



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Τομέας Βιομηχανικών Διατάξεων και Συστημάτων
Αποφασεων

**Αναπαράσταση Γνώσης σε
Πρακτικές Αξιολόγησης Έργων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μαρία – Μαλεβή Π. Δουρούκα

Επιβλέπων: Δημήτριος Ασκούνης
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2011



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
 ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
 ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
 Τομέας Βιομηχανικών Διατάξεων και Συστημάτων
 Αποφασεων

**Αναπαράσταση Γνώσης σε
 Πρακτικές Αξιολόγησης Έργων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μαρία – Μαλεβή Π. Δουρούκα

Επιβλέπων: Δημήτριος Ασκούνης
 Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

.....
 Δημήτριος Ασκούνης
 Επίκουρος Καθηγητής Ε.Μ.Π.

.....
 Ιωάννης Ψαρράς
 Καθηγητής Ε.Μ.Π.

.....
 Βασίλειος Ασημακόπουλος
 Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2011

.....

Μαρία – Μαλεβή Π. Δουρούκα

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright©Μαρία – Μαλεβή Π. Δουρούκα 2011

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.

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια, ο αυξανόμενος όγκος των διαθέσιμων πληροφοριών δημιούργησε νέες ανάγκες διαχείρισης της γνώσης, τόσο στις επιστήμες όσο και στους τομείς της βιομηχανίας και των επιχειρηματικότητας. Τα συστήματα αναπαράστασης γνώσης έρχονται να καλύψουν τις ανάγκες αυτές, παρέχοντας τρόπους κατηγοριοποίησης, κατάταξης και ανάκτησης γνώσης που έχει ήδη αποκτηθεί. Ιδιαίτερα στον τομέα της διαχείρισης έργων, η γνώση αυτή είναι πολύτιμη, καθώς μπορεί να χρησιμεύσει μελλοντικά, είτε προτρεπτικά είτε ενθαρρυντικά, σε αποφάσεις για την υιοθέτηση τακτικών διαχείρισης.

Το σύστημα ROM είναι ένα σύστημα αξιολόγησης έργων που χρηματοδοτούνται από την Ευρωπαϊκή Ένωση. Μέσα από το σύστημα αυτό, παράγεται ένα πλήθος πληροφοριών σχετικά με την πορεία των έργων, υπό τη μορφή αναφορών αξιολόγησης που συντάσσονται από ειδικούς. Οι πληροφορίες αυτές συμπεριλαμβάνονται στις απαντήσεις των ειδικών, υπό τη μορφή ελεύθερου κειμένου, στις ερωτήσεις του συστήματος ROM και αφορούν τόσο σε παράγοντες που επηρεάζουν την βαθμολόγηση ενός έργου, όσο και σε συμπεράσματα – διδάγματα που προκύπτουν από την πορεία του.

Στόχος της παρούσας διπλωματικής εργασίας είναι ο σχεδιασμός ενός συστήματος για την αναπαράσταση της γνώσης που περιέχεται στις αναφορές αξιολόγησης του συστήματος ROM. Το σύστημα περιλαμβάνει τους παράγοντες και τα διδάγματα που εντοπίζονται, τις κατηγορίες κατάταξής τους, ανάλογα με τον τομέα του έργου στον οποίο αναφέρονται, τα στοιχεία εντοπισμού τους (id έργου, αριθμός ερώτησης), καθώς και δείκτες – βαθμολογίες όσον αφορά στη θετικότητα ή αρνητικότητά τους. Ως εφαρμογή, εξετάζεται ένα σύνολο αναφορών του συστήματος ROM. Τα στοιχεία που εντοπίζονται κατατάσσονται στις κατηγορίες που δημιουργήθηκαν και τους ανατίθεται κάποια βαθμολογία θετικότητας ή αρνητικότητας.

Λέξεις κλειδιά

Αναπαράσταση Γνώσης, Εξόρυξη Γνώσης, Διδάγματα, Εξαγωγή Πληροφοριών, Εξόρυξη Γνώμης, Βάση Γνώσης

ABSTRACT

In recent years, the continually increasing amount of available data has created new knowledge management needs, both in the scientific and in the professional fields. The knowledge representation systems have come to meet these needs, providing new means of knowledge categorization, segmentation and querying. This knowledge has been proved to be extremely valuable, particularly in project management processes, as it can serve as an incentive or a deterrent for the adoption of future management techniques.

The ROM system is a monitoring system designed for projects subsidized by the European Commission. Though the system processes, a plethora of information is produced, concerning the course of the projects, in the form of reviews conducted by experts. The information can be found in the free-text response of the experts on questions, including factors that affect the rating of a project and conclusions – lessons learned drawn up from the process.

The aim of the thesis is the design of a system representing the knowledge contained in the evaluation reports of ROM. The system includes identified factors and lessons learned, the categories of classification, depending on the area of they refer to, and indicators – scores of positivity or negativity of each entity. For its application, a set of reports is thoroughly examined. The extracted concepts are classified into the categories created and the scores of positivity or negativity are assigned to them.

Keywords

Knowledge representation, knowledge extraction, lessons learned, information extraction, opinion mining, knowledge – based system

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία αποτελεί την επισφράγιση της μέχρι τώρα πορείας μου στο Εθνικό Μετσόβιο Πολυτεχνείο και σηματοδοτεί την έναρξη μιας καινούργιας περιόδου, γι αυτό και αισθάνομαι την ανάγκη να αναγνωρίσω τη συμβολή κάποιων ανθρώπων στη πορεία αυτή.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της διπλωματικής εργασίας, κ. Δημήτριο Ασκούνη, για την ανάθεση της εργασίας, την ευκαιρία που μου έδωσε να γνωρίσω αυτό τον επιστημονικό τομέα και να αποκτήσω νέες γνώσεις μέσα από τη διαδικασία. Η συγκεκριμένη εργασία δεν θα ήταν φυσικά δυνατό να ολοκληρωθεί χωρίς την πολύτιμη καθοδήγηση και βοήθεια που έλαβα από τον υποψήφιο διδάκτορα Χρήστο Ντάνο, τον οποίο αισθάνομαι την υποχρέωση να ευχαριστήσω θερμά για τη διάθεσή του και την ουσιαστική βοήθειά του σε κάθε σημείο που αυτή χρειάστηκε.

Αναμφίβολα, ξεχωριστό μέρος των φοιτητικών μου χρόνων αποτέλεσαν οι φίλοι μου, τους οποίους θα ήθελα να ευχαριστήσω θερμά που ήταν δίπλα μου και με δίδαξαν όλοι, με τη σειρά τους, πράγματα. Ξεχωριστά, θα ήθελα να ευχαριστήσω τη Γιώτα, τον Μιχάλη, τη Φαίη, τη Μαρία και την Ίριδα, που με στήριξαν σε όλες τις δυσκολίες και έκαναν τις ωραίες στιγμές ακόμη ομορφότερες.

Τέλος, και περισσότερο από κάθε άλλον, θα ήθελα να ευχαριστήσω την οικογένειά μου και ιδιαίτερα τους γονείς μου, Ελένη και Τάκη, και τους παππούδες μου, Κική και Βαγγέλη, οι οποίοι μου δίδαξαν, μεταξύ άλλων, την υπομονή και την επιμονή, στοιχεία που διαμόρφωσαν τη διαδρομή μου. Η αγάπη τους και η στήριξή τους με βοήθησε να φτάσω εδώ σήμερα και είναι αυτά που μου δίνουν τη δύναμη να κυνηγήσω τα όνειρά μου.

Μαρία – Μαλεβή Δουρούκα
Ιούνιος 2011

Στην Ελένη, τον Τάκη,
την Κική και τον Βαγγέλη
που πάντα μοιράζονταν τη γνώση τους μαζί μου

«If only HP knew what HP knows, we would be three times more productive»
Lew Platt, Hewlett-Packard CEO

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ



ΑΚΡΩΝΥΜΙΑ

Ακρωνύμιο	Περιγραφή Ακρωνυμίου
AAC	Adverb - Adjective Combination
APPL	Academy of Program and Project Leadership
BCS	Background Conclusion Sheets
CRF	Conditional Random Fields
DM	Data Mining
EC	European Commission
FSM	Finite State Machines
IE	Inform

	ation Extracti on
KD	Knowle dge Discov ery
KEFIR	Key Finding Report er
KOMIS	Knowle dge Optimi zation Meta- Inform ation System
LF	Logical Frame work
LLIS	Lesson s Learne d Inform ation System
MEM M	Maxim um Entrop y Marko v Model
MISP	Munici pal Infrastr ucture Support Progra

	mme
ML	Machine Learning
MR	Monitoring Report
NLP	Natural Language Processing
NSA Knowledge Base	National Security Agency Knowledge Base
OLAP	Optical Analytical Processing
OO	Overall Objective
OVI	Objectively Verifiable Indicator
PBMA Knowledge System	Process-Based Mission Assurance Knowledge

	Manag ement System
PP	Project Purpos e
ROM	Results Orienta ted Monito ring
sc(adj)	score of adjecti ve
sc(adv)	score of adverb
SKICAT	Sky Image Catalog ing and Analysi s Tool
SMART	Specific Measur able Attaina ble Releva nt Time - Bound
TASA	Teleco mmuni cation Networ k Alarm Sequen ce Analyz er

TGRW	Things Gone Right/ Wrong
TM	Text Mining

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

ΚΕΦΑΛΑΙΟ 1^ο: ΕΙΣΑΓΩΓΗ

1.1 Το αντικείμενο και ο στόχος της διπλωματικής εργασίας

Αναμφίβολα, ζούμε στην εποχή της γνώσης. Ο όγκος των διαθέσιμων δεδομένων τα τελευταία χρόνια είναι ολοένα και αυξανόμενος. Οι πηγές γνώσης ποικίλουν σε μορφή και τομείς εφαρμογής, αλλά γενικά περιλαμβάνουν εγχειρίδια, γράμματα, περιλήψεις, αναφορές, άρθρα, και σε επιχειρησιακό επίπεδο, απαντήσεις σε πελάτες, ανταγωνιστική γνώση (competitor intelligence) και γνώση που προέρχεται από εργασιακές διαδικασίες. Ένα ευρύ φάσμα τεχνολογιών χρησιμοποιείται για να εφαρμόσει τη διαχείριση γνώσης σε συστήματα: e-mail, βάσεις δεδομένων (databases) και αποθήκες δεδομένων (data warehouses), συστήματα υποστήριξης, περιηγητές (browsers) και μηχανές αναζήτησης, διαδίκτυο, συστήματα βασισμένα στη γνώση (knowledge – based systems) και «έξυπνοι πράκτορες γνώσης» (intelligent agents).

Η εξόρυξη γνώσης είναι ένας νέος και γρήγορα αναπτυσσόμενος τομέας, που χρησιμοποιεί στοιχεία από την τεχνητή νοημοσύνη, τα μαθηματικά, την στατιστική για να εξάγει τη γνώση που βρίσκεται σε μορφή δεδομένων. Οι μεθοδολογικές προσεγγίσεις για την εξόρυξη γνώσης σχεδιάζονται με σκοπό να εκμεταλλευτούν τα χαρακτηριστικά και τις δομές του τομέα στον οποίο αναφέρονται, γι αυτό και η διαδικασία εξόρυξης γνώσης βρίσκει εφαρμογή σε πλήθος περιπτώσεων, όπως ανάλυση νοθιών, ασφάλεια, ανάλυση πελατών και προϊόντων.

Ένας από τους κυριότερους τομείς εφαρμογής της διαχείρισης γνώσης σε επιχειρησιακό επίπεδο, σχετίζεται με τη γνώση που έχει ήδη αποκτηθεί από πρωθυστερες ενέργειες. Πιο συγκεκριμένα, η γνώση που αφορά σε τακτικές που υιοθετήθηκαν σε παλαιότερες περιπτώσεις και στις επιδράσεις που είχαν, εάν καταχωρηθεί και διαχειριστεί με κατάλληλο τρόπο, μπορεί μελλοντικά να χρησιμεύσει διδακτικά για τη λήψη αποφάσεων σχετικά με παρόμοιες ενέργειες.

Το σύστημα ROM είναι ένα σύστημα αξιολόγησης έργων που χρηματοδοτούνται από την Ευρωπαϊκή Ένωση. Η διαδικασία της αξιολόγησης παράγει ένα πλήθος αναφορών, στις οποίες οι εμπειρογνώμονες – επιβλέποντες των έργων καλούνται να βαθμολογήσουν και να απαντήσουν με ελεύθερο κείμενο σε ερωτήσεις σχετικά με την πορεία των έργων, τους παράγοντες που την επηρέασαν καθώς και διδάγματα που προέκυψαν από τη διαδικασία. Μέσω του συστήματος ROM, συνεπώς, δημιουργείται ένα πλήθος αναφορών που περιλαμβάνουν ιδιαίτερα χρήσιμες πληροφορίες, των οποίων όμως η ανάκτηση είναι μια επίπονη διαδικασία, στην παρούσα μορφή της. Μέσα σε αυτό το σύνολο αναφορών περιέχεται γνώση σχετικά με τακτικές που οδηγούν σε επιτυχία ή αποτυχία ενός έργου, καθώς και με ενέργειες που μπορούν να γίνουν ώστε να διορθωθούν κάποιες αδυναμίες ή να βελτιωθούν κάποια θετικά στοιχεία του έργου. Γίνεται, επομένως, σαφής η χρησιμότητα ενός συστήματος που θα περιλαμβάνει όλες αυτές τις πληροφορίες, προσφέροντας έτσι τη δυνατότητα εύκολης, ταχείας και αποτελεσματικής ανάκτησής τους.

Στόχος της παρούσας διπλωματικής εργασίας είναι ο σχεδιασμός ενός συστήματος αναπαράστασης της γνώσης που περιέχεται στις αναφορές του συστήματος ROM. Μέσα από ένα σύνολο σχεδιαστικών αποφάσεων και μεθοδολογιών, παρουσιάζεται η δομή του συστήματος, τα στοιχεία και οι πληροφορίες που θα περιέχει, και γίνεται μία δοκιμαστική

εφαρμογή του σε κάποιες από τις αναφορές έργων. Σκοπός είναι ο σχεδιασμός του συστήματος να ανταποκρίνεται στα δεδομένα του συστήματος ROM, και τα αποτελέσματα που θα παρουσιάζει να είναι στα απαιτούμενα, για τη χρήση του, επίπεδα γενίκευσης και λεπτομέρειας. Ιδιαίτερη έμφαση δίνεται στις πληροφορίες που αφορούν στα διδάγματα που προκύπτουν από τις αναφορές των έργων, καθώς η χρησιμότητα του συστήματος έγκειται στη διασύνδεση των ενεργειών που επηρεάζουν την πορεία ενός έργου με τους τρόπους αποφυγής ή υιοθέτησης των ενεργειών αυτών, ανάλογα με το αν είναι αρνητικές ή θετικές αντίστοιχα. Επιπλέον, έγινε η προσπάθεια απονομής βαρών στις έννοιες που εντοπίζονται, έτσι ώστε ο χρήστης να μπορεί να αντιληφθεί τον βαθμό θετικότητας ή αρνητικότητας των εννοιών. Αρχικά, το σύστημα θα χρησιμοποιείται χειροκίνητα, με τις καταχωρήσεις στις κατηγορίες του να γίνονται από ειδικούς. Απώτερος σκοπός είναι η μελλοντική αυτοματοποίηση της καταχώρησης εννοιών στο σύστημα, από σύστημα αυτόματης εξόρυξης από ελεύθερο κείμενο. Τα χειροκίνητα αποτελέσματα, σε αυτή την περίπτωση, θα μπορούν να χρησιμεύσουν για «εκπαίδευση» (training) της εκμάθησης μηχανής (ML), και για σύγκριση με τα αυτόματα.

1.2 Η δομή της διπλωματικής

Η δομή της διπλωματικής περιλαμβάνει τόσο θεωρητικά όσο και πρακτικά στοιχεία, καθώς κρίθηκε απαραίτητη η αρχική ανάλυση σε θεωρητικό επίπεδο των πρακτικών ζητημάτων.

Αρχικά, στο δεύτερο κεφάλαιο γίνεται μια εισαγωγή στον τομέα της διαχείρισης γνώσης, παραθέτοντας βασικά στοιχεία και αρχές σχετικά με τον κλάδο και τις διαδικασίες που ακολουθούνται. Γίνονται αναφορές στις έννοιες της Εξόρυξης Δεδομένων και της Εξαγωγής Πληροφοριών και ερευνάται η ευρύτερη συμβολή τους στη διαχείριση γνώσης. Επιπλέον, μελετάται ο τομέας της Αναπαράστασης Γνώσης, δίνοντας έμφαση στην ιεράρχηση των εννοιών, και παρουσιάζονται κάποια βασικά στοιχεία της θεωρίας των Οντολογιών. Στο κεφάλαιο 2.5 παρουσιάζονται κάποια επιλεγμένα παραδείγματα συστημάτων αναπαράστασης γνώσης, που προέκυψαν έπειτα από σύντομη βιβλιογραφική έρευνα για συστήματα με παρόμοιους σκοπούς με το παρόν.

Το τρίτο κεφάλαιο αναφέρεται στον τομέα της διαχείρισης έργων. Αρχικά, δίνονται κάποια βασικά στοιχεία του κλάδου και των διαδικασιών επίβλεψης και αξιολόγησης έργων. Ιδιαίτερη έμφαση δίνεται στον ρόλο της αναπαράστασης γνώσης στη διαχείριση έργων, καθώς και στην αξία των διδαγμάτων, και έτσι γίνεται περισσότερο εμφανής η χρησιμότητα του συστήματος. Επιπλέον, λαμβάνοντας υπόψη τον τομέα εφαρμογής του συστήματος, παρουσιάζεται και αναλύεται ο τρόπος αξιολόγησης έργων του συστήματος ROM. Δίνονται παραδείγματα αναφορών και αναλύονται τα κριτήρια βαθμολόγησης έργων σύμφωνα με το σύστημα, καθώς ήταν καθοριστικός παράγοντας στις σχεδιαστικές αποφάσεις που ελήφθησαν.

Στο τέταρτο κεφάλαιο, παρουσιάζεται η μεθοδολογία σχεδιασμού του συστήματος. Κάθε βήμα αναλύεται διεξοδικά, και συμπεριλαμβάνει όλες τις σχεδιαστικές αποφάσεις καθώς και τους λόγους που οδήγησαν στη λήψη τους. Στο κεφάλαιο 4.3.3 παρουσιάζονται αναλυτικά οι κατηγορίες που επιλέχθηκαν για το σύστημα και η ιεράρχησή τους. Στο κεφάλαιο 4.3.6 αναλύεται η μεθοδολογία που ακολουθήθηκε για την βαθμολόγηση των εννοιών και δίνονται διάφορα παραδείγματα υπολογισμού τους.

Το πέμπτο κεφάλαιο περιλαμβάνει την εφαρμογή του συστήματος, η οποία δίνεται υπό μορφή πίνακα στο Παράρτημα Α1. Παρουσιάζονται συνοπτικά τα στοιχεία που περιλαμβάνει το σύστημα και μελετάται ένα σύνολο αναφορών έργων. Οι παράγοντες που εντοπίζονται στα έργα καταχωρούνται στο σύστημα και τους αναθέτονται βαθμολογίες, σύμφωνα με τη μεθοδολογία που αναπτύχθηκε.

Στο έκτο κεφάλαιο συγκεντρώνονται τα συμπεράσματα από την εφαρμογή του συστήματος. Αναλύεται η αποτελεσματικότητα παρουσίασης των αποτελεσμάτων, βάσει των εννοιών που εντοπίστηκαν στις αναφορές. Επίσης, αξιολογείται η μεθοδολογία βαθμολόγησης των εννοιών και εντοπίζονται περιθώρια βελτίωσής της. Τέλος, στο έκτο κεφάλαιο παρουσιάζονται μελλοντικές ενέργειες σχετικά με το σύστημα που αφορούν στην εξέλιξη και αυτοματοποίησή του.

ΚΕΦΑΛΑΙΟ 2^ο: ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ ΑΠΟ ΕΛΕΥΘΕΡΟ ΚΕΙΜΕΝΟ

2.1 Γενικά για την Εξόρυξη Γνώσης

Είναι γεγονός πως η σύγχρονη τεχνολογία υπολογιστών και αποθήκευσης συνεχώς τείνει να ξεπερνά την ικανότητα των ανθρώπων να μπορούν να χρησιμοποιούν ουσιαστικά τον τεράστιο όγκο των διαθέσιμων δεδομένων. Χωρίς μια εναρμονισμένη προσπάθεια για ανάπτυξη τεχνικών εξεύρεσης γνώσης, οι χρήστες στερούνται μεγάλο μέρος της αξίας των δεδομένων που συλλέγονται και αποθηκεύονται.

Ο απώτερος σκοπός της εξεύρεσης γνώσης (Knowledge Discovery, KD) είναι η εξαγωγή υπονοούμενων, πρωθύστερα άγνωστων και πιθανόν χρήσιμων πληροφοριών από δεδομένα (Frawley W.J, 1991). Λόγω της συνεχούς αύξησης του όγκου των διαθέσιμων ηλεκτρονικών δεδομένων, οι τεχνικές αυτόματης εξαγωγής γνώσης γίνονται όλο και περισσότερο απαραίτητες για τη διαχείριση του πλήθους των πληροφοριών που είναι καταχωρημένες σε πληροφοριακά συστήματα. Επιπλέον, οι συνήθεις τεχνικές Εξόρυξης Δεδομένων (Data Mining, DM) είναι σχεδιασμένες πρωτίστως για την εξαγωγή πληροφοριών από δομημένες βάσεις δεδομένων. Συνεπώς, είναι προφανής η ανάγκη για ανάπτυξη ειδικών τεχνικών Εξόρυξης Κειμένου (Text Mining, TM) για την εξαγωγή σημαντικών πληροφοριών που συμπεριλαμβάνονται σε ελεύθερο κείμενο.

Ο τομέας της εξόρυξης δεδομένων και της εξεύρεσης γνώσης είναι άρρηκτα συνδεδεμένος με τις βάσεις δεδομένων. Η εξεύρεση γνώσης επιτρέπει την σημαντική αύξηση του μεγέθους των βάσεων δεδομένων από τις οποίες μπορούν να εξαχθούν χρήσιμα συμπεράσματα, κάνοντάς τες περισσότερο φιλικές προς τους χρήστες, και έτσι διευκολύνοντας τη διαχείριση και χρήση των τεράστιων όγκων δεδομένων. Από τις αρχές της επιστήμης της πληροφορικής, οι βάσεις δεδομένων ήταν εξαιρετικά χρήσιμες. Οποιοσδήποτε μπορούσε εύκολα να καταχωρήσει εγγραφές σημαντικών μεταφορών, να τις αναζητήσει αργότερα χωρίς καμία δυσκολία και να λάβει αποφάσεις βασιζόμενες στα δεδομένα που βρίσκονταν στη βάση. Μέσω της ιστοσελίδας Wal-mart, για παράδειγμα, πραγματοποιούνται καθημερινά περίπου 20 εκατομμύρια μεταφορών. Το σύστημα παρακολούθησης της γης της NASA, στο τέλος της προηγούμενης δεκαετίας, αναμενόταν να παράγει 50GB δεδομένων εικόνας ανά ώρα. Η γνώση που αποθηκεύεται με ολόένα και ταχύτερους ρυθμούς, γίνεται όλο και λιγότερο κατανοητή καθώς «χάνεται» μέσα στα gigabytes των καταχωρήσεων. Κανένας άνθρωπος δε μπορεί να χρησιμοποιήσει τόσο μεγάλο πλήθος δεδομένων με αποτελεσματικό τρόπο, ώστε να καταλάβει τα βασικά στοιχεία που χρειάζονται για να μπορέσει να πάρει λογικές αποφάσεις. Η πληροφορία, αν και αποθηκευμένη, γίνεται όλο και λιγότερο χρήσιμη καθώς συνεχώς ερχόμαστε αντιμέτωποι με δυσκολίες σχετικά με το πώς να την ανακτήσουμε και να τη κάνουμε διαθέσιμη σε μια κατανοητή μορφή. Για παράδειγμα, αν το αποτέλεσμα μιας αναζήτησης σε βάση δεδομένων αποτελείται από χιλιάδες καταχωρήσεις, τα δεδομένα που ανακτήθηκαν έχουν μηδαμινή χρηστική αξία.

Η περιοχή της εξόρυξης γνώσης εμφανίστηκε σαν μια προσπάθεια να αντιμετωπιστούν αυτά τα ζητήματα. Ο όρος εμφανίστηκε για πρώτη φορά το 1989, από τους Frawley, Piatetsky, Shapiro και Matheus:

Η εξεύρεση γνώσης σε βάσεις δεδομένων είναι μια διαδικασία εντοπισμού έγκυρων, νέων, δυνητικά χρήσιμων και απολύτως κατανοητών μοτίβων σε δεδομένα.

Ο όρος «μοτίβο» (pattern) χρησιμοποιείται εδώ με την ευρεία έννοια του. Μπορεί να περιλαμβάνει σχέσεις, συσχετίσεις, περιγραφές γεγονότων κ.κ. (Krzysztof J. Cios, 1998)

Υπάρχουν πολλοί τρόποι ορισμού του στόχου της εξόρυξης δεδομένων και των διαδικασιών εξεύρεσης γνώσης, και εξαρτώνται κυρίως από τη γενικότητα του προβλήματος και τα αναμενόμενα αποτελέσματα. Μπορούμε να διακρίνουμε διάφορα επίπεδα εξόρυξης δεδομένων:

- **Αμιγής εξόρυξη δεδομένων:** Αυτό είναι ίσως το πιο πιθανό σενάριο. Η αναζήτηση του χρήστη μπορεί να περιγραφεί γενικά με την πρόταση «Πες μου κάτι ενδιαφέρον για τα δεδομένα μου». Δεν υπάρχουν περιορισμοί στο σύστημα και δεν υπάρχουν ενδείξεις σχετικά με τι περιμένει ο χρήστης ή τι τύπος αναζήτησης τον ενδιαφέρει. Υπάρχει το ενδεχόμενο ανεύρεσης απρόσμενων (και γι αυτό και πολύ ενδιαφερόντων) μοτίβων αλλά ενέχεται και ο κίνδυνος της αναπαραγωγής ήδη γνωστών μοτίβων που δεν προσφέρουν καμία νέα πληροφορία σχετικά με τα δεδομένα.
- **Κατευθυνόμενη εξόρυξη δεδομένων:** Εδώ το σημείο εστίασης του ενδιαφέροντος είναι περισσότερο ορισμένο. Ο χρήστης, για παράδειγμα, ενδιαφέρεται να μάθει τα προφίλ κάποιων συγκεκριμένων πελατών. Σε σύγκριση με τον προηγούμενο τύπο, η διαδικασία περιλαμβάνει ένα πρόβλημα συμπερασμάτων.
- **Βελτίωση υποθέσεων και δοκιμές:** Σε αυτή την περίπτωση ο χρήστης γίνεται ακόμη πιο συγκεκριμένος, δίνοντας κάποιες υποθέσεις και αναμένοντας από το σύστημα να τις αξιολογήσει, να δοκιμάσει τα αποτελέσματα και, σε περίπτωση που δεν ισχύουν οι υποθέσεις, να τα τροποποιήσει και να τα εξελίξει. Ένα παράδειγμα αυτού του τύπου αναζήτησης θα μπορούσε να είναι «Πιστεύω ότι υπάρχει συσχέτιση μεταξύ των πωλήσεων υπολογιστών και των πωλήσεων εξοπλισμού βίντεο, τους τελευταίους μήνες, ισχύει;»

Ο Fayyad (Usama Fayyad, 1996) επικεντρώθηκε στην εξόρυξη γνώσης και εντόπισε την αλληλεπίδραση και φιλικότητα της διαδικασίας με τον χρήστη. Κάτι τέτοιο είναι άλλωστε αναμενόμενο, καθώς ο τελικός χρήστης είναι εκείνος που θα πρέπει να δεχτεί ή να απορρίψει τα αποτελέσματα που προκύπτουν από τη διαδικασία της εξαγωγής γνώσης. Όταν ορίζουμε τη διαδικασία της εξαγωγής γνώσης, θα πρέπει να δώσουμε έμφαση στον επαναληπτικό της χαρακτήρα όπου ωστόσο η αλληλεπίδραση με τον χρήστη παίζει ένα πρωταρχικό ρόλο. (Krzysztof J. Cios, 1998) Οι όροι «εξόρυξη δεδομένων» και «εξόρυξη γνώσης» είναι πολύ συχνά εναλλασσόμενοι μεταξύ τους.

