσταση όπου οι χρήστες βλέπουν βίντεο ή εικόνες με την υψηλότερη πιθανότητα και αυτό μπορεί να χαρακτηριστεί ως κατάσταση «**Εξέταση των Ενδείξεων (Clues Examination)**». Σε αυτήν την κατάσταση, οι χρήστες πατούν τυχαία κλικ με την υψηλότερη πιθανότητα, πράγμα που σημαίνει ότι ψάχνουν για περισσότερες ενδείξεις. Αυτό μπορεί να αντιστοιχιστεί στην κατάσταση «**Σάρωση για ενδείξεις (Scanning for Clues)**».

5.4.2.2 Μοντέλα με βάση την Προσωπικότητα

Η εικόνα 40 απεικονίζει το άθροισμα των τιμών των λογαριθμικών πιθανοτήτων σε σχέση με τον αριθμό των κρυφών καταστάσεων για κάθε συστάδα. Όπως απεικονίζεται μπορούμε να δούμε ότι το κριτήριό μας έχει τη βέλτιστη τιμή όταν ο αριθμός των κρυφών καταστάσεων είναι 5 για τη συστάδα 1, 6 για τη συστάδα 2, 5 για τη συστάδα 3, 6 για τη συστάδα 4 και 4 για τη συστάδα 5.

Όπως βλέπουμε στην εικόνα 41 ένα κοινό χαρακτηριστικό των τεσσάρων από τις πέντε συστάδες είναι η κατάσταση στην οποία οι χρήστες εκτελούν ερωτήματα με την υψηλότερη πιθανότητα, κατάσταση «**Διεξαγωγή Ερωτημάτων (Query)**». Οι καταστάσεις «**Διεξαγωγή Ερωτημάτων (Query)**», «**Εξέταση των Ενδείξεων (Clues Examination)**» και «**Καταγραφή Ενδείξεων (Clues Capturing)**» είναι γενικά παρούσες σε όλες τις συστάδες όπως επίσης φαίνεται στην εικόνα 39. Η κατάσταση «**Σάρωση για ενδείξεις (Scanning for Clues)**» εμφανίζεται σε όλες τις συστάδες εκτός από τη

συστάδα 2 και η κατάσταση «Συνδυασμός Ενδείξεων (Clues combination)» εμφανίζεται μόνο στη συστάδα 2. Υπάρχει επίσης μία κατάσταση «Αδράνεια (Idle



Εικόνα 40 Άθροισμα της Λογαριθμικής Πιθανότητας των HMM παραμέτρων.

state)» η οποία αντιστοιχεί στον κωδικό 13 (που δεν είναι ενέργεια παρά η κατάσταση κατά την οποία ο χρήστης παραμένει ανενεργός ενώ βρίσκεται στη συνεδρία για πάνω από 1 λεπτό).

Cluster 1	Q	SH	VI	WV	TC	Cluster 3	Q	SH	VI	TC	LC		
N=1	0.96					N=1	0.99						
N=2		0.54				N=2		0.55					
N=3			0.56			N=3			0.24				
N=4				0.4		N=4				0.49			
N=5					0.38	N=5					0.48		
Cluster 2	Q	WV	TC	IC	EC	I	Cluster 4	Q	SH	VI	WV	TC	LC
N=1	0.98						N=1	0.98					
N=2		0.49					N=2		0.62				
N=3			0.55				N=3			0.58			
N=4				0.42			N=4				0.53		
N=5					0.35		N=5					0.56	
N=6						0.66	N=6						0.35
				Cluster 5	SH	VI	TC	IC					
				N=1	0.34								
				N=2		0.62							
				N=3			0.44						
				N=4				0.46					

Εικόνα 41 Οι υψηλότερες τιμές Πιθανοτήτων Εκπομπών για κάθε κατάσταση και συστάδα.

Επίσης, λαμβάνοντας υπόψιν τις υπόλοιπες πιθανότητες κάθε ενέργειας σε κάθε κατάσταση φαίνεται ότι οι χρήστες της συστάδας 3 (που είναι οι πιο νευρωτικοί) είναι αυτοί που αλληλεπιδρούν περισσότερο με άρθρα από ότι οι χρήστες των συστάδων 2,4 και 5, οι οποίοι μπορούν να χαρακτηριστούν ως πιο οπτικοί τύποι με υψηλότερες πιθανότητες να δουν μία εικόνα ή να παρακολουθήσουν ένα video. Επιπρόσθετα, οι χρήστες της συστάδας 4 έχουν σχεδόν όμοια πιθανότητα να αλληλεπιδράσουν με όρους και άρθρα Ιστού, εικόνες και βίντεο. Λεπτομερέστερα, οι χρήστες των συστάδων 1,2 και 4 είναι αυτοί οι οποίοι έχουν μεγαλύτερη πιθανότητα να παρακολουθήσουν ή να προσθέσουν ένα βίντεο στη συλλογή τους. Εμφανίζουν επίσης υψηλή πιθανότητα να προσθέσουν έναν όρο στις συλλογές τους, πιθανόν επειδή αυτή είναι η βασική λειτουργικότητα του εργαλείου ή επειδή το να αλληλοεπιδρά κανείς με όρους είναι πιο φιλικό και άμεσο σε μία συνεδρία διερευνητικής περιήγησης. Μόνο στη συστάδα 2, η οποία αναπαρίσταται από μετριοπαθείς χρήστες ως προς τις βαθμολογίες χαρακτηριστικών προσωπικότητας παρουσιάζεται υψηλότερη πιθανότητα για διερεύνηση των καταγεγραμμένων αντικειμένων. Είναι αξιοσημείωτο ότι σε καμία συστάδα δεν προτιμώνται οι ενέργειες σάρωσης στοιχείων, είτε διαβάζοντας μία περίληψη ενός άρθρου είτε σχετικών tweets, που περιλαμβάνουν ένα συγκεκριμένο όρο. Ίσως αυτό να οφείλεται στη φύση των συγκεκριμένων χρηστών, οι οποίοι είναι περισσότερο προσανατολισμένοι στην ολοκλήρωση της εργασίας στο συντομότερο χρόνο.

Αν και η ανάλυση των μοντέλων αποκάλυψε 4,5 ή 6 πιθανές κρυφές καταστάσεις – αναλόγως της συστάδας – πίσω από την διερευνητική συμπεριφορά ενός χρήστη, υπάρχει ένδειξη ότι αποκαλύπτονται βασικά δύο γενικοί τύποι καταστάσεων: (i) Η κατάσταση στην οποία ο χρήστης είναι σε μία ροή ενεργειών , όπου κάθε ενέργεια ενεργοποιεί μία άλλη και τελικά προκύπτει μία επιμηκυμένη συνεδρία ενεργειών και (ii) η κατάσταση στην οποία ο χρήστης είτε δεν ενδιαφέρεται για το εργαλείο, τις δυνατότητες που παρέχει ή τα αποτελέσματα που επιστρέφει, το οποίο αναπόφευκτα οδηγεί σε μικρότερες συνεδρίες και τελικά πιο αφηρημένες

διερευνήσεις χωρίς στόχο, πράγμα το οποίο απεικονίζεται ειδικά στην κατάσταση 6 της συστάδας 2.

5.4.3 Παραγωγή Προτάσεων

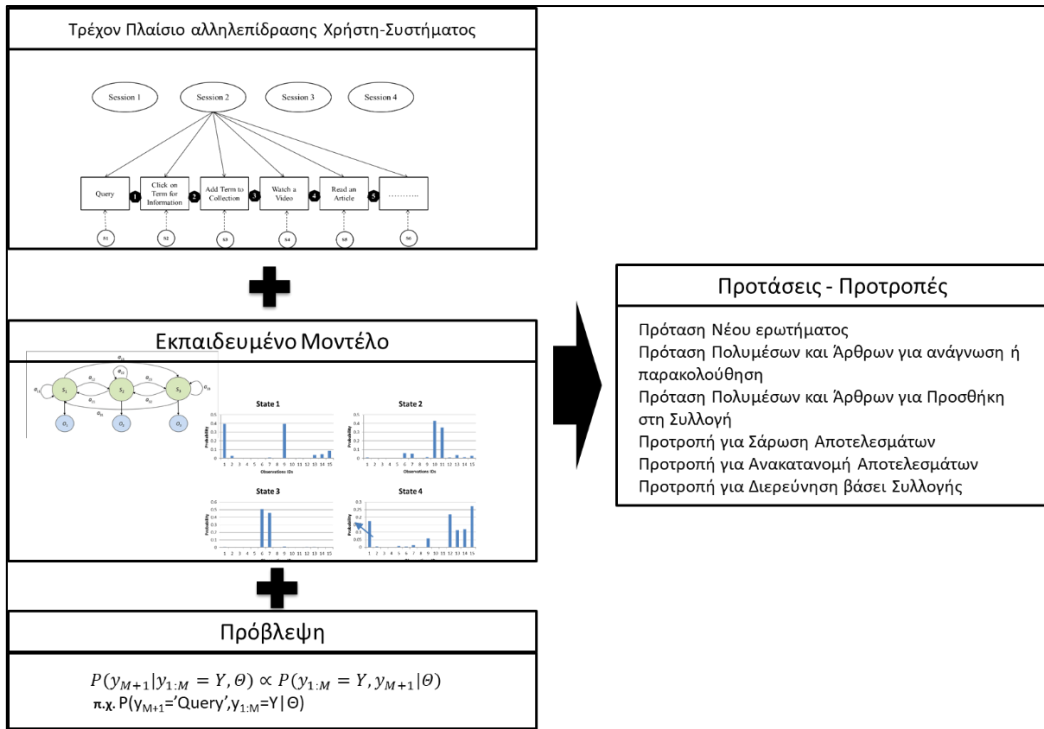
Οι προτάσεις που παράγονται με την προτεινόμενη μέθοδο αξιοποιούν τα μοντέλα που περιγράφηκαν στην αμέσως προηγούμενη ενότητα και κατά συνέπεια είναι δύο τύπων:

- Προτάσεις βάσει του Γενικευμένου μοντέλου
- Προτάσεις βάσει των μοντέλων Προσωπικότητας

5.4.3.1 Προτάσεις βάσει Γενικευμένου Μοντέλου

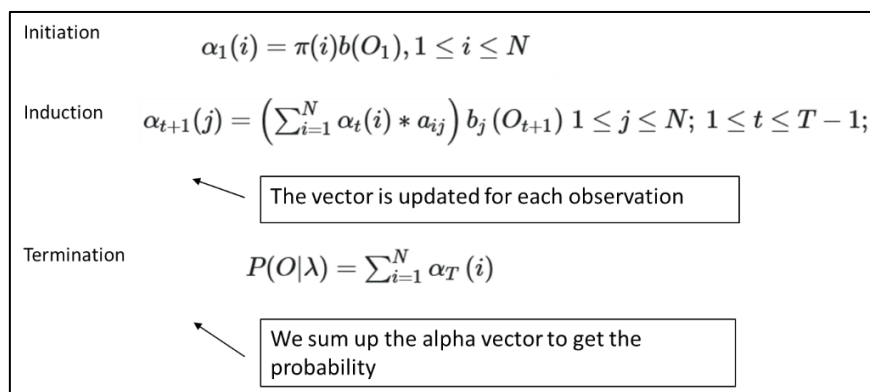
Παρακάτω συνοψίζονται τα βήματα της μεθοδολογίας παραγωγής προτάσεων βάσει του Γενικευμένου μοντέλου (Εικόνα 42):

1. Παράγεται το Γενικευμένο Κρυφό Μαρκοβιανό Μοντέλο (βλ. προηγούμενη ενότητα)
2. Συλλέγεται η τρέχουσα αλληλουχία αλληλεπιδράσεων του χρήστη με το σύστημα και **τροφοδοτούμε**
3. τον αλγόριθμο πρόβλεψης (Εικόνα 43) της πιθανότερης επόμενης ενέργειας αλληλεπίδρασης του χρήστη με το σύστημα (αλγόριθμος Προώθησης-Forward algorithm, βλ. Παράρτημα 5).



Εικόνα 42 Παραγωγή Προτάσεων βάσει Γενικευμένου Μοντέλου

Εφόσον ο στόχος μας είναι να προσδιορίσουμε την πιθανότητα μίας δοσμένης ακολουθίας παρατηρήσεων, αυτό που συμβαίνει στον αλγόριθμο προώθησης είναι ότι διαιρούμε την αλληλουχία σε υπο-ακολουθίες, υπολογίζουμε τις πιθανότητες και τις αποθηκεύουμε σε πίνακες για μετέπειτα χρήση. Η πιθανότητα της αρχικής ακολουθίας προκύπτει από το συνδυασμό των πιθανοτήτων των υπο-ακολουθιών.



Εικόνα 43 Αλγόριθμος Προώθησης (Forward Algorithm)

Όπως αναφέρθηκε και στην ενότητα 5.3.3, αναλόγως του τύπου της ενέργειας που προβλέπεται από τον αλγόριθμο ως πιο πιθανή, παρέχονται προτάσεις ή προτροπές. Οι προτάσεις αξιοποιούν το σημασιολογικό πλαίσιο της συνεδρίας, καθώς πρόκειται για προτάσεις με περιεχόμενο (για παράδειγμα πρόταση ενός ερωτήματος) ενώ οι προτροπές προτρέπουν το χρήστη να ακολουθήσει μία ενέργεια που δε σχετίζεται με περιεχόμενο (για παράδειγμα η ενέργεια του να διερευνηθεί περαιτέρω η συλλογή των αντικειμένων που ο χρήστης έχει προσθέσει στη συλλογή του). Σχετικά παραδείγματα αναφέρονται παρακάτω στην ενότητα 5.4.4.

5.4.3.2 Προτάσεις βάσει Μοντέλων Προσωπικότητας

Η περίπτωση αυτή έχει ομοιότητες αλλά και διαφορές σε σχέση με την περίπτωση του γενικευμένου μοντέλου.