Τα κύρια χαρακτηριστικά της εξόρυξης γνώσης σχετίζονται τόσο με τον τομέα όσο και με τον τρόπο εφαρμογής της. Η εξόρυξη γνώσης ασχολείται με προβλήματα που σχετίζονται με τον μεγάλο όγκο δεδομένων και με τη διαχείριση μεγάλων αρχείων, προσφέροντας τη δυνατότητα για εύκολη και αποτελεσματική πρόσβαση σε αυτά.

Επιπλέον, βασίζεται σε μια διεπιστημονική έρευνα και χρησιμοποιεί ένα πλήθος τεχνολογιών πληροφορικής, οι οποίες είναι ή μπορούν να θεωρηθούν συμπληρωματικές όταν χρησιμοποιούνται σε αυτά τα πλαίσια. Το αποτέλεσμα της συνύπαρξης διαφορετικών επιστημονικών τομέων, είναι καθοριστικό για την πρόοδο σε αυτή την περιοχή ενδιαφέροντος και βοηθάει στην ανάπτυξη χρήσιμων συστημάτων. Όλα αυτά καθιστούν την εξόρυξη γνώσης πολύ διαφορετική από επιστήμες όπως η στατιστική, η μάθηση μηχανής (Machine Learning, ML) κτλ, αν και η ίδια ενέχει στοιχεία από όλες αυτές. Τέλος, η εξόρυξη γνώσης βασίζεται στην αλληλεπίδραση ανθρώπου – συστήματος. Είναι δύσκολο να επιτευχθεί μια απολύτως αυτόματη διαδικασία εξόρυξης γνώσης, καθώς κάτι τέτοιο θα προϋπέθετε πως το σύστημα είναι ικανό να αναγνωρίζει τις προθέσεις του χρήστη. Για το λόγο αυτό είναι σημαντικός ο ρόλος των διάφορων εργαλείων μορφοποίησης και οπτικοποίησης που θα καλύπτουν την ανάγκη κατασκευής αποτελεσματικών εφαρμογών για την επικοινωνία ανθρώπου – υπολογιστή. Η αλληλεπίδραση συμβαίνει σε διάφορα επίπεδα και βοηθά το σύστημα εξόρυξης γνώσης να διαλέγει ανάμεσα στην πληθώρα των επιλογών αναζήτησης. (Krzysztof J. Cios, 1998)

2.2 Η διαδικασία της Εξόρυξης Γνώσης

Η εξεύρεση χρήσιμων μοτίβων σε δεδομένα είναι γνωστή με διαφορετικά ονόματα (συμπεριλαμβανομένου και αυτού της «εξόρυξης δεδομένων») σε διάφορες επιστημονικές κοινότητες, πχ. εξαγωγή γνώσης, εξεύρεση πληροφοριών, εξόρυξη πληροφοριών, αρχειοθέτηση δεδομένων και επεξεργασία μοτίβων δεδομένων. Ο όρος «εξόρυξη δεδομένων» είναι ο πιο συχνά χρησιμοποιούμενος από τους στατιστικολόγους, όσους ασχολούνται με βάσεις δεδομένων και προσφάτως στον χώρο των πληροφοριακών συστημάτων και επιχειρήσεων. Στην παρούσα διπλωματική εργασία, χρησιμοποιείται ο όρος Εξόρυξη Γνώσης, ο οποίος αναφέρεται στην συνολική διαδικασία της ανακάλυψης χρήσιμης γνώσης από δεδομένα. Η εξόρυξη δεδομένων είναι ένα ξεχωριστό βήμα αυτής της διαδικασίας – η εφαρμογή συγκεκριμένων αλγορίθμων για την εξαγωγή μοτίβων από δεδομένα. Τα επιπλέον βήματα της διαδικασίας Εξόρυξης Γνώσης, όπως η προετοιμασία των δεδομένων, η συλλογή και επιλογή δεδομένων, η ενσωμάτωση κατάλληλης πρωθύτερης γνώσης και η κατάλληλη ερμηνεία των αποτελεσμάτων της εξόρυξης, διασφαλίζουν πως προκύπτει χρήσιμη γνώση από τα δεδομένα. Η τυφλή εφαρμογή μεθόδων εξόρυξης δεδομένων, που συχνά αναφέρεται ως «βύθιση δεδομένων» (data dredging) στη βιβλιογραφία, μπορεί να οδηγήσει στην εξαγωγή ασήμαντων μοτίβων.

Η Εξόρυξη Γνώσης έχει εξελιχθεί, και εξακολουθεί να εξελίσσεται, παίρνοντας στοιχεία από πολλούς επιστημονικούς τομείς, όπως οι βάσεις δεδομένων, η μάθηση μηχανής (ML), η αναγνώριση μοτίβων, στατιστική, τεχνητή νοημοσύνη και λογική με αβεβαιότητα, απόκτηση γνώσης σε εξελιγμένα συστήματα, οπτικοποίηση δεδομένων, εξόρυξη πληροφοριών και υπολογιστικούς αλγορίθμους υψηλού επιπέδου. Τα λογισμικά Εξόρυξης Γνώσης συμπεριλαμβάνουν και συνδυάζουν θεωρίες, αλγορίθμους και μεθόδους από όλους αυτούς τους τομείς.

Οι θεωρίες και τα εργαλεία των βάσεων δεδομένων προσφέρουν τις κατάλληλες δομές για καταχώρηση, πρόσβαση και επεξεργασία δεδομένων. Η «αποταμίευση» δεδομένων (data warehousing), ένας πρόσφατα διαδεδομένος όρος, σχετίζεται με την παρούσα τάση που επικρατεί στον τομέα των επιχειρήσεων για συλλογή και «καθαρισμό» (cleaning) δεδομένων που αφορούν συναλλαγές, με σκοπό να είναι διαθέσιμα για online ανάλυση και υποστήριξη αποφάσεων. Μια δημοφιλής προσέγγιση ανάλυσης αποθηκευμένων δεδομένων ονομάζεται Online Analytical Processing (OLAP). Τα εργαλεία της OLAP επικεντρώνονται σε πολυδιάστατη ανάλυση των δεδομένων, η οποία υπερέρχει της καθιερωμένης για επεξεργασία δεδομένων SQL, όσον αφορά στην σύνοψη των υπολογισμών (computing summaries) και στην ανάλυση σε πολλές διαστάσεις. Παρόλο που τα σύγχρονα εργαλεία OLAP στοχεύουν στην ανάλυση διαδραστικών δεδομένων, αναμένεται πως στο μέλλον θα περιλαμβάνουν περισσότερα στοιχεία αυτόματης εξεύρεσης δεδομένων.

Οι τομείς που σχετίζονται με την εξαγωγή μοτίβων από δεδομένα – όπως η στατιστική αναγνώριση μοτίβων, η εφαρμοσμένη στατιστική, η μάθηση μηχανής (ML) και τα νευρωνικά δίκτυα – έδωσαν την ώθηση για έρευνα στα πρώιμα στάδια της Εξόρυξης Γνώσης. Η Εξόρυξη Γνώσης βασίζεται κατά μεγάλο μέρος σε μεθόδους που προκύπτουν από τους κλάδους αυτούς για την ανεύρεση μοτίβων σε δεδομένα, όποτε απαιτείται η εξόρυξη δεδομένων στη διαδικασία. Ένα ερώτημα που προκύπτει άμεσα είναι: «Σε τι διαφέρει η Εξόρυξη Γνώσης από τις επιστήμες αυτές;» Η Εξόρυξη Γνώσης επικεντρώνεται στην συνολική διαδικασία για ανακάλυψη γνώσης που βρίσκεται σε δεδομένα, η οποία

περιλαμβάνει το πώς αποθηκεύονται ή προσπελάζονται τα δεδομένα, πώς μπορούν να προσαρμοστούν οι αλγόριθμοι σε μεγάλες βάσεις δεδομένων ώστε να τρέχουν αποδοτικά, πώς μπορούν να ερμηνευτούν και να παρουσιαστούν τα αποτελέσματα και πώς μπορεί να μοντελοποιηθεί και υποστηριχτεί η διαδικασία αλληλεπίδρασης ανθρώπου – μηχανής. Η εξόρυξη γνώσης δίνει ιδιαίτερη βαρύτητα στην ανεύρεση κατανοητών μοτίβων, που να μπορούν να ερμηνευτούν ως χρήσιμη και ενδιαφέρουσα γνώση. Η ευελιξία και η κλιμάκωση των αλγορίθμων μοντελοποίησης, ώστε να μπορούν να εξυπηρετήσουν μεγάλο όγκο δεδομένων, είναι επίσης αντικείμενο ενδιαφέροντος.

Η επιστήμη της στατιστικής διαθέτει πολλά κοινά στοιχεία με την Εξόρυξη Γνώσης. Η εξαγωγή συμπερασμάτων από δεδομένα έχει ένα βασικό στατιστικό χαρακτήρα. Η στατιστική προσφέρει μια γλώσσα και ένα πλαίσιο για ποσοτικοποίηση του αβέβαιου αποτελέσματος, όταν κάποιος προσπαθεί να εντοπίσει γενικά μοτίβα από ένα ειδικό δείγμα κάποιου συνόλου. Ο όρος της εξόρυξης δεδομένων είχε αρνητική σημασία στην στατιστική τη δεκαετία του 1960, όταν και πρωτοεμφανίστηκαν οι τεχνικές ανάλυσης δεδομένων που βασίζονταν σε υπολογιστές. Η ανησυχία που είχε προκύψει τότε αφορούσε το γεγονός πως, εάν κάποιος ψάξει αρκετά σε οποιοδήποτε σύνολο δεδομένων, ακόμη και αν αυτά έχουν προκύψει τυχαία, μπορεί να εντοπίσει μοτίβα τα οποία εμφανίζονται σημαντικά από στατιστική άποψη αλλά στην ουσία δεν είναι. Αυτό το ζήτημα είναι ιδιαίτερα σημαντικό για την Εξόρυξη Γνώσης. Τα τελευταία χρόνια έχει υπάρξει αρκετή πρόοδος στην κατανόηση τέτοιων θεμάτων στην στατιστική. Επομένως, η εξόρυξη δεδομένων είναι μια θεμιτή διαδικασία στα πλαίσια που είναι σωστά εφαρμοσμένη. Η εξόρυξη γνώσης μπορεί επίσης να θεωρηθεί πως περιλαμβάνει μια ευρύτερη άποψη της μοντελοποίησης από την στατιστική, με στόχο να προσφέρει εργαλεία για την αυτοματοποίηση, στον βαθμό που αυτό είναι δυνατό, ολόκληρης της διαδικασίας της ανάλυσης δεδομένων, συμπεριλαμβανομένης και της τέχνης του στατιστικολόγου της επιλογής υποθέσεων.

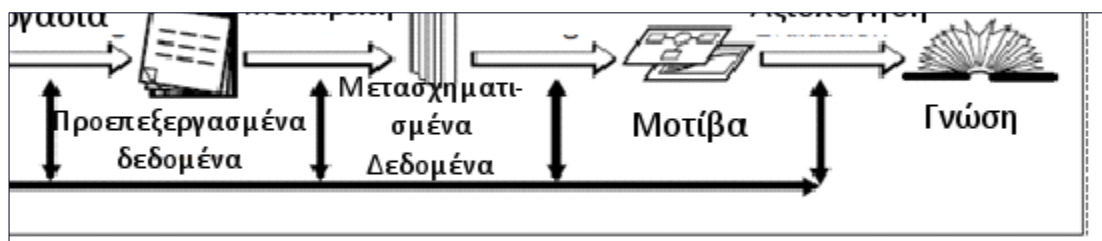
Στην συνέχεια θα παρουσιάσουμε ένα προτεινόμενο πλαίσιο διαδικασίας Εξόρυξης Γνώσης, που συνδυάζει διεπιστημονικές διαδικασίες και ακολουθεί τον ορισμό της Εξόρυξης Γνώσης κατά Fayyad (Usama Fayyad, 1996).

Σύμφωνα με τον ορισμό της διαδικασίας Εξόρυξης Γνώσης (Usama Fayyad, 1996), πρόκειται για μια διαδικασία εντοπισμού έγκυρων, νέων, δυνητικά χρήσιμων και απόλυτα κατανοητών μοτίβων δεδομένων. Ο όρος «μοτίβο» χρησιμοποιείται με την ευρύτερη έννοια του και περιλαμβάνει μοντέλα ή δομές δεδομένων. Σε αυτό τον ορισμό, τα δεδομένα αποτελούν ένα σύνολο γεγονότων (π.χ. περιπτώσεις (cases) σε μια βάση δεδομένων), και μοτίβο είναι μια έκφραση που περιγράφει ένα υποσύνολο δεδομένων, ή ένα μοντέλο που μπορεί να εφαρμοστεί σε αυτό το υποσύνολο. Ο όρος «διεργασία» (process) υπονοεί την ύπαρξη πολλών βημάτων, συμπεριλαμβανομένων αυτών της προετοιμασίας δεδομένων, της αναζήτησης μοτίβων, την αξιολόγηση γνώσης, τα οποία επαναλαμβάνονται σε πολλά στάδια. Η διαδικασία χαρακτηρίζεται από μεγάλο βαθμό πολυπλοκότητας, με την έννοια ότι πηγαίνει πέρα από το επίπεδο του υπολογισμού ποσοτήτων κλειστού τύπου αλλά πρέπει να περιλαμβάνει την αναζήτηση για δομές, μοντέλα, μοτίβα, ή παραμέτρους. Τα μοτίβα που εντοπίζονται πρέπει να είναι έγκυρα για νέα δεδομένα, με κάποιο βαθμό βεβαιότητας. Επιπλέον, υπάρχει η προσδοκία τα μοτίβα να είναι καινούργια, τουλάχιστον για τα σύστημα και κατά προτίμηση και για τον χρήστη, και δυνητικά χρήσιμα για τον χρήστη ή για τον σκοπό του έργου. Τέλος, τα μοτίβα πρέπει να είναι κατανοητά, αν όχι άμεσα, μετά από κάποια επεξεργασία.

Σύμφωνα με τον ορισμό κατά Fayyad, μπορούν να οριστούν επίσης μετρήσιμα μεγέθη για την αξιολόγηση των μοτίβων που εξάγονται. Σε πολλές περιπτώσεις, είναι πιθανό να ορίζονται μέτρα αβεβαιότητας, π.χ. εκτιμώμενη ακρίβεια ταξινόμησης, ή χρησιμότητας, π.χ. κέρδος χάρη στις εγκυρότερες προβλέψεις ή στην επιτάχυνση της ανταπόκρισης ενός συστήματος. Όροι όπως η καινοτομία ή η ευκολία στην κατανόηση είναι περισσότερο υποκειμενικοί. Σε ορισμένες περιπτώσεις, η ευκολία στην κατανόηση μπορεί να επιτευχθεί μέσω της απλότητας, για παράδειγμα η χρησιμοποίηση ενός αριθμού bits για την περιγραφή ενός μοτίβου. Μια ενδιαφέρουσα έννοια, η ύπαρξη ενδιαφέροντος, χρησιμοποιείται συνήθως ως ένα συνολικό μέτρο της αξίας του μοτίβου, που συνδυάζει την αξιολόγηση της εγκυρότητας, καινοτομίας και απλότητάς του. Οι λειτουργίες αξιολόγησης ενδιαφέροντος οριοθετούνται, ή μπορούν να διαπιστωθούν εμμέσως μέσα από το σύστημα Εξόρυξης Γνώσης, και εφαρμόζονται στα εξαγόμενα μοντέλα ή μοτίβα.

Η εξόρυξη δεδομένων είναι ένα βήμα της διαδικασίας Εξόρυξης Γνώσης, που αποτελείται από την απαρίθμηση των μοτίβων ή μοντέλων δεδομένων, και ενδέχεται να υπόκειται σε κάποια αποδεκτά όρια όσον αφορά στην υπολογιστική απόδοση. Δεδομένου ότι τα μοτίβα σε ένα πεπερασμένο σύνολο δεδομένων μπορεί να είναι μη πεπερασμένα, και επειδή η απαρίθμησή τους περιλαμβάνει κάποιου είδους αναζήτηση σε μεγάλο χώρο, οι υπολογιστικοί περιορισμοί θέτουν αρκετά όρια όσον αφορά στον υποχώρο που μπορεί να εξερευνηθεί από έναν αλγόριθμο εξόρυξης δεδομένων.

Η διαδικασία Εξόρυξης Γνώσης (Usama Fayyad, 1996) παρουσιάζεται συνοπτικά στην Εικόνα . Για λόγους απλότητας, δεν παρουσιάζονται οι πιθανές επιστροφές ανάμεσα σε δύο βήματα της διαδικασίας (loops) που υπάρχουν, ούτε τα στοιχεία του συστήματος που χρησιμοποιούν τη γνώση για τη λήψη αποφάσεων ή την ανάληψη δράσεων.



Εικόνα : Η διαδικασία Εξόρυξης Γνώσης κατά Fayad

Η διαδικασία Εξόρυξης Γνώσης είναι διαδραστική και επαναληπτική, με πολλές από τις αποφάσεις να λαμβάνονται από τον χρήστη. Συμπεριλαμβάνει πολλά βήματα που συνοψίζονται σε:

1. **Κατανόηση του τομέα εφαρμογής:** περιλαμβάνεται σχετική πρωθύστερη γνώση και οι στόχοι της εφαρμογής.

2. **Δημιουργία ενός συνόλου δεδομένων – στόχων:** επιλέγεται ένα σύνολο δεδομένων ή ένα υποσύνολο μεταβλητών ή δείγματα δεδομένων, στα οποία θα επικεντρωθεί και θα εφαρμοστεί η διαδικασία.
3. **Εκκαθάριση δεδομένων και προεργασία:** περιλαμβάνει βασικές λειτουργίες, όπως η αφαίρεση ακραίων τιμών, η συλλογή των απαραίτητων πληροφοριών, η κατάστρωση στρατηγικών για τον χειρισμό ελλιπών πεδίων δεδομένων, τον καθορισμό της χρονολογικής σειράς πληροφοριών και γνωστών αλλαγών, καθώς και αποφάσεις σχετικά με ζητήματα της βάσης δεδομένων, όπως τύποι δεδομένων, διάταξη και χαρτογράφηση άγνωστων τιμών.
4. **Ελάττωση δεδομένων και σχεδιασμός:** περιλαμβάνει την ανεύρεση χρήσιμων χαρακτηριστικών για την αναπαράσταση δεδομένων, ανάλογα με τον σκοπό του έργου, και τη χρήση διαστατικής ελάττωσης (dimensionality reduction) ή μεθόδων μετατροπής για τη μείωση του αριθμού των υπό εξέταση μεταβλητών ή για την ανεύρεση αμετάβλητων αναπαραστάσεων για τα δεδομένα.
5. **Επιλογή της λειτουργίας εξόρυξης δεδομένων:** απόφαση για τον σκοπό του μοντέλου που θα προκύψει από τον αλγόριθμο εξόρυξης δεδομένων (π.χ. σύνοψη, κατάταξη, οπισθοδρόμηση (regression), ομαδοποίηση).
6. **Επιλογή αλγορίθμου / αλγορίθμων για την εξόρυξη δεδομένων:** επιλογή των μεθόδων που θα χρησιμοποιηθούν για την αναζήτηση μοτίβων στα δεδομένα και των κατάλληλων μοντέλων και παραμέτρων. Για παράδειγμα, μοντέλα για απόλυτα δεδομένα είναι διαφορετικά από τα μοντέλα για διανύσματα ή πραγματικούς αριθμούς. Περιλαμβάνεται επίσης και η επιλογή συγκεκριμένης μεθόδου εξόρυξης δεδομένων που να ανταποκρίνεται στα γενικά κριτήρια της συνολικής διαδικασίας Εξόρυξης Δεδομένων. Ο χρήστης, για παράδειγμα, μπορεί να ενδιαφέρεται περισσότερο για την κατανόηση του μοντέλου παρά για τις δυνατότητες πρόβλεψης που παρέχει.
7. **Εξόρυξη Δεδομένων:** Αναζήτηση για μοτίβα που ενδιαφέρουν σε μια συγκεκριμένη μορφή αναπαράστασης ή σε ένα σύνολο τέτοιων αναπαραστάσεων. Περιλαμβάνονται κανόνες κατάταξης, δέντρα αποφάσεων, μοντελοποίηση ακολουθιών, ανάλυση εξαρτήσεων και σειρών.
8. **Ερμηνεία:** Στο στάδιο αυτό ερμηνεύονται τα μοντέλα που έχουν βρεθεί. Υπάρχει επίσης το ενδεχόμενο επιστροφής σε προηγούμενα βήματα, καθώς και της οπτικοποίησης των εξαγόμενων μοτίβων, αφαιρώντας τα περιττά ή άσχετα μοτίβα, και μεταφράζοντας τα χρήσιμα σε όρους κατανοητούς από τους χρήστες.

9. **Χρήση της γνώσης που προέκυψε:** ενσωμάτωση της γνώσης στην απόδοση του συστήματος, ανάληψη δράσεων που βασίζονται σε αυτή τη γνώση ή απλή αρχειοθέτησή της και αναφορά της στα ενδιαφερόμενα μέρη, καθώς και έλεγχός της για την επίλυση ενδεχόμενων συγκρούσεων με προηγούμενη (εξαγόμενη ή όχι) γνώση.

Τα πρώτα χρόνια της ανάπτυξης τέτοιων διαδικασιών, συνηθιζόταν η επικέντρωση στο βήμα της εξόρυξης δεδομένων. Ωστόσο τα υπόλοιπα στάδια της διαδικασίας είναι εξίσου, αν όχι περισσότερο, σημαντικά για την επιτυχή εφαρμογή της διαδικασίας Εξόρυξης Γνώσης. (Frawley W.J, 1991) Η μεθοδολογία αυτή είναι αυτή που θα χρησιμοποιηθεί στην παρούσα διπλωματική καθώς χαρακτηρίζεται από πληρότητα και εκβάθυνση στους τομείς ενδιαφέροντος που αφορούν τους σκοπούς της διπλωματικής εργασίας.

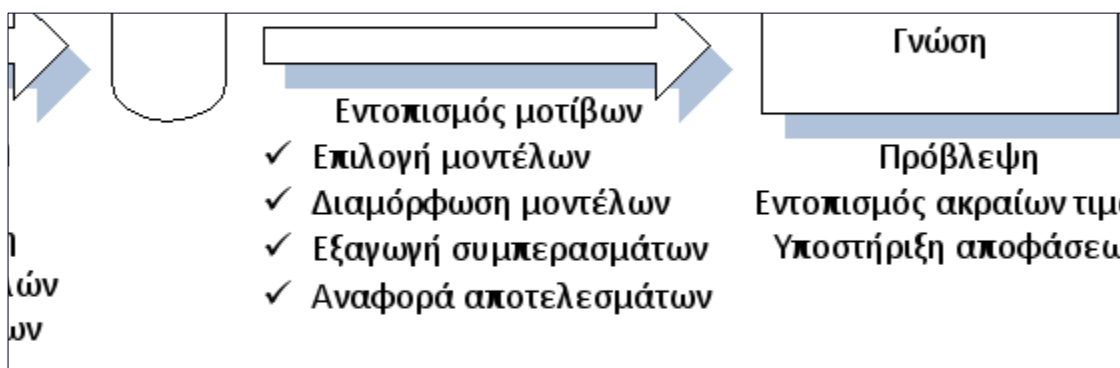
2.2 Η Εξαγωγή Πληροφοριών

Το πρώτο βήμα προς την Εξόρυξη Γνώσης είναι η Εξαγωγή Πληροφοριών από τα δεδομένα προς ανάλυση. Εξαγωγή Πληροφοριών (Information Extraction, IE) είναι το όνομα που δίνεται σε όποια διαδικασία επιλεκτικά δομεί και συνδυάζει δεδομένα, τα οποία αναφέρονται ρητά ή υπονοούνται, σε ένα ή περισσότερα κείμενα. Το τελικό αποτέλεσμα της εξαγωγής πληροφοριών ποικίλει, σε κάθε περίπτωση ωστόσο μπορεί να διαμορφωθεί ώστε να έχει τη μορφή μιας βάσης δεδομένων (Wilks, 1996).

Η αυτόματη εξαγωγή πληροφοριών από μη δομημένες πηγές έχει ανοίξει νέους ορίζοντες στα θέματα αναζήτησης, οργάνωσης και ανάλυσης δεδομένων. Ο τομέας της εξαγωγής πληροφοριών έχει τις ρίζες του στη διεργασία φυσικής γλώσσας (Natural Language Processing, NLP). Το έναυσμα για την στροφή του ενδιαφέροντος προς αυτούς τους τομείς, έδωσαν αρχικά προβλήματα όπως η αναγνώριση καταχωρήσεων ονομάτων, όπως ονόματα ανθρώπων και η οργάνωση άρθρων ειδήσεων. Καθώς ο όγκος των δεδομένων γινόταν όλο και πιο μαζικός και ο παγκόσμιος ιστός, η χρήση του οποίου γινόταν όλο και περισσότερο διαδεδομένη, προσέφερε πρόσβαση σε πλήθος δομημένων ή μη δεδομένων, άρχισαν να εμφανίζονται νέες εφαρμογές για την εξαγωγή δομών. Πλέον, το ενδιαφέρον επικεντρώνεται στη μετατροπή των προσωπικών υπολογιστών σε δομημένες βάσεις δεδομένων, της γνώσης που περιέχεται σε επιστημονικά άρθρα σε δομημένες εγγραφές και στην αξιοποίηση του διαδικτύου για δομημένες αναζητήσεις ερωτημάτων (Sarawagi, 2008).

Ο τομέας της εξαγωγής πληροφοριών αναφέρεται στην αυτόματη εξαγωγή δομημένων πληροφοριών, όπως καταχωρήσεις, σχέσεις μεταξύ καταχωρήσεων και χαρακτηριστικά που τις περιγράφουν, από μη δομημένες πηγές. Με τον τρόπο αυτό παρέχεται η δυνατότητα πολύ περισσότερων τύπων αναζητήσεων στις πολυπληθείς αδόμητες πηγές, σε σχέση με μια πιθανή αναζήτηση μόνο με λέξεις – κλειδιά. Στην περίπτωση συνύπαρξης δομημένων και μη δεδομένων, η εξαγωγή πληροφοριών δίνει τη δυνατότητα συγχώνευσης των δύο τύπων έτσι ώστε να μπορούν να γίνουν αναζητήσεις στο σύνολό τους. (Sarawagi, 2008)

Η συμμετοχή της εξαγωγής πληροφοριών στη διαδικασία εξόρυξης γνώσης φαίνεται στο παρακάτω σχήμα (Εικόνα).



Εικόνα : Η Εξαγωγή Πληροφοριών και η Εξόρυξη Δεδομένων στη Διαδικασία Εξόρυξης Γνώσης

Η διαδικασία της επιλογής των πληροφοριών προς ανάλυση περιλαμβάνει πέντε κύριες υπο-διαδικασίες (McCallum, 2008).

- **Κατάτμηση (Segmentation):** Είναι η διαδικασία κατά την οποία ορίζονται τα αρχικά και τελικά όρια των αποσπασμάτων του κειμένου που θα γεμίσουν μια βάση δεδομένων.
- **Ταξινόμηση – κατάταξη (Classification):** Μέσω της διαδικασίας αυτής καθορίζεται ποιο πεδίο της βάσης δεδομένων αποτελεί τον σωστό προορισμό για κάθε τομέα κειμένου. Για παράδειγμα, σε μια περίπτωση μελέτης κειμένων που αναφέρονται σε εκπαιδευτικά θέματα, «Εισαγωγή στην Βιβλιοθηκονομία» ανήκει στο πεδίο «Όνομα μαθήματος», «Το μάθημα αυτό αναφέρεται...» ανήκει στο πεδίο «Περιγραφή μαθήματος». Συχνά, η κατάτμηση και η ταξινόμηση γίνονται ταυτόχρονα.
- **Συσχέτιση (Associate):** Στο στάδιο αυτό καθορίζονται ποιοι τομείς θα πρέπει να ανήκουν στην ίδια καταχώρηση. Στην περίπτωση των εκπαιδευτικών θεμάτων, για παράδειγμα, μερικά μαθήματα ενδέχεται να περιγράφονται από πολλές παραγράφους ενός κειμένου και άλλα από μόνο μία. Η εξαγωγή θα πρέπει να καθορίζει ποιες αναφορές σε τομείς από ποιες παραγράφους αναφέρονται στο ίδιο μάθημα. Το στάδιο της συσχέτισης απαιτεί περισσότερο λεπτομερή ανάλυση, για παράδειγμα, στην περίπτωση εξαγωγής καταγραφών σε άρθρα εφημερίδων σχετικά με τις συναντήσεις εμπορικών διαπραγματεύσεων. Εκεί θα έπρεπε κατόπιν να μπορεί να εντοπιστεί ποιος πρωθυπουργός συναντήθηκε με ποιον ομόλογό του και ανάμεσα σε ποιες χώρες έγινε η διαπραγμάτευση. Κάτι τέτοιο απαιτεί λεπτομερή γλωσσολογικά στοιχεία σχετικά με τις σχέσεις και τις συσχετίσεις. Αυτό το βήμα συχνά είναι γνωστό ως «εξαγωγή σχέσεων» στην περίπτωση που δύο καταχωρήσεις συσχετίζονται. Η «εξαγωγή σχέσεων» δεν εμφανίζεται συχνά στα λογισμικά του εμπορίου για εξαγωγή γνώσης.
- **Εξομάλυνση (Normalize):** Στο στάδιο αυτό η πληροφορία διαμορφώνεται κατάλληλα ώστε να μπορεί να είναι συγκρίσιμη. Για παράδειγμα, οι ώρες για κάποιο μάθημα μπορεί να δίνονται σε 24ωρη μορφή ενώ κάποιες άλλες σε 12ωρη, αλλά η αναζήτηση θα πρέπει να μπορεί να εντοπίσει τέτοιων μορφών υπερκαλύψεις.

- **Εξάλειψη των διπλών καταχωρήσεων (Deduplicate):** Αποκλεισμός επιπλέον καταχωρήσεων που περιγράφουν – αναφέρονται στο ίδιο νόημα. Συνήθως τα λογισμικά που πραγματοποιούν εξομάλυνση και εξάλειψη καταχωρήσεων είναι ξεχωριστά από αυτά που κάνουν την κατάτμηση και ταξινόμησή τους.

Ορισμένες απλές εφαρμογές της εξαγωγής γνώσης μπορούν να υλοποιηθούν με απλές κανονικές εκφράσεις (regular expressions). Η εξαγωγή από περισσότερο σύνθετες πηγές κειμένου, που όμως εξακολουθούν να παρουσιάζουν μια ομοιομορφία ως προς τη δομή τους, μπορεί να επιτευχθεί με ακρίβεια με χειροκίνητους προγραμματιστικούς κανόνες. Για παράδειγμα, εάν θέλουμε να εξάγουμε τίτλους βιβλίων και ονόματα συγγραφέων από το Amazon.com, μπορούμε να βασιστούμε στο γεγονός πως εμφανίζονται με μια ακριβή και συνεπή μορφοποίηση, όπως το ότι ο τίτλος είναι ακριβώς κάτω από μια μπλε μπάρα, με έντονα γράμματα, το όνομα του συγγραφέα βρίσκεται ως υπερσύνδεσμος από κάτω, ακολουθώντας τη λέξη «by». Έτσι θα μπορούσαμε να γράψουμε έναν κώδικα που θα μας εξυπηρετούσε χρησιμοποιώντας αυτή την ομοιομορφία. Ωστόσο, αν θέλαμε επίσης να συλλέξουμε πληροφορίες από άλλες ιστοσελίδες που επίσης αφορούν στις πωλήσεις βιβλίων, θα έπρεπε να γράψουμε νέους κανόνες για καθεμία από αυτές, τους οποίους πιθανώς να χρειαζόταν να ξαναγράψουμε κάθε φορά που κάποια από τις σελίδες άλλαζε τη μορφοποίησή της.

Στις περιπτώσεις που δεν υπάρχει ανθρώπινο δυναμικό διαθέσιμο για αυτό το επίπεδο σύνταξης και ανασύνταξης κανόνων, ή όταν δεν υπάρχουν αξιόπιστα ή διαθέσιμα στοιχεία μορφοποίησης, η εξαγωγή πληροφοριών πρέπει να βασιστεί στην ίδια τη γλώσσα, δηλαδή στις λέξεις, την σύνταξή τους και τη γραμματική, ίσως σε συνδυασμό με κάποια υπάρχοντα αδύναμα στοιχεία μορφοποίησης. Η προσέγγιση της χειροκίνητης σύνταξης κανόνων έχει επιτυχώς χρησιμοποιηθεί από πολλές εταιρείες, ωστόσο, καθώς τα μοτίβα της γλώσσας γίνονται όλο και πιο λεπτομερή και οι κανόνες αλληλεπιδρούν μεταξύ τους, η σύνταξη κανόνων μπορεί να γίνει πολύ σύνθετη. Πολύ συχνά τέτοια συστήματα περιλαμβάνουν χιλιάδες κανόνες, με λεπτομερείς διασυνδέσεις μεταξύ τους, που κάνουν ιδιαίτερα εύκολη την εμφάνιση σφαλμάτων στον καθορισμό κανόνων.

Την τελευταία δεκαετία, το ενδιαφέρον περιστρέφεται γύρω από την χρήση στατιστικών μεθόδων και μεθόδων εκμάθησης μηχανής (ML) για την εξαγωγή πληροφοριών. Αυτές οι μέθοδοι συντάσσουν αυτόματα κανόνες ή παραμέτρους, προκειμένου να μεγιστοποιήσουν την απόδοση σε ένα σύνολο παραδειγμάτων κειμένων που έχει σωστά επισημανθεί με το χέρι. Με άλλα λόγια, αντί να εξάγουμε χειροκίνητα σύνθετους κανόνες, δείχνουμε στη μηχανή πώς να χειριστεί κάποια συγκεκριμένα παραδείγματα πραγματοποιώντας μόνοι μας τη διαδικασία εξαγωγής πληροφοριών. Η μηχανή, μέσα από αυτά τα παραδείγματα, παράγει μόνη της τους κατάλληλους κανόνες και τις παραμέτρους. Για σύνθετα ζητήματα εξαγωγής πληροφοριών μπορεί να απαιτείται μεγάλος αριθμός παραδειγμάτων, της τάξεως των εκατοντάδων ή χιλιάδων, αλλά η επισήμανση (labeling) των δεδομένων είναι συχνά σημαντικά ευκολότερη από τη χειροκίνητη σύνταξη κανόνων, και δεν απαιτεί τα ίδια επίπεδα εξειδίκευσης. Σε πολλές περιπτώσεις, η εκμάθηση μηχανής επιτυγχάνει ιδιαίτερα υψηλότερη ακρίβεια από τις χειροκίνητες μεθόδους.

2.2.1 Μοντέλα Εξαγωγής Πληροφοριών

Οι μηχανές πεπερασμένων καταστάσεων (Finite State Machines, FSM) είναι το κυρίαρχο μοντέλο εξαγωγής πληροφοριών, τόσο στον τομέα της βιομηχανίας όσο και στον τομέα της έρευνας. Οι πρώτες προσπάθειες επικεντρώθηκαν στον χειροκίνητο συντονισμό τέτοιων μηχανών αλλά οι πιο πρόσφατες έρευνες αφορούν σε finite state machines των οποίων οι παράμετροι ορίζονται μέσω ML, με περισσότερο συχνά τα «κρυμμένα μοντέλα Markov» (hidden Markov models: (D. Bikel, 1999) (Leek, 1997) (D. Freitag, 2000) (S. Ray, 2001))

Τα συγκεκριμένα μοντέλα διαθέτουν παραμέτρους για την πιθανότητα μετάβασης καταστάσεων, και παρέχουν τη δυνατότητα υπολογισμού της πιθανότητας που έχει το μοντέλο να παράγει μια συγκεκριμένη ακολουθία καταστάσεων, δεδομένης μια συγκεκριμένης ακολουθίας παρατηρήσεων. Όταν χρησιμοποιούνται για εξαγωγή πληροφοριών, οι παράμετροι είναι συνήθως λέξεις της φυσικής γλώσσας και οι καταστάσεις σχετίζονται με διαφορετικούς τομείς εξαγωγής. Για παράδειγμα, για να εξάγει ονόματα ανθρώπων, το μοντέλο μπορεί να έχει δύο καταστάσεις, μία για «ονόματα-ανθρώπων» και μία για «άλλα». (Andrew McCallum, 2003)

Ένα μειονέκτημα των μοντέλων αυτών είναι πως, λόγω του γεννητικού χαρακτήρα τους μέσω της παρατήρησης ακολουθιών, έχουν περιορισμένη ικανότητα στο να αναπαραστήσουν ανεξάρτητα ή επικαλυπτόμενα στοιχεία της ακολουθίας. Με άλλα λόγια, καθώς οι παρατηρήσεις παράγονται από το μοντέλο, το μοντέλο θα πρέπει να αναπαριστά οποιεσδήποτε συσχετίσεις μεταξύ στοιχείων, προκειμένου να τις αναπαράγει. Σε περιπτώσεις που υπάρχουν πολλά επικαλυπτόμενα στοιχεία ή σύνθετες εξαρτήσεις ανάμεσά τους, ή υπάρχει η ανάγκη για συλλογή στοιχείων σε διάφορα επίπεδα λεπτομέρειας, του παρελθόντος ή του παρόντος, αυτός ο τρόπος σχεδιασμού είναι απαγορευτικά δύσκολος και σε πολλές περιπτώσεις αδύνατος. (Andrew McCallum, 2003)

Τα μοντέλα «μέγιστης εντροπίας» Markov (Maximum entropy Markov models, MEMMs) και τα μοντέλα «τυχαίων πεδίων υπό όρους» (Conditional Random Fields, CRFs) είναι δύο πιθανοτικά μοντέλα καταστάσεων. Λόγω αυτής ακριβώς της ιδιότητάς τους, του στοιχείου των πιθανοτήτων, μπορούν να συμπεριλάβουν τυχαία στοιχεία, και γι αυτό έχουν γνωρίσει ιδιαίτερη εμπορική επιτυχία. Εξάγουν πληροφορίες από λίστες συχνών ερωτήσεων με διπλή ακρίβεια σε σχέση με τα «κρυμμένα μοντέλα» Markov και με περισσότερη ακρίβεια από άλλα μοντέλα και παραμένουν υποσχόμενος τομέας έρευνας. (Andrew McCallum, 2003)

2.3 Η Εξόρυξη Δεδομένων – Data Mining

Ο καθένας από εμάς θα έχει συναντήσει περιπτώσεις όπου χρειάστηκε να χρησιμοποιήσει την εξόρυξη δεδομένων, δηλαδή να αναλύσει κάποιου είδους δεδομένα. Η εξόρυξη δεδομένων δίνει τη δυνατότητα εντοπισμού μοτίβων, πρόβλεψης μελλοντικών δεδομένων και χρησιμεύει στη λήψη αποφάσεων βάσει πληροφοριών και στοιχείων που περιέχονται σε μεγάλες βάσεις δεδομένων. Από τις αρχές της δεκαετίας του '90, οι τεχνικές εξόρυξης δεδομένων έχουν εξελίξει ιδιαίτερα εμπορικές και κυβερνητικές δραστηριότητες, καθώς καθιστούν δυνατή τη λήψη αποφάσεων με ακρίβεια, διαδικασία εξεχούσης σημασίας σε τομείς όπως ο βιομηχανικός έλεγχος, η ανίχνευση σφαλμάτων και η διαχείριση πελατειακών σχέσεων. (Andrew McCallum, 2003)

Υπάρχει ήδη μεγάλος όγκος δεδομένων διαθέσιμος σε μορφή βάσεων δεδομένων, με πεδία και καταχωρήσεις, αλλά επίσης υπάρχει μεγάλη ποσότητα σημαντικών πληροφοριών που είναι διαθέσιμες μόνο σε κείμενα φυσικής γλώσσας, όπως σε ιστοσελίδες, δημοσιεύσεις, κυβερνητικές αναφορές και άλλα. Προκειμένου να γίνει με ακρίβεια η εξόρυξη δεδομένων, θα πρέπει πρώτα αυτά τα δεδομένα να είναι οργανωμένα σε μορφή βάσης δεδομένων, κάτι που γίνεται στο βήμα της εξόρυξης πληροφοριών. Η διαδικασία της εξόρυξης δεδομένων συμμετέχει στη διαδικασία εξόρυξης γνώσης, έχοντας ως είσοδο τις πληροφορίες που έχουν εξαχθεί κατά το προηγούμενο βήμα και εντοπίζοντας μοτίβα στις πληροφορίες αυτές. Το στάδιο αυτό περιλαμβάνει την επιλογή του μοντέλου data mining που θα ακολουθηθεί, τη διαμόρφωσή του ώστε να αντιστοιχεί στις παραμέτρους που έχουν επιλεγεί, την εξαγωγή συμπερασμάτων και την αναφορά των αποτελεσμάτων (McCallum, 2008).

Κάθε διαδικασία εξόρυξης μοτίβων καθορίζεται από ένα σύνολο βασικών αρχών, οι οποίες επιτρέπουν στον χρήστη να επικοινωνεί αλληλεπιδραστικά με το σύστημα ανακάλυψης γνώσης με σκοπό να κατευθύνει τη διαδικασία ή να εξετάζει να αποτελέσματα από διαφορετικές γωνίες ή σε διαφορετικό βάθος. (Jiawei Han, 2006)

- *Το σύνολο των σχετικών δεδομένων προς εξόρυξη:* Αυτό καθορίζει το μέρος της βάσης δεδομένων ή το σύνολο των δεδομένων που ενδιαφέρουν τον χρήστη και περιλαμβάνει χαρακτηριστικά της βάσης που τον αφορούν ή διαστάσεις ενδιαφέροντος.
- *Το είδος της γνώσης που θα εξαχθεί:* Αυτό καθορίζει τις διαδικασίες εξόρυξης δεδομένων που θα εκτελεστούν, όπως ο χαρακτηρισμός (characterization), η διαχώριση (discrimination), η ανάλυση συσχέτισης (association/correlation analysis), η κατάταξη (classification), η πρόβλεψη (prediction), η ανάλυση ακραίων τιμών (outlier analysis) ή η ανάλυση εξέλιξης (evolution analysis).
- *Το γνωστικό υπόβαθρο που θα χρησιμοποιηθεί στην διαδικασία εξόρυξης γνώσης:* Η γνώση σχετικά με τον τομέα εφαρμογής της διαδικασίας είναι χρήσιμη για την καθοδήγηση των επιμέρους διαδικασιών και για την αξιολόγηση των μοτίβων που θα βρεθούν. Οι ιεραρχίες εννοιών (concept hierarchies) είναι ένα χρήσιμο εργαλείο σε αυτό το σημείο, καθώς επιτρέπουν την εξαγωγή δεδομένων σε πολλαπλά επίπεδα απροσδιοριστίας (multiple levels of abstraction).
- *Το μέτρο και το κατώτατο όριο ενδιαφέροντος:* Τα στοιχεία αυτά ενδέχεται να χρησιμοποιηθούν για να καθοδηγήσουν τη διαδικασία εξόρυξης ή, μετά την εξεύρεση γνώσης, για να αξιολογήσουν τα μοτίβα που θα προκύψουν. Διαφορετικά

είδη γνώσης μπορεί να έχουν διαφορετικά μέτρα ενδιαφέροντος. Για παράδειγμα, στη διαδικασία κατά την οποία εντοπίζονται κανόνες συσχέτισης (association rules), οι κανόνες αξιολογούνται βάσει της στήριξης (support) και της αξιοπιστίας (confidence) τους. Οι κανόνες που έχουν τιμές στήριξης και αξιοπιστίας χαμηλότερες από το κατώτατο όριο που έχει οριστεί από τον χρήστη, θεωρούνται μη ενδιαφέροντες.

- *Η αναμενόμενη αναπαράσταση για την οπτικοποίηση των μοτίβων:* Αυτό αναφέρεται στον τρόπο παρουσίασης των μοτίβων, ο οποίος μπορεί να συμπεριλαμβάνει πίνακες, γραφήματα, διαγράμματα και δέντρα αποφάσεων.

Υπάρχει πλήθος αλγορίθμων και τεχνικών για τη διαδικασία εξόρυξης δεδομένων, των οποίων η αναφορά και ανάλυση δεν εκπίπτει στα πλαίσια αυτής της διπλωματικής εργασίας.

2.4 Η Αναπαράσταση της Γνώσης

Το ζήτημα της αναπαράστασης της γνώσης προκαθορίζει την απόδοση οποιουδήποτε συστήματος εξεύρεσης γνώσης. Πιο αναλυτικά, ο σχεδιαστής θα πρέπει να λάβει υπόψη του τους τύπους δεδομένων και να προσδιορίσει τον ρόλο της αποδόμησης (granularization) στις προσπάθειες εξεύρεσης γνώσης. (Krzysztof J. Cios, 1998)

Οι καταχωρήσεις δεδομένων στις βάσεις δεδομένων μπορεί να έχουν διάφορους τύπους. Γενικά, μπορούμε να τις διαχωρίσουμε σε αριθμητικές και συμβολικές. Στην πρώτη κατηγορία συμπεριλαμβάνονται συνήθως νούμερα, διανύσματα, δυσδιάστατοι πίνακες. Οι συμβολικές οντότητες χρησιμοποιούνται για να περιγράψουν κάποιες ποιοτικές μεταβλητές. Εκτός από αυτά, μπορεί κανείς να συναντήσει σύνθετους τύπους δεδομένων, οι οποίοι οδηγούν σε μια συγκεκριμένη ιεράρχηση εννοιών. (Krzysztof J. Cios, 1998)

Το κύριο ζήτημα από την οπτική γωνία της ανακάλυψης γνώσης είναι το ερώτημα πώς να χειριστεί κανείς αυτούς τους τύπους δεδομένων. Αρχικά, θα πρέπει να δοθεί προσοχή στο πρόβλημα σύγκρισης δύο μερών δεδομένων, υπολογίζοντας την απόσταση, αριθμητική ή νοηματική, ανάμεσά τους. Όταν αυτό αφορά σε αριθμητικές τιμές, είναι μια εύκολη υπόθεση: μπορεί να χρησιμοποιηθούν οποιαδήποτε μέτρα απόστασης, Ευκλείδεια, Minkowski, Tschebyschev κτλ. Το πρόβλημα απαιτεί περισσότερη προσοχή όταν πρόκειται για την ιεραρχία δύο οντοτήτων. (Krzysztof J. Cios, 1998)

Επιπλέον, η κατανόηση των αριθμών από τους ανθρώπους έχει κάποια όρια, γι αυτό το ενδιαφέρον κινείται γύρω από νέες προσεγγίσεις. Αντί της χρήσης πολλών αριθμών, παράγεται ένα συνάθροισμα. Για παράδειγμα, όταν κανείς εξετάζει έναν χάρτη καιρού, δεν κοιτά συγκεκριμένες θερμοκρασίες, αλλά περισσότερο διαμορφώνει ένα πιο αόριστο, ανώτερου-επιπέδου σχέδιο της έννοιας των υψηλών θερμοκρασιών σε κάποια περιοχή της χώρας. Με άλλα λόγια, η πληροφορία αποδομείται και παρουσιάζεται με τη μορφή ενός υψηλότερου επιπέδου αοριστίας (aggregation – συνάθροιση). Ομοίως, όλες οι σχέσεις μεταξύ των μεταβλητών ενδιαφέροντος ποσοτικοποιούνται με παρόμοιο τρόπο. Όταν εξετάζουμε τα δεδομένα με σκοπό την ανακάλυψη μοτίβων, τα εξετάζουμε από μια συγκεκριμένη νοηματική απόσταση που αποτελείται από κομμάτια γνώσης συγκεκριμένης λεπτομέρειας. Γενικότερα, με τον όρο «αποδόμηση πληροφορίας» εννοείται μια μορφή αντιστοίχισης αριθμητικών δεδομένων σε μια νοηματική οντότητα. Το πώς καθορίζεται αυτό το παράθυρο γνώσης, αποτελεί σημαντικό θεωρητικό αλλά και πρακτικό ζήτημα. (Krzysztof J. Cios, 1998)

Σε τέτοιου είδους ζητήματα έρχεται να ανταποκριθεί η ανάπτυξη οντολογιών για την αναπαράσταση γνώσης. Οι ταξινομίες ή ιεραρχίες εννοιών είναι ζωτικής σημασίας για κάθε σύστημα βασισμένο στη γνώση, δηλαδή για οποιοδήποτε σύστημα χρησιμοποιεί τη γνώση σχετικά με τον τομέα με τον οποίο ασχολείται. Σε διαδικασίες που συμπεριλαμβάνουν διεργασία φυσικής γλώσσας (NLP), εξαγωγή πληροφοριών (IE) και γενικότερα στα συστήματα εξόρυξης, μπορεί να ωφεληθεί κανείς από μια ταξινομία που παρέχει πληροφορίες σε διαφορετικά επίπεδα λεπτομέρειας. Γενικότερα, είναι προφανές πως οποιοδήποτε είδος συντακτικής και σημασιολογικής αλληλεπίδρασης, θα είναι περισσότερο ακριβές εάν είναι διαθέσιμη μια ιεραρχία εννοιών. (Philipp Cimiano, 2002)

2.4.2 Η ιεράρχηση εννοιών

Η ιεράρχηση εννοιών καθορίζει την ακολουθία των χαρτογραφήσεων από έννοιες χαμηλότερου επιπέδου αοριστίας προς έννοιες υψηλότερου επιπέδου, περισσότερο γενικές. Για παράδειγμα, στην ιεράρχηση εννοιών αναφορικά με την *τοποθεσία*, ως τιμές πόλεων είναι το Βανκούβερ, Τορόντο, Νέα Υόρκη και Σικάγο. Κάθε πόλη ωστόσο μπορεί να αναλυθεί σε περιοχή ή κράτος στο οποίο ανήκει. Για παράδειγμα το Βανκούβερ στη Βρετανική Κολούμπια, το Σικάγο στο Ιλινόις. Οι περιοχές και τα κράτη μπορούν κατόπιν να χαρτογραφηθούν στην χώρα όπου ανήκουν, όπως ο Καναδάς ή οι ΗΠΑ. (Jiawei Han, 2006)

Πολλές ιεραρχίες εννοιών υπονοούνται σε ένα σχήμα βάσης δεδομένων. Στο παράδειγμα της τοποθεσίας, αυτή θα περιγράφεται από τα χαρακτηριστικά αριθμός, οδός, πόλη, περιοχή ή κράτος, ταχυδρομικός κώδικας και χώρα. Αυτά τα χαρακτηριστικά είναι συνδεδεμένα από μια συνολική σειρά, σχηματίζοντας μια ιεραρχία ιδεών της μορφής: «αριθμός < οδός < πόλη < περιοχή ή κράτος < ταχυδρομικός κώδικας < χώρα». Εναλλακτικά, τα χαρακτηριστικά μιας διάστασης μπορεί να οργανώνονται σε μια μερική σειρά, σχηματίζοντας ένα δίκτυο. Ένα παράδειγμα τέτοιας οργάνωσης είναι για τη διάσταση χρόνος που βασίζεται στα χαρακτηριστικά μέρα, εβδομάδα, μήνας, εποχή και χρόνος. Δεδομένου ότι η εβδομάδα συχνά μπορεί να ανήκει σε δύο διαφορετικούς μήνες, συχνά δεν θεωρείται ως χαμηλότερο επίπεδο αοριστίας από τον μήνα. Ωστόσο θεωρείται χαμηλότερο επίπεδο από τον χρόνο, καθώς ο χρόνος περιέχει σχεδόν 52 εβδομάδες. Θα είναι λοιπόν: «μέρα < { μήνας < εποχή ; εβδομάδα} < χρόνος. (Jiawei Han, 2006)

Μια ιεραρχία εννοιών που θέτει σε απόλυτη ή μερική σειρά τα χαρακτηριστικά σε μια βάση δεδομένων, ονομάζεται σχήμα ιεραρχίας (schema hierarchy). Τα συστήματα εξόρυξης γνώσης θα πρέπει να προσφέρουν στον χρήστη την ελαστικότητα να διαμορφώνει τις προκαθορισμένες ιεραρχίες σύμφωνα με τις ανάγκες του. Για παράδειγμα, οι χρήστες μπορεί να ενδιαφέρονται να καθορίσουν ένα δημοσιονομικό έτος που ξεκινάει την 1^η Απριλίου ή ένα ακαδημαϊκό έτος που ξεκινάει την 1^η Σεπτεμβρίου. (Jiawei Han, 2006)

Οι ιεραρχίες εννοιών μπορεί επίσης να καθορίζονται από τη διακριτοποίηση ή ομαδοποίηση τιμών για μια συγκεκριμένη διάσταση ή χαρακτηριστικό, μιλώντας έτσι για ένα σύνολο-ομαδοποίηση ιεραρχιών. Μια ολική ή μερική κατάταξη μπορεί να καθοριστεί ανάμεσα σε ομάδες τιμών. Επιπλέον, μπορεί να υπάρχουν περισσότερες από μία ιεραρχίες εννοιών για ένα συγκεκριμένο χαρακτηριστικό ή μια διάσταση, που να βασίζονται στις διαφορετικές οπτικές γωνίες του χρήστη. Για παράδειγμα, ο χρήστης μπορεί να προτιμά να οργανώνει τις τιμές ως φθηνές, μέτριες και ακριβές αντί να αναφέρεται σε σύνολα τιμών ({X...Y}). (Jiawei Han, 2006)

Οι ιεραρχίες εννοιών μπορεί να παρέχονται χειροκίνητα από τους χρήστες του συστήματος, τους ειδικούς του τομέα ή τους μηχανικούς γνώσης ή μπορεί να παράγονται αυτόματα βασιζόμενες σε στατιστική ανάλυση της κατανομής των δεδομένων. (Jiawei Han, 2006)

2.4.3 Οντολογίες

Αναμφίβολα, ο όρος «οντολογία» έχει βρει εφαρμογή σε πολλές διαφορετικές επιστήμες, ξεκινώντας από τη φιλοσοφία και συνεχίζοντας μέχρι τις σύγχρονες επιστήμες και την επιστήμη των υπολογιστών. Τις τελευταίες δεκαετίες, ο όρος έχει γίνει ιδιαίτερα

δημοφιλής στον τομέα της Διαχείρισης Γνώσης και έχουν υπάρξει πολλοί ορισμοί σχετικά με το τι είναι οντολογία, οι οποίοι και έχουν εξελιχθεί στον χρόνο. Ο πιο επικρατής, μέχρι σήμερα, ορισμός αναφέρει την οντολογία ως «an explicit specification of a conceptualization».(Gruber T. , 1993) Σύμφωνα δηλαδή με τον Gruber, «οντολογία» είναι ένα σύνολο αντιπροσωπευτικών αρχεγόνων με τα οποία μπορούμε να περιγράψουμε/μοντελοποιήσουμε έναν τομέα γνώσης.