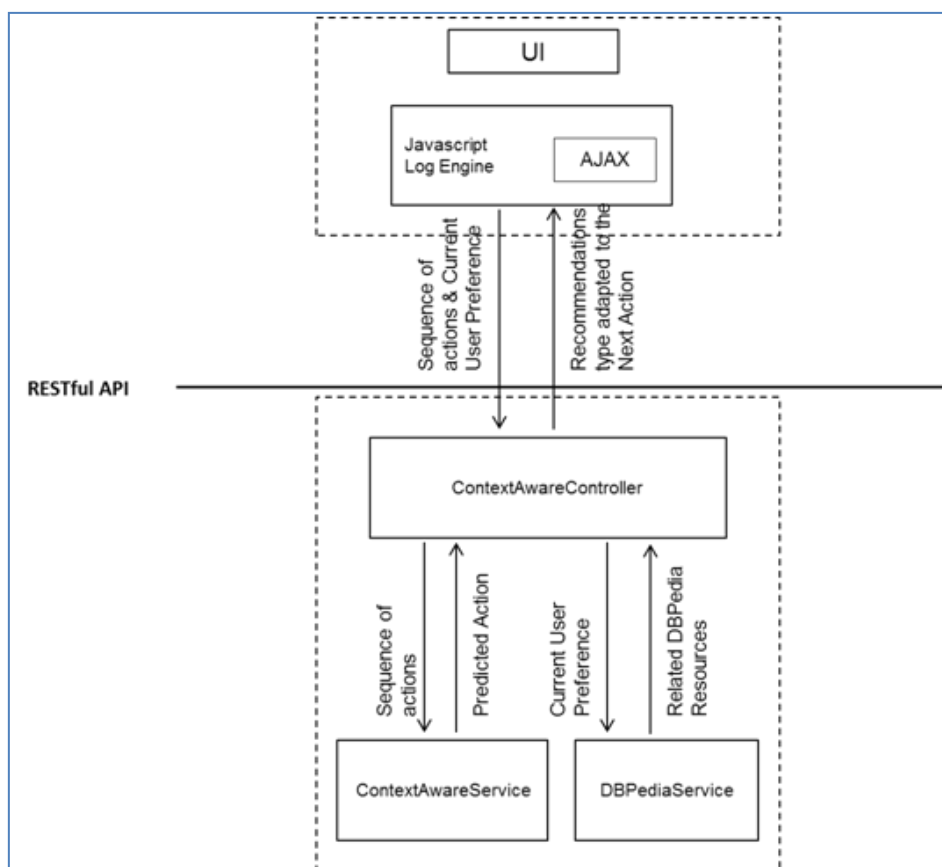
Καταρχήν για την πρόβλεψη και το είδος των προτάσεων ισχύει ό,τι περιγράφηκε στην ενότητα που αφορά την παραγωγή προτάσεων με βάση το γενικευμένο μοντέλο. Χρησιμοποιείται και πάλι ο αλγόριθμος πρόβλεψης. Η θεμελιώδης διαφορά έγκειται στην επιλογή του κατάλληλου Κρυφού Μαρκοβιανού μοντέλου προσωπικότητας, το οποίο θα αξιοποιηθεί από τον αλγόριθμο προώθησης. Πριν από την επιλογή του μοντέλου προηγούνται δύο βήματα. Τα στάδια της προτεινόμενης προσέγγισης παρατίθενται συνολικά:

1. Συλλογή Πληροφορίας Προσωπικότητας του Χρήστη βάσει ερωτηματολογίου Big Five.
2. Κατάταξη Χρήστη στην πλησιέστερη συστάδα χρηστών, όπως υπολογίστηκαν στην ενότητα 5.4.1.2, χρησιμοποιώντας ως μετρική το τετράγωνο της Ευκλείδειας απόστασης.
3. Επιλογή μοντέλου προσωπικότητας που αντιστοιχεί στη συστάδα που κατατάχθηκε ο χρήστης.
4. Παραγωγή Προβλέψεων τροφοδοτώντας τον Αλγόριθμο προώθησης με:
 - i. την τρέχουσα αλληλουχία αλληλεπιδράσεων του χρήστη με το σύστημα

- ii. το Κρυφό Μαρκοβιανό Μοντέλο προσωπικότητας που επιλέχθηκε στο προηγούμενο βήμα.

5.4.4 Αρχιτεκτονική Σχεδίαση & Παραδείγματα Προτάσεων

Δεδομένου ότι το CRUISE είναι η διεπαφή χρήστη στην οποία πραγματοποιήθηκαν οι δοκιμές μας, το Javascript UI του CRUISE ήταν υπεύθυνο για τη λήψη της ακολουθίας των ενεργειών των χρηστών που πραγματοποιούνταν σε πραγματικό χρόνο. Χρησιμοποιήσαμε το Log Engine, με τη διαφορά ότι μόνο η ακολουθία ενεργειών αποστέλλεται ως αντικείμενο JSON στο ContextAwareController, χωρίς τα μεταδεδομένα τους, δηλ. μόνο η ενέργεια 'NewQuery' χωρίς το περιεχόμενο του ερωτήματος. Το αντικείμενο JSON χρειάζεται μόνο να έχει την ακολουθία των ενεργειών-id σαν πληροφορίες.



Εικόνα 44 Αρχιτεκτονική Σχεδίαση Υπηρεσίας Παραγωγής Προτάσεων βάσει Πλαισίου Αλληλεπίδρασης

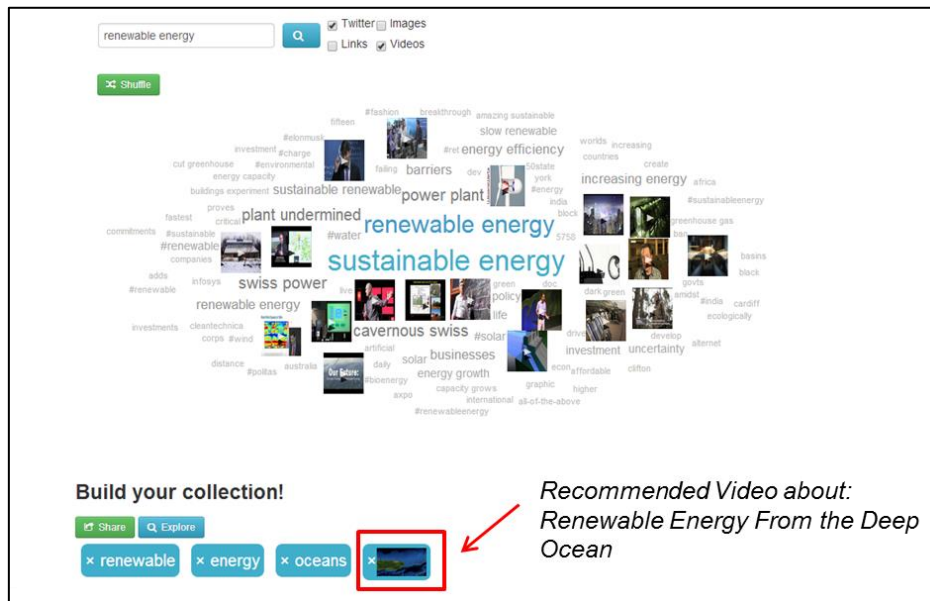
Αποστέλλεται στο ContextAwareController μέσω μιας κλήσης AJAX, η οποία ενεργοποιείται κάθε τρεις νέες ενέργειες που έχουν προστεθεί στην ακολουθία και ταυτόχρονα έχει προστεθεί ένα στοιχείο στην προσωπική συλλογή του χρήστη. Με αυτόν τον τρόπο, δεν συλλαμβάνεται μόνο η ακολουθία των ενεργειών αλλά και οι δυναμικές προτιμήσεις του χρήστη σύμφωνα με το περιεχόμενο του αντικειμένου που έχει προστεθεί στη συλλογή. Από τη στιγμή που η αλληλουχία και η σημασιολογική προτίμηση του χρήστη έχουν σταλεί στο ContextAwareController αναλαμβάνει την αποστολή πληροφοριών στο ContextAwareService για να φροντίσει για την πρόβλεψη της επόμενης ενέργειας. Αυτό συμβαίνει εφαρμόζοντας τον αλγόριθμο forward βάσει του εκπαιδευμένου μοντέλου HMM, την τρέχουσα ακολουθία ενεργειών και επίσης την τρέχουσα σημασιολογική προτίμηση του χρήστη που συνάγεται από αυτό που προσθέτει ο χρήστης στη συλλογή. Η υπηρεσία χρησιμοποιεί τη JAHMM, μια βιβλιοθήκη Java με αλγορίθμους σχετιζόμενους με Κρυφά Μαρκοβιανά Μοντέλα. Ο Controller αποστέλλει επίσης τη σημασιολογική προτίμηση του χρήστη στην υπηρεσία DBPedia προκειμένου να λάβει παρόμοιες έννοιες τις οποίες επιστρέφει η υπηρεσία DBPedia Lookup. Ο ContextAwareController καθορίζει τον τύπο της πρότασης σύμφωνα με την προβλεπόμενη επόμενη ενέργεια.

Για παράδειγμα, στην Εικόνα 45, όπου η προβλεπόμενη ενέργεια είναι 'New Query', ο χρήστης λαμβάνει μια πρόταση σχετικά με την πληκτρολόγηση ενός νέου ερωτήματος που σχετίζεται με αυτό που ο χρήστης έχει προσθέσει στη συλλογή.

Σε ένα δεύτερο παράδειγμα (Εικόνα 46), η προβλεπόμενη επόμενη ενέργεια είναι 'Add Video to Collection', οπότε το σύστημα προτείνει ένα βίντεο σχετικό με ό,τι έχει τοποθετήσει η ο χρήστης στη συλλογή του μέχρι εκείνη τη στιγμή.



Εικόνα 45 Συνδυάζοντας την προβλεπόμενη ενέργεια "Νέο ερώτημα" και την προτίμηση του χρήστη που συνάγεται από τη συλλογή, το σύστημα προτείνει στο χρήστη να αναζητήσει "Βιώσιμη Ενέργεια"



Εικόνα 46 Συνδυάζοντας την προβλεπόμενη ενέργεια "Add Video to collection" και το σημασιολογικό πλαίσιο "Ανανεώσιμη Ενέργεια και Ωκεανός", το σύστημα προτείνει ένα σχετικό βίντεο.

5.5 Αξιολόγηση

Τα υψηλά επίπεδα αλληλεπίδρασης, που αποτελούν αναπόσπαστο μέρος της διερευνητικής αναζήτησης, αποτελούν πρόκληση κατά την αξιολόγηση: υπάρχει πιθανότητα συγχύσεων από τα διαφορετικά χαρακτηριστικά του εργαλείου, είναι δύσκολο να μετρηθεί η επιθυμητή επίπτωση των αποτελεσμάτων και η δυναμική κόπωση περιορίζει την αξιολόγηση σε μικρό αριθμό αναθέσεων. Όλες αυτές οι ιδιότητες καθιστούν δύσκολη την επίτευξη της στατιστικής σημασίας που απαιτείται από μια σημαντική ποσοτική ανάλυση. Επιπλέον, μια ακόμα βασική συνιστώσα της διερευνητικής αναζήτησης είναι σιωπηρή και αναφέρεται στη δυνατότητα της διερευνητικής αναζήτησης να υποστηρίζει τους χρήστες να συναντούν επαρκή ποσότητα πληροφοριών και λεπτομερειών που θα τους βοηθήσουν να επιτύχουν το στόχο τους αντί να φτάσουν απλά σε ένα συγκεκριμένο σχετικό έγγραφο. Αυτοί οι παράγοντες περιορίζουν την εφαρμογή των παραδοσιακών μετρικών αξιολόγησης που εφαρμόζονται στην Ανάκτηση Πληροφορίας (IR), στην αξιολόγηση των συστημάτων διερευνητικής αναζήτησης. Γι' αυτό στην αξιολόγησή μας επικεντρωνόμαστε σε μετρικές που αναφέρθηκαν στην ενότητα 2.3.6, όπως η δέσμευση, ικανοποίηση, επιτυχία στόχου και δημιουργικότητα.

Λόγω της πολυπλοκότητας των συστημάτων διερευνητικής αναζήτησης όσον αφορά τα χαρακτηριστικά διεπαφής και την υποστήριξη αλληλεπίδρασης, θα χρειαζόταν μια ποικιλία μετρήσεων που θα χρησιμοποιούνταν ταυτόχρονα για την αξιολόγηση ενός συγκεκριμένου συστήματος. Ο στόχος της αξιολόγησης μας είναι να υπολογίσουμε τον αντίκτυπο των προτάσεων στην αποτελεσματικότητα της διερεύνησης, στα αποτελέσματα των εργασιών, καθώς και στην αίσθηση της συμμετοχής και της εμπλοκής των χρηστών.

5.5.1 Υποθέσεις

Όπως αναφέρθηκε στην εισαγωγή, η εργασία μας στοχεύει στην ανάλυση των τακτικών των ερευνητών, συνδέοντάς τους με τα χαρακτηριστικά της

προσωπικότητάς τους, και στη χρήση των αναλύσεων για τη δημιουργία προτάσεων για τη βελτίωση των δημιουργικών αποτελεσμάτων των χρηστών σε διερευνητικές εργασίες αναζήτησης. Η αξιολόγηση μας βοηθά να απαντήσουμε στο 3^ο και 4^ο ερευνητικό ερώτημα και τα υποερωτήματά τους.

- *Μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει το πλαίσιο αλληλεπίδρασης να ενισχύσει την αποτελεσματικότητα της διερεύνησης και τη συμμετοχή του χρήστη;*

Στοχεύουμε στο να αξιολογήσουμε εάν οι προτάσεις που προέρχονται από την ανάλυση του πλαισίου αλληλεπίδρασης βελτιώνουν το βαθμό στον οποίο εμπλέκονται οι χρήστες και τους προκαλεί θετικά συναισθήματα, το οποίο είναι σημαντική ένδειξη της αποτελεσματικότητας του συστήματος.

- *Μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει το πλαίσιο αλληλεπίδρασης να υποστηρίξει τη δημιουργικότητα;*

Στοχεύουμε να αξιολογήσουμε εάν οι προτάσεις που προέρχονται από την ανάλυση του πλαισίου αλληλεπίδρασης υποστηρίζουν την ποιότητα σχηματισμού ερωτημάτων και σαν αποτέλεσμα καθιστούν ικανούς τους χρήστες να αποκτήσουν υψηλότερο επίπεδο νόησης αναφορικά με το υπό διερεύνηση θέμα και τελικά να παραγάγουν πιο δημιουργικά αποτελέσματα.

- *Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει την προσωπικότητα να ενισχύσει την αποτελεσματικότητα (ή απόδοση) της διερεύνησης και τη συμμετοχή του χρήστη;*

Στοχεύουμε να αξιολογήσουμε εάν οι προτάσεις που βασίζονται στην προσωπικότητα βοηθούν το χρήστη να χρησιμοποιήσει αποτελεσματικά το σύστημα διερευνητικής αναζήτησης, για παράδειγμα αυξάνοντας τον αριθμό και την ποικιλία των ενεργειών.

- *Πώς μπορεί ένα σύστημα διερευνητικής αναζήτησης που ενσωματώνει την προσωπικότητα να υποστηρίξει τη δημιουργικότητα;*

Στοχεύουμε να αξιολογήσουμε εάν οι προτάσεις που, εκτός από την ανάλυση του πλαισίου αλληλεπίδρασης, ενσωματώνουν και την πληροφορία της προσωπικότητας είναι πιο αποτελεσματικές στο να υποστηρίξουν τους χρήστες να παράγουν περισσότερο δημιουργικά αποτελέσματα.

5.5.2 Σχεδιασμός Μελέτης

Η αξιολόγηση πραγματοποιήθηκε από 66 προπτυχιακούς και μεταπτυχιακούς φοιτητές της επιστήμης Υπολογιστών. Αναθέσαμε στο ένα τρίτο των συμμετεχόντων να εκτελέσουν δραστηριότητες διερευνητικής αναζήτησης με το σύστημα αναφοράς το οποίο δεν παρείχε προτάσεις ή προτροπές στη διάρκεια της διερευνητικής συνεδρίας, ενώ οι υπόλοιποι χωρίστηκαν σε δύο ομάδες, εκ των οποίων στη μία ανατέθηκε το εργαλείο που παρείχε προτάσεις βασισμένες στο πλαίσιο αλληλεπίδρασης και στην προσωπικότητα και στην άλλη το εργαλείο που παρείχε προτάσεις βάσει το γενικευμένου μοντέλου HMM. Σχεδιάσαμε μία διαδικασία αξιολόγησης μεταξύ υποκειμένων (between-subjects) κυρίως για να αποφύγουμε την επίδραση της μεταβίβασης (carryover effect). Δεδομένου ότι επιδιώξαμε να συγκρίνουμε την αποτελεσματικότητα των προτάσεων όσον αφορά τη διερευνητική συμπεριφορά των χρηστών και συγκεκριμένα όσον αφορά την ικανότητά τους να παράγουν πιο δημιουργικά αποτελέσματα, έπρεπε να μειώσουμε την πιθανότητα να έχουν διερευνήσει το ανατεθειμένο θέμα σε προηγούμενους γύρους της διαδικασίας. Επιπλέον, η ανάθεση της εργασίας στους συμμετέχοντες έγινε μόνο μια φορά, όπου έπρεπε να χρησιμοποιήσουν ένα από τα τρία συστήματα και να απαντήσουν σε ερωτηματολόγια που σχετίζονταν μόνο με αυτό, γεγονός που περιόρισε την επίδραση κούρασης που εντάσσεται και αυτή στην επίδραση μεταβίβασης. Επίσης, προκειμένου να μειωθεί η διακύμανση σφάλματος που συνήθως συνοδεύει αξιολογήσεις μεταξύ υποκειμένων, προσπαθήσαμε να μειώσουμε την ποικιλομορφία του δείγματος. Γι' αυτό οι συμμετέχοντες ανήκαν στην

ίδια ηλικιακή ομάδα, με παρόμοιο μορφωτικό υπόβαθρο, ίδια τάξη και εκτέλεσαν τα καθήκοντά τους στο ίδιο χρονικό διάστημα για να αποφύγουμε διαφοροποίηση στους περιβαλλοντικούς παράγοντες. Οι συμμετέχοντες κλήθηκαν αρχικά να συμπληρώσουν ένα ερωτηματολόγιο σχετικά με τα δημογραφικά τους στοιχεία και την εμπειρία αναζήτησης.