Εικόνα : Ο τομέας των οντολογιών

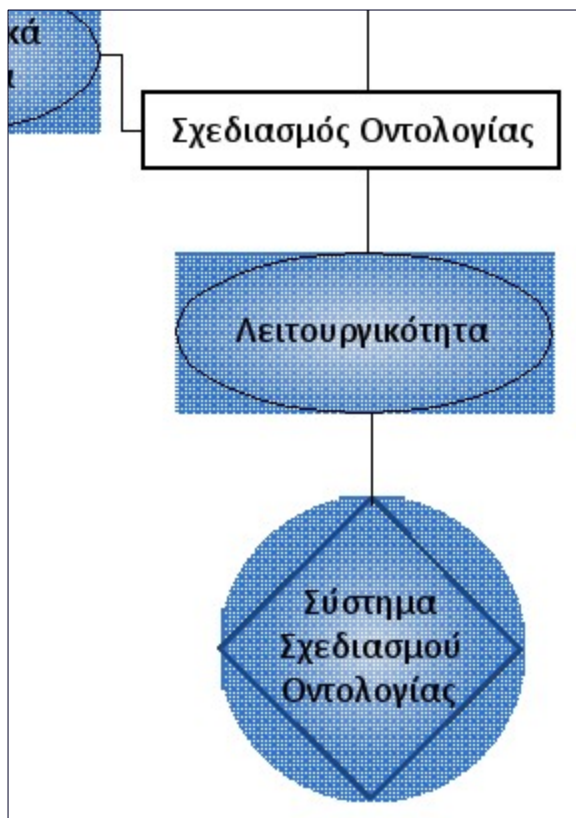
Τα στοιχεία της οντολογίας είναι συνήθως κατηγορίες (classes ή sets), ιδιότητες (attributes ή properties) και σχέσεις (relations) μεταξύ τους. Οι ορισμοί αυτών των αντιπροσωπευτικών αρχετύπων, περιλαμβάνουν πληροφορίες για την σημασία τους, καθώς και περιορισμούς για την συνεπή εφαρμογή τους. Στον τομέα των βάσεων δεδομένων, η οντολογία μπορεί να θεωρηθεί ως ένα level of abstraction των μοντέλων δεδομένων, ανάλογο με τα ιεραρχικά και σχεσιακά μοντέλα, αλλά με σκοπό τη μοντελοποίηση της γνώσης σχετικά με στοιχεία, τις ιδιότητες και στις σχέσεις τους. (Gruber T. , 2008)

Οι οντολογίες συνήθως περιγράφονται σε γλώσσες που επιτρέπουν ένα επίπεδο αοριστίας, πέρα από δομές δεδομένων. Γι αυτό και οι οντολογίες αναφέρονται να είναι σε «σημασιολογικό» επίπεδο, ενώ τα συστήματα δεδομένων είναι σε «λογικό» ή «φυσικό» επίπεδο. Δεν υπάρχουν κάποιες συγκεκριμένες γλώσσες για την περιγραφή μιας οντολογίας, και ένα πλήθος open source εργαλείων είναι διαθέσιμο για την δημιουργία της. (Gruber T. , 2008)

Προκειμένου να «χτιστεί» μια οντολογία, χρειάζονται 4 βασικά στάδια:

- Καθορισμός σκοπού

- Χτίσιμο της οντολογίας
 - Σύλληψη οντολογίας
 - Κωδικοποίηση οντολογίας
 - Ενοποίηση υπαρχουσών οντολογιών
- Αξιολόγηση
- Τεκμηρίωση



Εικόνα : Στοιχεία σχεδιασμού οντολογίας

Επιπλέον, η οντολογία πρέπει να περιλαμβάνει ένα σύνολο τεχνικών, μεθόδων και αρχών για κάθε στάδιο, καθώς επίσης και να καθορίζει τις σχέσεις που υπάρχουν ανάμεσα στα στάδια. (Gruber T. , 2008)

2.5 Επιλεγμένα παραδείγματα συστημάτων Εξόρυξης Γνώσης

Τα συστήματα εξόρυξης γνώσης είναι συστήματα που μπορούν να λειτουργήσουν σε δεδομένα του κανονικού κόσμου, και όχι μόνο σε θεωρητικό επίπεδο. Μια από τις περισσότερο δημοφιλείς και επιτυχημένες εφαρμογές των βάσεων δεδομένων είναι στον τομέα της διαφήμισης, όπου τόνοι πληροφοριών σχετικά με τις πελατειακές συμπεριφορές συλλέγονται και οι διαφημιστές ενδιαφέρονται στον εντοπισμό των προτιμήσεων των πελατών με σκοπό τη μελλοντική στόχευσή τους σε καμπάνιες. (Krzysztof J. Cios, 1998)

Η ανάπτυξη συστημάτων εξόρυξης γνώσης είναι μια σύνθετη διαδικασία η οποία περιλαμβάνει μια πληθώρα εργαλείων εξόρυξης και συνήθως εξαρτάται από τον τομέα εφαρμογής της και καθορίζεται από το περιβάλλον του χρήστη για τον οποίο προορίζεται.

Οι ακόλουθες σύντομες περιγραφές ήδη υπάρχοντων συστημάτων διευκρινίζουν τη φύση των προβλημάτων που αντιμετωπίζονται και υπογραμμίζουν τα ζητήματα σχεδιασμού που προκύπτουν κατά τη διαδικασία.

SKICAT (Sky Image Cataloging and Analysis Tool): Πρόκειται για ένα σύστημα αυτόματης μείωσης και ανάλυσης μεγάλου όγκου αστρονομικών δεδομένων. Το μέγεθος της βάσης δεδομένων ανέρχεται στα 3 terabytes εικόνων που απεικονίζουν περί τα 2 δισεκατομμύρια ουράνια αντικείμενα. Το ιδιαίτερο στοιχείο του συστήματος είναι προφανώς ο τεράστιος όγκος δεδομένων. Σκοπός του ήταν η κατηγοριοποίηση ουράνιων σωμάτων, των οποίων οι φωτογραφίες ήταν πολύ θαμπές για να αναγνωριστούν από τους αστρονόμους.

Health – KEFIR (Key Finding Reporter): Είναι ένα σύστημα εξόρυξης γνώσης που χρησιμοποιείται στην ιατρική φροντίδα ως ένα σύστημα έγκαιρης προειδοποίησης (Usama Fayyad, 1996). Το σύστημα επικεντρώνεται στην κατάταξη αποκλίσεων ανάλογα με το μέτρο του ενδιαφέροντος γεγονότων στους χρήστες. Στόχος του είναι η ανακάλυψη και αιτιολόγηση σημαντικών ευρημάτων σε μεγάλες βάσεις δεδομένων. Το κύριο εργαλείο που χρησιμοποιείται είναι οι αποκλίσεις, οι οποίες μετά την εντόπισή τους, κατατάσσονται με τη χρήση κάποιου μέτρου ενδιαφέροντος. Στη συνέχεια το σύστημα παράγει εξηγήσεις για τις πιο ενδιαφέρουσες από αυτές και μπορεί επίσης να δώσει συστάσεις για αντίστοιχες δράσεις.

TASA (Telecommunication Network Alarm Sequence Analyzer): Το σύστημα αναπτύχθηκε για την πρόβλεψη σφαλμάτων σε σύστημα επικοινωνίας. Ένα τυπικό δίκτυο παράγει εκατοντάδες συναγεργμούς την ημέρα. Το σύστημα TASA παράγει κανόνες όπως «Αν ένας συγκεκριμένος συνδυασμός συναγεργμών συμβεί μέσα σε κάποιο χρόνο, τότε ένας άλλος συναγεργμός άλλου τύπου θα συμβεί μέσα σε κάποιο χρόνο». Οι τιμές των χρόνων επιλέγονται από τον χρήστη, ο οποίος μπορεί κατόπιν να κατατάξει ή να ομαδοποιήσει τους κανόνες που προκύπτουν από το σύστημα.

R-MINI: Το σύστημα αυτό χρησιμοποιεί τόσο την ανάλυση απόκλισης όσο και τεχνικές κατηγοριοποίησης για να εξάγει χρήσιμη πληροφορία από πολυπληθείς τομείς (Usama Fayyad, 1996). Χρησιμοποιεί τη λογική για να παράγει κανόνες πλήρεις και συνεπείς. Οι παραγόμενοι κανόνες μπορούν στη συνέχεια να χρησιμοποιηθούν για την πρόβλεψη ενός συνόλου επιστροφών. Προφανώς, οι κανόνες πρέπει να αναπαράγονται περιοδικά καθώς νέα δεδομένα γίνονται διαθέσιμα.

DISCOTEX: Το σύστημα αυτό συνδυάζει μεθοδολογίες εξαγωγής πληροφοριών και εξόρυξης δεδομένων για να εξάγει κείμενο και για να βελτιώσει την απόδοση του ήδη υπάρχοντος συστήματος RAPIER, στο οποίο βασίστηκε. Οι κανόνες που εξάγονται από μια βάση δεδομένων χρησιμοποιούνται για την πρόβλεψη πρόσθετων πληροφοριών που ενδέχεται να εξαχθούν από μελλοντικά αρχεία, βελτιώνοντας έτσι την ανάκληση της εξαγωγής πληροφοριών. Για παράδειγμα, αν σε μια ανάρτηση για εύρεση εργασίας εξαχθούν τα: *Java-Programming_Languages* και *Graphic_areas*, το σύστημα ενδέχεται να εξετάσει την εξαγωγή του *PhotoShop_Applications* ως μια πρόσθετη καταχώρηση (Un Yong Nahm, 2000).

Μεγάλος αριθμός συστημάτων εξόρυξης γνώσης επικεντρώνεται σε online διαθέσιμες βάσεις δεδομένων και την εξόρυξη πληροφοριών από ελεύθερα κείμενα που είναι διαθέσιμα στο διαδίκτυο. Μια αξιοσημείωτη περίπτωση, που υπογραμμίζει την υπεροχή των συστημάτων εξόρυξης γνώσης έναντι των παραδοσιακών μέσων διαχείρισης γνώσης, είναι αυτή της ιστοσελίδας FlipDog.com.

Το 2000, το **FlipDog.com** λειτούργησε ως σελίδα αναζήτησης εργασίας και έκανε αίσθηση καθώς διέθετε διπλάσιες ευκαιρίες εργασίας από το Monster.com, που λειτουργούσε με μια διαθέσιμη βάση δεδομένων. Αυτό ήταν δυνατόν επειδή, αντί να συλλέγει πληροφορίες από εργοδότες που πλήρωναν για να αναρτηθούν στην ιστοσελίδα, το FlipDog εξήγαγε τις διαθέσιμες θέσεις από περισσότερες από 60.000 ιστοσελίδες εταιρειών, αυτόματα και απευθείας. Η διαδικασία περιελάμβανε τη συλλογή του τίτλου θέσης, περιγραφής, τοποθεσίας, ονόματος εταιρείας, στοιχεία επικοινωνίας κτλ και την τοποθέτηση των θέσεων εργασίας σε μια οντολογία. Η ιστοσελίδα αργότερα αγοράστηκε από το Monster.com. (McCallum, 2008)

Εκτός από τους ερευνητικούς σκοπούς, ιδιαίτερα σημαντική είναι η εφαρμογή της διαχείρισης γνώσης σε επιχειρησιακό επίπεδο. Καθώς οι οργανισμοί αποθηκεύουν ένα αυξανόμενο πλήθος πληροφοριών και γνώσης σε μορφή δεδομένων, οι προσπάθειες επικεντρώνονται στην αποτελεσματική διαχείριση της γνώσης αυτής. Ιστορικά, η επιχειρηματική γνώση αποθηκευόταν σε χαρτιά και στο μυαλό των ανθρώπων. Το χαρτί παρέχει περιορισμένη προσβασιμότητα και είναι δύσκολο να ανανεώνονται οι πληροφορίες που περιέχει. Επιπλέον, όταν οι εργαζόμενοι έφευγαν από τις θέσεις τους, έπαιρναν μαζί τους το μεγαλύτερο μέρος της γνώσης τους, και έτσι η ανακύκλωσή της δεν ήταν πάντα εφικτή. Επομένως, οι εταιρείες προχώρησαν σε αποθήκες δεδομένων και γνώσης για να βελτιώσουν την προσβασιμότητα, ανανεωσιμότητα και αρχειοθέτηση των δεδομένων και της γνώσης. Η γνώση μπορεί να προέλθει από εργασίες, αναφορές και μια μεγάλη γκάμα πηγών. Συνήθως οι βάσεις γνώσης που περιλαμβάνουν τέτοιου είδους πληροφορίες, περιέχουν εργασιακά έγγραφα, προτάσεις και σχετικά αρχεία. Επιπροσθέτως, οι βάσεις γνώσης σχεδιάζονται με σκοπό την παροχή συνέχειας και ιστορικού σε διαδικασίες και στην πελατειακή υποστήριξη. (Leary, 1998)

Lessons learned databases: Οι βάσεις διδαγμάτων (Lessons – Learned Databases) μπορεί να χρησιμοποιηθούν για την υποστήριξη λειτουργιών ή την παραγωγή πληροφοριών για την επιχείρηση γενικότερα. Για παράδειγμα, η **National Security Agency (NSA) Lessons Learned Knowledge Base** περιέχει τρεις τύπους διδαγμάτων: πληροφοριακά, επιτυχή και προβληματικά. Ένα πληροφοριακό δίδαγμα μπορεί να περιγράφει πώς ένας πελάτης της NSA μπορεί να αναλάβει κάποια καθήκοντα σε περίπτωση έκτακτης ανάγκης. Τα επιτυχή δίδαγματα αναφέρονται σε θετικές αντιμετώπισεις σε περιπτώσεις κρίσης. Τα

προβληματικά διδάγματα δίνουν παραδείγματα περιπτώσεων που δεν πήγαν καλά και πιθανούς τρόπους να λυθούν τα προβλήματα.

Ομοίως, η Ford Motor Co. έχει δημιουργήσει αυτό που η εταιρεία ονομάζει αρχεία **Things Gone Right/Wrong (TGRW)**.(Stewart, 1997) Τα αρχεία TGR περιλαμβάνουν πληροφορίες γεγονότων που διευκόλυναν την ολοκλήρωση στόχων ενώ τα TGW περιλαμβάνουν πληροφορίες για γεγονότα που εμπόδισαν να συμβεί κάτι τέτοιο. Γενικά, η συλλογή των TGR είναι πιο εύκολη, ειδικά αν η γνώση είναι αρχειοθετημένη, καθώς πολλοί υπάλληλοι αισθάνονται άβολα να συσχετίζονται με πράγματα που δεν πήγαν καλά. Οι TGRW βάσεις είναι ιδιαίτερα σημαντικές για τη δημιουργία αρχείων γεγονότων που χρειάζονται παρακολούθηση από τις μονάδες διαχείρισης έργων.

Best Practices: Οι βάσεις γνώσης αυτού του είδους περιλαμβάνουν γνώση για τις καλύτερες τακτικές. Συνήθως δημιουργούνται από την επισήμανση (bench-marking) δραστηριοτήτων σχεδιασμένων να στοχεύουν στο πιο αποτελεσματικό και αποδοτικό τρόπο για να γίνουν τα πράγματα. Για παράδειγμα, η General Motors Hughes Electronics διαθέτει την βάση «**Best Process Reengineering Database**». Κάθε καταχώρηση της βάσης συνοδεύεται από μια σύντομη περιγραφή και κάποια στοιχεία επικοινωνίας. Τις καταχωρήσεις της βάσης αποτελούν οι αλλαγές που έγιναν σε διαδικασίες του οργανισμού και οδήγησαν στη βελτίωση των διαδικασιών. Το να είναι διαθέσιμο το σύνολο των διαδικασιών αυτών μέσω της βάσης, αυξάνει την πιθανότητα να εντοπιστούν και να υιοθετηθούν αργότερα ή από άλλα μέρη του οργανισμού. (Leary, 1998)

Οι συμβουλευτικές εταιρείες είναι ανάμεσα τους πρώτους που ανέπτυξαν βάσεις καλύτερων τακτικών. Η Price Warehouse ήταν μία από αυτές, με την **Knowledge View**, η οποία είναι μια βάση καλύτερων τακτικών που επιτρέπει πολλαπλές οπτικές γωνίες της γνώσης, υπό τη βάση της βιομηχανίας, της διαδικασίας, του μέτρου απόδοσης κλπ. Βασίζεται σε μια οντολογία ενσωματωμένη σε ένα μοντέλο επιχείρησης που επικεντρώνεται σε διαδικασίες που οδηγούν στην δημιουργία κάποιας τιμής, για παράδειγμα «Παραγωγή Προϊόντων».(Leary, 1998)

Οι αναφορές ειδήσεων αποτελούν επίσης ένα μέσο ενσωμάτωσης εξωτερικών πληροφοριών σε μια επιχείρηση. Για παράδειγμα, η εταιρεία παροχής επαγγελματικών υπηρεσιών KPMG συνεργάστηκε με την Story Street Partners για να παρέχει ήδη φιλτραρισμένες, ταξινομημένες και αναζητούμενες πληροφορίες σχετικά με θέματα και εταιρείες ενδιαφέροντος των εργαζόμενων της KPMG.(Leary, 1998)

ΚΕΦΑΛΑΙΟ 3^ο: Η ΔΙΑΧΕΙΡΙΣΗ ΕΡΓΩΝ

3.1 Γενικά για τη διαχείριση και επίβλεψη έργων

Η διαχείριση έργων, όπως τη γνωρίζουμε σήμερα, ξεκίνησε να διευρύνεται μόλις μερικές δεκαετίες πριν. Παρόλα αυτά, η προσπάθεια για βελτίωση της παραγωγικότητας με χρήση των βέλτιστων πόρων, ήταν πάντα συνυφασμένη με την ανθρώπινη φύση και γι αυτό εξελισσόταν μαζί με αυτή. Ο αντίκτυπος των διαφορετικών μεθόδων διαχείρισης έργων είναι εμφανής σε τομείς όπως ο σχεδιασμός του έργου, ο έλεγχος του κόστους, του χρόνου και της ποιότητάς του. Μέρος της διαδικασίας διαχείρισης έργων αποτελούν τα στάδια της επίβλεψης και της αξιολόγησής τους.

Αξιολόγηση έργων ονομάζεται η περιοδική εκτίμηση της αποτελεσματικότητας, της απόδοσης, της επιρροής, της βιώσιμης ανάπτυξης και της σχετικότητας ενός έργου, σε σχέση με τους στόχους που έχουν θεσπιστεί για αυτό. Σε αντίθεση με την επίβλεψη του έργου η οποία πραγματοποιείται καθ' όλη τη διάρκεια του κύκλου του, η αξιολόγηση γίνεται σε συγκεκριμένες χρονικές στιγμές του κύκλου αυτού (έναρξη, μέσον, λήξη), γίνεται επομένως κατανοητό πως η επίβλεψη του έργου παρέχει τα βασικά στοιχεία για την αξιολόγησή του. (ETUI-REHS, 2004)

Monitoring – επίβλεψη έργων ονομάζεται η συστηματική και συνεχής συλλογή, ανάλυση και χρήση πληροφοριών με σκοπό τη διαχείριση και τη λήψη αποφάσεων, και συμπεριλαμβάνει μια ενδελεχή εξέταση πόρων και αποτελεσμάτων ενός έργου. Πρόκειται για μια ουσιώδη διαδικασία που πραγματοποιείται κατά τη διάρκεια εκτέλεσης ενός έργου, με σκοπό την άμεση διόρθωση των τυχόντων αποκλίσεων από τους στόχους του έργου και την εξασφάλιση της ολοκλήρωσής του. Μέσα από την επίβλεψη - παρακολούθηση, προκύπτουν δεδομένα τα οποία μπορούν να χρησιμοποιηθούν για την αξιολόγηση των έργων. (ETUI-REHS, 2004)

Σκοπός του monitoring είναι η επίτευξη της αποδοτικής και αποτελεσματικής εκτέλεσης μιας διαδικασίας. Συνεπώς, τα συστήματα επίβλεψης πρέπει να παρέχουν πληροφορίες στους σωστούς ανθρώπους τη σωστή στιγμή προκειμένου να τους βοηθήσουν να λάβουν αποφάσεις. Η διαδικασία της επίβλεψης θα πρέπει να εφιστά την προσοχή σε δυνάμεις και αδυναμίες κατά την εφαρμογή του έργου, δίνοντας τη δυνατότητα στους manager να βρουν λύσεις για να αντιμετωπίσουν προβλήματα και να προσαρμοστούν στις περιστάσεις έτσι ώστε να βελτιωθεί η απόδοση του έργου. Επιπλέον, η διαδικασία της επίβλεψης θα μπορούσε να θεωρηθεί ως ένα σύστημα έγκαιρης προειδοποίησης, το οποίο επιτρέπει την κατάλληλη παρέμβαση στην κατάλληλη στιγμή, αν το έργο δεν ακολουθεί το προγραμματισμένο πλάνο. (ETUI-REHS, 2004)

Οι πληροφορίες που παρέχονται σχετικά με την πρόοδο ενός έργου μέσω της διαδικασίας επίβλεψης, προκύπτουν από ένα πλήθος πηγών, όπως οικονομικές συμφωνίες, το πλάνο δράσης (Logframe) ενός έργου, το πρόγραμμα δραστηριοτήτων, μέσα από γραπτές τεκμηριώσεις και, κυρίως, μέσα από επισκέψεις στο χώρο πραγματοποίησης του έργου. Η κατάσταση ενός έργου μπορεί να αξιολογηθεί μόνο μετά από συνεντεύξεις με όλα τα συμμετέχοντα σε αυτό μέρη, συμπεριλαμβανομένων των δικαιούχων του.

Η επίβλεψη ενός έργου θα πρέπει να επικεντρώνεται στη συλλογή και ανάλυση πληροφοριών σχετικά με:

- Τη φυσική πρόοδο του έργου (παροχή δεδομένων, δραστηριότητες που αναλαμβάνονται και αποτελέσματα που λαμβάνονται) και την ποιότητα της διαδικασίας (π.χ. συμμετοχή των ενδιαφερόμενων μερών)
- Την οικονομική πρόοδο (προϋπολογισμός και δαπάνες)
- Την ανταπόκριση των ομάδων-στόχων (target groups) στις δραστηριότητες του έργου (π.χ. χρήση των υπηρεσιών, αλλαγές στη γνώση, στις συνήθειες και στις πρακτικές)
- Τους λόγους που μπορεί να οδηγήσουν σε απροσδόκητη ή δυσμενή ανταπόκριση από τις ομάδες-στόχους (target groups), και στις πράξεις που μπορούν να γίνουν για να αντιμετωπιστεί αυτό

Η επίβλεψη είναι ένα ουσιώδες μέρος της διαδικασίας Διαχείρισης Έργων και ένα διαχειριστικό εργαλείο εξέχουσας σημασίας. Είναι επομένως:

- Ένας τρόπος συλλογής πληροφοριών στηριζόμενος σε:
 - Γνώση που προκύπτει από τα έγγραφα που σχετίζονται με ένα έργο, την τρέχουσα κατάσταση και το γενικό περιβάλλον στο οποίο εξελίσσεται το έργο
 - Συνεντεύξεις με όλα τα συμμετέχοντα μέρη
- Ένα μέσο βελτίωσης της διαχείρισης έργου
- Ένας τρόπος ενημέρωσης όλων των συμμετεχόντων για την πρόοδο του έργου και τις δυσκολίες του (αν υπάρχουν)
- Ένας γρήγορος και αποτελεσματικός τρόπος παροχής σύντομων και ενημερωτικών αναφορών
- Μια επισκόπηση της εφαρμογής του έργου (Co-operation, Handbook for Results-Oriented Monitoring of EC External Assistance (projects and programmes), 2008)

Τόσο μέσα από την επίβλεψη όσο και μέσα από την αξιολόγηση του έργου, προκύπτουν ιδιαίτερα χρήσιμα συμπεράσματα, τα οποία αφορούν τόσο στην πορεία ενός έργου όσο και στους λόγους που οδήγησαν στην εξέλιξη του κάθε έργου.

3.2 Ο ρόλος της αναπαράστασης γνώσης στη διαχείριση έργων

Είναι γεγονός πως, μέσα από τις διαδικασίες αξιολόγησης έργων, μπορούν να δημιουργηθούν οι κατάλληλες συνθήκες για αλλαγή παραδοσιακών λειτουργιών μιας επιχείρησης. Αναμφίβολα, τα προβλήματα μάθησης στοιχείων από πλήθος έργων έχουν ευρύτερες επιπτώσεις για τις διαδικασίες μάθησης του οργανισμού και ως εκ τούτου, δεν προκαλεί έκπληξη το ότι η ανάπτυξη της ικανότητας διαχείρισης γνώσης στα διάφορα έργα θεωρείται ως μια πηγή ανταγωνιστικού πλεονεκτήματος ανάμεσα στις επιχειρήσεις. Ο κύκλος της διαχείρισης γνώσης περιλαμβάνει τα ακόλουθα βήματα: εντοπισμός και συλλογή της γνώσης, κοινοποίηση της γνώσης, εφαρμογή της και δημιουργία γνώσης. Ο εντοπισμός και η συλλογή της ζητούμενης γνώσης ακολουθείται από την κοινοποίησή της έτσι ώστε να μπορούν να την μοιραστούν και άλλοι. Στη συνέχεια, η εφαρμογή της γνώσης και η διαμόρφωσή της ανάλογα με κάθε περίπτωση δημιουργεί νέα γνώση, η οποία με την σειρά της θα περάσει από αυτόν το κύκλο κ.ο.κ. (Jay Liebowitz, 2003)

Ο διαμοιρασμός της γνώσης (knowledge sharing), είναι άμεσα συνδεδεμένος με την αναπαράστασή της. Στόχος είναι η δημιουργία και γαλούχηση μιας κοινής κουλτούρας γνώσης. Κάτι τέτοιο απαιτεί την ανάπτυξη ικανοτήτων σχετικά με την γνώση και, όταν αναφερόμαστε σε επιχειρησιακό επίπεδο, αναζητείται η υποστήριξη της στρατηγικής της εταιρείας μέσω της διευκόλυνσης της επιτυχούς επικοινωνίας σε όλη την επιχείρηση. Η αναπαράσταση γνώσης στον τομέα των έργων, διευκολύνει και διαχειρίζεται τη μνήμη της εταιρείας. Τα μέσα για να συμβεί αυτό περιλαμβάνουν ιστοσελίδες, intranet, και κοινότητες ενδοεπιχειρησιακής επικοινωνίας (practice/chat rooms). Το σύνολο αυτών των κινήσεων μπορεί να δημιουργήσει μετέπειτα διδάγματα από τη διαχείριση κάθε μεγάλου έργου, καθώς και να αυξήσει τον αριθμό και την ποιότητά τους. Επιπλέον, μέσα από την αναπαράσταση και τον διαμοιρασμό της γνώσης αυξάνονται οι πιθανότητες ευρύτερης και αποτελεσματικότερης χρήσης επιτυχημένων τακτικών διαχείρισης έργων. (Jay Liebowitz, 2003)

Στο σημείο αυτό τίθεται το ερώτημα: τι αξία έχουν όλα αυτά τα πλαίσια διαχείρισης γνώσης και τα εργαλεία αναπαράστασης στη διαχείριση ενός έργου; Με άλλα λόγια, πώς μπορούν οι τεχνικές διαχείρισης και αναπαράστασης γνώσης να βοηθήσουν στις καθημερινές δραστηριότητες ενός διαχειριστή έργου; Σύμφωνα με τους White και Fortune (D. White, 2002) στην έρευνά τους σχετικά με τις παρούσες πρακτικές διαχείρισης έργων, οι τρεις καθοριστικοί παράγοντες στη διαχείριση έργων είναι: ξεκάθαροι στόχοι, υποστήριξη από τα ανώτερα διοικητικά στελέχη και επαρκή κεφάλαια (επαρκείς πόροι). Η διαχείριση γνώσης μπορεί να βελτιώσει την επικοινωνία και να παρέχει περισσότερη χρήσιμη πληροφορία στον διαχειριστή του έργου και την ομάδα του. Με αυτό τον τρόπο, οι στόχοι μπορούν να είναι καλύτερα διατυπωμένοι και κοινοποιημένοι στα μέλη της ομάδας. Επιπλέον, τα συστήματα διαχείρισης γνώσης μπορούν να δημιουργηθούν για να εξασφαλίσουν έναν καλύτερο διαμοιρασμό αρχείων καλύτερων τακτικών, διδαγμάτων, μεθοδολογιών διαχείρισης έργων και συστημάτων, παραδειγμάτων αναφορών πακέτων, λόγων λήψεων στρατηγικών αποφάσεων και παρόμοιων στοιχείων. Για παράδειγμα, στην NASA Goddard, έχει δημιουργηθεί για τον σκοπό αυτό μια ιστοσελίδα διαχείρισης έργων, που διαθέτει ένα είδος «αποθήκης» αρχείων για να βοηθά τα διάφορα προγράμματα πτήσεων να βλέπουν τι έχουν κάνει άλλοι. Τα αρχεία αυτά συνδέονται σε κάθε στάδιο της εξέλιξης του έργου. Ένα άλλο παράδειγμα συστήματος διαχείρισης γνώσης στην περίπτωση της NASA είναι το Process-Based Mission Assurance (PBMA) Knowledge Management

System, που περιλαμβάνει διδάγματα, καλύτερες τακτικές, στιγμιότυπα βίντεο ειδικών, online κοινότητες και άλλα χαρακτηριστικά που μπορούν να βοηθήσουν στη διαχείριση του έργου. Παράλληλα, η Academy of Program and Project Leadership (APPL) της NASA λειτουργεί ως εκπαιδευτικός σύμβουλος στην προώθηση του διαμοιρασμού της γνώσης ανάμεσα στους διαχειριστές έργων. Η APPL εμπλέκεται με πρωτοβουλίες όπως: διδασκαλία και διαχείριση προγράμματος ανάπτυξης περιβάλλοντος διαχείρισης έργων, δημιουργία φόρουμ ανταλλαγής γνώσης διαχειριστών έργων σε διάφορα κέντρα της NASA, μελέτη περιπτώσεων αποτυχημένων και επιτυχημένων αποστολών της NASA, και άλλα. (Jay Liebowitz, 2003)

Η διαχείριση γνώσης μπορεί επίσης να χρησιμοποιηθεί σαν καινοτομία για τη δημιουργία νέων ιδεών προϊόντων και υπηρεσιών. Οι οργανισμοί που βασίζονται στις υπηρεσίες τους στην ανάληψη έργων, οφείλουν να επωφεληθούν από τον εγγενή καινοτόμο χαρακτήρα των εργασιών ενός έργου. Δεδομένου ότι τα έργα αφορούν χαρακτηριστικά στην ανάπτυξη νέων προϊόντων και διαδικασιών, είναι προφανείς οι ευκαιρίες για ανάδειξη καινοτόμων ιδεών και μάθηση διατομεακών θεμάτων, στοιχεία που ενισχύουν την ικανότητα καινοτομίας ενός οργανισμού και τις δυνατότητές του. (Mike Bresnen, 2003) Σύμφωνα με την έρευνα του Majchrzak σχετικά με την επαναχρησιμοποίηση της καινοτομίας (Majchrzak A, 2001), έχουν εντοπιστεί οχτώ παράγοντες που επηρεάζουν την μεταφορά γνώσης για επαναχρησιμοποίηση:

- Εμπειρίες σχετικά με τις διαφορές στις αποδόσεις έργων
- Απαιτήσεις έργων σχετικά με τη μείωση κινδύνων
- Προσωπική δεκτικότητα της εξέτασης ευρέος φάσματος γνώσης για την επίλυση προβλήματος
- Ευρείες βάσεις γνώσης με δυνατότητα εύκολης αναζήτησης
- Ομαδική και επιχειρησιακή κουλτούρα που ενθαρρύνει την επαναχρησιμοποίηση
- Δυνατότητα για γρήγορη αξιολόγηση της αξιοπιστίας και ευχρηστίας της επαναχρησιμοποιούμενης γνώσης
- Δυνατότητα για γρήγορη αξιολόγηση του βαθμού ανταπόκρισης της γνώσης σε συγκεκριμένο πρόβλημα
- Δυνατότητα για γρήγορη αξιολόγηση της εφαρμογής της επαναχρησιμοποιούμενης γνώσης

Για παράδειγμα, σχετικά με τον πρώτο παράγοντα, οι διαφορές στην απόδοση μπορούν να εκφραστούν με διάφορους τρόπους, όπως χρονικοί περιορισμοί, περιορισμοί προϋπολογισμού ή απόκλιση από τους στόχους. Η ύπαρξη τέτοιων στοιχείων έδωσε το κίνητρο στους χρήστες να αναζητούν κάθε φορά υπάρχουσα γνώση σχετικά με το θέμα που προκύπτει. Στο παράδειγμα της NASA, το σύστημα που λειτουργεί στο NASA Goddard Space Flight Center, NASA Lessons Learned Information System (LLIS), χρησιμοποιείται για τη διαχείριση και τον μετριασμό κινδύνου. Έτσι, από την έρευνα και την εκμάθηση τακτικών που ακολούθησαν άλλοι (επιτυχίες ή αποτυχίες) σε κάποια αντίστοιχη περιοχή του έργου, δημιουργείται και προσφέρεται γνώση που, αν επαναχρησιμοποιηθεί, μπορεί να βοηθήσει σημαντικά τον διαχειριστή του έργου.

Η διαχείριση γνώσης έχει και άλλες εφαρμογές στη διαχείριση έργων. Οι διαδικασίες ανταλλαγής γνώσης μπορούν να ενισχύσουν παράγοντες όπως η ηγεσία, η ομαδική εργασία, ο ρόλος της διαχείρισης έργων στην αλυσίδα εφοδιασμού και οι πολυπλοκότητες της γνώσης καθώς κινείται στην εφοδιαστική αλυσίδα. Για παράδειγμα, η καθοδήγηση (mentoring) και η σκίαση (shadowing) είναι τεχνικές διάδοσης της γνώσης που μπορούν να χρησιμοποιήσουν οι διαχειριστές έργων προκειμένου να μεταφέρουν γνώση για τον βελτιωμένο σχεδιασμό διαδοχής, ηγεσίας και ομαδικότητας. Εξίσου σημαντικά είναι ο ρόλος του διαχειριστή του έργου στην επιλογή δυναμικού, στη διαχείριση της απόδοσης και των συμβολαίων καθώς επίσης και η σημασία που αποδίδεται σε ιδιώτες με τις απαραίτητες γνώσεις, ικανότητες και σοφία. Η διαχείριση γνώσης θα πρέπει να αποτελεί αναπόσπαστο κομμάτι της στρατηγικής ανθρώπινου δυναμικού ενός οργανισμού προκειμένου να περιλαμβάνει διάφορα είδη γνώσης από τα ξεχωριστά άτομα και ομάδες. (Jay Liebowitz, 2003)

Οι λόγοι που μπορούν να περιορίσουν τη μάθηση στοιχείων από διάφορα έργα δεν είναι δύσκολο να εντοπιστούν. Οι εταιρείες έρχονται αντιμέτωπες με διάφορα προβλήματα που πρέπει να ξεπεράσουν σχετικά με την σύλληψη της γνώσης και την ανακύκλωσή της, που πηγάζουν από τη σχετικά ιδιοτελή, ιδιότυπη και πεπερασμένη φύση των εργασιών του έργου. Αναπόφευκτες ασυνέχειες συμβαίνουν στη ροή των πηγών –κυρίως προσωπικού και πληροφοριών- στον χρόνο και στον χώρο, από το ένα έργο στο άλλο. Η καταγραφή και η διάδοση της γνώσης και της μάθησης στα διάφορα έργα γίνεται συνεπώς ένα σημαντικό πρόβλημα, όπως και η αποφυγή της τάσης «επανεφεύρεσης του τροχού» στην αντιμετώπιση ενός προβλήματος που πρέπει να επιλυθεί. Διάφορες επιπλοκές εμφανίζονται στον τομέα των κατασκευών, ιδίως λόγω της περίπλοκης οργάνωσης του καταμερισμού εργασίας μεταξύ των επαγγελματικών και άλλων ομάδων που εμπλέκονται στη διαδικασία διαχείρισης των δομικών κατασκευών. Η διάσπαση αυτή έχει σημαντικές συνέπειες στις προσπάθειες για ανάπτυξη κοινής προοπτικής όσον αφορά στην καινοτομία, τη γνώση και τη μάθηση. Η υπερνίκηση των εμποδίων για την αποτελεσματική διαχείριση της γνώσης, γενικότερα, περιλαμβάνει μια σειρά παρεμβάσεων που αντανakλούν τους διάφορους τρόπους με τους οποίους η γνώση μπορεί να ενσωματωθεί στο πλαίσιο των οργανωτικών συστημάτων και διαδικασιών ή στο σύνολο των δεξιοτήτων και ικανοτήτων των ατόμων και των ομάδων (Mike Bresnen, 2003)

3.3 Η αξία των διδαγμάτων στη διαχείριση έργων

Είτε πρόκειται για τη διαχείριση ενός έργου είτε για ένα σύνολο αυτών, η αξία της κοινής χρήσης και επικοινωνίας των διδαγμάτων που προκύπτουν δεν θα πρέπει να υποτιμάται. Σε πολλές περιπτώσεις μπορεί να κάνει τη διαφορά ανάμεσα σε μια συνολική επιτυχία του έργου και στην έλλειψη σημαντικών στοιχείων που οδηγούν στην καθυστέρηση του έργου. Η επικοινωνία διδαγμάτων που θα προκύψουν από την επίβλεψη ενός έργου μπορεί να βεβαιώσει πως δεν θα επαναληφθούν λάθη που αφορούν στο κόστος της καθυστέρησης του έργου, της υπέρβασης του προϋπολογισμού και της μη ικανοποίησης των πελατών και άλλων εμπλεκομένων. (Seningen, 2004)

Ένας μεγάλος αριθμός διαχειριστών έργων δραστηριοποιείται στη βιομηχανία όπου τα έργα προς διαχείριση συχνά έχουν παρόμοια στοιχεία και ιδιότητες με πρωθύτερα έργα. Επομένως, παρόλο που το ακριβές έργο με όλες του τις διαδικασίες δεν μπορεί να επαναληφθεί, το αρχείο των διδαγμάτων που θα προκύψει θα είναι ιδιαίτερα χρήσιμο καθώς ενδέχεται να «σώσει» κάποιο μελλοντικό έργο, ή κάποιον μελλοντικό διαχειριστή έργου. (Seningen, 2004)

Κατά την διαδικασία διαχείρισης και επίβλεψης έργων, η ανανέωση των αρχείων διδαγμάτων καθώς και η έμφαση στα συμπεράσματα που προκύπτουν μέσα από αυτές τις διαδικασίες, μπορεί να βοηθήσει σε δύο σημαντικά σημεία (Seningen, 2004):

- 1) Αποφυγή επανάληψης ίδιων σφαλμάτων
- 2) Βελτίωση της πιθανότητας ισορροπίας του «τριπλού τριγώνου»¹ έτσι ώστε να μην υπάρχουν υπερβάσεις κόστους ή προγράμματος σε μελλοντικά έργα.

Ορισμένα από τα λάθη που αναφέρονται στα διδάγματα και κατά συνέπεια μπορούν να προληφθούν, συμπεριλαμβάνουν ζητήματα σχετικά με την παράδοση εξοπλισμού, την αποδοχή του έργου και του προϋπολογισμού του, την οριζόντια ή κάθετη επικοινωνία, την εξέταση του έργου, την τεχνική υποστήριξη, την εκπαίδευση, και η λίστα μπορεί να συνεχιστεί επ' άπειρο. (Seningen, 2004)

Καθίσταται επομένως σαφής η ανάγκη για συλλογή και καταχώρηση των διδαγμάτων κάπου όπου να μπορούν να συγκριθούν, και να διαμορφωθούν ανάλογα με τις δράσεις στις οποίες αναφέρονται. Κάτι τέτοιο μπορεί να γίνει σε κάποιου είδους βάση διδαγμάτων έτσι ώστε η καταχώρηση των δεδομένων που προκύπτουν να γίνεται με μία συνεπή μορφή σε ένα κεντρικό σύστημα. Οι βάσεις διδαγμάτων είναι συχνά δομικά στοιχεία των συστημάτων διαχείρισης γνώσης εταιρειών παγκόσμιας εμβέλειας, και πολλά παραδείγματά τους μπορούν να βρεθούν και στον παγκόσμιο ιστό.

¹ Το «Τριπλό Τρίγωνο» (γνωστό και ως «Τριπλό Εμπόδιο») της Διαχείρισης Έργων είναι ένα μοντέλο εμποδίων της διαχείρισης έργων, αναφορικά με τον χρόνο, το κόστος και τα αγαθά (υπηρεσίες ή προϊόντα) που απαιτούνται για την ολοκλήρωση ενός έργου. Χρησιμοποιείται συχνά για να αναδείξει πως η επιτυχία της διαχείρισης ενός έργου μπορεί να μετρηθεί από την ικανότητα της ομάδας να διαχειριστεί το έργο έτσι ώστε τα αναμενόμενα αποτελέσματα να παραχθούν με ταυτόχρονη διαχείριση χρόνου και κόστους (Michael W. Newell, 2004)

3.4 Το σύστημα ROM

3.4.1 Γενικά

Η Ευρωπαϊκή Ένωση, συμπεριλαμβανομένων των κρατών μελών της και της Ευρωπαϊκής Κομισιόν, έχει αναλάβει την επιδότηση ενός συνόλου έργων παγκοσμίως. Υπεύθυνη για την διεξαγωγή των εξωτερικών αυτών προγραμμάτων και έργων βοήθειας είναι η Γενική Διεύθυνση της Ευρωπαϊκής Ένωσης, EuropeAid, της οποίας στόχος είναι η παροχή αποδοτικής και αποτελεσματικής βοήθειας. Προκειμένου να εξασφαλίσει τη συνοχή, τη συμπληρωματικότητα και το συντονισμό στην εκτέλεση των προγραμμάτων βοήθειας, η EuropeAid συνεργάζεται με ένα πλήθος εταιρών και διαθέτει πλήθος μηχανισμών. Το σύστημα ROM (Result Orientated Monitoring) είναι ένας τέτοιος μηχανισμός, ο οποίος βασίζεται σε τακτικές αξιολογήσεις τρεχόντων έργων από ανεξάρτητους παράγοντες, προκειμένου να εκτιμήσει την ποιότητα και την πρόοδο ενός έργου.

Το σύστημα ROM εφαρμόζεται για περισσότερο από 10 χρόνια και καλύπτει όλες τις χώρες που εκπέτουν στη δραστηριότητα των εξωτερικών προγραμμάτων της Ένωσης. Στόχοι του Results Orientated Monitoring είναι:

- Η παροχή ανεξάρτητων αξιολογήσεων σχετικά με την επίδοση των έργων, με έμφαση στα αποτελέσματά τους, τα οποία συμπεριλαμβάνουν για το σύστημα ROM τις συνέπειες και επιπτώσεις των έργων.
- Η παροχή συστάσεων / υποδείξεων προς τους ενδιαφερόμενους για τα έργα
- Η δημιουργία συνοπτικών δεδομένων για αναφορά στους επόπτες της EuropeAid και στο Ευρωπαϊκό Κοινοβούλιο (Co-operation, Handbook for Results-Oriented Monitoring of EC External Assistance (projects and programmes), 2008)

Το σύστημα ROM παρέχει αναλυτικές πληροφορίες και συστάσεις για την απόδοση των έργων Ανάπτυξης, βασιζόμενο στα πέντε κριτήρια αξιολόγησης της Επιτροπής Αναπτυξιακής Βοήθειας (Development Assistance Committee), τα οποία είναι: Σχετικότητα (Relevance), Αποδοτικότητα (Efficiency), Αποτελεσματικότητα (Effectiveness), Αντίκτυπος (Impact) και Αειφορία (Sustainability).

Το είδος των πληροφοριών που παρέχονται από το σύστημα ROM είναι ιδιαίτερο από πολλές απόψεις. Αρχικά, το σύστημα προσφέρει μια πλήρη κάλυψη, που περιλαμβάνει όλες τις περιοχές, τους θεματικούς τομείς, και ένα ευρύ φάσμα εταιρών υλοποίησης των έργων. Τα τελευταία χρόνια, το σύστημα ROM έχει καλύψει κάθε περιοχή και σχεδόν κάθε χώρα που εκπέττει στις εξωτερικές δραστηριότητες της Κομισιόν, προσφέροντας ετησίως 1.600 αναφορές για 1.400 έργα και προγράμματα, συνολικής αξίας περίπου 10 δισεκατομμυρίων ευρώ. Επιπλέον, παρόλο που το πρόγραμμα είχε σχεδιαστεί αρχικά για να παρέχει στήριξη σε μεμονωμένα σχέδια/ προγράμματα, το γεγονός πως το ROM χαρακτηρίζεται από ομοιομορφία ως προς την προσέγγιση και τη μορφή, σε όλους τους τομείς, επιτρέπει την σύγκριση μεταξύ παρεμβάσεων σε βαθμό που δεν είναι δυνατός μέσω άλλου εργαλείου διαχείρισης. Επιπροσθέτως, το σύστημα ROM προσφέρει αναλύσεις του έργου σε διαφορετικές χρονικές περιόδους (εκ των υστέρων ή τρέχουσες αναλύσεις) και, σε πολλές περιπτώσεις, προσφέρει παρατηρήσεις στο ίδιο έργο (διαδικασία επανα-επίβλεψης) γεγονός που επιτρέπει την δυναμική κατανόηση της απόδοσης του έργου στο πέρασμα του χρόνου.

Υπάρχει μια αυξανόμενη ζήτηση από πλευράς της Ευρωπαϊκής ένωσης και των ενδιαφερόμενων μερών, για εκμετάλλευση της αφθονίας και συγκρισιμότητας των πληροφοριών που παρέχονται από το σύστημα, με σκοπό την εξαγωγή συμπερασμάτων ή διδαγμάτων σχετικά με την Αναπτυξιακή Συνεργασία της Ευρωπαϊκής Ένωσης αλλά και με τις προσεγγίσεις της συνεργασίας εν γένει. Ωστόσο, οι ευκαιρίες που προσδίδει το σύστημα ROM από αυτή την άποψη δεν έχουν ακόμα αξιοποιηθεί πλήρως.

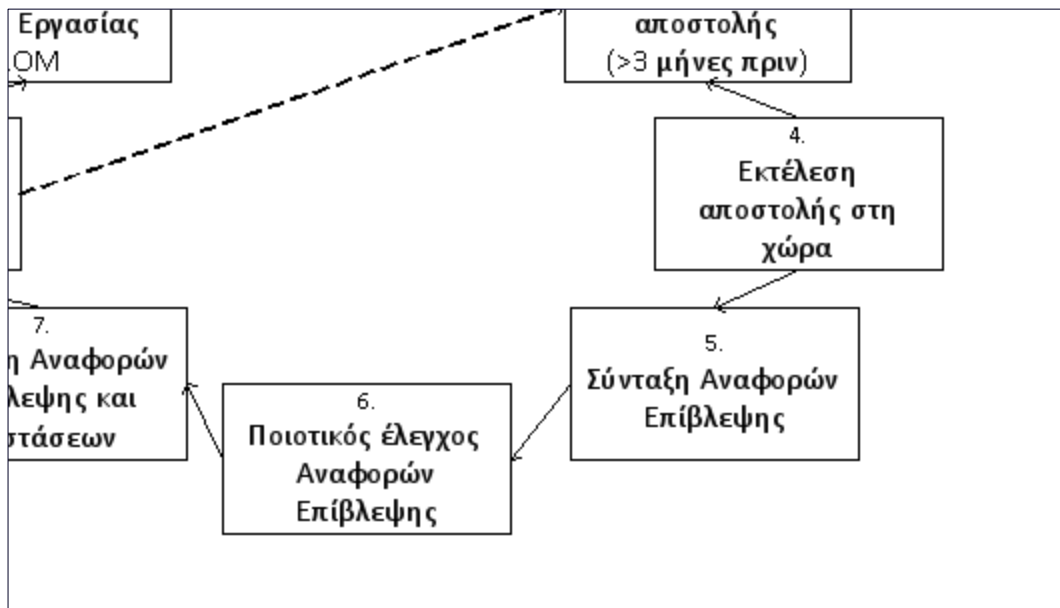
Μέχρι σήμερα, οι πληροφορίες του ROM έχουν χρησιμοποιηθεί πρωτίστως για υποστήριξη διαχείρισης σε μεμονωμένο επίπεδο ή σε επίπεδο προγραμμάτων, ενώ η χρήση του για εξαγωγή γενικών συμπερασμάτων έχει επιχειρηθεί μόνο εν μέρει. Ορισμένες προσπάθειες έχουν γίνει για την εξαγωγή συμπερασμάτων σε συγκεκριμένα θέματα μέσω του συντονισμού ROM, δηλαδή στην κυβέρνηση, το περιβάλλον ή την αειφορία, καθώς και μέσω κοινών σεμιναρίων. Επιπλέον, οι ανάδοχοι του ROM έχουν εκδώσει θεματικές εκθέσεις που περιορίζονται σε γεωγραφικές περιοχές, ωστόσο έχοντας περιορισμένα δείγματα και χωρίς στόχο για δημιουργία μιας συστηματικής μεθοδολογίας.

Οι ετήσιες αναφορές του συστήματος ROM ανταποκρίθηκαν στο σύνολο των αναγκών πληροφόρησης της ΕΕ, μετατρέποντας τις επιδόσεις των επιμέρους εκθέσεων σε αριθμητικές τιμές. Οι αριθμοί αυτοί χρησιμοποιούνται για τον μετέπειτα υπολογισμό του μέσου όρου των βαθμολογιών ROM για περιφέρειες, τομείς και θέματα, που ακολουθείται από μια σύγκριση μεταξύ της αύξησης ή μείωσης της ποσοτικής τιμής των αποτελεσμάτων. Η εστίαση σε αριθμητικά αποτελέσματα δίνει συγκρίσιμα στοιχεία, δεν μπορεί ωστόσο να απαντήσει στο ερώτημα τι κάνει κάποια έργα να αποδίδουν καλά και κάποια άλλα όχι.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανασκόπηση των παραγόντων που επηρεάζουν την πορεία ενός έργου και η δημιουργία μιας βάσης γνώσης που θα περιλαμβάνει όλους αυτούς τους παράγοντες καθώς και σχετικά διδάγματα, έτσι όπως προκύπτουν από τη μελέτη και ανάλυση των αναφορών αξιολόγησης του συστήματος ROM.

3.4.2 Διαδικασία επίβλεψης και αξιολόγησης

Η διαδικασία επίβλεψης έργων αποτελείται από οχτώ στάδια τα οποία παρουσιάζονται στο ακόλουθο σχήμα. Προκειμένου να επιτευχθεί μία σαφής εικόνα της διαδικασίας, είναι απαραίτητο ο επόπτης του έργου να κατανοεί τις οικονομικές και τεχνικές πτυχές του έργου καθώς και τον ρόλο του σε κάθε στάδιο της διαδικασίας. (Co-operation, Handbook for Results-Oriented Monitoring of EC External Assistance (projects and programmes), 2008) Καθένα από τα βήματα, προϋποθέτει ένα σύνολο διαδικασιών που πρέπει να ακολουθούνται. Σε περίπτωση που κάποιο έργο κριθεί απαραίτητο να προβεί σε μια σειρά από διορθώσεις, προβλέπεται επιστροφή στο βήμα 3, και επανάληψη της διαδικασίας επίβλεψης από το στάδιο αυτό.



Εικόνα : Διαδικασία επίβλεψης συστήματος ROM

Εστιάζοντας περισσότερο στα σημεία ενδιαφέροντος της συγκεκριμένης μεθοδολογίας, μετά τον καθορισμό του χρονοδιαγράμματος επισκέψεων των έργων (missions), πραγματοποιούνται επισκέψεις σε κάθε έργο από τους πραγματογνώμονες και έπειτα από μία σειρά δομημένων συνεντεύξεων και επιθεωρήσεων, συντάσσονται οι αναφορές αξιολόγησης των έργων (MR, BCS)(Ντάνος, 2010)

Οι αξιολογήσεις των έργων πραγματοποιούνται υπό τη μορφή συμβολαίων που αναφέρονται σε πακέτα έργων (Lots) είτε βάσει θεματολογίας, είτε βάση γεωγραφικής συσχέτισης, είτε και των δύο. Η διαδικασία αξιολόγησης ακολουθεί μια συγκεκριμένη μεθοδολογία και τα αποτελέσματά της προκύπτουν με βάση τα εξής κριτήρια:

- **Συνάφεια και Ποιότητα Σχεδιασμού του Έργου (Relevance and Quality of Project Design):** Ο βαθμός κατά τον οποίο το έργο διατηρεί τη συνάφειά του με τους στόχους που έχει αρχικά ορίσει, όπως και την διατήρηση της χρησιμότητας του ως προς την κάλυψη των γενικών αναγκών στις οποίες στόχευε, όσο και η ποιότητα του σχεδιασμού βάσει του οποίου επιδιώκει να τις καλύψει.
- **Αποδοτικότητα (Efficiency):** Ο βαθμός κατά τον οποίο το έργο διαχειρίζεται αποτελεσματικά τους πόρους που του παρέχονται και το κατά πόσο πραγματοποιείται ορθή διαχείριση του έργου, των παραδοτέων του και της συνεργασίας των εμπλεκόμενων φορέων.
- **Αποτελεσματικότητα (Effectiveness):** Ο βαθμός κατά τον οποίο το έργο επιτυγχάνει τους επιμέρους στόχους που έχει ορίσει, όπως και το κατά πόσο εμφανίζεται ικανό να ολοκληρώσει επιτυχώς το σύνολο των στόχων του, λαμβάνοντας υπόψη και την προσαρμογή του σε πιθανώς μεταβαλλόμενο περιβάλλον και την εμφάνιση νέων κινδύνων.

- **Προοπτικές Αντίκτυπου (Impact Prospects):** Ο βαθμός κατά τον οποίο το έργο ενδέχεται να έχει θετικό ή αρνητικό αντίκτυπο, έμμεσα ή και άμεσα στις ομάδες στόχους του έργου, ή σε οποιοδήποτε άλλο παράγοντα.
- **Δυνητική Αειφορία (Potential Sustainability):** Ο βαθμός κατά τον οποίο το έργο ενδέχεται να εξακολουθήσει να προσφέρει θετικό αντίκτυπο στο μέλλον βάσει της οικονομικής του βιωσιμότητας, της συνεχιζόμενης αξιοποίησής του από τις ομάδες στόχους και της πολιτικής στήριξης που ενδέχεται να είναι απαραίτητη για τη συνέχισή του.