Οι άνδρες συμμετέχοντες αποτέλεσαν το 68.9% του συνολικού αριθμού συμμετεχόντων, η μέση ηλικία ήταν $M=23$ με τυπική απόκλιση $SD=2.5$ και η μέση εμπιστοσύνη στην αναζήτηση ήταν 4 στις 5 και ιδιαίτερα στη διερευνητική αναζήτηση ($M=4.1$). Τα τρία συστήματα συγκρίθηκαν έμμεσα μέσω των επιδόσεων των συμμετεχόντων α) στην απόδοση στην κύρια εργασία αναζήτησης, β) στην αποτελεσματική χρήση του συστήματος μέσω των καταγεγραμμένων δεδομένων κάθε συνεδρίας και γ) στα ερωτηματολόγια που εξέτασαν τις υποκειμενικές αντιδράσεις των συμμετεχόντων. Τα δεδομένα καταγραφής κατέγραψαν τις ενέργειες χρήστη και τις εξόδους συστήματος: για παράδειγμα, τα ερωτήματα των χρηστών, τα κλικς στο σύστημα αναφοράς αλλά και στα πειραματικά συστήματα, τα στοιχεία που συλλέχθηκαν, το χρόνο που δαπανήθηκε για την εργασία αναζήτησης, την αδράνεια κατά τη διάρκεια της περιόδου διερεύνησης, τις προτάσεις που παρέχονται από το πειραματικό σύστημα κατά τη διάρκεια μιας συνεδρίας και την άμεση ανατροφοδότηση από τον χρήστη σχετικά με τις προτάσεις που παρέχει το πειραματικό σύστημα.

5.5.3 Πρωτόκολλο Μελέτης

Προκειμένου να ελεγχθούν οι υποθέσεις μας, ζητήσαμε από κάθε συμμετέχοντα να πραγματοποιήσει μία συνεδρία αναζήτησης. Σε κάθε συνεδρία, χρησιμοποιήθηκε είτε ένα από τα πειραματικά είτε το βασικό σύστημα για την ολοκλήρωση της ανάθεσης της οποίας η περιγραφή απεικονίζεται στον Πίνακα 29. Πριν από την εκκίνηση της ανάθεσης, οι συμμετέχοντες κλήθηκαν να συμπληρώσουν ένα ερωτηματολόγιο σχετικά με την ηλικία, το φύλο και την εμπειρία τους σε εργασίες διερευνητικής αναζήτησης. Στη συνέχεια πραγματοποιήθηκε δεκάλεπτη επίδειξη και

πρακτική άσκηση επάνω στο σύστημα και αμέσως μετά οι χρήστες ξεκίνησαν τις συνεδρίες τους. Δεν ήταν αυστηρά περιορισμένες χρονικά, καθώς ένας χρονικός περιορισμός θα παραβίαζε τις βασικές αρχές μιας διερευνητικής συνεδρίας αναζήτησης. Τους ζητήθηκε να σημειώσουν όσο το δυνατόν περισσότερες ιδέες.

Πίνακας 29 Περιγραφή Πειράματος

Experiment Description
Choose one of the following questions and note down all the ideas you found or the ideas you identified when using the tool pertaining to the selected question: 1. Report and describe as many ways as possible in which 'Artificial Intelligence' could improve everyday life and business. 2. Report and describe as many ways as possible that '3D Printing' can assist you in everyday life and business.

Στο τέλος κάθε συνεδρίας αναζήτησης, οι συμμετέχοντες κλήθηκαν να συμπληρώσουν ένα ερωτηματολόγιο μετά τη μελέτη. Όλοι οι συμμετέχοντες απάντησαν σε ερωτήσεις σχετικά με την υποκειμενική τους εμπειρία στο σύστημα σύμφωνα με την κλίμακα δέσμευσης χρηστών (User Engagement Scale) που πρότειναν οι O'Brien και Toms (2013). Οι συμμετέχοντες, στους οποίους είχε ανατεθεί κάποια από τις εκδόσεις του εργαλείου με υπηρεσία προτάσεων, κλήθηκαν να αξιολογήσουν ξεχωριστά τη λειτουργικότητα των προτάσεων απαντώντας σε 9 από τις 15 ερωτήσεις που συνθέτουν τη σύντομη έκδοση του υποδείγματος ερωτήσεων ResQue, το οποίο έχει προταθεί από τους Chen και Pu (2010).

5.6 Αποτελέσματα

5.6.1 Η Επίδραση των Προτάσεων στην Απόδοση της Διερευνητικής Αναζήτησης

Αναλύσαμε την καταγραφή των συνεδριών αναζήτησης ανά εργαλείο, για να εκτιμήσουμε τη συμπεριφορά αναζήτησης και κυρίως την επίδραση του στοιχείου των προτάσεων στην απόδοση της αναζήτησης (Πίνακας 30). Εφαρμόσαμε t-test

Πίνακας 30 Αξιολόγηση της συμπεριφοράς του χρήστη ως προς την αναζήτηση στα τρία διαφορετικά εργαλεία

Feature	CRUISE (Personality HMM)		CRUISE (Global HMM)		CRUISE (No Recommendation)	
	Mean	StDev	Mean	StDev	Mean	StDev
#Actions	24.3	12.7	33.1	17.4	24.8	18.2
#Queries	3.70	2.25	4.38	3.53	4.29	4.45
#AddToCollectionActions	4.35	3.16	3.82	2.72	3.45	2.60
#ExploreCollectionActions	4.54	4.13	3.74	2.91	1.54	2.37
Duration in minutes	18.9	9.78	21.4	10.7	20.2	14.2

The significant features (with $p < .05$, two-tailed) are typeset in **bold**.

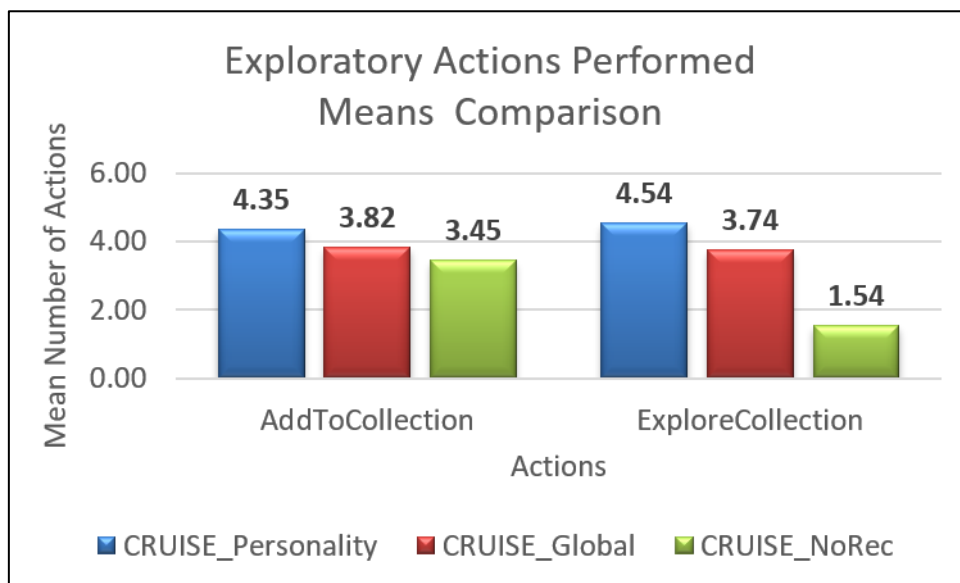
ανεξάρτητων δειγμάτων για να αξιολογήσουμε τη σημασία των αποτελεσμάτων. Η επιλογή του συστήματος δεν επηρεάζει τη μέση τιμή του αριθμού των ενεργειών και των ερωτημάτων, εάν και στο σύστημα όπου παρέχονταν προτάσεις βάσει πλαισίου και προσωπικότητας, η τυπική απόκλιση είναι η χαμηλότερη και από τα τρία εργαλεία. Η ενέργεια ‘Add to Collection’ φαίνεται να έχει κυρίαρχη θέση στα αποτελέσματα στο σύστημα με τις προτάσεις βάσει προσωπικότητας ($M_{pers} = 4.35$, $SD_{pers} = 3.16$) σε σύγκριση με το σύστημα που παράγει προτάσεις βάσει του γενικευμένου μοντέλου ($M_{global} = 3.82$, $SD_{global} = 2.72$), $t(21) = 2.11$, $p = .056$ και το σύστημα χωρίς προτάσεις ($M_{norec} = 3.45$, $SD_{norec} = 2.60$), $t(21) = 2.09$, $p < .05$. Όμοια, η ενέργεια ‘Explore Collection’ έχει κυρίαρχη θέση στο σύστημα με τις προτάσεις βάσει προσωπικότητας ($M_{pers} = 4.54$, $SD=4.13$), ιδίως συγκριτικά με το σύστημα χωρίς

προτάσεις ($M_{\text{norec}} = 1.54$, $SD_{\text{norec}} = 2.36$), $t(21) = 2.10$, $p < .05$. Αυτό μπορεί να αποδοθεί στις καλά κατευθυνόμενες προτεινόμενες ενέργειες αυτού του εργαλείου, που ωθούν τους χρήστες να επιλέξουν το κουμπί 'Explore Collection' πιο συχνά απ' ό,τι θα έκαναν χωρίς αυτές τις προτάσεις. Σαφώς, το γεγονός ότι οι χρήστες που χρησιμοποίησαν το σύστημα με βάση την προσωπικότητα πρόσθεσαν περισσότερα στοιχεία (δηλ. εικόνες, όροι, άρθρα ή βίντεο) στη συλλογή τους, δικαιολογεί την αυξημένη μέση τιμή για την ενέργεια 'Explore Collection'.

Οι υψηλότερες μέσες τιμές για τις ενέργειες 'Add to Collection' και 'Explore Collection' υποδηλώνουν υψηλή πιθανότητα να δεσμευτεί ο χρήστης σε μια επαναληπτική ακολουθία ενεργειών όπως 'Query-Scan-Add toCollection-Explore Collection-Scan-AddtoCollectionExplore Collection'. Επιπρόσθετα, αποφασίσαμε να εξετάσουμε έναν άλλο δείκτη της συμπεριφοράς των χρηστών, τον χρόνο που αφιέρωσαν χρησιμοποιώντας το σύστημα. Η «διάρκεια» της διερευνητικής συνεδρίας τους δεν φαίνεται να επηρεάζεται σημαντικά ούτε από την ύπαρξη του χαρακτηριστικού προτάσεων στο εργαλείο ούτε από τις προτάσεις που βασίζονται στην προσωπικότητα κάθε χρήστη. Συγκεκριμένα, τα αποτελέσματα καταδεικνύουν ότι τόσο στο περιβάλλον χωρίς προτάσεις όσο και στο περιβάλλον με προτάσεις που βασίζονται στο Γενικευμένο μοντέλο HMM, οι συμμετέχοντες έχουν παρόμοια διάρκεια μέσης συνεδρίας. Αυτό εξηγείται πιθανότατα από το γεγονός ότι, παρόλο που οι συμμετέχοντες δεν πραγματοποίησαν την έρευνα υπό αυστηρό χρονικό περιορισμό, το πείραμα πραγματοποιήθηκε ανάμεσα σε δύο προγραμματισμένα μαθήματα, τα οποία θα μπορούσαν ενδεχομένως να επιβάλουν κάποια πίεση χρόνου στους συμμετέχοντες.

Συνολικά, φαίνεται ότι τα συστήματά που βασίζονται σε προτάσεις βοηθούν τους χρήστες να κάνουν αποτελεσματική χρήση του συστήματος διερευνητικής αναζήτησης, ενισχύοντας τον αριθμό και την ποικιλία των ενεργειών με το σύστημά προτάσεων που βασίζεται στην προσωπικότητα να έχει μικρή προτεραιότητα έναντι αυτού που βασίζεται στο γενικευμένο μοντέλο. Τα παραπάνω αποτελέσματα

ανταποκρίνονται θετικά στο πρώτο ερευνητικό ερώτημα που τέθηκε στο συγκεκριμένο κεφάλαιο.



Εικόνα 47 Διάγραμμα που συγκρίνει τους μέσους όρους του αριθμού των διερευνητικών ενεργειών «Προσθήκης Ευρήματος στη Συλλογή» και «Διερεύνησης Συλλογής»

5.6.2 Η Επίδραση των Προτάσεων Στο Αποτέλεσμα της Εργασίας

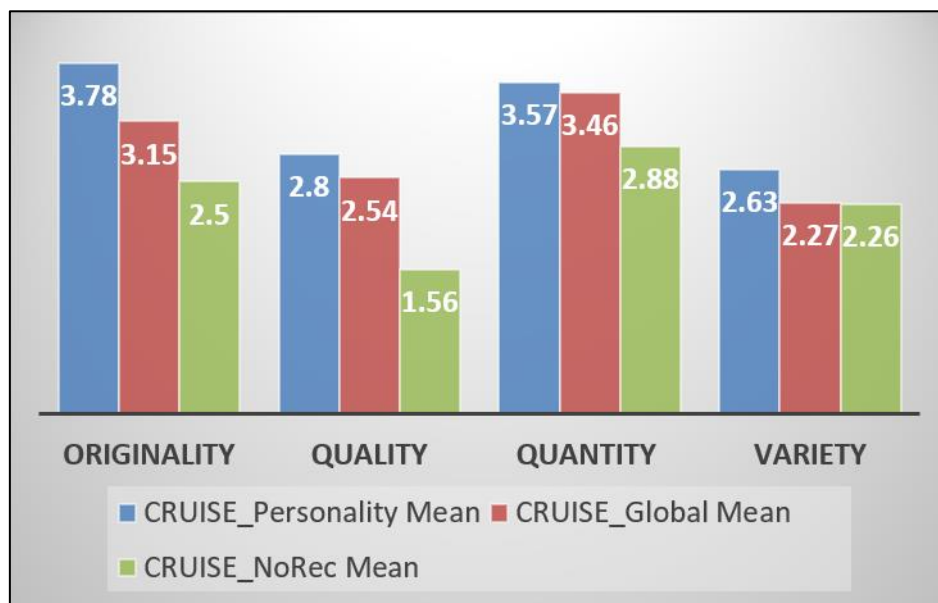
Προκειμένου να αξιολογηθεί η δυνατότητα καθενός από τα τρία συστήματα να υποστηρίξουν το δημιουργικό αποτέλεσμα, υιοθετήσαμε τα ακόλουθα κριτήρια δημιουργικότητας: πρωτοτυπία, ποιότητα, ποσότητα και ποικιλία (Shah, Smith και Vargas-Hernandez, 2003). Σύμφωνα με τους συγγραφείς, η «πρωτοτυπία» αντιπροσωπεύει πόσο ασυνήθιστη, απροσδόκητη και καινοτόμος είναι η ιδέα αν συγκριθεί με άλλες ιδέες, η «ποιότητα» συναρτάται με τη σκοπιμότητα μιας ιδέας, την αντιληπτή συγγένειά της με την ερώτηση και πόσο σαφώς εκφράζεται ή διατυπώνεται. Η «ποσότητα» είναι ο αριθμός ιδεών και η «ποικιλία» ορίζεται ως ένα μέτρο του διερευνηθέντος χώρου λύσης κατά τη διάρκεια της διαδικασίας δημιουργίας ιδεών. Μετά την ολοκλήρωση του πειράματος, ζητήσαμε από δύο καθηγητές να αξιολογήσουν την πρωτοτυπία και την ποιότητα των ιδεών. Προκειμένου να εκτιμηθεί η πρωτοτυπία, υιοθετήσαμε τη μέθοδο που περιγράψαμε στο κεφάλαιο 4. Ο Πίνακας 31 και η Εικόνα 48 συνοψίζουν τα περιγραφικά στατιστικά

σχετικά με τα προαναφερθέντα κριτήρια δημιουργικότητας και στα τρία περιβάλλοντα.