Η βασική παράμετρος για την βαθμολόγηση κάθε έργου είναι το κατά πόσο ακολουθήθηκε το αρχικό σχέδιο. Οι πληροφορίες και τα αποτελέσματα που προκύπτουν από τη διαδικασία συγκεντρώνονται στα εξής παραδοτέα:

- Αναφορές Αξιολόγησης (Monitoring Reports, MR) και Φύλλα Συμπερασμάτων Ιστορικού (Background Conclusion Sheets, BCS), σύμφωνα με την τυποποιημένη μορφοποίηση τους, τα οποία παραδίδονται εντός συγκεκριμένων χρονοδιαγραμμάτων και βάσει συγκεκριμένων προτύπων ποιότητας. Το μήκος ενός MR εξαρτάται από το τι μπορεί να αναφερθεί σχετικά με το έργο, και μπορεί να είναι έως δύο σελίδες, το οποίο ωστόσο δεν είναι υποχρεωτικό.
- Οριζόντιες και ειδικές αναφορές (μεταξύ των οποίων την σύνθεση και ανάλυση των αποτελεσμάτων του ROM, είτε βάσει γεωγραφικής, είτε τομεακής κατάτμησης και αναφορών συστάσεων), εφόσον αυτά ζητηθούν από την αρμόδια υπηρεσία της Κομισιόν.
- Παρουσίαση των ευρημάτων στους ενδιαφερόμενους, μέσα από σεμινάρια και συναντήσεις εργασίας, στις Βρυξέλλες ή στην περιοχή που αφορούν τα έργα και επί τόπου βοήθεια για την συνέχιση των έργων. (Ντάνος, 2010)

Στο τέλος κάθε έτους, συντάσσεται η Ετήσια Αναφορά, η οποία παρουσιάζει τα αποτελέσματα της εφαρμογής του συστήματος ROM για το συγκεκριμένο σύνολο έργων που ορίζονται από το συμβόλαιο για το συγκεκριμένο έτος και παρέχει μία συνολική εικόνα της επίδοσης της Εξωτερικής Βοήθειας της Κομισιόν (EC External Assistance) για αυτό το σύνολο έργων, ενώ περιγράφεται και ο συνολικός αντίκτυπος της σε διάφορα οριζόντια θέματα υψηλής προτεραιότητας. Παράλληλα, διάφορες αναφορές οι οποίες αφορούν σε οριζόντια θέματα μπορεί να ζητηθούν από την Κομισιόν. Οι αναφορές αυτές μπορεί να εξετάζουν χαρακτηριστικά υποσυνόλων των έργων βάση του γεωγραφικού τους προσδιορισμού, βάσει των στρατηγικών τους στόχων ή άλλων κοινών χαρακτηριστικών.

Η βάση δεδομένων του συστήματος ROM βασίζεται στην online φόρτωση των MR και των BCS. Η προώθηση των παραγόμενων εγγράφων (BCS, MR, σύνοψη έργου) σε ηλεκτρονική μορφή στο Γραφείο των Βρυξελλών είναι ευθύνη του επιβλέποντος, όπως επίσης και η προώθηση εγγράφων όπως οικονομικές συμφωνίες, πλάνο δραστηριοτήτων έργου κλπ. Το σύνολο αυτών των εγγράφων στη συνέχεια φορτώνεται στη βάση δεδομένων CRIS, έπειτα από ενδεδειγμένο ποιοτικό έλεγχο.

3.4.3 Αναφορές αξιολόγησης έργων στο σύστημα ROM

Τα κυριότερα στοιχεία του ROM είναι η Αναφορά Επίβλεψης (Monitoring Report, MR) και το Φύλλο Συμπερασμάτων Ιστορικού (Background Conclusion Sheet, BCS), τα οποία μπορούν να χρησιμοποιηθούν από τους ενδιαφερόμενους προκειμένου να συντελέσουν στην επιτυχία του έργου. Το MR αποτελεί μία συνοπτική έκδοση του BCS, ενώ μετά το πέρας όλων των επισκέψεων, σε περίπτωση που κάποιο έργο εκτελείται σε περισσότερες από μία χώρες, συντάσσονται οι αντίστοιχες οριζόντιες αναφορές για το σύνολο κάθε διακρατικού. Οι αναφορές του συστήματος ROM περιέχουν τη συνολική αξιολόγηση ενός έργου, σε συνδυασμό με σχετικές εξηγήσεις και συστάσεις για βελτιώσεις. Από αυτές τις αναφορές προκύπτουν ιδιαίτερα χρήσιμα συμπεράσματα αναφορικά με την πορεία ενός έργου, τόσο όσον αφορά την πρόοδό του αλλά κυρίως όσον αφορά τους λόγους που οδήγησαν στην πρόοδο ή μη του έργου.

Το BCS που συντάσσεται για κάθε έργο είναι ένα αρχείο του Microsoft Excel που περιλαμβάνει αρκετά φύλλα και έχει τυποποιημένη μορφή ώστε να διευκολύνεται η συμπλήρωσή του. Καθένα από τα πρώτα πέντε φύλλα περιέχει την εκτίμηση του έργου αναφορικά με τα κριτήρια βάσει των οποίων γίνεται η αξιολόγηση στο σύστημα ROM (Ποιότητα Σχεδιασμού, Αποδοτικότητα, Αποτελεσματικότητα, Προοπτικές Αντίκτυπου, Δυνητική Αειφορία). Υπάρχουν δύο επιπλέον φύλλα που περιλαμβάνουν πληροφορίες για οριζόντια και διατομεακά θέματα καθώς και ένα ακόμη φύλλο στο οποίο αναφέρονται τα ονόματα των συνεντευξιζόμενων.

BACKGROUND CONCLUSION SHEET - ongoing projects						
3. EFFECTIVENESS TO DATE						
Project title	Many Voices against Poverty: Building a Diverse Constituency for Development					
CRIS contract N°	120765	Monitoring report reference N°	MR-124240	Date	30/9/2009	
CRIS decision N°	017-216	Monitor	Marc Raemaekers			
The contribution made by the project's results (as in "outcomes") to the achievement of the project purpose.						
PERFORMANCE CONCLUSION						
Prime Issues	a=4	b=3	c=2	d=1	Always to be added	
3.1 How well is the project achieving its planned results?					Weighting	50%
Have the planned results to date been achieved?	In spite of the good collaboration with the schools and the existing NGO networks, the quantitative targets (50000 students reached) are not likely to be realised. This is partly due to (a) an unclear formulation in the project proposal, (b) the trickle down effect of the school-based activities and (c) little contact between the participating schools. Continuity is not facilitated by the fact that a large part of the target group is dispersed at the end of every school year and, in spite of a degree of continuity provided by links with Universities and other institutions of Higher Education and links that ex students maintain with their schools/colleges, the project faces a considerable challenge at the start of every year, mobilising the new schools.					
Are the OVI/targets for the PP appropriate and are they being reported against?	The quality of the outputs (website, presentations, educational printed material etc.) is high and the collaboration with field NGO (OXFAM, Action Aid etc.) enhances it considerably. The target groups are taking the initiative in the use of the services/outcomes of the project, but the typology of the					

Εικόνα : Παράδειγμα φύλλου αναφοράς BCS - effectiveness to date

Όπως φαίνεται και στο παραπάνω σχήμα, η δομή κάθε φύλλου περιλαμβάνει τα κύρια θέματα / ερωτήματα (υποκριτήρια) που πρέπει να απαντηθούν, υπο-ερωτήματα καθώς και βαθμολογία για κάθε ερώτηση. Τα υπο-ερωτήματα έχουν ως σκοπό να καθοδηγήσουν τον πραγματογνώμονα για τη σύνταξη απάντησης προς το κύριο ερώτημα, και ενδέχεται να μην απαντώνται καθόλου στο κείμενο της αναφοράς. Υπάρχει επίσης η πιθανότητα ο πραγματογνώμονας να περιλαμβάνει στην αναφορά του στοιχεία που δεν

αναφέρονται σε κανένα υπο-ερώτημα αλλά που θεωρεί βασικά για την αξιολόγηση του έργου.

Σε κάθε φύλλο επίσης περιλαμβάνεται και η βαθμολογία κάθε ερώτησης, η οποία μπορεί να πάρει τέσσερις διακριτές τιμές: a, b, c, d που αντιστοιχούν στα αριθμητικά 4, 3, 2, 1 αντίστοιχα. Τα κριτήρια βαθμολόγησης παρουσιάζονται στον Πίνακα 1.

Βαθμός	Αριθμητικό	Ποιοτικό	Περιγραφή
a	4	Πολύ καλό	Η κατάσταση θεωρείται ιδιαίτερα ικανοποιητική. Η πορεία του έργου είναι καλύτερα από το μέσο όρο και ενδεχομένως αποτελεί σημείο αναφοράς για καλές πρακτικές. Τα διδάγματα εστιάζουν στην εφαρμογή των καλών πρακτικών και σε άλλες δραστηριότητες.
b	3	Καλό	Η κατάσταση θεωρείται ικανοποιητική, ωστόσο υπάρχει περιθώριο για βελτιώσεις. Τα διδάγματα είναι χρήσιμα αλλά όχι κρίσιμα για τη συνέχιση των δραστηριοτήτων του έργου.
c	2	Προβλήματα	Υπάρχουν προβλήματα, τα οποία θα πρέπει να αντιμετωπιστούν διαφορετικά η συνολική επίδοση του έργου ενδέχεται να επηρεαστεί αρνητικά. Οι απαραίτητες βελτιώσεις, ωστόσο, δεν απαιτούν μεγάλες μετατροπές στη στρατηγική υλοποίησης.
d	1	Σημαντική Ανεπάρκεια	Υπάρχουν ανεπάρκειες, οι οποίες δεν είναι τόσο σοβαρές ώστε να οδηγήσουν στην αποτυχία του έργου εάν δεν αντιμετωπιστούν. Σημαντικές μετατροπές πρέπει να πραγματοποιηθούν στην στρατηγική υλοποίησης.

Πίνακας : Οι βαθμολογίες του συστήματος ROM

Οι επιβλέποντες θα πρέπει να έχουν υπόψη ότι οι βαθμολογίες αναφέρονται μόνο στα κύρια ζητήματα και όχι στα υπο-ζητήματα κάθε κατηγορίας. Στη συνέχεια, θα πρέπει να κρίνουν βάσει του κειμένου των σχολίων τους που γράφεται στην διαθέσιμη περιοχή για κάθε ερώτηση. Ουσιαστικά δεν υπάρχει η επιλογή της μεσαίας βαθμολογίας, «στον στόχο, όπως είχε σχεδιαστεί» αλλά οι επιβλέποντες θα πρέπει να αποφανθούν εάν το έργο είναι καλύτερο ή χειρότερο από αυτή (a, b, c, d). Μια συγκεντρωτική βαθμολογία θα πρέπει να δίδεται σε κάθε περίπτωση, εκτός από έργα που θεωρούνται πολύ πρώιμα για να συμβεί κάτι τέτοιο. Σε αυτή την περίπτωση ο πραγματογνώμονας μπορεί να χρησιμοποιεί την επιλογή N/A, η οποία ωστόσο θα πρέπει να συνοδεύεται από επεξηγηματικό κείμενο που να την δικαιολογεί, ειδικά αν πρόκειται για επαναληπτική διαδικασία επίβλεψης του έργου.

Μεγάλες διακυμάνσεις και αποκλίσεις στις βαθμολογίες των ερωτήσεων, ιδιαίτερα αν είναι σημαντικές (όπως «b» με «d» ή αντίστροφα) πρέπει να εξηγούνται στο κείμενο της αναφοράς. Σε κάθε ερώτηση αντιστοιχεί επίσης και ένα βάρος (weighting) που αναφέρεται στη σχετική σημαντικότητα της ερώτησης ως προς την πορεία του έργου. Από την βαθμολόγηση όλων των επιμέρους ερωτήσεων και λαμβάνοντας υπόψη τα βάρη τους, υπολογίζεται αυτόματα ένας γενικός μέσος όρος βαθμολογίας για κάθε κριτήριο, που εμφανίζεται στο κουτί “overall conclusion” στο κάτω μέρος κάθε φύλλου.

Οι πραγματογνώμονες **δε** θα πρέπει:

- Να βαθμολογούν με «**d**» ώστε να επιστήσουν την προσοχή σε κάποιο θέμα που θεωρούν σημαντικό, αλλά το οποίο δεν δικαιολογεί τέτοιο βαθμό.
- Να βαθμολογούν με «**a**» λόγω συμπάθειας στον τρόπο προσέγγισης του έργου ή της συνάφειάς του.
- Να μεταβάλλουν τις βαθμολογίες των κυρίων θεμάτων με σκοπό να προκύψει ο επιθυμητός συνολικός βαθμός.

Με βάση την αναφορά και τη βαθμολογία κάθε ερώτησης, στο κάτω μέρος κάθε φύλλου περιλαμβάνονται από τον πραγματογνώμονα διδάγματα που προέκυψαν από την επίβλεψη του έργου ή/και συστάσεις για τη βελτίωση της πορείας του έργου, καθώς και το από ποια μέρη θα πρέπει να προέλθουν οι προτεινόμενες δράσεις. Ο πραγματογνώμον οφείλει να δίνει προτεραιότητα στις δράσεις που θεωρεί περισσότερο σημαντικές, και ο τρόπος σύνταξής τους να είναι συνοπτικός και άμεσος. Τα Διδάγματα αποτελούν ίσως το σημαντικότερο σημείο της αναφοράς, από τη στιγμή που δε λειτουργούν απολογιστικά στα έργα στα οποία αναφέρονται, αλλά προληπτικά, καθώς από αυτά προκύπτουν οι συστάσεις στις οποίες πρέπει η Κομισιόν και οι φορείς των έργων να δώσουν τη μεγαλύτερη έμφαση, είτε για τα ήδη υπάρχοντα έργα, είτε για τα νέα έργα της επόμενης χρονιάς. Παράγοντες που οδήγησαν στην επιτυχία ή όχι ενός έργου μπορεί, ενδεχομένως σε μικρότερο βαθμό, να δημιουργήσουν αντίστοιχες ενδείξεις στα υπόλοιπα έργα. Επομένως η διαχείριση των διδασμάτων και συστάσεων που προκύπτουν από το σύνολο των αναφορών αποτελεί ένα ισχυρό όπλο καθώς μπορεί να αποτρέψει ή να προωθήσει αρνητικές ή θετικές αντίστοιχα εξελίξεις στην πορεία ενός έργου.

ΚΕΦΑΛΑΙΟ 4^ο : ΜΕΘΟΔΟΛΟΓΙΑ

4.1 Εισαγωγή

Η απόκτηση γνώσης είναι η διαδικασία που εξάγει τη γνώση από μία πηγή και την ενσωματώνει σε ένα σύστημα βασισμένο στη γνώση (knowledge-based system) το οποίο λύνει κάποιο πρόβλημα. Είτε η εξόρυξη γίνεται από κάποιο άτομο είτε από κάποιο αυτόματο πρόγραμμα, τελικά πρέπει να κωδικοποιείται σε κάποια μορφή αναπαράστασης γνώσης. Συνεπώς, για την απόκτηση γνώσης είναι ιδιαίτερα σημαντικά τόσο το στάδιο της εξόρυξης όσο και το στάδιο της καταχώρησης και οπτικοποίησης της.

Ένα από τα πρώτα προβλήματα που θα έπρεπε να λάβουμε υπόψη μας κατά την ανάπτυξη της μεθοδολογίας αναπαράστασης γνώσης, ήταν το πρόβλημα της αλληλεπίδρασης (interaction problem): Η αναπαράσταση γνώσης είναι έντονα συνδεδεμένη με τη φύση του προβλήματος που καλούμαστε να αντιμετωπίσουμε καθώς και με τη συμπερασματική στρατηγική που θα εφαρμοστεί στη γνώση. Το πρόβλημα της αλληλεπίδρασης δεν είναι κάτι καινούργιο. Ο Minsky, αναφέρει πως «τα πραγματικά και διαδικαστικά περιεχόμενα πρέπει να είναι ιδιαίτερα συνδεδεμένα για να εξηγούν την εμφανή δύναμη και ταχύτητα των πνευματικών δραστηριοτήτων» (Minsky, 1975). Ο Marr έχει σημειώσει ότι «το πώς αναπαριστούνται οι πληροφορίες μπορεί να επηρεάσει σημαντικά το πόσο εύκολο είναι να κάνουμε διάφορα πράγματα με αυτές» (Marr, 1982). Στην ανάπτυξη της δικής μας μεθοδολογίας, η γνώση θα αναπαρασταθεί με τέτοιο τρόπο ώστε να εξυπηρετεί τις ανάγκες που καλείται να καλύψει.

Ο πυρήνας της μελέτης ως προς την ανάλυση ήταν τα έργα και οι ποιοτικές πληροφορίες που περιείχαν οι αναφορές τους (MR, BCS) σύμφωνα με το σύστημα ROM. Επιλέχθηκαν σκοπίμως έργα με ιδιαίτερα καλές βαθμολογίες («a») ή ιδιαίτερα άσχημες («d»), καθώς θεωρήθηκε πως σε αυτές τις αναφορές θα διαφαίνονται εντονότερα και πιο ξεκάθαρα οι παράγοντες που επηρέασαν την βαθμολόγησή τους.

Η ανάπτυξη της μεθοδολογίας βασίστηκε τόσο σε βιβλιογραφικές έρευνες, όσο και ήδη επιτυχώς εφαρμοζόμενες μεθοδολογίες αναπαράστασης γνώσης σε συστήματα εξόρυξης γνώσης. Τα βήματα της διαδικασίας σκιαγραφήθηκαν από τον οδηγό για κατασκευή συστημάτων αναπαράστασης γνώσης που έχει συντάξει ομάδα του πανεπιστημίου Stanford, καθώς θεωρήθηκε ο πιο πλήρης και αντιπροσωπευτικός για το εγχείρημά μας, χαρακτηρίζεται από σαφήνεια και βασίζεται σε εμπειριστωμένη βιβλιογραφία. (Natalya F. Noy, 2000) Επιπλέον ιδιαίτερη βάση δόθηκε στην μεθοδολογία που αναπτύχθηκε από μέρους της Κομισιόν σχετικά με την εξεύρεση αιτιών που σχετίζονται με την αποτελεσματικότητα έργων ανάπτυξης του συστήματος ROM. Πρόκειται για μεθοδολογία που σχετίζεται με το αντικείμενο μελέτης της εργασίας και συνεπώς σχετίζεται άμεσα με ορισμένα σημεία της. (Co-operation, Causes underlying Effectiveness and Impact of EC Development Projects, 2009)

4.2 Το Μεθοδολογικό Πλαίσιο

Στον τομέα της αναπαράστασης γνώσης και των οντολογιών, είναι κοινά αποδεκτό πως δεν υπάρχει ένας μοναδικός σωστός τρόπος ή μεθοδολογία για την ανάπτυξη συστημάτων αναπαράστασης γνώσης. Αυτό συμβαίνει γιατί κάθε σύστημα έχει τον δικό του χαρακτήρα και θα πρέπει να εξυπηρετεί διαφορετικές κάθε φορά ανάγκες. Γι αυτό το λόγο, η ανάπτυξη της μεθοδολογίας περιέχει μια σειρά από σχεδιαστικές αποφάσεις. Προκειμένου να υπάρχει μια βάση αναφοράς στη λήψη των αποφάσεων αυτών, τέθηκαν και ακολουθήθηκαν ορισμένοι βασικοί κανόνες.

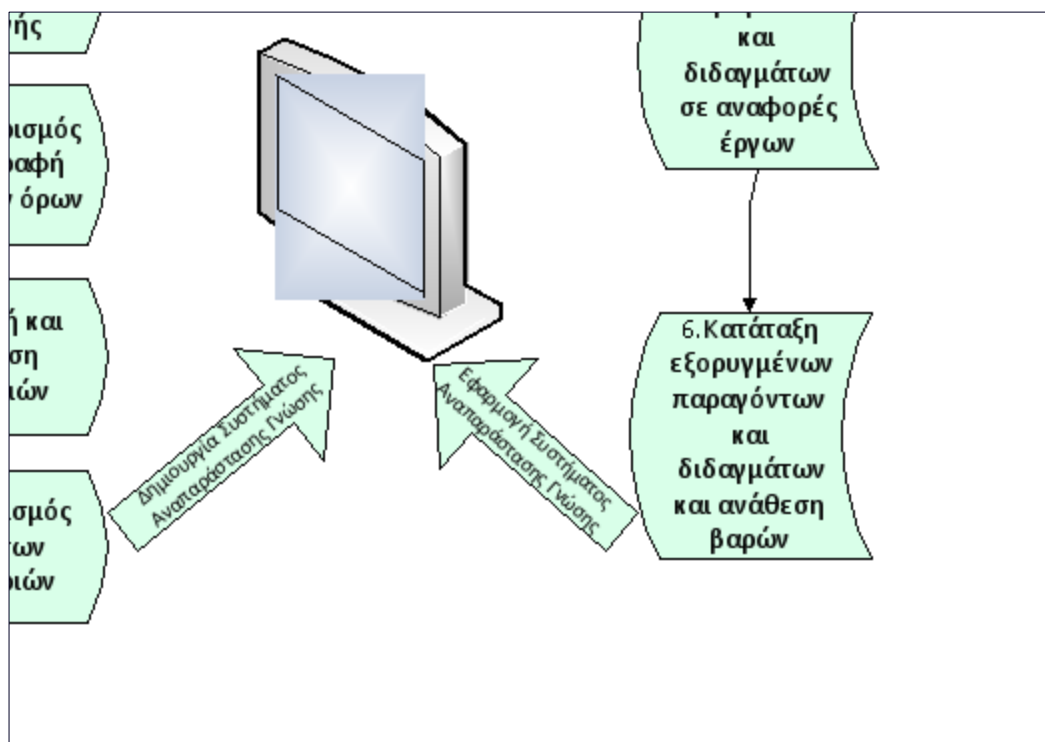
- 1) *Δεν υπάρχει μόνο ένας σωστός τρόπος αναπαράστασης ενός τομέα, υπάρχουν πάντα πιθανές εναλλακτικές λύσεις. Η καλύτερη λύση πάντα εξαρτάται από την εφαρμογή που θέλουμε να πετύχουμε και τις προεκτάσεις που προσδοκούμε.*

- 2) *Η ανάπτυξη ενός συστήματος αναπαράστασης γνώσης είναι αναγκαστικά μια επαναληπτική διαδικασία.*

- 3) *Οι έννοιες του συστήματος θα πρέπει να είναι κοντά στα αντικείμενα (φυσικά ή λογικά) και στις σχέσεις στον τομέα ενδιαφέροντος. Στην περίπτωση μας θα είναι προτάσεις που θα περιγράφουν τον τομέα.*

Η μελλοντική χρήση του συστήματος αναπαράστασης γνώσης και το επιθυμητό επίπεδο λεπτομέρειας είναι οδηγοί στη λήψη πολλών από τις σχεδιαστικές αποφάσεις. Ένα αποτελεσματικό σύστημα αναπαράστασης γνώσης θα πρέπει να είναι σχεδιασμένο και δομημένο σύμφωνα με τις ανάγκες του χρήστη της γνώσης. Μετά από τον αρχικό σχεδιασμό του συστήματος, ακολουθείται αξιολόγησή του και λύση προβλημάτων, προκειμένου να αντιπροσωπεύει όσο το δυνατόν περισσότερο την πραγματικότητα. Άλλωστε, η διαδικασία επαναληπτικού σχεδιασμού συνεχίζεται καθ' όλη τη διάρκεια ζωής του συστήματος αναπαράστασης γνώσης.

Τα βήματα ανάπτυξης της μεθοδολογίας φαίνονται στο ακόλουθο σχήμα (Εικόνα). Τα τέσσερα πρώτα αναφέρονται στον σχεδιασμό του συστήματος και τα δύο τελευταία αφορούν στην εισαγωγή εννοιών σε αυτό. Στα επόμενα κεφάλαια αναλύεται διεξοδικά κάθε βήμα, επισημαίνοντας και αιτιολογώντας τις σχεδιαστικές αποφάσεις που λαμβάνονταν κάθε φορά. Επιπλέον, παραδίδεται ο αναλυτικός τρόπος υπολογισμού βαρών, ο οποίος χρησιμοποιείται κατά την εφαρμογή του συστήματος, στο βήμα 6. Απώτερος σκοπός της διπλωματικής είναι η αυτόματη πραγματοποίηση των βημάτων 5 και 6 με χρήση τεχνικών εκμάθησης μηχανής. Όπως θα δούμε και από την εφαρμογή σε αναφορές έργων, σε αυτή την περίπτωση η αυτόματη καταχώρηση θα είναι δυνατόν να πραγματοποιηθεί στο σύστημα που σχεδιάστηκε.



Εικόνα : Βήματα μεθοδολογίας ανάπτυξης του συστήματος Αναπαράστασης Γνώσης

4.3 Μεθοδολογική προσέγγιση

Στη συνέχεια παρουσιάζονται αναλυτικά τα βήματα ανάπτυξης της μεθοδολογίας. Σε κάθε στάδιο αναφέρονται τα ερωτήματα που χρειάστηκε να απαντηθούν και οι αποφάσεις που έπρεπε να ληφθούν ώστε το σύστημα να είναι λειτουργικό και να εξυπηρετεί τους αναμενόμενους σκοπούς.

4.3.1 Καθορισμός του τομέα και του πεδίου εφαρμογής της γνώσης

Αρχικό και ουσιώδες βήμα της διαδικασίας αναπαράστασης γνώσης είναι ο καθορισμός του τομέα και του πεδίου εφαρμογής της γνώσης που θα προκύψει. Προκειμένου να γίνουν σαφείς οι σκοποί που θα εξυπηρετεί το σύστημα, χρειάστηκε να απαντηθούν μερικές βασικές ερωτήσεις.

➤ *Ποιον τομέα θα καλύψει το σύστημα αναπαράστασης γνώσης;*

Ο τομέας στον οποίο θα αναφέρονται τα στοιχεία του συστήματος αποτελείται από ένα σύνολο παραγόντων και διδαγμάτων που εντοπίζονται στις αναφορές αξιολόγησης, μέσω του συστήματος ROM, έργων της Ευρωπαϊκής Ένωσης. Τόσο οι παράγοντες όσο και τα διδάγματα καταχωρούνται υπό μορφή ολοκληρωμένων προτάσεων, καθώς και με τον ακριβή τρόπο με τον οποίο αναφέρονται στο κείμενο της αξιολόγησης.

➤ *Ποια θα είναι η χρήση του συστήματος αναπαράστασης γνώσης;*

Στην τελική του μορφή το σύστημα θα πρέπει να παρέχει στον χρήστη τη δυνατότητα να επιλέγει οποιονδήποτε συνδυασμό κατηγοριών παραγόντων επιθυμεί (οικονομικούς, σχεδιασμού κτλ) και να παρουσιάζει ως αποτελέσματα παράγοντες που υπόκεινται σε αυτές τις κατηγορίες και επηρεάζουν τις αξιολογήσεις των έργων καθώς και πιθανούς τρόπους αντιμετώπισης ή βελτίωσης των φαινομένων, εάν υπάρχουν. Τα αποτελέσματα που θα προκύπτουν θα βασίζονται σε έργα των οποίων οι αναφορές έχουν ήδη μελετηθεί. Ωστόσο σκοπός είναι το σύστημα να μπορεί να χρησιμοποιηθεί αργότερα για σύγκριση αποτελεσμάτων σε σχέση με αυτόματα συστήματα εντοπισμού τέτοιων παραγόντων, τα οποία ίσως να μπορούν επίσης να προβλέψουν επιπλέον στοιχεία και να εξάγουν αυτόματα συμπεράσματα.