Πίνακας 31 Αξιολόγηση των Δημιουργικών Αποτελεσμάτων από τους ειδήμονες

Metric	CRUISE (Personality HMM)		CRUISE (Global HMM)		CRUISE (No Recommendation)	
	Mean	StDev	Mean	StDev	Mean	StDev
Originality	3.78	2.15	3.15	1.93	2.5	1.67
Quality	2.80	1.21	2.54	1.48	1.56	1.20
Quantity	3.57	2.54	3.46	2.43	2.88	2.12
Variety	2.63	1.58	2.27	1.91	2.26	1.52

The significant features (with $p < .05$, two-tailed) are typeset in **bold**.



Εικόνα 48 Απεικόνιση των μέσων όρων των βαθμολογιών αναφορικά με τα κριτήρια αξιολόγησης ενός δημιουργικού αποτελέσματος ή ιδέας στα τρία διαφορετικά εργαλεία

Παρατηρούμε ότι τα τρία εργαλεία δεν οδηγούν σε σημαντικές διαφορές ως προς την ποικιλία ($p = .073$), παρατήρηση που πιθανώς εξηγείται από το γεγονός ότι ανεξάρτητα από το χρησιμοποιούμενο σύστημα οποιοσδήποτε χρήστης έχει τη δυνατότητα να συναντήσει όλες τις πιθανές ερμηνείες του ερωτήματός του στα στοιχεία του νέφους λέξεων.

Πίνακας 32 Συμφωνία μεταξύ καθηγητών για την αξιολόγηση τόσο της ποιότητας όσο και της πρωτοτυπίας

#Raters	# Agreements	Kappa	Standard Error	CI 95%
2	110	78.62%	3.98%	68.30% to 82.25%

Όσον αφορά την συμφωνία μεταξύ των δύο καθηγητών, ο Πίνακας 32 απεικονίζει το συνολικό αριθμό αξιολογήσεων στις οποίες συμφώνησαν, την αντίστοιχη στατιστική Kappa, την τυπική τιμή σφάλματος καθώς και το εύρος του διαστήματος εμπιστοσύνης. Ο Cohen (1960) πρότεινε το αποτέλεσμα *kappa* να ερμηνεύεται ως εξής: οι τιμές ≤ 0 δεν δείχνουν συμφωνία οι 0.01-0.20 από καθόλου ως ελάχιστη συμφωνία, 0.21-0.40 ως αποδεκτές, 0.41-0.60 ως μέτριες, 0.61-0.80 ουσιαστικές και 0.81-1.00 ως σχεδόν τέλεια συμφωνία (McHugh, 2012). Παρατηρούμε ότι με 66 ονομαστικά αντικείμενα και σύμφωνα με τα προτεινόμενα όρια από τον Cohen, η τιμή Kappa 78,62% θεωρείται σημαντική, αν και το εύρος διαστήματος εμπιστοσύνης είναι μάλλον ευρύ για την τιμή *kappa* που πρέπει να θεωρηθεί ιδιαίτερα ακριβής και επηρεάζεται κυρίως από τη μη δυαδική φύση των αξιολογήσεων του καθηγητή. Παρόλα αυτά, δεδομένης της φύσης της μελέτης και του γεγονότος ότι οι βαθμολογητές ακολούθησαν μια κλίμακα 5-Likert, είναι αποδεκτό ένα ποσοστό διαφωνίας 21.38%.

Τα αποτελέσματα δείχνουν ότι οι προτάσεις μπορούν να βελτιώσουν το αποτέλεσμα της ανατεθειμένης εργασίας αναφορικά με την πρωτοτυπία, ποιότητα και ποσότητα, πράγμα το οποίο επιβεβαιώνει τη δεύτερη ερευνητική υπόθεση που τέθηκε στο συγκεκριμένο κεφάλαιο της Διατριβής, ενώ αντίθετα δεν υπάρχει ένδειξη

ότι οι προτάσεις με βάση την προσωπικότητα βοηθούν περισσότερο προς αυτή την κατεύθυνση σε σχέση με αυτές που βασίζονται στο γενικευμένο μοντέλο.

5.6.3 Επίδραση των Προτάσεων στη Δέσμευση του Χρήστη

Για να αξιολογήσουμε τη δέσμευση ή εμπλοκή του χρήστη, υιοθετήσαμε την Κλίμακα Δέσμευσης Χρήστη (User Engagement Scale-UES) από O' Brien & Toms (2013). Η κλίμακα UES μετρήθηκε σε μία κλίμακα 5-point Likert scale και αποτελείται από 31 στοιχεία και έξι υπο-κλίμακες: εστιασμένη προσοχή (FA), αισθητή εμπλοκή (FI), καινοτομία (NO), ανθεκτικότητα (EN), αισθητική (AE) και αντιληπτή χρηστικότητα (PUs). Σχεδιάσαμε ένα ερωτηματολόγιο μετά την εκπόνηση της ανάθεσης, με βάση αυτά τα στοιχεία που αναδιατυπώθηκαν σύμφωνα με το πλαίσιο.

Ο Πίνακας 33 δείχνει τιμές μέσης και τυπικής απόκλισης για κάθε παράγοντα του UES. Όσον αφορά την Αισθητική και την Αντιληπτή Χρηστικότητα, σημειώνουμε ότι η

Πίνακας 33 Αξιολόγηση εργαλείου με την Κλίμακα Δέσμευσης Χρήστη

Feature	CRUISE (Personality HMM)		CRUISE (Global HMM)		CRUISE (No Recommendation)	
	Mean	StDev	Mean	StDev	Mean	StDev
Aesthetics (AE)	2.93	0.69	3.17	0.88	2.96	0.71
Focused Attention (FA)	2.90	0.84	2.88	0.84	2.85	0.85
Felt Involvement (FI)	3.87**	0.55	3.79**	0.65	3.46	0.78
Perceived Usability (PUs)	2.64	0.41	2.65	0.45	2.70	0.51
Novelty (NO)	3.91*	0.75	3.74	0.80	3.55	0.77
Endurability (EN)	3.70*	0.47	3.54*	0.60	3.35	0.69

The significant features are typeset in **bold**.

* $p < .05$ (two-tailed)

** $p < .01$ (two-tailed)

μέση τιμή και για τα τρία εργαλεία είναι πολύ κοντά, γεγονός που εξηγείται από το γεγονός ότι η διεπαφή χρήστη παρέμεινε η ίδια και για τα τρία εργαλεία. Το εργαλείο που βασίζεται στην προσωπικότητα ($M_{pers} = 3.87$, $SD_{pers} = 0.55$) βαθμολογήθηκε υψηλότερα στον παράγοντα Αισθητή Εμπλοκή σε σύγκριση με το εργαλείο χωρίς προτάσεις ($M_{norec} = 3.46$, $SD_{norec} = 0.76$), $t(21) = 2.10$, $p < 0.01$ καθώς επίσης και το CRUISE Γενικευμένου Μοντέλου ($M_{global} = 3.79$, $SD_{global} = 0,65$) συγκριτικά με το σύστημα χωρίς προτάσεις $t(21) = 2.10$ $p <.01$. Το CRUISE που λαμβάνει υπόψιν και την προσωπικότητα σημείωσε υψηλότερη βαθμολογία στην Αντοχή ($M_{pers} = 3.70$, $SD_{pers} = 0.47$) σε σύγκριση με το βασικό σύστημα ($M_{norec} = 3.35$, $SD_{norec} = 0.69$), $t(21) = 2.10$, $p = 0.015$ που πιθανόν να οφείλεται στις προτάσεις που ήταν σχετικές με την πρόθεση των χρηστών. Ομοίως, το CRUISE Γενικευμένου μοντέλου σε σύγκριση με το βασικό σύστημα επιτυγχάνει σημαντική διαφορά, $t(21) = 2.09$, $p <.05$. Δε βρέθηκε σημαντική στατιστική διαφορά από τη σύγκριση του CRUISE που λαμβάνει υπόψιν την προσωπικότητα σε σχέση με το Γενικευμένο CRUISE όσον αφορά την Αντοχή. Η Καινοτομία αποδεικνύεται ισχυρότερη στο CRUISE με βάση την προσωπικότητα σε σύγκριση με το βασικό σύστημα, $t(21) = 2.10$, $p <.05$, ενώ δε βρέθηκε σημαντική στατιστική διαφορά κατά τη σύγκριση του Γενικευμένου CRUISE με το βασικό σύστημα. Ενδεχομένως, αυτό μπορεί να εξηγηθεί από την ύπαρξη των προτάσεων και από τη γενική ικανοποίηση που προκαλούν οι προτάσεις που βασίζονται στην προσωπικότητα.

Τα παραπάνω αποτελέσματα επιβεβαιώνουν ότι η ύπαρξη προτάσεων που προκύπτουν από την ανάλυση του πλαισίου αλληλεπίδρασης μπορεί να βελτιώσει το βαθμό εμπλοκής των χρηστών και την εμπειρία θετικών συναισθημάτων, τα οποία απαντούν θετικά στην τρίτη ερευνητική μας ερώτηση που τέθηκε στο συγκεκριμένο κεφάλαιο.

5.6.4 Αξιολόγηση των Χαρακτηριστικών των Προτάσεων

Για την αξιολόγηση των χαρακτηριστικών των προτάσεων, οι συμμετέχοντες απάντησαν σε ένα ερωτηματολόγιο μετά την εκπόνηση της κύριας ανάθεσης, για το

οποίο υιοθετήσαμε τμήματα του ερωτηματολογίου ResQue, ένα σύνολο ερωτήσεων που χρησιμοποιήθηκαν ευρέως για την αξιολόγηση των συστημάτων προτάσεων (Πίνακας 34). Το σύστημα προτάσεων με βάση την προσωπικότητα σκοράρει υψηλότερα σε σχέση με τη συνάφεια ($M_{pers} = 4.10$ and $SD_{pers} = 0.94$) συγκριτικά με το σύστημα προτάσεων του γενικευμένου μοντέλου ($M_{global} = 3.57$ and $SD_{global} = 0.93$) $t(21) = 2.11$, $p = 0.058$, αν και χωρίς υψηλή στατιστική σημαντικότητα. Επιπλέον περισσότεροι συμμετέχοντες επιβεβαίωσαν ότι «οι περισσότερες προτάσεις ταίριαζαν με τις προθέσεις τους» βαθμολογώντας υψηλότερα το σύστημα προτάσεων με βάση την προσωπικότητα ($M_{pers} = 3.91$ and $SD_{pers} = 0.98$), ενώ η αντίστοιχες μέσες τιμές για το σύστημα προτάσεων του γενικευμένου μοντέλου ήταν ($M_{global} = 3.24$ and $SD_{global} = 1.04$), $t(21) = 2.10$, $p < 0.05$. Τα επίπεδα ενδιαφέροντος

Πίνακας 34 Αξιολόγηση των προτάσεων βάσει κλίμακας βαθμολογίας από το 1 ως το 5 για τα συστήματα προτάσεων που βασίζονται στο Γενικευμένο HMM και στο HMM που βασίζεται και στη προσωπικότητα.

Criterion	Personality-based HMM Recommender		Global HMM Recommender	
	Mean	StDev	Mean	StDev
Most of the recommendations were relevant.	4.1	0.94	3.57	0.93
Most of the recommendations matched my intentions.	3.91	0.98	3.24	1.04
The tool helped me discover interesting aspects	3.86	0.45	3.71	0.58
The items recommended to me were diverse.	3.71	1.05	3.19	0.87
I understood why the items were suggested to me.	3.81	0.99	3.85	0.8
I found it easy to inform the system if I dislike/like the suggested action.	3.81	0.98	3.62	1.24
The layout of the suggestions interface is adequate.	3.48	1.11	3.33	0.80

Significant criteria ($p < .05$, two-tailed) are typeset in **bold**.

αναφορικά με το περιεχόμενο εκτιμήθηκαν παρόμοια μεταξύ των δύο εργαλείων, ενώ κατά την εκτίμηση της ποικιλίας στις προτάσεις, το σύστημα προτάσεων με βάση την προσωπικότητα διατήρησε υψηλότερη θετική αντίδραση από τους συμμετέχοντες ($M_{pers} = 3.71$ and $SD_{pers} = 1.05$), $t(21) = 2.11$, $p < 0.05$. Αναφορικά με την επαρκή επεξήγηση των προτάσεων, αλλά και την επάρκεια της διεπαφής, και οι δύο προσεγγίσεις παραγωγής προτάσεων βαθμολογήθηκαν όμοια. Έγινε προσπάθεια να αξιολογηθεί η αντιληπτή χρησιμότητα με τη χρήση της ακόλουθης δήλωσης του ερωτηματολογίου: «Το βρήκα εύκολο να ενημερώσω το σύστημα εάν μου αρέσει ή όχι η προτεινόμενη ενέργεια». Τα αποτελέσματα έδειξαν ότι και οι δύο προσεγγίσεις είχαν παρόμοια επίδραση στο χρήστη, το οποίο ίσως εξηγείται από το γεγονός ότι και στα δύο συστήματα οι προτάσεις παρουσιάζονταν με τον ίδιο τρόπο.

5.7 Συμπεράσματα

Στη συγκεκριμένη ενότητα παρουσιάστηκε το δεύτερο μέρος της προτεινόμενης προσέγγισης της Διατριβής. Ο αρχικός στόχος ήταν να διερευνηθεί εάν το πλαίσιο αλληλεπίδρασης και τα χαρακτηριστικά της προσωπικότητας μπορούν να αποτελέσουν μέσο για την ενίσχυση της αποτελεσματικότητας των συστημάτων διερευνητικής αναζήτησης μέσω της δημιουργίας και παρουσίασης στους χρήστες προτάσεων που θα τους βοηθήσουν στις αναζητήσεις τους. Η ανάλυση του πλαισίου αλληλεπίδρασης και των χαρακτηριστικών προσωπικότητας βασίστηκε σε Κρυφά Μαρκοβιανά Μοντέλα. Η Διατριβή συμβάλλει στο πεδίο της διερευνητικής αναζήτησης, α) περιγράφοντας τον τρόπο με τον οποίο ένας χρήστης μπορεί να ταξινομηθεί σε συγκεκριμένες ομάδες χρηστών που παρουσιάζουν παρόμοιες συμπεριφορές αναζήτησης με βάση τα χαρακτηριστικά της προσωπικότητάς τους. Επωφελούμαστε από το γεγονός ότι οι ομάδες χρηστών ακολουθούν παρόμοιες τακτικές διερεύνησης για να επιτύχουν τους στόχους αναζήτησης, προκειμένου να προβλεφθεί η επόμενη ενέργεια χρήστη, δεδομένης μιας μερικής ακολουθίας ενεργειών στην τρέχουσα συνεδρία αναζήτησης. Με την περαιτέρω σύζευξη των προβλέψεων ενεργειών με το πλαίσιο αλληλεπίδρασης, προτείνουμε μια καινοτόμο

προσέγγιση προτάσεων που βοηθά τους χρήστες στη διαδικασία διερεύνησής τους παρέχοντας σχετικές προτάσεις και προτροπές προς πιθανώς ανεξερεύνητες πτυχές του υπό έρευνα θέματος.

Η δουλειά μας έχει δείξει ότι τα HMM παρέχουν ένα χρήσιμο εργαλείο για την ανάλυση του πλαισίου αλληλεπίδρασης και των χαρακτηριστικών προσωπικότητας. Επιπλέον, υπάρχουν σαφείς ενδείξεις ότι τα χαρακτηριστικά αλληλεπίδρασης και τα χαρακτηριστικά προσωπικότητας ενισχύουν την απόδοση διερευνητικής αναζήτησης, την ποιότητα των αποτελεσμάτων αναζήτησης καθώς και την αντιληπτή συμμετοχή του χρήστη. Η απόκτηση πρόσβασης σε μια μεγαλύτερη βάση χρηστών, με δεδομένα αλληλεπίδρασης και στοιχεία προσωπικότητων θα μας οδηγούσαν σε ακριβέστερα HMMs και, ως εκ τούτου, θα ενίσχυε περαιτέρω την αποτελεσματικότητα των παραγόμενων προτάσεων. Ένας περιορισμός της δουλειάς μας σχετίζεται με την ανάλυση της αλληλεπίδρασης αλλά και των στοιχείων προσωπικότητας που συλλέγονται από το Amazon Mechanical Turk. Αν και το δείγμα δεδομένων δεν ήταν ασήμαντο, η ανάλυση που πραγματοποιήσαμε βασίστηκε στην υπόθεση ότι δεν υπήρχαν κρυφοί παράγοντες που θα μπορούσαν να επηρεάσουν την απόδοση των Εργαζομένων και, συνεπώς, το παραγόμενο γενικευμένο μοντέλο και το μοντέλο που βασίζεται στα χαρακτηριστικά της προσωπικότητας.

Η έρευνά μας έχει επιπτώσεις τόσο για τους ερευνητές όσο και για τους επαγγελματίες. Έχουμε δείξει ότι τα συστήματα προτάσεων μπορούν να διαδραματίσουν ένα ρόλο στη βελτίωση της αποτελεσματικότητας των διερευνητικών συστημάτων. Είναι προφανές ότι η προσέγγισή μας είναι περιορισμένη, με την έννοια ότι εστιάζεται στην εξαγωγή προτάσεων μόνο με βάση τα χαρακτηριστικά αλληλεπίδρασης και τα χαρακτηριστικά της προσωπικότητας των χρηστών. Επιπλέον, δεν δημιουργεί εξατομικευμένες προτάσεις. Περαιτέρω έρευνα μπορεί να διερευνήσει άλλα χαρακτηριστικά των χρηστών, πέρα από την προσωπικότητα, και να εξετάσει πώς μπορούν να ληφθούν υπόψη για την ανάπτυξη συστημάτων προτάσεων για διερευνητική αναζήτηση. Επιπλέον, τα προφίλ των χρηστών θα μπορούσαν να χρησιμοποιηθούν στην ανάπτυξη συστημάτων

προτάσεων για διερευνητική αναζήτηση. Οι επαγγελματίες μπορούν να επωφεληθούν από τους αλγορίθμους και τις μεθόδους που παρουσιάζονται σε αυτό το έγγραφο, προκειμένου να αναπτύξουν συστήματα προτάσεων για άλλα διερευνητικά συστήματα ενδιαφέροντος, πέρα από το CRUISE που χρησιμοποιήσαμε στη μελέτη μας. Επιπλέον, η συγκεκριμένη χρήση των HMMs για τη μοντελοποίηση των αλληλεπιδράσεων των χρηστών και την πρόβλεψη των αλληλεπιδράσεων που πιθανόν να εκτελεί ο επόμενος χρήστης μπορεί εύκολα να εφαρμοστεί σε άλλα συστήματα πληροφοριών που καταγράφουν αλληλεπιδράσεις χρηστών, όπως ηλεκτρονικά καταστήματα ή workflow συστήματα.

6 Συμπεράσματα και Επόμενα Βήματα

Στο συγκεκριμένο κεφάλαιο παρουσιάζουμε τα γενικά συμπεράσματα της μελέτης μας. Επίσης, περιγράφουμε τις δυνητικές μελλοντικές επεκτάσεις της συγκεκριμένης ερευνητικής προσπάθειας.

6.1 Συμπεράσματα

Στην παρούσα διδακτορική διατριβή πραγματοποιήθηκε επισκόπηση της βιβλιογραφίας, που σχετίζεται με τα μοντέλα ζήτησης πληροφορίας, παρουσιάζοντας προσεγγίσεις αφενός προσανατολισμένες στο σύστημα αφετέρου με επίκεντρο το χρήστη. Δόθηκε έμφαση στο μοντέλο τη Διερευνητικής Αναζήτησης ειδικότερα, καθώς και στο ρόλο της Αναζήτησης πληροφορίας στην υποστήριξη της Δημιουργικότητας.

Με οδηγό τα κενά που εντοπίστηκαν κατά τη βιβλιογραφική μελέτη, η διατριβή προτείνει μία μεθοδολογία διερεύνησης, η οποία αξιολογείται ως προς την επίδρασή της στην παραγωγή δημιουργικών ιδεών. Η προτεινόμενη μεθοδολογία υλοποιείται σε ένα πληροφοριακό σύστημα Διερευνητικής Αναζήτησης. Η μεθοδολογία περιλαμβάνει την εφαρμογή τεχνικών και αλγορίθμων για την ενίσχυση της Εύνοιας των Τυχαίων Ανακαλύψεων, όπως η αλληλεπίδραση και οπτικοποίηση σε επίπεδο διεπαφής, η διαφοροποίηση των αποτελεσμάτων, η εμφάνιση αποτελεσμάτων ποικίλων τύπων, η διήθηση απροσδόκητης (σε σχέση με την πηγή από την οποία προέρχεται) πληροφορίας από κοινωνικά μέσα δικτύωσης, με σκοπό τη μελέτη της επίδρασης στη δημιουργική διαδικασία και το δημιουργικό αποτέλεσμα. Ακόμη, σχεδιάστηκαν και υλοποιήθηκαν συστήματα προτάσεων με τη χρήση Κρυφών Μαρκοβιανών Μοντέλων, που βασίζονται σε καταγεγραμμένες αλληλεπιδράσεις χρηστών με το σύστημα, με στόχο την καλύτερη καθοδήγηση των χρηστών στη διάρκεια της Διερευνητικής Αναζήτησης και τη μελέτη της επίδρασής τους στη Δημιουργικότητα. Όμοια, προτείνεται η εκπαίδευση διαφορετικών μοντέλων ανά

ομάδες προσωπικότητας, προκειμένου οι προτάσεις να προσαρμόζονται στην προσωπικότητα του χρήστη, βασιζόμενοι στην υπόθεση ότι χρήστες με παρόμοια χαρακτηριστικά ακολουθούν παρόμοιες τακτικές αναζήτησης και διερεύνησης.

Η προσέγγισή μας αξιολογήθηκε σε πραγματικές συνθήκες από ένα οργανισμό, καθώς και σε εργαστηριακές συνθήκες από φοιτητές και ερευνητές έπειτα από ανάθεση δημιουργικών εργασιών. Προέκυψε ότι και οι τρεις προαναφερθέντες παράγοντες, ο σχεδιασμός με στόχο την υποστήριξη της εύνοιας των τυχαίων ανακαλύψεων, η αξιοποίηση του πλαισίου αλληλεπίδρασης χρήστη-συστήματος, καθώς και ο συνδυασμός του τελευταίου με την προσωπικότητα του χρήστη, βελτιώνουν τη διαδικασία Διερευνητικής Αναζήτησης, λόγω της καλύτερης καθοδήγησης του χρήστη και επιτυγχάνουν τη μεγαλύτερη ικανοποίησή του. Αναφορικά με τη Δημιουργικότητα, η χρήση αλγορίθμων και τεχνικών για την υποστήριξη της Εύνοιας των Τυχαίων Ανακαλύψεων συμβάλλει θετικά τόσο στη Δημιουργική Διαδικασία όσο και στο Αποτέλεσμα. Το Πλαίσιο Αλληλεπίδρασης επιδρά θετικά στο Δημιουργικό Αποτέλεσμα, ενώ η ενσωμάτωση και της Προσωπικότητας, υποβοηθά μεν τη διερευνητική διαδικασία, δε φαίνεται όμως να συνεισφέρει σημαντικά στο δημιουργικό αποτέλεσμα.

6.2 Μελλοντική Έρευνα

Η βασική συνεισφορά της Διατριβής θα μπορούσε να επεκταθεί μελλοντικά σε επίπεδο Μεθοδολογίας, Υλοποίησης και Αξιολόγησης ως εξής:

- Σε επίπεδο διεπαφής, το CRUISE θα μπορούσε να εμπλουτιστεί με λειτουργικότητα καταγραφής ιστορικού αναζήτησης, με τρόπο που να επιτρέπει στο χρήστη εύκολα να επιστρέψει σε προηγούμενα στάδια του μονοπατιού αναζήτησης.
- Στην κατεύθυνση της εύνοιας των τυχαίων ανακαλύψεων, η μέθοδός μας αξιολογήθηκε συνολικά ως προς την επίδρασή της στη διερεύνηση, τις τυχαίες ανακαλύψεις και τη δημιουργικότητα. Προτείνεται η μεμονωμένη αξιολόγηση των παραγόντων που συνθέτουν το πλαίσιο για την υποστήριξη της Εύνοιας των Τυχαίων

Ανακαλύψεων (π.χ. οπτικοποίηση, διαφοροποίηση αποτελεσμάτων, ποικιλία πηγών, αλγόριθμος απροσδόκητης πληροφορίας από μέσα κοινωνικής δικτύωσης), προκειμένου να αξιολογηθεί και να ποσοτικοποιηθεί η συνεισφορά του καθενός παράγοντα ως προς τη βελτίωση της Διερεύνησης, την υποστήριξη της ίδιας της Εύνοιας των Τυχαίων Ανακαλύψεων και της Δημιουργικότητας.

- Η μέθοδός μας δεν περιλαμβάνει βρόγχο ανατροφοδότησης του μοντέλου πρόβλεψης για την παραγωγή προτάσεων, καθώς απουσίαζε η μακρόχρονη χρήση της εφαρμογής που υλοποιεί την πρότασή μας από σταθερούς χρήστες σε πραγματικές συνθήκες διερεύνησης. Με τη συλλογή Ιστορικού αλληλεπίδρασης χρηστών-συστήματος σε βάθος χρόνου σε πραγματικές συνθήκες Διερευνητικής Αναζήτησης, θα ήταν εφικτό να παραχθούν Κρυφά Μαρκοβιανά Μοντέλα που θα μπορούσαν να γενικευθούν και να προκύψουν από αυτά ασφαλέστερα συμπεράσματα είτε αναφορικά με το είδος των κρυφών καταστάσεων είτε αναφορικά με τις προβλέψεις επόμενης ενέργειας για ακριβέστερες και περισσότερο στοχευμένες προτάσεις.

- Η προσθήκη λειτουργικότητας που προάγει τη συνεργατικότητα θα ήταν μία ενδιαφέρουσα επέκταση της προτεινόμενης προσέγγισης και συστήματος. Κάποια παραδείγματα, θα μπορούσε να είναι: α) ο διαμοιρασμός ενός κοινού χώρου, όπου θα απεικονίζονται μονοπάτια αναζήτησης ή ευρήματα άλλων χρηστών, καθώς και β) η εφαρμογή μεθόδων που αξιοποιούν τις προτιμήσεις παρόμοιων χρηστών για την παραγωγή προτάσεων.

Δημοσιεύσεις και Ανακοινώσεις

Στα πλαίσια της παρούσας εργασίας έχουν πραγματοποιηθεί οι παρακάτω δημοσιεύσεις και ανακοινώσεις.

Δημοσιεύσεις σε Περιοδικά

[j1] Taramigkou, M., Apostolou, D., & Mentzas, G. (2018). Leveraging exploratory search with personality traits and interactional context. *Information Processing & Management*, 54(4), 609-629. (**5-Year Impact Factor: 3.295**)

[j2] Taramigkou, M., Apostolou, D., & Mentzas, G. (2017). Supporting creativity through the interactive exploratory search paradigm. *International Journal of Human-Computer Interaction*, 33(2), 94-114. (**5-Year Impact Factor: 1.905**)

[j3] Apostolou, D., Zachos, K., Maiden, N., Agell, N., Sanchez-Hernandez, G., Taramigkou, M., ... & Wippoo, M. (2016). Facilitating creativity in collaborative work with computational intelligence software. *IEEE Computational Intelligence Magazine*, 11(2), 29-40. (**5-Year Impact Factor: 5.857**)

[j4] Mitropoulos, N., Taramigkou, M., & Apostolou, D. (2015). Exploiting Readily Available Information to Support Everyday Creativity on the Move. *IxD&A*, 24, 88-99.

Ανακοινώσεις σε Συνέδρια

[c1] Taramigkou, M., Bothos, E., Apostolou, D., & Mentzas, G. (2013, June). Fostering serendipity in online information systems. In *2013 International Conference on Engineering, Technology and Innovation (ICE) & IEEE International Technology Management Conference* (pp. 1-10). IEEE.

[c2] Taramigkou, M., Bothos, E., Christidis, K., Apostolou, D., & Mentzas, G. (2013, October). Escape the bubble: Guided exploration of music preferences for serendipity and novelty. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 335-338). ACM.

[c3] Taramigkou, M., Paraskevopoulos, F., Bothos, E., Apostolou, D., & Mentzas, G. (2013, September). Designing Computational Systems for Serendipity in Learning. In *European Conference on Technology Enhanced Learning* (pp. 639-640). Springer, Berlin, Heidelberg.

[c4] Paraskevopoulos, F., Taramigkou, M., Bothos, E., Apostolou, D., & Mentzas, G. (2014, February). Creative user centric inspirational search. In *Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces* (pp. 25-28). ACM.

[c5] Taramigkou, M., Paraskevopoulos, F., Bothos, E., Apostolou, D., & Mentzas, G. (2014, June). Leveraging user inspiration with microblogging-driven exploratory search. In *International Conference on Advanced Information Systems Engineering* (pp. 238-249). Springer, Cham.

[c6] Star, K., Paraskevopoulos, F., Taramigkou, M., Apostolou, D., Schot, M., & Mentzas, G. N. (2015, June). A playful affinity space for creative research. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (pp. 107-110). ACM.

Βιβλιογραφία

Acosta, A. (2012). Using Serendipity to Advance Knowledge Building Activities. *Ontario Institute for Studies in Education, University of Toronto, Canada*.

Adamopoulos, P., & Tuzhilin, A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4), 54.

Agrawal, R., Gollapudi, S., Halverson, A., & Leong, S. Diversifying search results. In *Proc. WSDM 2009*, ACM Press (2009), 5-14.

Akiyama, T., Obara, K., & Tanizaki, M. (2010, September). Proposal and evaluation of serendipitous recommendation method using general unexpectedness. In *Proc. Workshop on the Practical Use of Recommender Systems, Algorithms and Technologies* (pp. 3-10).

Al-Samarraie, H., Eldenfria, A., & Dawoud, H. (2017). The impact of personality traits on users' information-seeking behavior. *Information Processing & Management*, 53(1), 237-247.

Amabile, T. M. Creativity and innovation in organizations. *Harvard Business School* (1996), 1-15.

Åman, P., & Liikkanen, L. A. (2017). . *International Journal of Human-Computer Interaction*, 33(3), 165-179. Baltrunas, L. Context-aware collaborative filtering recommender systems. PhD thesis, Free University of Bozen-Bolzano, 2011

Andolina, S., Klouche, K., Ruotsalo, T., Floréen, P., & Jacucci, G. (2018). Querytogether: Enabling entity-centric exploration in multi-device collaborative search. *Information Processing & Management*, 54(6), 1182-1202.

André, P., Teevan, J., & Dumais, S. T. (2009, April). From x-rays to silly putty via Uranus: serendipity and its role in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2033-2036). ACM.