➤ *Σε τι είδους ερωτήσεις θα πρέπει να απαντά η πληροφορία που θα παρέχεται από το σύστημα αναπαράστασης γνώσης;*

Η βασική ερώτηση που θα αντιστοιχεί στην αναζήτηση πληροφοριών στο σύστημα αναπαράστασης γνώσης, μπορεί να περιγραφεί ως εξής: «Τι προβλήματα/θετικά στοιχεία έχουν εντοπιστεί που να σχετίζονται με τα εξής κριτήρια αξιολόγησης έργων και ποιοι τρόποι αντιμετώπισης ή ενίσχυσής τους έχουν σημειωθεί;». Τα κριτήρια θα επιλέγονται κάθε φορά από τον χρήστη. Το σύστημα θα παρουσιάζει τόσο τους αρνητικούς όσο και τους θετικούς παράγοντες που σχετίζονται με κάθε κατηγορία, διαχωρίζοντας ωστόσο το αν ωφελούν ή ανακόπτουν την πορεία του έργου. Επίσης, στην περίπτωση που έχει εντοπιστεί κάποιος τρόπος αντιμετώπισης αρνητικού φαινομένου ή ενίσχυσης θετικού φαινομένου, το σύστημα θα παραπέμπει τον χρήστη και σε αυτή την αναφορά.

4.3.2 Προκαθορισμός και καταγραφή σημαντικών όρων του συστήματος αναπαράστασης γνώσης

Στο στάδιο αυτό καταγράφηκε ένα μέρος σημαντικών παραγόντων που αναμένονταν να προκύψουν κατά την μετέπειτα διαδικασία εξόρυξης γνώσης. Το βήμα αυτό ήταν χρήσιμο διότι έδωσε μια πρώτη εικόνα του είδους των παραγόντων που προκύπτουν, των περισσότερο συχνά εμφανιζόμενων καθώς και μια γενική αίσθηση του επιπέδου λεπτομέρειας τους, στοιχεία που βοήθησαν στα μετέπειτα βήματα του καθορισμού και της ιεράρχησης των κατηγοριών. Στο αρχικό αυτό σημείο δεν δόθηκε έμφαση στην επικάλυψη παραγόντων (σε περίπτωση που αναφέρονταν σε ίδιες έννοιες), ούτε στις διασυνδέσεις μεταξύ παραγόντων και διδαγμάτων. Για το βήμα μελετήθηκαν 3 αναφορές έργων, με αρνητικές και θετικές βαθμολογίες, στις οποίες εντοπίστηκαν 23 παράγοντες που επηρέασαν την πορεία τους και 4 σχετικά διδάγματα που προέκυψαν από την επίβλεψη των έργων αυτών.

Τα επόμενα δύο βήματα -της ανάπτυξης και ιεράρχησης των κατηγοριών καθώς και του καθορισμού των ιδιοτήτων τους- είναι στενά συνυφασμένα μεταξύ τους, γι αυτό και πολλές από τις διαδικασίες τους πραγματοποιήθηκαν ταυτόχρονα. Αξίζει να σημειωθεί πως αυτά τα δύο βήματα είναι τα περισσότερο σημαντικά στον σχεδιασμό του συστήματος αναπαράστασης γνώσης.

4.3.3 Επιλογή και ιεράρχηση κατηγοριών

Στο βήμα αυτό ήταν ιδιαίτερα απαραίτητο να προσαρμόσουμε το σύστημα αναπαράστασης γνώσης όσο το δυνατόν περισσότερο στα δεδομένα της αξιολόγησης μέσω του συστήματος ROM. Θεωρήθηκε επίσης σκόπιμο να βασίσουμε πολλές από τις κατηγορίες του συστήματος σε αυτές που προέκυψαν από την αναπτυσσόμενη μεθοδολογία για αυτόματο εντοπισμό παραγόντων και διδαγμάτων στις αναφορές του συστήματος ROM, καθώς προβλέπεται η μελλοντική ταυτόχρονη χρήση των δύο συστημάτων (Assistance, 2009). Κατά τη μελέτη αυτή, εντοπίστηκαν 47 μεταβλητές για τα προβληματικά έργα και 43 για όσα έχουν καλή απόδοση, μερικές από τις οποίες είναι ίδιες και για την περίπτωση καλής και για την περίπτωση κακής απόδοσης. Για την ανάπτυξη της παρούσας μεθοδολογίας, από αυτό το σύνολο μεταβλητών αφαιρέθηκαν όσες αναφέρονταν στη σύνδεση κριτηρίων, καθώς θεωρήθηκε πως κάτι τέτοιο θα μπορεί να γίνει αντιληπτό μέσω της χρήσης του συστήματος. Οι μεταβλητές αυτές αποτέλεσαν τη βάση για την επιλογή κατηγοριών.

#	Variables	Characteristics and explanatory causes				
		for Good Performance			for Poor performance	
		Code	Crucial	BCS part	Code	Crucial
A	Conception and Formulation					
1	Strategic/apparent relevance	A.1	2	1.1	A.1	2
2	Practical/real relevance	A.2	2	1.1/1.3	A.2	2
3	Overambitious objectives	-	-	1.2	A.3	3

4	Adequate logframe	A.3		1.2/1.2	A.4	
5	Appropriate OVIs (PP, OO levels)	A.4		1.2	A.5	
6	Appropriate intervention logic	A.5		1.2	A.6	
7	Proper risks and assumptions	A.6		1.2	A.7	4
8	Appropriate Analysis Stage	A.7	1	1.1/1.2/1.3/1.4	A.8	1
9	Appropriate implementation set-up to ensure outputs/outcomes	A.8		1.2	A.9	
10	Acceptable Internal Monitoring System for activities / outcomes	A.9/A.10		1.3 (2.1/2.2/2.4/6.1/7.3)	A.10 / A.11	5
11	Use of Logframe as a management and reporting tool	A.11		1.2 (2.2)	A.12	
12	Use of OVIs	A.12		1.2 (2.3)	A.13	
B	Implementation					
1	Consideration of external factors re: the timing of activities	-	-	2.2	B.1	
2	Consideration of other external factors	-	-	2.4 (3.2/4.1)	B.8	
3	Understanding of EC procedures	B.1		2.1	B.2	
4	Appropriate Reporting	B.2		2.4 (7.3)	B.3	
5	Appropriate provision of Inputs	B.3		2.1	B.4	
6	Proper implementation of Activities	B.3		2.2	-	-
7	Appropriate management	B.6	3	2.1/2.2/2.4 (1.3)	B.5	
8	Effective communication among project partners	B.4	6	2.4	B.6	6
9	Effective communication between the EC and its partners	-	-	2.4	B.7	6
10	Appropriate quality of Human Resources	B.5		2.1 (6.2.1)	B.9	
C	Outputs into outcomes (EFFECTS)					
1	Link of effectiveness and external factors	C.2		3.2	C.1	
2	Link of effectiveness and achievement of the PP	C.3		3.2	C.2	
3	Association of effectiveness and OVIs	C.1		3.1	-	-
4	Association of effectiveness with access to project benefits	-	-	3.1	C.3	
5	Association of effectiveness with use of project benefits	C.4	5	3.1	C.4	
D	Transformation of outcomes into WIDER EFFECTS					
1	Association of impact with External Factors	D.2		4.1	D.1	
2	Association of impact and achievement of the PP	D.9		4.1 (3.2)	D.2	

3	Distinction between degree of effectiveness and impact	D.8		4 + 3	D.3	
4	Follow up of risks and assumptions	D.4		3.2	D.4	
5	Association of impact and OVIs	D.1		4.1	-	-
6	Association of impact and PP's contribution to the OO	D.3		4.1 + 3.2	-	-
7	Relation of impact with other actions/interventions	D.5		4.2	D.5	
8	Coordination among donors	D.6		4.1	D.6	
9	Appropriate modelling or replication	D.7		4.1	D.7	
10	Avoidance of negative impact	-	-	4.2	D.8	
11	EC visibility	D.10		6.3	D.9	
E	CROSSCUTTING issues					
1	Ownership/Participation	E.1		5.2 (6.2)	E.1	
2	Technical capacity of the implementing partner(s)	E.2	4	5.4	E.2	
3	Administrative capacity of the implementing partner(s)	E.3	4	5.4	E.3	
4	Project adaptation to arisen problems or opportunities	E.4	6	5.3	E.4	6
5	Absorption capacity of local institutions	-	-	5.4	E.5	
6	Effective suggested and granted extensions	E.5/E.6		2.4	E.6/E.7	

Πίνακας : Το σύνολο των μεταβλητών όπως προέκυψε από την μελέτη για ανάπτυξη μεθοδολογίας αυτόματης εξόρυξης γνώσης (Assistance, 2009)

Για την ιεράρχηση των κατηγοριών, χρησιμοποιήθηκε η συνδυαστική προσέγγιση ανάπτυξης κλάσεων ιεραρχιών των Uschold και Gruninger (M. Uschold, 1996). Συγκεκριμένα, έχοντας σαν δεδομένες τις μεταβλητές, τις γενικοποιήσαμε υπό κάποιες περισσότερο ευρείες έννοιες και στη συνέχεια τις συγκεκριμενοποιήσαμε σε κάποιες λιγότερο γενικές.

Δεδομένου του χαρακτήρα του συστήματος αναπαράστασης γνώσης που σχετίζεται άμεσα με το σύστημα ROM, καθώς και της μετέπειτα χρήσης του συστήματος, που σχετίζεται με τις αναφορές του συστήματος ROM και τη δομή τους, θεωρήθηκε απαραίτητο να λάβουμε υπόψη τα στοιχεία αυτά κατά την κατηγοριοποίηση και ιεράρχηση. Συγκεκριμένα, οι 5 βασικές και γενικές κατηγορίες επιλέχθηκαν να είναι τα 5 βασικά κριτήρια αξιολόγησης του συστήματος ROM (Quality of Project Design, Efficiency, Effectiveness, Impact, Sustainability). Μετά την κατηγοριοποίηση των μεταβλητών σε καθένα από αυτά τα βασικά κριτήρια (down-top προσέγγιση), ανάλογα με το σε ποιο αναφέρονταν, κρίθηκε απαραίτητη μια επιπλέον κατηγοριοποίηση, βάσει υπο-κριτηρίων, που θα σχετιζόταν άμεσα με τις ερωτήσεις της φόρμας αξιολόγησης του συστήματος ROM. Στις ερωτήσεις αυτές καλούνται να απαντήσουν οι εμπειρογνώμονες με ελεύθερο κείμενο κατά τη διαδικασία επίβλεψης και από τις απαντήσεις τους θα προκύψουν τα συμπεράσματά μας. Έτσι έγινε μια επιπλέον κατάταξη των μεταβλητών στις υποκατηγορίες των κύριων κριτηρίων (top-down φάση της προσέγγισης), προσθέτοντας ένα ακόμα επίπεδο αοριστίας στο σύστημα αναπαράστασης γνώσης. Για τις ανάγκες τις συγκεκριμένης

μεθοδολογίας κρίθηκε πως δεν είναι απαραίτητο επιπλέον επίπεδο αοριστίας, δεδομένου μάλιστα ότι τα υπο-ερωτήματα των ερωτήσεων έχουν γενικότερα καθοδηγητικό χαρακτήρα και δεν είναι υποχρεωτική η απάντηση των συγκεκριμένων υπο-ερωτημάτων στην σύνταξη αναφοράς του συστήματος ROM.

Αναμφίβολα, μια επιπλέον κλάση θα έπρεπε να αποτελούν τα διδάγματα. Δεδομένου του χαρακτήρα και του περιεχομένου του συστήματος προς σχεδιασμό, τα διδάγματα έπαιξαν ρόλο και στο μέρος των σχεδιαστικών αποφάσεων. Με δεδομένο ότι περισσότερες από μία έννοιες μπορεί να έχουν ένα δίδαγμα, τα διδάγματα επιλέχθηκαν ως ένα επίπεδο αφαίρεσης αντίστοιχο των υποκριτηρίων. Ένα συχνό λάθος σε συστήματα αναπαράστασης γνώσης είναι η δομή της βάσης σύμφωνα με τον τρόπο που παρέχονται οι πληροφορίες. Τα διδάγματα θα μπορούσαν να κατηγοριοποιηθούν ανά έργο, για παράδειγμα, σε ομάδες συμπερασμάτων που προέκυψαν από τις αναφορές ενός έργου. Ωστόσο, κάτι τέτοιο θα ήταν μηδενικής ή λίγης αξίας για κάποιον που αναζητά πρακτικές βελτίωσης έργων γενικά. Ο χρήστης του συστήματος δεν χρειάζεται να γνωρίζει την ιστορία προηγούμενων έργων ή το ποια διδάγματα εντοπίστηκαν σε ποια έργα. Αυτό που αναζητά και θα μπορεί να βρει μέσω του συστήματος θα είναι διδάγματα αναφορικά σε κάποιες δραστηριότητες ή τομείς. Η γενική μορφή της καταχώρησης των διδαγμάτων συμπεριλαμβάνει το δίδαγμα, το εμπλεκόμενο μέρος που μπορεί να προβεί σε κάποια διορθωτική/βελτιωτική δράση, και το περιεχόμενο της δράσης, έτσι ώστε να είναι σαφές και απλό στην χρήση.

Το σχεδιάγραμμα ιεράρχησης των κλάσεων παρουσιάζεται στο επόμενο σχήμα



Εικόνα : Σχεδιάγραμμα επιπέδων ιεράρχησης κλάσεων συστήματος

Η σχέση ανάμεσα σε κλάσεις - υποκλάσεις

Η ιεραρχία των κλάσεων αντιπροσωπεύει την εξής σχέση: Αν η κλάση A είναι υποκλάση της B, τότε κάθε στοιχείο της B είναι επίσης στοιχείο και της A. Ένας τρόπος περιγραφής της σχέσης αυτής είναι με τον όρο «είναι ένα είδος». Για παράδειγμα, η ύπαρξη επαρκούς πλαισίου σχεδιασμού, είναι ένα είδος κριτηρίου του σχεδιασμού του έργου. Μια υποκλάση μιας κλάσης αντιπροσωπεύει μια έννοια που είναι «είδος» της έννοιας που αντιπροσωπεύει η υπερ-κλάση της.

Η σχέση ανάμεσα σε κλάσεις – υποκλάσεις είναι κληρονομική (μεταβατική – transitivity). Εάν η B είναι υποκλάση της A και η C είναι υποκλάση της B, τότε η C είναι υποκλάση της A. Για παράδειγμα, καθορίσαμε την κλάση Efficiency και στη συνέχεια την υποκλάση της Input_Management. Κατόπιν ορίσαμε την κλάση Provision_of_inputs ως υποκλάση της Inputs_Management. Η κληρονομικότητα ανάμεσα στις κλάσεις σημαίνει πως η κλάση Provision_of_inputs είναι επίσης υποκλάση της Efficiency. Χρησιμοποιώντας έναν άλλον όρο που αναφέρεται στην βιβλιογραφία (Natalya F. Noy, 2000), μπορούμε να κάνουμε διαχωρισμό και σε άμεσες ή όχι κλάσεις, θεωρώντας ως άμεση υποκλάση την κοντινότερη υποκλάση στην κλάση: δεν υπάρχουν άλλες κλάσεις ανάμεσα σε μια κλάση και στην άμεση υποκλάση της. Θα μπορούσαμε έτσι να πούμε πως η κλάση Provision_of_inputs είναι άμεση υποκλάση της Inputs_Management αλλά όχι άμεση υποκλάση της Efficiency.

«Αδέρφια» στην ιεραρχία κλάσεων

Με τον όρο «αδέρφια» στην ιεραρχία, εννοούμε τις κλάσεις που είναι άμεσες υποκλάσεις της ίδιας κλάσης. Όλα τα «αδέρφια» στην ιεραρχία, με εξαίρεση αυτά που είναι στη ρίζα, είναι στο ίδιο επίπεδο γενικότητας. Για παράδειγμα, η Quality_of_HR και η Monitoring_of_inputs έχουν ένα τέτοιο είδος σχέσης, όχι όμως και με την Inputs_Management, η οποία αντιπροσωπεύει μια περισσότερο γενική έννοια. Τα 5 βασικά κριτήρια, που αποτελούν και τον κύριο διαμοιρασμό του τομέα, δεν χρειάζεται να έχουν το ίδιο επίπεδο λεπτομέρειας.

Για την κατηγοριοποίηση των κλάσεων, έπρεπε επίσης να αναρωτηθούμε σχετικά με τον αριθμό των υποκλάσεων που θα είχε μια κλάση. Για να αποφευχθούν προβλήματα δομής του συστήματος, ακολουθήσαμε δύο βασικούς κανόνες (Natalya F. Noy, 2000):

Εάν μια κλάση έχει μόνο μια άμεση υποκλάση, ενδέχεται να υπάρχει σχεδιαστικό πρόβλημα ή το σύστημα αναπαράστασης γνώσης να μην είναι ολοκληρωμένο.

Εάν υπάρχουν περισσότερες από δώδεκα υποκλάσεις για μια κλάση, τότε ίσως να χρειάζονται πρόσθετες κατηγοριοποιήσεις.

Με άλλα λόγια, μια κλάση μπορεί να έχει από δύο έως δώδεκα υποκλάσεις, αλλιώς δεν έχει καμία.

4.3.4 Καθορισμός ιδιοτήτων κατηγοριών

Οι κλάσεις από μόνες τους δεν παρέχουν αρκετή πληροφορία ώστε να μπορούν να απαντηθούν οι ερωτήσεις του πρώτου βήματος. Επομένως, κρίθηκε απαραίτητη η περιγραφή της εσωτερικής δομής των εννοιών. Έτσι, ορισμένες μεταβλητές εξειδικεύτηκαν περαιτέρω, ανάλογα με τον τρόπο εμφάνισής τους. Για παράδειγμα η μεταβλητή

Project_Benefits περιέχει τα slots *Access* και *Use*, καθώς θεωρείται διαφορετικός παράγοντας η πρόσβαση στα οφέλη ενός έργου από τη χρήση τους. Στο βήμα αυτό χρησιμοποιήθηκε η γνώση που προέκυψε από το βήμα 2, καθώς μπορέσαμε να εντοπίσουμε τέτοιου είδους διακρίσεις.

Μια επιπλέον διαχώριση σε αυτό το σημείο έγινε σε αρνητικούς ή θετικούς παράγοντες. Η διαχώριση αυτή έγινε στο τελευταίο επίπεδο, καθώς θεωρούμε πως, λόγω κληρονομικότητας, τα υπο-κριτήρια που είναι πάνω από μία αρνητική έννοια ορίζονται βάσει της αρνητικής τους πλευράς. Η έννοια της κληρονομικότητας είναι παρούσα σε όλη τη δομή της οντολογίας, καθώς όλες οι υποκατηγορίες κάποιας κατηγορίας κληρονομούν τις ιδιότητες της κατηγορίας.

Με βάση όλα τα παραπάνω, δημιουργήθηκε ο ακόλουθος πίνακας, στον οποίο παρουσιάζονται τόσο οι κατηγορίες όσο και η ιεράρχησή τους.

M1	RELEVANCE / QUALITY OF PROJECT DESIGN
M11	Level of relevance
I.	Alignment to Policies
II.	Alignment to Target groups' needs
III.	Strategic relevance
IV.	Practical relevance
M12	Appropriateness of the Intervention Logic
I.	Adequate Logframe
II.	OVI's
a.	<i>SMART</i>
b.	<i>Used</i>
III.	Logical Objectives
IV.	Proper Risks
V.	Appropriate Analysis Stage
M13	Support by Stakeholders
M14	Mainstream of cross-cutting issues
M2	EFFICIENCY
M21	Inputs Management
I.	Provision of Inputs
a.	<i>On time</i>
b.	<i>On cost</i>
II.	Quality of HR
III.	Monitoring of inputs
IV.	Appropriate reporting
V.	Understanding of EC procedures
M22	Activity Timelines
I.	Consideration of external factors regarding the timing of activities
II.	Use of Timescale
III.	Proper implementation of activities
IV.	Use of Logframe
M23	Outputs Achievement
I.	Likelihood of PP to be achieved
II.	Reflection of outputs through OVI's
III.	Contribution of outputs to intended results
M24	Partner Contribution & Involvement

I.	Effective communication among partners
II.	Effective communication among EC & partners
M3	EFFECTIVENESS
M31	Results attainment
I.	To-date
II.	Project benefits
a.	<i>Accessible</i>
b.	<i>Used</i>
III.	Quality of OVI's
M32	Project Purpose achievement
M4	IMPACT
M41	Contribution to the Overall Objective
I.	Direct Impact
II.	External factors
III.	Coordination among donors
IV.	Appropriate modeling or replication
M42	Wider effects
I.	Indirect impact
II.	EC visibility
III.	Continuation
IV.	Avoidance of negative impact
V.	Other actions/interventions
M5	SUSTAINABILITY
M51	Economic Viability
M52	Local issues
I.	Local ownership / participation
II.	Adaptation / Absorption
M53	Policy Support
M54	Institutional Capacity Building
I.	Technical capacity
II.	Administrative capacity
III.	Absorption capacity

Πίνακας : Οι κατηγορίες του συστήματος αναπαράστασης γνώσης κι η ιεράρχησή τους

Τα επόμενα δύο βήματα της μεθοδολογίας αναφέρονται στην καταχώρηση της γνώσης στο σύστημά μας και συχνά οι διαδικασίες τους πραγματοποιούνταν ταυτόχρονα.

4.3.5 Ανάλυση σε βάθος αναφορών αξιολόγησης έργων και εντοπισμός παραγόντων και διδαγμάτων

Στο στάδιο αυτό μελετήθηκαν ελεύθερα κείμενα των εμπειρογνομόνων που παρουσιάζονται στις αναφορές του συστήματος ROM. Η αναζήτηση των αναφορών των έργων έγινε μέσα από το σύστημα KOMIS (Knowledge Optimization Meta-Information System), που πρόκειται για online βάση που περιλαμβάνει όλες τις αναφορές του συστήματος ROM και διαθέτει επιλογές αναζήτησης. Για την επιλογή των αναφορών έργων προς εξέταση χρησιμοποιήθηκε το μη-πιθανοτικό μοντέλο του δείγματος αποφάσεων

(judgment sample), δηλαδή επιλέχθηκαν κάποια είδη «εκπροσώπων» των αναφορών βάσει κάποιων κριτηρίων (Deming, 1990). Τα κριτήρια ήταν:

1. Ιδιαίτερα υψηλή ή χαμηλή βαθμολογία (a ή d), είτε στο σύνολο του έργου είτε σε επιμέρους ερωτήματα της αξιολόγησής τους.
2. Επαρκής και σαφής αιτιολόγηση στο ελεύθερο κείμενο της βαθμολογίας κάθε ερώτησης

Το πρώτο κριτήριο επιλέχθηκε καθώς, λόγω της κλίμακας της βαθμολογίας, θεωρήθηκε πως έργα με βαθμολόγηση b ή c δεν θα έχουν σαφείς διαφοροποιήσεις και τα στοιχεία που θα αναφέρονται στις αξιολογήσεις τους δεν θα έχουν ιδιαίτερα έντονο θετικό ή αρνητικό αντίκτυπο. Δεδομένου του χαρακτήρα του συστήματος αναπαράστασης γνώσης που βασίζεται στη δομή των αναφορών έργων και στα ερωτήματά τους, η ύπαρξη ιδιαίτερα υψηλής ή χαμηλής βαθμολογίας ακόμη και σε μέρος των ερωτήσεων μας εξυπηρετεί τους σκοπούς της εργασίας μας.

Το δεύτερο κριτήριο σχετίζεται με τον τρόπο συγγραφής του ελεύθερου κειμένου από τους εμπειρογνώμονες. Σε πολλές περιπτώσεις, η ανάλυση των απαντήσεων περιείχε ασάφειες ως προς τη διατύπωση, επομένως δεν θα μπορούσαν να εξαχθούν ασφαλή συμπεράσματα σχετικά με τη γνώμη του ειδικού στο θέμα. Επιπλέον, πολλές φορές η βαθμολογία δεν συνοδευόταν από την αιτιολόγησή της, με αποτέλεσμα τέτοιου είδους αναφορές να μην μπορούν να χρησιμοποιηθούν καθόλου στη μελέτη μας.

Ιδιαίτερα σημαντικό ρόλο επίσης έπαιξε και η ύπαρξη ελεύθερου κειμένου στο τμήμα της αναφοράς που αναφέρεται στα διδάγματα. Η εξαγωγή των διδαγμάτων από τις αναφορές έγινε είτε με εντοπισμό τους στο σημείο της αναφοράς που προορίζεται για διδάγματα είτε συμπερασματικά μέσα από το ευρύτερο ελεύθερο κείμενο του ειδικού. Πολλές φορές, παρόλο που ο ειδικός αιτιολογεί τη βαθμολογία του παραθέτοντας μια σειρά παραγόντων που επηρεάζουν το έργο, δεν απαντά σχετικά με τα διδάγματα στον ειδικό χώρο της αναφοράς που προορίζεται για αυτό τον λόγο. Σε αυτές τις περιπτώσεις, αναζητήθηκαν διδάγματα στο σώμα της απάντησης της ερώτησης που υπονοούνταν ή δηλώνονταν σαφώς στο κομμάτι αυτό και υπήρξε προσπάθεια εξαγωγής συμπερασμάτων από το ελεύθερο κείμενο του ειδικού. Ωστόσο, αξίζει να σημειωθεί πως η παρουσία διδαγμάτων είναι σχετικά σπάνια, συγκρινόμενη με το πλήθος των παραγόντων που επηρεάζουν τη βαθμολόγηση ενός έργου. Αυτό συνδέεται με το γεγονός πως περισσότεροι από ένας παράγοντες μπορεί να αντιστοιχούν σε ένα δίδαγμα, ωστόσο ενδέχεται να υποδεικνύει και την αμέλεια των ειδικών να συμπεριλαμβάνουν συχνά προτεινόμενες διορθωτικές κινήσεις στις αναφορές τους.

Κατά τον εντοπισμό τόσο των παραγόντων όσο και των διδαγμάτων ιδιαίτερη προσοχή δόθηκε στο επίπεδο γενικότητας που μας ενδιέφερε για τους σκοπούς του συστήματός μας. Αυτό σημαίνει πως παράγοντες που αφορούσαν στην πορεία ενός συγκεκριμένου έργου δεν μας ενδιαφέρουν αν δεν μπορούν να γενικευτούν ώστε να βρίσκουν εφαρμογή σε περισσότερα έργα. Για παράδειγμα, στην αναφορά του έργου VRBAS, Joint Tourism Development of the Vrbas Valley, που ασχολείται με την τουριστική ανάπτυξη σε περιοχή των Βαλκανίων, αναφέρεται ως προς τις επιπτώσεις του έργου (ερώτηση 4.1) (Elma Balic, 2009): «The government's commitment to support the tourism

sector creates the foundation for direct impact prospects». Η κίνηση, επομένως, της κυβέρνησης για στήριξη στον τουριστικό τομέα δημιουργεί καλές προοπτικές άμεσου αντίκτυπου για το έργο. Ωστόσο, αυτό συμβαίνει προφανώς επειδή το έργο αφορά στον τομέα του τουρισμού. Δεν θα μπορούσε, επομένως, ο παράγοντας να γενικευτεί και να εξαχθεί το συμπέρασμα πως έργα που συμβαίνουν σε χώρες των οποίων η κυβέρνηση υποστηρίζει τον τουρισμό έχουν ευνοϊκές συνθήκες δημιουργίας άμεσου αντίκτυπου, γιατί προφανώς κάτι τέτοιο θα ήταν αβάσιμο.