André, P., Teevan, J., & Dumais, S. T. (2009, October). Discovery is never by chance: designing for (un) serendipity. In *Proceedings of the seventh ACM conference on Creativity and cognition* (pp. 305-314). ACM.

Apiola, M., Lattu, M., & Pasanen, T. A. (2012). Creativity-Supporting Learning Environment---CSLE. *ACM Transactions on Computing Education (TOCE)*, 12(3), 1-25.

Aragon, C. R., & Williams, A. (2011, May). Collaborative creativity: a complex systems model with distributed affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1875-1884).

Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: a psychometric view. *Behavioural brain research*, 214(2), 143-156.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.

Baldonado, M., and Winograd, T. Sensemaker: an information-exploration interface support-ing the contextual evolution of a user's interests. In *Proc. SIGCHI 1997*, 11-18.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, 34(1), 215-220.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online review*, 13(5), 407-424.

Baum, L. E.; Petrie, T. (1966). "Statistical Inference for Probabilistic Functions of Finite State Markov Chains". *The Annals of Mathematical Statistics* 37 (6): 1554-1563.

Baum, L. E.; Eagon, J. A. (1967). "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology". *Bulletin of the American Mathematical Society* 73 (3): 360.

Baum, L. E.; Sell, G. R. (1968). "Growth transformations for functions on manifolds". *Pacific Journal of Mathematics* 27 (2): 211–227.

Baum, L. E.; Petrie, T.; Soules, G.; Weiss, N. (1970). "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". *The Annals of Mathematical Statistics* 41: 164

Bennett, P. N., White, R. W., Chu, W., Dumais, S. T., Bailey, P., Borisyuk, F., & Cui, X. (2012, August). Modeling the impact of short-and long-term behavior on search personalization. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval* (pp. 185-194). ACM.

Benford, S., Taylor, I., Brailsford, D., Koleva, B., Craven, M., Fraser, M., ... & Greenhalgh, C. (1999). Three dimensional visualization of the world wide web. *ACM Computing Surveys (CSUR)*, 31(4es), 25.

Bernstein, M. S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E. H. Eddi: interactive topic-based browsing of social status streams. In *Proc. UIST 2010*, ACM Press (2010), 303-312.

Björneborn, L. (2008). Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research*, 13(4), 13-4.

Björneborn, L. (2010). Design dimensions enabling divergent behaviour across physical, digital, and social library interfaces. *Persuasive Technology*, 143-149.

Boden, M.A. (1990). *The creative mind: myths and mechanisms*, 2nd ed. Routledge, London; New York

Boden, M. A. (Ed.). (1996). *Dimensions of creativity*. MIT Press.

Bordino, I., Mejova, Y., & Lalmas, M. (2013, October). Penguins in sweaters, or serendipitous entity search on user-generated content. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (pp. 109-118). ACM.

Boyce, B. R., Meadow, C. T., & Kraft, D. H. (1994). *Measurement in information science*. Academic Pr.

Broder, A., Fontoura, M., Josifovski, V., and Riedl, L. (2007). A semantic approach to contextual advertising. *Proc. SIGIR*, 559-566.

Broder, A., Churchill, E. F., Hearst, M., Pell, B., Raghavan, P., & Tomkins, A. (2010, April). Search is dead!: long live search. In *Proceedings of the 19th international conference on World wide web* (pp. 1337-1338). ACM.

Brown, V., Tumeo, M., Larey, T. S., & Paulus, P. B. (1998). Modeling cognitive interactions during group brainstorming. *Small Group Research*, 29(4), 495-526.

Burghardt, M., Heckner, M., & Wolff, C. (2012). The Many Ways of Searching the Web Together: A Comparison of Social Search Engines. *Library and Information Science*, 4, 19-46.

Burkell, J., Quan-Haase, A., & Rubin, V. L. (2012, February). Promoting serendipity online: recommendations for tool design. In *Proceedings of the 2012 iConference* (pp. 525-526). ACM.

Capannini, G., Nardini, F. M., Perego, R., & Silvestri, F. Efficient diversification of web search results. *VLDB Endowment* 4, 7 (2011), 451-459.

Carbonell, J., & Goldstein, J. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proc. SIGIR 1998*, ACM Press (1998), 335-336.

Caruso, J. C., & Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurement*, 57(4), 637-654.

Celeux, G., & Durand, J. B. (2008). Selecting hidden Markov model state number with cross-validated likelihood. *Computational Statistics*, 23(4), 541-564.

Chang, J. C., Hahn, N., Perer, A., & Kittur, A. (2019, March). SearchLens: composing and capturing complex user interests for exploratory search. In *IUI* (pp. 498-509).

Chen, A., Context-aware collaborative filtering system: predicting the user's preferences in ubiquitous computing. In *CHI EA*, pages 1110–1111, 2005

Chen, L., & Pu, P. (2010). A user-centric evaluation framework of recommender systems. In *Proceedings of the ACM RecSys 2010 workshop on user-centric evaluation of recommender systems and their interfaces (UCERSTI)*. CEUR workshop proceedings (Vol. 612, pp. 14–21). CEUR-WS.org.

Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. Short and tweet: experiments on recommending content from information streams. In *Proc. CHI 2010*, ACM Press (2010), 1185-1194.

Chen, L., Wu, W., & He, L. How personality influences users' needs for recommendation diversity? In *CHI 2013*, (pp. 829-834). ACM.

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4), 21.

Chirita, P. A., Nejdl, W., Paiu, R., & Kohlschütter, C. (2005, August). Using ODP metadata to personalize search. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 178-185). ACM.

Chirita, P. A., Firan, C. S., & Nejdl, W. (2006, November). Summarizing local context to personalize global web search. In Proceedings of the 15th ACM international conference on Information and knowledge management (pp. 287-296). ACM.

Clarke, C. L., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., Büttcher, S., & MacKinnon, I. (2008, July). Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 659-666). ACM.

Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20:37-46, 1960.

Cohen J. Weighted kappa: nominal scale agreement with provision for scale and disagreement or partial credit. *Psychol. Bull.* 70:213-20. 1968.

Cory, K. A. (1997). Discovering hidden analogies in an online humanities database. *Computers and the Humanities*, 31(1), 1-12.

Csikszentmihalyi, M. (1997). *Flow and the psychology of discovery and invention*. HarperPerennial, New York, 39.

Daoud, M., Tamine-Lechani, L., Boughanem, M., & Chebaro, B. (2009, March). A session based personalized search using an ontological user profile. In Proceedings of the 2009 ACM symposium on Applied Computing (pp. 1732-1736). ACM.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

Denecke, K. Diversity-Aware Search: New Possibilities and Challenges for Web Search. *Library and Information Science* 4 (2012), 139-162.

de Gemmis, M., Lops, P., Semeraro, G., & Musto, C. (2015). An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, 51(5), 695-717.

Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management*, 2(2), 36-46.

Devendorf, L., O'Donovan, J., & Hollerer, T. TopicLens: An Interactive Recommender System based on Topical and Social Connections. In Proc. RecSys 2012, 41.

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological bulletin*, 136(5), 822.

Donaldson, J. J., Conover, M., Markines, B., Roinestad, H., & Menczer, F. (2008, June). Visualizing social links in exploratory search. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia* (pp. 213-218). ACM.

Dong, A., Zhang, R., Kolari, P., Bai, J., Diaz, F., Chang, Y., ... & Zha, H. (2010, April). Time is of the essence: improving recency ranking using twitter data. In *Proceedings of the 19th international conference on World wide web* (pp. 331-340). ACM.

Dormann, C., & Lindgaard, G. (2004, November). Developing innovative systems: creative ideation. In *Proceedings of the symposium on interaction: Systems, practice and theory, Sydney, Australia*.

Dourish, P. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1):19–30, 2004

e Cunha, M. P., Clegg, S. R., & Mendonça, S. (2010). On serendipity and organizing. *European Management Journal*, 28(5), 319-330.

Edlund, J. E., Sagarin, B. J., Skowronski, J. J., Johnson, S. J., & Kutter, J. (2009). Whatever happens in the laboratory stays in the laboratory: The prevalence and

prevention of participant crosstalk. *Personality and Social Psychology Bulletin*, 35(5), 635-642.

Ellis, D. 1993. Modeling the information seeking patterns of academic researchers: a ground theory approach. *Library Quarterly*. 63, 4 (1993), 469–486.

Elsweiler, D. et al. 2009. Evaluating Personal Information Management Behaviour using Markov Models. . *Wissensmanagement 2009*, 280–288.

Erdelez, S. (1997, August). Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an international conference on Information seeking in context* (pp. 412-421). Taylor Graham Publishing.

Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F., & Cantador, I. (2015). Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction*, 1-35.

Ferwerda, B., Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2016, August). The influence of users' personality traits on satisfaction and attractiveness of diversified recommendation lists. In *4 th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) 2016* (p. 43).

Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59(3), 321-340.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS)*, 23(2), 147-168.

Fu, W.-T. (2009). A Semantic Imitation Model of Social Tagging, *Proceedings of the IEEE conference on Social Computing*: 66–72.

Fu, W.-T., Kannampalill, T. G., & Kang, R. (2010). Facilitating exploratory search by model-based navigational cues. In Proceedings of the ACM International conference on Intelligent User Interface. 199-208.

Fuhr, N. 2008. A probability ranking principle for interactive information retrieval. *Information Retrieval*. 11, 3 (Feb. 2008), 251–265.

Garber, S. R., & Grunes, M. B. (1992, June). The art of search: A study of art directors. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 157-163). ACM.

Gauch, S., Chaffee, J., & Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems: An international Journal*, 1(3, 4), 219-234.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010, September). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 257-260). ACM.

Glăveanu, V. P. (2010). Paradigms in the study of creativity: Introducing the perspective of cultural psychology. *New ideas in psychology*, 28(1), 79-93.

Glowacka, D., Ruotsalo, T., Konuyshkova, K., Kaski, S., & Jacucci, G. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proc. IUI 2013, ACM Press (2013)*, 117-128

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002, March). Eye tracking in web search tasks: design implications. In Proceedings of the 2002 symposium on Eye tracking research & applications (pp. 51-58).

Gollapudi, S., & Sharma, A. An axiomatic approach for result diversification. In *Proc. WWW 2009, ACM Press (2009)*, 381-390.

Golovchinsky, G., Adcock, J., Pickens, J., Qvarfordt, P., & Back, M. (2008). Cerchiamo: a collaborative exploratory search tool. *Proceedings of Computer Supported Cooperative Work (CSCW)*, 8-12.

Gwizdka, J. (2008). Revisiting search task difficulty: Behavioral and individual difference measures. *Proceedings of the American Society for Information Science and Technology*, 45(1), 1-12.

Han, K., Shih, P. C., & Carroll, J. M. (2014). Local News Chatter: augmenting community news by aggregating hyperlocal microblog content in a tag cloud. *International Journal of Human-Computer Interaction*, 30(12), 1003-1014.

Hariri, N., B. Mobasher, and R. Burke. Context-aware music recommendation based on latent topic sequential patterns. In *ACM RecSys*, pages 131-138, 2012.

Hassan, A., Jones, R., & Klinkner, K. L. (2010, February). Beyond DCG: user behavior as a predictor of a successful search. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 221-230).

Hassan, A., Song, Y., & He, L. W. (2011, October). A task level metric for measuring web search satisfaction and its application on improving relevance estimation. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 125-134).

Havre, S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. Interactive visualization of multiple query results. In *Proc. IEEE Symposium on Information Visualization 2001*, 105.

He, J., Qvarfordt, P., Halvey, M., & Golovchinsky, G. (2016). Beyond actions: Exploring the discovery of tactics from user logs. *Information Processing & Management*, 52(6), 1200-1226.

Henzinger, M., Chang, B., Milch, B., and Brin, S. (2005). Query-Free News Search. *World Wide Web: Internet and Web Information Systems*, 8(2):101-126.

Herceg, P. M., Allison, T. B., Belvin, R. S., & Tzoukermann, E. (2018). Collaborative exploratory search for information filtering and large-scale information triage. *Journal of the Association for Information Science and Technology*, 69(3), 395-409.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(2), 174-196.

Horvitz, E. (1998, November). Continual computation policies for utility-directed prefetching. In *Proceedings of the seventh international conference on Information and knowledge management* (pp. 175-184).

Hu, R., & Pu, P. Enhancing collaborative filtering systems with personality information. In *Proc. RecSys 2011* (pp. 197-204). ACM.

Hu, R., W. Dou, and J. Liu. A context-aware collaborative filtering approach for service recommendation. In *CSC*, pages 148–155, 2012.

Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008, September). Introducing serendipity in a content-based recommender system. In *Hybrid Intelligent Systems, 2008. HIS'08. Eighth International Conference on* (pp. 168-173). IEEE.

Iijima, C., Kimura, M., & Yamaguchi, T. (2010). Implementing an image search system with integrating social tags and DBpedia. *Knowledge-Based and Intelligent Information and Engineering Systems*, 264-272.

Kammerer, Y., Nairn, R., Pirolli, P., & Chi, E. H. (2009, April). Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 625-634). ACM.

Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. *International Journal of Human-Computer Interaction*, 30(3), 177-191.

Kangasrääsio, A., Glowacka, D., & Kaski, S. (2015, March). Improving controllability and predictability of interactive recommendation interfaces for exploratory search. In Proceedings of the 20th international conference on intelligent user interfaces (pp. 247-251). ACM.

Kanuka, H., & Nocente, N. (2003). Exploring the effects of personality type on perceived satisfaction with web-based learning in continuing professional development. *Distance Education*, 24(2), 227-244.

Kelly, D., & Teevan, J. (2003, September). Implicit feedback for inferring user preference: a bibliography. In *ACM SIGIR F Kelly, D., & Belkin, N. J. (2004, July). Display time as implicit feedback: understanding task effects. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 377-384). ACM.orum (Vol. 37, No. 2, pp. 18-28). ACM.*

Kerne, A., Smith, S. M., Koh, E., Choi, H., & Graeber, R. (2008). An experimental method for measuring the emergence of new ideas in information discovery. *Intl. Journal of Human-Computer Interaction*, 24(5), 460-477.

Kim, Y., Hassan, A., White, R. W., & Zitouni, I. (2014, February). Modeling dwell time to predict click-level satisfaction. In Proceedings of the 7th ACM international conference on Web search and data mining (pp. 193-202).

Khalili, A., Van den Besselaar, P., & de Graaf, K. A. (2018, June). FERASAT: A Serendipity-Fostering Faceted Browser for Linked Data. In European Semantic Web Conference (pp. 351-366). Springer, Cham.

Kiseleva, J., Lam, H. T., Pechenizkiy, M., & Calders, T. (2013, December). Predicting current user intent with contextual markov models. In *2013 IEEE 13th International Conference on Data Mining Workshops (pp. 391-398). IEEE.*

Kiseleva, J., Crestan, E., Brigo, R., & Dittel, R. (2014, November). Modelling and detecting changes in user satisfaction. In Proceedings of the 23rd ACM International

Conference on Conference on Information and Knowledge Management (pp. 1449-1458). ACM.

Koh, E., Kerne, A., & Hill, R. (2007, July). Creativity support: information discovery and exploratory search. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 895-896). ACM.

Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, 51(4), 472-477(2011).

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802-5805.

Kraft, R., Mahoul, F., and Chang, C.C. (2005). Y!Q: contextual search at the point of inspiration. *Proc. CIKM*, 816-823.

Kuhlthau, C.C. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*. 42, 5 (Jun. 1991), 361-371.

Kules, B. (2005). Supporting creativity with search tools. *Creativity Support Tools*, 53-64.

Leong, T. W., Vetere, F., & Howard, S. (2012). Experiencing coincidence during digital music listening. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(1), 1-19.

Lieberman, H. (1995). Letizia: An agent that assists web browsing. *IJCAI* (1), 1995, 924-929.

Liebling, D., Bennett, P., White, R. (2012) Anticipatory Search: Using Context to Initiate Search, SIGIR'12, August 12-16, 2012, Portland, Oregon, USA, pp. 1035-1036

Liu, F., Yu, C., & Meng, W. (2002, November). Personalized web search by mapping user queries to categories. In Proceedings of the eleventh international conference on Information and knowledge management (pp. 558-565). ACM.

Liu, F., Yu, C., & Meng, W. (2004). Personalized web search for improving retrieval effectiveness. *IEEE Transactions on knowledge and data engineering*, 16(1), 28-40.

Liu, X. and Karl Aberer. 2013. SoCo: a social network aided context-aware recommender system. In Proceedings of the 22nd international conference on World Wide Web (WWW '13). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 781-802.

Ma, H., H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In Proceedings of the 17th ACM conference on Information and knowledge management, 2008.

Ma, H., D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In Proceedings of the fourth ACM international conference on Web search and data mining, 2011.

Ma, C., & Zhang, B. (2018). A New Query Recommendation Method Supporting Exploratory Search Based on Search Goal Shift Graphs. *IEEE Transactions on Knowledge and Data Engineering*, 30(11), 2024-2036.

MacCrimmon, K. R., & Wagner, C. (1994). Stimulating ideas through creative software. *Management Science*, 40(11), 1514-1532.

Maher, M. L. (2012, May). Computational and collective creativity: Who's being creative?. In ICC (pp. 67-71).

Makri, S., & Blandford, A. (2012). Coming across information serendipitously-Part 1: A process model. *Journal of Documentation*, 68(5), 684-705.

Makri, S., & Blandford, A. (2012). Coming across information serendipitously-Part 2: A classification framework. *Journal of Documentation*, 68(5), 706-724.

Marchionini, G. 1995. Information seeking in electronic environments. Cambridge University Press.

Marchionini, G. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), pp. 41–46(2006a). doi:10.1145/1121949.1121979

Marchionini, G. (2006b). Toward human-computer information retrieval. *Bulletin of the American Society for Information Science and Technology*, 32(5), 20-22.

Marchionini, G., & White, R. (2007). Find what you need, understand what you find. *International Journal of Human [x02013] Computer Interaction*, 23(3), 205-237.

Marie, N., Gandon, F., Ribière, M., & Rodio, F. (2013, September). Discovery hub: on-the-fly linked data exploratory search. In *Proceedings of the 9th International Conference on Semantic Systems* (pp. 17-24). ACM.

McCay-Peet, L., & Toms, E. G. (2010, August). The process of serendipity in knowledge work. In *Proceedings of the third symposium on Information interaction in context* (pp. 377-382). ACM.

McCay-Peet, L., & Toms, E. G. (2011). Measuring the dimensions of serendipity in digital environments. *Information Research*, 16(3).

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2), 175-215.

Melo, R., & Carvalhais, M. (2013). The Design of Horacle: Inducing Serendipity on the Web. In *xCoAx 2013: Proceedings of the first conference on Computation, Communication, Aesthetics and X* (pp. 183-192).

Millen, D., Yang, M., Whittaker, S., & Feinberg, J. (2007). Social bookmarking and exploratory search. In Proc. ECSCW 2007, 21-40.

Mishra, N., White, R. W., Leong, S., & Horvitz, E. (2014, July). Time-critical search. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (pp. 747-756). ACM.

Murakami, T., Mori, K., & Orihara, R. (2007). Metrics for evaluating the serendipity of recommendation lists. In *New frontiers in artificial intelligence* (pp. 40-46). Springer Berlin Heidelberg.

Nagpal, A., Hangal, S., Joyee, R. R., & Lam, M. S. Friends, romans, countrymen: lend me your URLs. using social chatter to personalize web search. In Proc. CSCW 2012, ACM Press (2012), 461-470.

Natarajan, N., Shin, D., Dhillon, I., S., Which App Will You Use Next? Collaborative Filtering with Interactional Context, RecSys'13, October 12–16, 2013, Hong Kong, China.

Nguyen, L., & Shanks, G. (2009). A framework for understanding creativity in requirements engineering. *Information and software technology*, 51(3), 655-662.

Nie, D., Guan, Z., Zhu, T. Your Search Behavior and Your Personality. In Proc. ICPCA/SWS 2014.

Nijstad, B.A., & Stroebe, W. (2006). How the group affect the mind: a cognitive model of idea generation in groups. *Personality and Social Psychology Bulletin*, 10(3), 186-213.

Nilan, M. S. (1985). *Structural constraints and situational information seeking: a test of two predictors in a sense-making context* (Doctoral dissertation).

O'Brien, H. L., & Toms, E. G. (2013). Examining the generalizability of the User Engagement Scale (UES) in exploratory search. *Information Processing & Management, 49*(5), 1092-1107.

O'Connor, B., Krieger, M., & Ahn, D. TweetMotif: Exploratory Search and Topic Summarization for Twitter. In Proc. ICWSM 2010.

Osborn, A. (1992). The creative trend in education. Source Book for Creative Problem Solving, Creative Education Foundation Press, New York, NY, 39-62.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk.

Parise, S., Whelan, E., & Todd, S. (2015). How Twitter Users Can Generate Better Ideas. *MIT Sloan Management Review, 56*(4), 21.

Parnes, S. J. (1967). Creativity behavior guidebook. New York: Charles Scribners Sons.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review, 106*(4), 643.

Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (Vol. 5, pp. 2-4).

Plucker, J. A., & Beghetto, R. A. (2004). Why Creativity Is Domain General, Why It Looks Domain Specific, and Why the Distinction Does Not Matter.

Pretschner, A., & Gauch, S. (1999, November). Ontology based personalized search. In Proceedings 11th International Conference on Tools with Artificial Intelligence (pp. 391-398). IEEE.

Pu, P., Chen, L., and Hu, R. A user-centric evaluation framework for recommender systems. In Proc. RecSys 2011, 157–164.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.

L. Rao. Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links. TechCrunch, Sept. 14, 2010.

Rentfrow, P. J., Goldberg, L. R., & Zilca, R. (2011). Listening, watching, and reading: The structure and correlates of entertainment preferences. *Journal of personality*, 79(2), 223-258.

Roberts, R. M. (1989). Serendipity: Accidental discoveries in science. *Serendipity: Accidental Discoveries in Science*, by Royston M. Roberts, pp. 288. ISBN 0-471-60203-5. Wiley-VCH, June 1989., 1.

Rocchio, J. J. (1971). Relevance feedback in information retrieval. *The Smart retrieval system-experiments in automatic document processing*. Joachims, T. (2002, July). Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 133-142). ACM

Santos, R. L., Macdonald, C., & Ounis, I. Exploiting query reformulations for web search result diversification. In Proc. WWW 2010, ACM Press (2010), 881-890.

Sauro, J., & Dumas, J. S. (2009, April). Comparison of three one-question, post-task usability questionnaires. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1599-1608).

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1994). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind hypothesis.

Sellen, A. J., Murphy, R., & Shaw, K. L. (2002, April). How knowledge workers use the web. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 227-234). ACM.

Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design studies*, 24(2), 111-134.

Shah, C., Hendahewa, C., & González-Ibáñez, R. (2015). Rain or shine? forecasting search process performance in exploratory search tasks. *Journal of the Association for Information Science and Technology*.

Shen, X., Tan, B., & Zhai, C. (2005, August). Context-sensitive information retrieval using implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 43-50). ACM.

Shen, X., Tan, B., & Zhai, C. (2005, October). Implicit user modeling for personalized search. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 824-831). ACM.

Shneiderman, B., Byrd, D., & Croft, W. B. (1998). Sorting out searching: A user-interface framework for text searches. *Communications of the ACM*, 41(4), 95-98.

Shneiderman, B. (1999, October). User interfaces for creativity support tools. In Proceedings of the 3rd conference on Creativity & cognition (pp. 15-22).

Shneiderman, B. (2000). Creating creativity: user interfaces for supporting innovation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1), 114-138.

Singh, M., Cafarella, M. J., & Jagadish, H. V. (2016). DBExplorer: Exploratory Search in Databases. In EDBT (pp. 89-100).

Speretta, M., & Gauch, S. (2005, September). Personalized search based on user search histories. In Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence (pp. 622-628). IEEE Computer Society.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Information Processing & Management*, 34(5), 599-621

Sriram, S., Shen, X., & Zhai, C. (2004, July). A session-based search engine. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 492-493). ACM.

Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004, May). Adaptive web search based on user profile constructed without any effort from users. In Proceedings of the 13th international conference on World Wide Web (pp. 675-684). ACM.

Sun, T., Zhang, M., & Mei, Q. (2013, June). Unexpected Relevance: An Empirical Study of Serendipity in Retweets. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Tan, B., Shen, X., & Zhai, C. (2006, August). Mining long-term search history to improve search accuracy. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 718-723). ACM.

Taramigkou, M., Bothos, E., Christidis, K., Apostolou, D., & Mentzas, G. (2013, October). Escape the bubble: Guided exploration of music preferences for serendipity and novelty. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 335-338). ACM.

Taramigkou, M., Apostolou, D., & Mentzas, G. (2017). Supporting Creativity through the Interactive Exploratory Search Paradigm. *International Journal of Human-Computer Interaction*, 33(2), 94-114.

Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. The perfect search engine is not enough: a study of orienteering behavior in directed search. In Proc. of SIGCHI 2004, 415–422.

Teevan, J., Dumais, S. T., & Horvitz, E. (2005, August). Personalizing search via automated analysis of interests and activities. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 449-456). ACM.

Teevan, J., Dumais, S. T., & Horvitz, E. (2010). Potential for personalization. ACM Transactions on Computer-Human Interaction (TOCHI), 17(1), 4.

Thudt, A., Hinrichs, U., & Carpendale, S. (2012, May). The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1461-1470).

Vallet, D., & Castells, P. Personalized diversification of search results. In Proc. SIGIR 2012, ACM Press (2012) 841-850.

Van Laere, O., Bordino, I., Mejova, Y., & Lalmas, M. (2014, November). DEESSE: entity-Driven Exploratory and sERendipitous Search SystEm. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (pp. 2072-2074). ACM.

Veale, T. (2011, June). Creative language retrieval: A robust hybrid of information retrieval and linguistic creativity. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 278-287). Association for Computational Linguistics.

Waitelonis, J., & Sack, H. Towards exploratory video search using linked data. *Multimedia Tools and Applications* 59, 2 (2012), 645-672.

Wallas, G. (1926). *The art of thought*.

Wang, K. and Gloy, N. 2010. Inferring Search Behaviors Using Partially Observable Markov (POM) Model. *Web Search and Data Mining (2010)*, 211–220.

Wang, H. C. (2008). Modeling idea generation sequences using Hidden Markov Models. In *The Annual Meeting of the Cognitive Science Society* (pp. 107-112).

Warr, A., & O'Neill, E. (2005, April). Understanding design as a social creative process. In *Proceedings of the 5th conference on Creativity & cognition* (pp. 118-127).

West, R., White, R. W., & Horvitz, E. (2013, July). Here and there: Goals, activities, and predictions about location from geotagged queries. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 817-820). ACM.

White, R. W., Muresan, G., & Marchionini, G. (2006, December). Report on ACM SIGIR 2006 workshop on evaluating exploratory search systems. In *Acm Sigir Forum* (Vol. 40, No. 2, pp. 52-60). ACM.

White, R., Jose, J., van Rijsbergen, C., & Ruthven, I. (2004). A simulated study of implicit feedback models. *Advances in Information Retrieval*, 311-326.

White, R. W., Bailey, P., & Chen, L. (2009, July). Predicting user interests from contextual information. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 363-370). ACM.

White, R. W., Chu, W., Hassan, A., He, X., Song, Y., & Wang, H. (2013, May). Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1411-1420). ACM.

White, R. W. (2016). *Interactions with search systems*. Cambridge University Press.

Wiggins, G. A. (2006). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, 19(7), 449-458.

Williamson, K. (1998). Discovered by chance: The role of incidental information acquisition in an ecological model of information use. *Library & Information Science Research*, 20(1), 23-40.

Woodman, R. W., Sawyer, J. E., & Griffin, R. W. Toward a theory of organizational creativity. *Academy of management review* (1993), 293-321.

Wu, W., Chen, L., & He, L. Using personality to adjust diversity in recommender systems. In *Proc. Hypertext and Social Media 2013*, (pp. 225-229). ACM.

Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., & Li, H. (2010, July). Context-aware ranking in web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 451-458). ACM.

Xie, I. and Joo, S. 2010. Transitions in Search Tactics During the Web-Based Search Process. *Journal of the American Society for Information Science*. 61, 11 (2010), 2188–2205.

Yang, X., H. Steck, and Y. Liu. Circle-based recommendation in online social networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012.

Yue, Z., Shuguang Han, and Daqing He. 2014. Modeling search processes using hidden states in collaborative exploratory web search. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*. ACM, New York, NY, USA, 820-830.

Zachos, K., Apostolou, D., Paraskevopoulos, F., Ientsek, S., Maiden, N., Brown, A., & Mentzas, G. (2018, July). Creative Information Exploration in Journalism. In *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-7). IEEE.

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*,1(1-4), 43-52.

Zhang, Y., Chen, W., Wang, D., & Yang, Q. (2011, August). User-click modeling for understanding and predicting search-behavior. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1388-1396). ACM.

Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012, February). Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 13-22). ACM

Zilincik, M., Navrat, P., & Koskova, G. (2013). Exploratory search on Twitter utilizing user feedback and multi-perspective microblog analysis. *PloS one*, 8(11), e78857.

Παράρτημα 1: Ερωτηματολόγιο εστιασμένο στην υπηρεσία παροχής προτάσεων από το Κοινωνικό Δίκτυο του Χρήστη (Εργαστηριακό Πείραμα)

Αριθμός Ερώτησης	Διατύπωση	Τύπος
1	I felt supported in exploring the search space related to my needs.	Κλίμακα Likert
2	I felt in control of exploring the search space	Κλίμακα Likert
3	Most of the recommendations were relevant	Κλίμακα Likert
4	Most of the recommendations matched my intentions	Κλίμακα Likert
5	The tool helped me discover interesting aspects	Κλίμακα Likert
6	The items recommended to me were diverse	Κλίμακα Likert
7	I understood why the items were recommended to me	Κλίμακα Likert
8	I found it easy to inform the system if I dislike/like the recommended item.	Κλίμακα Likert
9	The layout of the recommender interface is adequate.	Κλίμακα Likert

Παράρτημα 2: Ερωτηματολόγια Αθροιστικής Αξιολόγησης

Ερωτηματολόγιο Αξιολόγησης CRUISE χωρίς υπηρεσία προτάσεων		
Αριθμός Ερώτησης	Διατύπωση	Τύπος
1	Most of the results the tool provided were useful and relevant to my need	Κλίμακα Likert
2	Some of the results were novel to me.	Κλίμακα Likert
3	Some of the results were surprising discoveries that I wasn't expecting to find.	Κλίμακα Likert
4	Using the tool increased my knowledge about the subject I was investigating.	Κλίμακα Likert
5	There were times when my imagination was aroused.	Κλίμακα Likert
6	Doing the task I found items I could share with my friends.	Κλίμακα Likert
7	I felt confident I would be inspired with some of the results.	Κλίμακα Likert
8	I became familiar with the tool quickly.	Κλίμακα Likert
9	I think that I would use this tool again	Κλίμακα Likert
10	I will tell my friends about this tool.	Κλίμακα Likert

Ερωτηματολόγιο Αξιολόγησης CRUISE με υπηρεσία προτάσεων από το Κοινωνικό Δίκτυο του χρήστη		
Αριθμός Ερώτησης	Διατύπωση	Τύπος
1	After signing-in I found more useful and relevant results to my need than before	Κλίμακα Likert
2	After signing-in I found more items I was not aware of than before	Κλίμακα Likert
3	After signing-in I found more results I wasn't expecting to find than before	Κλίμακα Likert
4	After signing-in I found more items I could share with my friends than before.	Κλίμακα Likert
5	Overall allowing the system to create my profile provided a better experience.	Κλίμακα Likert
6	The transparency of how the user profile is built gave me more confidence in the tool.	Κλίμακα Likert

Παράρτημα 3: Ερωτηματολόγια Αξιολόγησης Κεφαλαίου 5

Ερωτηματολόγιο πριν από την ανάθεση διερευνητικής εργασίας		
Αριθμός Ερώτησης	Διατύπωση	Τύπος
1	What is the main reason for searching the web?	Multiple Choice
2	How often do you use a search engine for a lookup task?	Κλίμακα Likert
3	How often do you use a search engine for an exploratory search task?	Κλίμακα Likert

4	Have you ever used a Search Engine for Inspiration?	Yes/No
5	How would you rate your search experience?	Κλίμακα Likert

Προσαρμογή User Engagement Scale ερωτηματολογίου το οποίο απαντήθηκε από τους συμμετέχοντες μετά την ολοκλήρωση της ανατεθειμένης εργασίας

Αριθμός Ερώτησης	Διατύπωση	Τύπος
1	I felt discouraged while using CRUISE	Κλίμακα Likert
2	I felt frustrated while using CRUISE	Κλίμακα Likert
3	I felt annoyed with using CRUISE	Κλίμακα Likert
4	This search experience did not work out the way I had imagined at first place.	Κλίμακα Likert
5	I could not do some of the things I needed to do using CRUISE.	Κλίμακα Likert
6	I found CRUISE confusing to use.	Κλίμακα Likert
7	Using CRUISE was mentally taxing.	Κλίμακα Likert
8	This search experience was demanding.	Κλίμακα Likert
9	I felt in control of the searching experience.	Κλίμακα Likert
10	I felt interested in my searching tasks.	Κλίμακα Likert
11	The content of CRUISE incited my curiosity.	Κλίμακα Likert
12	The tool helped me discover interesting aspects.	Κλίμακα Likert
13	My search experience was fun.	Κλίμακα Likert
14	I felt involved in the searching tasks	Κλίμακα Likert
15	My search experience was rewarding.	Κλίμακα Likert
16	I would recommend CRUISE to my friends.	Κλίμακα Likert
17	I was really drawn into my searching task.	Κλίμακα Likert
18	I consider my search experience a success.	Κλίμακα Likert

19	Searching using CRUISE was worthwhile.	Κλίμακα Likert
20	The screen layout of CRUISE appealed to my visual senses.	Κλίμακα Likert
21	The CRUISE interface is aesthetically appealing.	Κλίμακα Likert
22	The CRUISE interface is attractive.	Κλίμακα Likert
23	I liked the graphics and images used by CRUISE.	Κλίμακα Likert
24	I was so involved in my searching task that I lost track of time.	Κλίμακα Likert
25	The time I spent searching just slipped away.	Κλίμακα Likert
26	I lost myself in this searching experience.	Κλίμακα Likert
27	I blocked out things around me when I was using CRUISE.	Κλίμακα Likert
28	I was absorbed in my searching task.	Κλίμακα Likert
29	Most of the suggested actions were relevant.	Κλίμακα Likert
30	Most of the suggested actions matched my intentions.	Κλίμακα Likert
31	The actions suggested to me were diverse.	Κλίμακα Likert
32	I understood why the items were suggested to me.	Κλίμακα Likert
33	I found it easy to inform the system if I dislike/like the suggested action.	Κλίμακα Likert
34	The layout of the suggestions interface is adequate.	Κλίμακα Likert

Παράρτημα 4: Επισκόπηση των διαθέσιμων εργαλείων με επίκεντρο τη δημιουργικότητα, την έμπνευση και την ανακάλυψη από πόρους Ιστού

Μηχανές Αναζήτησης, Διεπαφές Περιήγησης, Συστήματα Προτάσεων	Περιγραφή	Κύριο χαρακτηριστικό
http://creativesear.ch/	Μηχανή αναζήτησης ανάμιξης ποικίλου περιεχομένου από Wikipedia, Google, News, Youtube, Flickr, i-tunes, Amazon. Δημιουργήθηκε το 2009 ως μέρος διαγωνισμού D&DA (μη κερδοσκοπικού οργανισμού για τη δημιουργικότητα)	Ποικιλία πηγών πληροφορίας
https://duckduckgo.com/	Εμπορική μηχανή αναζήτησης με αποτελέσματα αναζήτησης από ποικίλες πηγές	Ποικιλία πηγών πληροφορίας
http://search-cube.com/	Προσέγγιση οπτικής αναζήτησης του διαδικτύου για εικόνες, βίντεο και ιστοσελίδες. Αναζήτηση περιστροφή και προβολή αποτελεσμάτων σε μια διεπαφή τρισδιάστατου κύβου.	Ενδιαφέρουσα Οπτικοποίηση
http://www.yossarianlives.com/	Οπτικοποιημένη και μεταφορική μηχανή αναζήτησης. Οι αλγόριθμοι της επιστρέφουν ανόμοια μεν αποτελέσματα, που όμως πιθανά έχουν μεταφορική σχέση με το ερώτημα του χρήστη.	Καινοτομία & Απροσδόκητες συσχετίσεις
http://www.cluuz.com/	Μηχανή αναζήτησης που εμφανίζει αποτελέσματα μέσω σημασιολογικών γράφων συσταδοποίησης, εξαγωγή εικόνων και νέφους επισημάνσεων. Το Cluuz ψάχνει στις σελίδες, εξάγει σημαντικούς όρους και εικόνες, τα ομαδοποιεί και τελικά τους δίνει σε	Οπτικοποίηση

	μορφή διαγράμματος (σημασιολογικός γράφος) αλλά και σε μορφή νέφους επισημάνσεων.	
http://spezify.com/	Το Spezify είναι ένα εργαλείο οπτικής αναζήτησης που προάγει την εύνοια τυχαίων ανακαλύψεων έναντι της συνάφειας αποτελεσμάτων. Τα αποτελέσματα βασίζονται σε APIs από διαφορετικούς παρόχους περιεχομένου όπως Amazon, Facebook, Soundcloud και Vimeo, παράγοντας ένα εκλεκτικό μίγμα από αντικείμενα εκτεθειμένα σε μορφή κρυμμένων πολύτιμων πετραδιών.	Ποικιλία πηγών πληροφορίας
http://search.carrot2.org/	Το Carrot είναι μία Ανοιχτού Κώδικα Μηχανή Ομαδοποίησης Αποτελεσμάτων Αναζήτησης. Αυτόματα οργανώνει μικρές συλλογές εγγράφων (αποτελεσμάτων αναζήτησης και όχι μόνο) σε θεματικές κατηγορίες.	Οπτικοποίηση
http://www.mashpedia.com/	Επιστρέφει περιεχόμενο από διάφορες online πηγές σε πραγματικό χρόνο, το συγκεντρώνει και παρουσιάζει μια οργανωμένη προοπτική για κάθε θέμα, η οποία περιλαμβάνει: πληροφορία από Wikipedia, πρόσφατα νέα, βιβλία, βίντεο, εικόνες, κοινωνικές απαντήσεις, μηνύματα twitter και σχετικές σελίδες Facebook.	Ποικιλία πηγών πληροφορίας
http://www.helioid.com/	Μια μηχανή αναζήτησης που δείχνει τι κρύβεται πίσω από την πρώτη	Οπτικοποίηση

	σελίδα των αποτελεσμάτων και βοηθάει τον χρήστη να βρει πράγματα ακόμη και όταν δεν είναι σίγουρος τι ψάχνει παρουσιάζοντας κατηγορίες που δημιουργούνται από το κείμενο των αποτελεσμάτων αναζήτησης για να επαναπροσδιορίσετε την αναζήτησή σας ή να αναλύσετε τα συγκεκριμένα θέματα ενδιαφέροντος.	
http://search.yippy.com/	Η μηχανή αναζήτησης Yippy, γνωστή και ως Clusty, ομαδοποιεί τα αποτελέσματα σε θεματικές, ή "clouds" για καλύτερη αναζήτηση και ανακάλυψη.	Οπτικοποίηση
http://www.twitscoop.com	Το Twitscoop είναι ένα εργαλείο απεικόνισης σε πραγματικό χρόνο το οποίο επιτρέπει στους χρήστες να «εξορύσσουν τη ροή σκέψεων» που παρέχει το Twitter. Ο αλγόριθμος τους κόβει κάθε Αγγλικό μη επαναλαμβανόμενο tweet σε κομμάτια ("tags"), και τα ταξινομεί ανάλογα με το πόσο συχνά χρησιμοποιούνται σε σχέση με την κανονική χρήση.	Οπτικοποίηση
http://cloud.li/	Ανακαλύπτοντας εξέχοντα hashtags και αναδυόμενες τάσεις που σχετίζονται με οποιοδήποτε θέμα. Παρέχει μια λίστα λέξεων-κλειδιών και όρων που χρησιμοποιούνται συνήθως σε συνδυασμό με έναν συγκεκριμένο όρο ερωτήματος.	Περιεχόμενο από Κοινωνικά Μέσα

	Εξετάζει μόνο τα τελευταία tweets, οπότε για ένα καυτό θέμα παρέχει ένα συνεχώς μεταβαλλόμενο σύννεφο.	
http://www.filtize.com/	Πραγματικού χρόνου αναζήτηση του Twitter μέσω της επεξεργασίας δεδομένων, το φιλτράρισμα και την ενισχυμένη οπτικοποίηση και πλοήγηση. Τα δεδομένα επεξεργάζονται και εμπλουτίζονται στον Filtize Server χρησιμοποιώντας τεχνολογίες επεξεργασίας φυσικής γλώσσας.	Οπτικοποίηση
http://www.jasondavies.com/wordcloud/	Οπτικοποίηση των αποτελεσμάτων μοντελοποίησης θέματος με τη χρήση σύννεφων λέξεων. Αναπαράσταση όχι μόνο των λέξεων που αποτελούν ένα θέμα, αλλά και της αναλογίας αυτών των λέξεων.	Οπτικοποίηση
http://www.nextanalytics.com/twitter-analytics-excel-twitter-tag-cloud/	Ένα εργαλείο ανάλυσης που αναλύει το περιεχόμενο του Twitter από το Χρονολόγιο ενός ατόμου	Περιεχόμενο από Κοινωνικά Μέσα
http://twahpic.cloudapp.net/S4.aspx/	Έργο επίδειξης Έρευνας Microsoft που συνδυάζει ημι-εποπτευόμενη μοντελοποίηση θεμάτων με ροές twitter για να επιτρέπει στους ανθρώπους να διερευνήσουν τι αφορά ένα σύνολο tweets (από έναν χρήστη, που ταιριάζει σε ένα ερώτημα).	Οπτικοποίηση βάσει θεμάτων
http://www.chacha.com/	Μηχανή αναζήτησης που παρέχει δωρεάν, σε πραγματικό χρόνο	Κοινωνική Δικτύωση

	απαντήσεις σε κάθε ερώτηση, μέσω της ιστοσελίδας της, με χρήση μηνυμάτων κειμένου ή χρησιμοποιώντας μία από τις εφαρμογές της εταιρείας για κινητό.	
https://www.quora.com/	Ιστοσελίδα ερωτήσεων και απαντήσεων που δημιουργήθηκε, επεξεργάστηκε και οργανώθηκε από την κοινότητα χρηστών της. Η Quora συγκεντρώνει ερωτήσεις και απαντήσεις σε θέματα. Οι χρήστες μπορούν να συνεργαστούν με επεξεργασία ερωτήσεων και προτείνοντας αλλαγές στις απαντήσεις άλλων χρηστών.	Κοινωνική Δικτύωση
http://socialmention.com/	Το Social Mention είναι μια κοινωνική πλατφόρμα αναζήτησης και ανάλυσης μέσων που συγκεντρώνει το περιεχόμενο που δημιουργείται από το χρήστη από όλο το σύμπαν σε μια ενιαία ροή πληροφοριών .	Ποικιλία περιεχομένου & περιεχόμενο από κοινωνικά μέσα
https://delicious.com/	Κοινωνική υπηρεσία web bookmarking για την αποθήκευση, κοινή χρήση και ανακάλυψη σελιδοδεικτών ιστού.	Κοινωνική Συνεργασία
http://en.reddit.com/	Ιστοσελίδα κοινωνικών νέων και ψυχαγωγίας, όπου οι εγγεγραμμένοι χρήστες υποβάλλουν περιεχόμενο είτε με τη μορφή συνδέσμου ή με ανάρτηση κείμενου. Έπειτα άλλοι χρήστες υπερψηφίζουν ή καταψηφίζουν την ανάρτηση, ταξινομώντας τη ανάλογα και	Κοινωνική Συνεργασία

	ορίζοντας έτσι τη θέση της στις σελίδες του ιστότοπου και την πρώτη σελίδα του.	
https://www.stumbleupon.com/	Μηχανή Ανακάλυψης που βρίσκει και συνιστά περιεχόμενο ιστού στους χρήστες του. Οι δυνατότητές του επιτρέπουν στους χρήστες να ανακαλύπτουν και να αξιολογούν ιστοσελίδες, φωτογραφίες και βίντεο που είναι εξατομικευμένα σύμφωνα με τις προτιμήσεις και τα ενδιαφέροντά τους, χρησιμοποιώντας αρχές ομοτίμων και κοινωνικής δικτύωσης.	Κοινωνική Συνεργασία και Εξατομικευμένες Προτάσεις
http://www.eurekster.com/	Οι χρήστες μπορούν να δημιουργήσουν και να προσαρμόσουν την πύλη αναζήτησης σε οποιοδήποτε θέμα και να μοιραστούν και να διανείμουν το widget κοινωνικής αναζήτησης για να αναπτύξουν μια κοινότητα ενδιαφερομένων χρηστών. Με κάθε αναζήτηση, το swiki γίνεται πιο συναφές και σημαντικό για την κοινότητα των χρηστών και πιο πολύτιμο για το swikibuide. Εμφανίζει επίσης τα πιο συχνά ερωτήματα στο σύννεφο ετικετών..	Κοινωνική Συνεργασία – Εξατομίκευση - Οπτικοποίηση
http://pipl.com	Μηχανή αναζήτησης προσώπων	Κοινωνικός παράγοντας
http://www.google.com/cse/	Εξατομικευμένη μηχανή αναζήτησης	Εξατομίκευση

http://www.amazon.com	Υπηρεσία συστάσεων βασισμένη στο συνεργατικό φιλτράρισμα για προϊόντα (βιβλία, λογισμικό, κλπ.).	Εξατομίκευση και Κοινωνικά Δίκτυα
http://www.lastfm.com/	Υπηρεσία συστάσεων βασισμένη στο συνεργατικό φιλτράρισμα για μουσική.	Εξατομίκευση και Κοινωνικά Δίκτυα
http://www.movielens.org/	Υπηρεσία συστάσεων βασισμένη στο συνεργατικό φιλτράρισμα για ταινίες.	Εξατομίκευση και Κοινωνικά Δίκτυα

Παράρτημα 5: Κώδικας Matlab Αλγορίθμου Προώθησης

```
function p=pr_hmm(o,a,b,pi)
%INPUTS:
%O=Given observation sequence labelled in numerics
%A(N,N)=transition probability matrix
%B(N,M)=Emission matrix
%pi=initial probability matrix
%Output
%P=probability of given sequence in the given model
n=length(a(1,:));
T=length(o);
%it uses forward algorithm to compute the probability
for i=1:n %it is initialization
    m(1,i)=b(i,o(1))*pi(i);
end
for t=1:(T-1) %recursion
    for j=1:n
        z=0;
        for i=1:n
            z=z+a(i,j)*m(t,i);
        end
        m(t+1,j)=z*b(j,o(t+1));
    end
end
p=0;
for i=1:n %termination
    p=p+m(T,i);
end
```