Προφανώς, δεδομένου πως οι ειδικοί αξιολογούν το έργο και με βάση τον χαρακτήρα του, οι περισσότεροι από τους παράγοντες που εντοπίστηκαν δεν είναι γενικοί, στη μορφή με την οποία βρίσκονται στο κείμενο. Πολύ συχνά ωστόσο η γενίκευση ήταν εφικτή μετά από κατάλληλη «διαμόρφωση» και «φιλτράρισμα» του κειμένου του εμπειρογνώμονα. Για παράδειγμα, στην αναφορά του έργου «Municipal Infrastructure Support Programme (MISP)», (Sinisa Bralic, 2010) στην ερώτηση 1.1 αναφέρεται: «The ongoing CARDS Municipal Infrastructure Support Programme (MISP) project is fully in line with both national and EC main development strategies in the concerned field». Προφανώς, εάν παραληφθεί η ονομασία του έργου, το γεγονός πως το έργο είναι σε συμφωνία με τις εθνικές και ευρωπαϊκές στρατηγικές του τομέα, είναι ένας γενικός παράγοντας που επηρεάζει τη βαθμολογία του έργου και άρα μπορεί να συμπεριληφθεί στο σύστημά μας.

4.3.6 Κατάταξη των εξορυγμένων παραγόντων και διδαγμάτων στο σύστημα αναπαράστασης γνώσης

Μετά τον εντοπισμό των παραγόντων και διδαγμάτων στο σύνολο των έργων που μελετήθηκαν, ή και πολλές φορές ταυτόχρονα με αυτό, ακολούθησε η κατάταξη των εξορυγμένων στοιχείων στο σύστημα αναπαράστασης γνώσης.

Παράγοντες που περιλαμβάνουν στοιχεία από περισσότερες από μία κατηγορίες καταχωρούνται σε όλες τις κατηγορίες στις οποίες αναφέρονται, το ίδιο και τα διδάγματα. Για παράδειγμα, στο έργο «NEWADA - Network of Danube Waterway Administrations», στην ερώτηση 3.2 αναφέρεται: «there were no negative impacts on target groups». Επομένως το στοιχείο αυτό θα καταχωρηθεί στην κατηγορία Avoidance of negative Impacts (M42iv). Στην ίδια ερώτηση όμως, αναφέρεται: «it is likely that the PP will be achieved during the timeframe of the project». Ο παράγοντας αυτός θα πρέπει να καταχωρηθεί τόσο στην κατηγορία Likelihood of PP to be achieved (M23i) όσο και στο Use of Timescale (M22ii).

Τα βάρη των εκφράσεων

Είναι κατανοητό πως οι παράγοντες και τα διδάγματα που εντοπίζονται στις αναφορές του συστήματος ROM, εμπεριέχουν διαβαθμίσεις στις έννοιες τους. Για παράδειγμα, οι εκφράσεις «it is a good response» και «it is a very good response», αποτελούν μια απλή περίπτωση εκφράσεων με διαφορετικό «βαθμό δύναμης» (degree of expressions). Καθίσταται επομένως σαφής η ανάγκη για ποσοτικοποίηση τέτοιου είδους διαφοροποιήσεων, με τρόπο που να προσδίδει σε κάθε έκφραση το «βάρος» που τις αρμόζει. Κάτι τέτοιο εκπίπτει στον ερευνητικό τομέα της ανάλυσης συναισθημάτων (Sentiment Analysis), έναν αναπτυσσόμενο επιστημονικό κλάδο που ασχολείται με την ανάλυση εκφράσεων και της «δύναμής» τους και βρίσκει εφαρμογές σε πολλές περιπτώσεις, όπως σε εταιρείες που ενδιαφέρονται για την άποψη των πελατών για τα

προϊόντα τους, σε πολιτικούς που αναλύουν τον τρόπο που τους παρουσιάζουν τα μέσα κ.ο.κ.

Το ιδιαίτερο στοιχείο των αναφορών του συστήματος ROM είναι πως περιέχουν υποκειμενικές εκφράσεις. Για τον καθορισμό της «έντασης» των εκφράσεων αυτών, η ανάλυση επικεντρώνεται σε συγκεκριμένα μέρη του λόγου, όπως επίθετα, ρήματα και ουσιαστικά. Οι πιο πρόσφατες προσεγγίσεις συμπεριλαμβάνουν και τη χρήση επιρρημάτων, ενώ ακόμη περισσότερο ακριβή αποτελέσματα μπορούν να προκύψουν από την ανάλυση των συνδυασμών επιρρημάτων-επιθέτων (Adverb-Adjective Combination, AAC), τεχνική η οποία κρίθηκε και καταλληλότερη για να εφαρμοστεί στη συγκεκριμένη μεθοδολογία. Πιο συγκεκριμένα, βασιστήκαμε στην ανάπτυξη μεθοδολογίας για ανάλυση συναισθημάτων που χρησιμοποιήθηκε στο σύστημα OASYS, online σύστημα που παρουσιάζει τις απόψεις αναγνωστών πάνω σε κάποια κείμενα (<http://oasys.umiacs.umd.edu/oasys/>) (Farah Benamara, 2006). Ακολουθήθηκαν τόσο τα βήματα της μεθοδολογίας και υιοθετήθηκαν τα καθορισμένα αξιώματα, έχοντας ωστόσο κάποιες διαφοροποιήσεις στην ανάθεση τιμών-δεικτών σε επιρρήματα και επίθετα, για την οποία βασιστήκαμε και στην κατηγοριοποίηση και ανάθεση τιμών κατά Chklovski (Chklovski, 2006) και στο σύστημα SentiWordNet 3.0.

Για την ανάλυση των επιθετικών – επιρρηματικών εκφράσεων, αρχικά εξετάζεται η «ένταση» των επιρρημάτων σε γλωσσολογικό επίπεδο και καθορίζονται τα αξιώματα σύμφωνα με τα οποία τους ανατίθεται η βαθμολογία σε κλίμακα από 0 έως 1. Στην ανάλυση μας λαμβάνουμε υπόψη τα επιρρήματα που υποδεικνύουν την ένταση με την οποία κάποιο γεγονός συμβαίνει (very, extremely, absolutely, hardly, precisely, really). Η βαθμολογία 1 σημαίνει πως το επίρρημα επιβεβαιώνει απολύτως το επίθετο, ενώ η βαθμολογία 0 σημαίνει πως το επίρρημα δεν επηρεάζει καθόλου το επίθετο. Τα επιρρήματα κατηγοριοποιήθηκαν ως εξής:

1. Επιρρήματα επιβεβαίωσης - AFF: περιλαμβάνονται επίθετα όπως *absolutely, certainly, exactly, totally*
2. Επιρρήματα αμφιβολίας - DOU: περιλαμβάνονται επίθετα όπως *possibly, roughly, apparently, seemingly*
3. Επιρρήματα έντονης εντατικοποίησης - STR: περιλαμβάνονται επιρρήματα όπως *astronomically, exceedingly, extremely, immensely*
4. Επιρρήματα μερικής εντατικοποίησης - WEA: περιλαμβάνονται επιρρήματα όπως *barely, scarcely, weakly, slightly*
5. Επιρρήματα άρνησης και ελαχιστοποίησης - MIN: περιλαμβάνονται επιρρήματα όπως *hardly*.

Για την ανάθεση βαθμολογιών στα επιρρήματα που ανήκουν σε καθεμία από τις παραπάνω κατηγορίες, εκτός από την κατηγορία 5, ορίστηκαν τα εξής αξιώματα:

A1. Η βαθμολογία κάθε επιρρηματος μερικής εντατικοποίησης και κάθε επιρρηματος αμφιβολίας είναι μικρότερη ή ίση από τη βαθμολογία επιρρηματος έντονης εντατικοποίησης.

A2. Η βαθμολογία κάθε επιρρηματος μερικής εντατικοποίησης και κάθε επιρρηματος αμφιβολίας είναι μικρότερη ή ίση από τη βαθμολογία επιρρηματος επιβεβαίωσης.

Το πρώτο αξίωμα προκύπτει λογικά καθώς, για παράδειγμα, η πρόταση «The financial sustainability is slightly secured» περιγράφει ένα λιγότερο έντονο γεγονός από την πρόταση «The financial sustainability is highly secured». Ομοίως, το δεύτερο αξίωμα είναι επίσης λογικό, καθώς για παράδειγμα η πρόταση «The financial sustainability is slightly secured» περιγράφει ένα λιγότερο έντονο γεγονός από την πρόταση «The financial sustainability is absolutely secured». Αξίωμα για τη σύγκριση βαθμολογιών των επιρρημάτων των κατηγοριών 2 και 4 δεν ορίστηκε, καθώς η μεταξύ τους σχέση δεν είναι πάντα ίδια.

Η κατηγορία των επιρρημάτων άρνησης ή ελαχιστοποίησης περιλαμβάνει επιρρήματα όπως το «hardly», που ουσιαστικά ανατρέπουν τη βαθμολογία του επιθέτου. Για παράδειγμα, στην πρόταση «The implementation was hardly good», το «hardly» ελαχιστοποιεί την θετική βαθμολογία της πρότασης «The implementation was good». Αντίστοιχα, η χρήση του επιρρήματος «hardly» στην έκφραση «hardly bad», αυξάνει την βαθμολογία της πρότασης, καθώς το «bad» είναι επίθετο αρνητικής σημασίας.

Βάσει αυτών των αρχών, δόθηκαν σε κάθε κατηγορία επιρρημάτων βαθμολογίες από 0 έως 1. Για τις βαθμολογίες αυτές, χρησιμοποιήθηκε η βάση SentiWordNet 3.0 (<http://sentiwordnet.isti.cnr.it/>), που πρόκειται για μία λεξιλογική βάση, σχεδιασμένη ειδικά για την υποστήριξη κατηγοριοποίησης συναισθημάτων και εφαρμογών εξαγωγής γνώμης (opinion mining applications) (Stefano Baccianella, 2008). Η βάση στηρίζεται στην αυτόματη επισήμειωση (annotating) των περιεχομένων του WordNet, ανάλογα με τη θετικότητα, αρνητικότητα ή ουδετερότητά τους. Μερικά αντιπροσωπευτικά παραδείγματα βαθμολογιών χαρακτηριστικών επιρρημάτων κάθε κατηγορίας, σύμφωνα με το SentiWordNet 3.0, βρίσκονται στον παρακάτω πίνακα.

Κατηγορία Επιρρημάτων	Βαθμολογία (sc(adv))	Χαρακτηριστικό Επίρρημα
Επιρρήματα Επιβεβαίωσης	0.75	Certainly
Επιρρήματα Αμφιβολίας	0.25	Likely
Επιρρήματα Έντονης Εντατικοποίησης	0.875	Exceedingly
Επιρρήματα Μερικής Εντατικοποίησης	0.125	Barely

Πίνακας : Παραδείγματα βαθμολογιών κατηγοριών επιρρημάτων (sc(adv))

Για τη χρήση του SentiWordNet 3.0, λαμβανόταν κάθε φορά υπόψη η βαθμολογία του επιρρήματος ανάλογα με τη χρήση του στην πρόταση. Για παράδειγμα, η αναζήτηση του επιρρήματος «good» δίνει 27 αποτελέσματα, ανάλογα με το αν χρησιμοποιείται σαν επίθετο, ουσιαστικό ή επίρρημα και ανάλογα με το περιεχόμενο της πρότασης. Στην έκφραση «the plan was pretty good organized», για παράδειγμα, επιλέχθηκε η βαθμολογία 0.625 για το «good», καθώς χρησιμοποιείται ως επίρρημα και έχει την έννοια «in a good or proper or satisfactory manner or to a high standard». Σε κάποια άλλη έκφραση ωστόσο, όπως «good performance», το «good» λειτουργεί σαν επίθετο και θα πρέπει να έχει τη βαθμολογία +0.875, καθώς έχει την έννοια «having desirable or positive qualities especially those suitable for a thing specified». Κατά την εφαρμογή του προγράμματος θα δειχθεί πως η διαδικασία αυτή δίνει λογικά αποτελέσματα τιμών.

The screenshot shows the SentiWordNet 3.0 interface. At the top, there is a search bar with the word 'successful' entered and a 'Search!' button. Below the search bar, the word 'successful' is displayed in large red letters, followed by 'ADJECTIVE' in red. A triangular sentiment plot is shown on the left, with a yellow dot indicating the word's position. To the right of the plot, the word 'successful#1' is listed with its ID '02331262'. Below this, the definition is provided: 'having succeeded or being marked by a favorable outcome; "a successful architect"; "a successful business venture"'. At the bottom left of the plot area, the sentiment scores are shown: 'P: 0.625 O: 0.375 N: 0'. A 'Feedback!' button is located at the bottom right of the plot area. At the very bottom of the page, the copyright '2010 © SentiWordNet' and a link 'Index :: Contact us' are visible.

Εικόνα : Παράδειγμα αναζήτησης βαθμολογίας του επιθέτου successful στο SentiWordNet 3.0

Σε συνδυασμό με τις βαθμολογίες των επιρρημάτων που αναφέρθηκαν, υπάρχουν και οι βαθμολογίες των επιθέτων των εκφράσεων, οι οποίες κυμαίνονται σε μία κλίμακα από -1 έως 1. Χρησιμοποιώντας αυτές τις δύο κλίμακες ([0..1] για τα επιρρήματα και [-1..1] για τα επίθετα), είναι δυνατή η ποσοτική αναπαράσταση των εννοιών. Με εξαίρεση την κατηγορία 5 των επιρρημάτων, όλα τα επιρρήματα ενδυναμώνουν την έννοια του επιθέτου, το θέμα είναι σε ποιο βαθμό. Η βαθμολογία 0 έως 1 αντιπροσωπεύει αυτή την ενδυνάμωση. Αντίθετα, στα επίθετα η βαθμολογία κυμαίνεται από -1 έως 1, επειδή μπορούν να είναι είτε θετικά είτε αρνητικά. Στην περίπτωση των επιθέτων, θα ήταν χρήσιμη η ανάλυση των περισσότερο συχνά εμφανιζόμενων επιθέτων, και η αξιολόγηση της έντασής τους μέσω ερωτηματολογίων, δεδομένου πως ένα επίθετο μπορεί να είναι περισσότερο θετικό από ένα άλλο, που επίσης έχει θετική έννοια. Για τους σκοπούς της παρούσας διπλωματικής, κρίθηκε πως επαρκεί η βαθμολογία που δίνεται στα επίθετα μέσω της βάσης του SentiWordNet 3.0, κατά απόλυτη τιμή, και στη συνέχεια η χρήση του κατάλληλου πρόσημου, ανάλογα με το αν πρόκειται για επίθετο με θετική ή αρνητική έννοια. Έτσι, για τα θετικά επίθετα λαμβάνεται υπόψη η βαθμολογία θετικότητας του SentiWordNet 3.0 (P) με θετικό πρόσημο, ενώ για τα αρνητικά επίθετα λαμβάνεται υπόψη η βαθμολογία αρνητικότητας του SentiWordNet 3.0 (N) με αρνητικό πρόσημο. Παραδείγματα βαθμολογιών επιθέτων μέσω του SentiWordNet 3.0 δίνονται στον ακόλουθο πίνακα.

Επίθετο	Βαθμολογία SentiWordNet 3.0	sc(adj)
good	0.875	+0.875
appropriate	1	+1
coherent	0.75	+0.75
bad	0.625	-0.625

Πίνακας - Παραδείγματα Βαθμολογιών Επιθέτων (sc(adj))

Στο αγγλικό κείμενο, μια επιρρηματική-επιθετική έκφραση έχει τη μορφή:

<επίρρημα><επίθετο>

ή <επίρρημα_i, επίρρημα_j><επίθετο>

όπου το επίρρημα_i μπορεί να είναι επίρρημα αμφιβολίας ή έντονης εντατικοποίησης ενώ το επίρρημα_j μπορεί να είναι επίρρημα έντονης ή μερικής εντατικοποίησης, για παράδειγμα *very very good, possibly less productive* κ.τ.λ. Άλλοι συνδυασμοί δεν συναντώνται συχνά, όπως επίσης ούτε περιπτώσεις τριπλών επιρρημάτων. Για την ανάθεση βαθμολογιών στις εκφράσεις ($f(adv, adj)$), ορίστηκαν επίσης κάποια αξιώματα (Farah Benamara, 2006):

1. Επιρρήματα επιβεβαίωσης και έντονης εντατικοποίησης:

- AAC-1: Εάν $sc(adj) > 0$ και $adv \in [AFF \cup STR] \Rightarrow f(adv, adj) \geq sc(adj)$.
- AAC-2: Εάν $sc(adj) < 0$ και $adv \in [AFF \cup STR] \Rightarrow f(adv, adj) \leq sc(adj)$.

2. Επιρρήματα μερικής εντατικοποίησης:

- AAC-3: Εάν $sc(adj) > 0$ και $adv \in WEA \Rightarrow f(adv, adj) \leq sc(adj)$.
- AAC-4: Εάν $sc(adj) < 0$ και $adv \in WEA \Rightarrow f(adv, adj) \geq sc(adj)$.

3. Επιρρήματα ελαχιστοποίησης:

- AAC-5: Εάν $sc(adj) > 0$ και $adv \in MIN \Rightarrow f(adv, adj) \leq sc(adj)$.
- AAC-6: Εάν $sc(adj) < 0$ και $adv \in MIN \Rightarrow f(adv, adj) \geq sc(adj)$.

Στα αξιώματα AAC-1 και AAC-2, η ερμηνεία είναι πως τα επιρρήματα επιβεβαίωσης ή έντονης εντατικοποίησης, ενδυναμώνουν τη θετικότητα των θετικών επιθέτων και την αρνητικότητα των αρνητικών. Έτσι, το *very* ενδυναμώνει τη θετικότητα του *good*, κάνοντας τη βαθμολογία του *very good* μεγαλύτερη από τη βαθμολογία του *good* και ομοίως ενδυναμώνει την αρνητικότητα του *bad*, κάνοντας τη βαθμολογία του *very bad* να είναι χαμηλότερη από τη βαθμολογία του *bad*. Σύμφωνα με το αξίωμα AAC-3, ένα θετικό επίθετο θα καταλήξει με χαμηλότερη βαθμολογία αν εντοπιστεί με επίρρημα μερικής εντατικοποίησης. Για παράδειγμα το *barely good*, έχει χαμηλότερη βαθμολογία από το *good*. Το αξίωμα AAC-4 λέει πως ένα αρνητικό επίθετο έχει μεγαλύτερη ένταση όταν χρησιμοποιείται με επίρρημα μερικής εντατικοποίησης καθώς το *barely bad* έχει θετικότερη έννοια από το *bad*. Τέλος, τα αξιώματα AAC-5 και AAC-6 δείχνουν πως τα επιρρήματα ελαχιστοποίησης αντιστρέφουν την πολικότητα του επιθέτου. Επίσης, η πολικότητα της βαθμολογίας αντιστρέφεται σε περίπτωση που στην έκφραση υπάρχει άρνηση. Για παράδειγμα, στην έκφραση «*it was not really bad*», υπολογίζεται η βαθμολογία του «*really bad*» (-0.64) και στη συνέχεια δίνεται στην έκφραση η αντίθετη βαθμολογία (+0.64).

Σύμφωνα με όλα τα παραπάνω, οι βαθμολογίες των εκφράσεων διαμορφώνονται βάσει της επόμενης μεθόδου:

- Εάν $adv \in [AFF \cup STR]$, τότε:

$$f(adv, adj) = sc(adj) + (1 - sc(adj)) \cdot sc(adv), \text{ εάν } sc(adj) > 0 \text{ (1) και}$$

$f_{adv, adj} = sc_{adj} - (1 - sc_{adj}) \cdot sc(adv)$, εάν $sc_{adj} < 0$ (2).

- Εάν $adv \in [WEA \cup DOU]$, τότε:

$f_{adv, adj} = sc_{adj} - (1 - sc_{adj}) \cdot sc(adv)$, εάν $sc_{adj} > 0$ (3) και

$f_{adv, adj} = sc_{adj} + (1 - sc_{adj}) \cdot sc(adv)$, εάν $sc_{adj} < 0$ (4).

Έτσι για παράδειγμα, η έκφραση «certainly organized» θα έχει βαθμολογία:

$f_{certainly, organised} = 1 + 1 - 1 \cdot 0.75 = 1$ ενώ η έκφραση «probably successful» θα έχει

βαθμολογία: $f_{probably, successful} = 0.625 - 1 - 0.625 \cdot 1 = 0.25$.

Αξίζει να αναφερθεί πως οι όροι «επίθετο» και «επίρρημα» χρησιμοποιήθηκαν με την ευρεία έννοια του χαρακτηριστικού και του προσδιορισμού, αντίστοιχα. Για παράδειγμα, στην πρόταση «The team showed considerable capacity», η λέξη «capacity» συμμετείχε στον καθορισμό του βάρους της πρότασης ως επίθετο και η λέξη «considerable» ως το επίρρημα που το διαμορφώνει, καθώς είναι το χαρακτηριστικό που

προσδιορίζει την έννοια. Έτσι, η βαθμολογία της πρότασης προκύπτει: $f_{considerable}$,

$capacity = 0.875 + 1 - 0.875 \cdot 1 = 1$, ενισχύει δηλαδή την θετικότητα τις «ικανότητας» της

ομάδας. Αντίθετα, για την πρόταση «The team showed little capacity», η βαθμολογία θα

είναι: $f_{\text{little, capacity}} = 0.875 - 1 - 0.875 \cdot 0.375 = 0.828$. Επομένως, με τη χρήση του

SentiWordNet 3.0 μπορέσαμε να γενικεύσουμε τη μεθοδολογία του συστήματος OASYS και για εκφράσεις επιθέτου – ουσιαστικού, χρησιμοποιώντας τα ίδια αξιώματα και κανόνες.

ΚΕΦΑΛΑΙΟ 5^ο: ΕΦΑΡΜΟΓΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

Κατά την εφαρμογή του Συστήματος Αναπαράστασης Γνώσης, μελετήθηκαν αναλυτικά οι αναφορές 4 έργων, που ανακτήθηκαν μέσω του συστήματος KOMIS. Εντοπίστηκαν συνολικά 63 παράγοντες που επηρέασαν την αξιολόγηση των έργων αυτών και 8 σχετικά διδάγματα που προέκυψαν από την επίβλεψή τους. Τα έργα που επιλέχθηκαν είχαν ιδιαίτερα καλές γενικές βαθμολογίες (α) ή βαθμολογία α σε κάποια από τα επιμέρους κριτήρια αξιολόγησης. Επίσης μελετήθηκε και η περίπτωση ερωτήσεων με ιδιαίτερα αρνητική βαθμολόγηση, από όπου και προέκυψαν τα περισσότερα από τα στοιχεία των διδαγμάτων.

Τα στοιχεία που παρουσιάζονται στο σύστημα σχετικά με τα δεδομένα που προέκυψαν είναι:

1. Παράγοντας που επηρέασε τη βαθμολόγηση του έργου
2. Κατηγορία – κριτήριο της αξιολόγησης όπου αναφέρεται ο παράγοντας
3. Αύξοντας αριθμός αναφοράς όπου εντοπίστηκε (BCSiD)
4. Αριθμός ερώτησης στην απάντηση της οποίας περιλαμβάνεται
5. Βάρος παράγοντα, όπως υπολογίστηκε στην ενότητα 4.3.6
6. Εάν υπάρχει, σχετικό δίδαγμα

Με βάση όλα τα παραπάνω, προκύπτει ο πίνακας παραγόντων – διδαγμάτων που παρουσιάζεται στο Παράρτημα Α1, ο οποίος περιέχει όλα τα στοιχεία που θα πρέπει να καταχωρηθούν στη βάση δεδομένων κατά την υλοποίηση του προγράμματος. Η πρώτη στήλη περιλαμβάνει τους εντοπισμένους παράγοντες. Στη δεύτερη στήλη αναφέρονται οι κωδικοί των κατηγοριών στις οποίες κατατάσσονται οι παράγοντες (βλ. Πίνακας 3). Για να γίνει περισσότερο σαφής η κατάταξη των εντοπισμένων παραγόντων και διδαγμάτων σε κατηγορίες, στο Παράρτημα Α2 παρατίθεται συγκεντρωτικός πίνακας των όρων που χρησιμοποιούνται στο σύστημα ROM και των ερμηνειών τους. Οι στήλες 3 και 4 του πίνακα του παραρτήματος Α1 δίνουν τον αύξοντα αριθμό της αναφοράς BCS του συστήματος KOMIS, στην οποία εντοπίστηκε ο παράγοντας και τον αριθμό της ερώτησης της αναφοράς, αντίστοιχα. Τα στοιχεία αυτά χρησιμεύουν στον εύκολο εντοπισμό των παραγόντων, έτσι ώστε αργότερα να μπορούν να χρησιμοποιηθούν τα ίδια κείμενα ως είσοδοι για την εξόρυξη από αυτόματο σύστημα. Ο αριθμός της ερώτησης, επιπλέον, υποδεικνύει το βάρος της ερώτησης σύμφωνα με το σύστημα ROM, γι αυτό και δεν υπάρχει ξεχωριστή στήλη που να περιέχει την τιμή αυτή. Το βάρος της ερώτησης θα χρησιμεύσει στη μελλοντική εφαρμογή του συστήματος με χρήση αυτοματοποιημένων τεχνικών.

Στο σημείο αυτό ελέγχθηκε, επίσης, η περίπτωση ύπαρξης του ίδιου παράγοντα σε παραπάνω από μία απαντήσεις. Για παράδειγμα, ο παράγοντας «The project design can lead to a positive impact» θα μπορούσε να περιέχεται τόσο στις ερωτήσεις που αφορούν στο σχεδιασμό του έργου (π.χ. 1.1, 1.2) όσο και στις ερωτήσεις σχετικά με την επίδραση του έργου (π.χ. 3.1). Σε αυτή την περίπτωση, καταχωρούνται και οι δύο αριθμοί των ερωτήσεων όπου εντοπίζεται ο παράγοντας.

Επιπλέον, αξίζει να σημειωθεί πως η καταχώρηση μπορεί να γίνει σε οποιοδήποτε από τα επίπεδα ιεράρχησης των κατηγοριών. Έτσι, για παράδειγμα, ο παράγοντας «The LF contains suitable OVIs», θα πρέπει να καταχωρηθεί στο τελευταίο επίπεδο λεπτομέρειας, M12iia (SMART OVIs) καθώς δίνει την πληροφορία της φύσης των δεικτών που χρησιμοποιούνται. Αντίθετα, ο παράγοντας «The project design can lead to a positive impact», θα καταχωρηθεί στη γενική κατηγορία M3 (Impact), καθώς στην έννοια εμπεριέχονται όλες οι υποκατηγορίες που την αποτελούν. Με άλλα λόγια, αν ένας παράγοντας μπορεί να καταχωρηθεί σε όλες τις υποκατηγορίες κάποιας κατηγορίας, καταχωρείται απευθείας στην κατηγορία. Παράγοντες που μπορούν να καταχωρηθούν σε παραπάνω από μία υποκατηγορίες, καταχωρούνται σε όλες.

Τα βάρη (στήλη 5) που ανατίθενται σε κάθε παράγοντα, υπολογίζονται μέσω του συστήματος SentiWord 3.0. Παράγοντες που δεν έχουν επιθετικές – επιρρηματικές εκφράσεις, παίρνουν απευθείας τη βαθμολογία που δίνει το σύστημα. Η βαθμολογία παραγόντων που εμπεριέχουν επιθετικές – επιρρηματικές εκφράσεις υπολογίζεται με βάση τις σχέσεις (1), (2), (3), (4) (Κεφ.4.3.6). Στο σημείο αυτό, τονίζεται και πάλι η σημαντικότητα της χρήσης του κατάλληλου όρου του SentiWordNet 3.0. Για παράδειγμα, ο παράγοντας «meet socioeconomic results» έχει μια βαθμολογία που θα καθοριστεί από την έκφραση «meet results». Αναζητώντας στο SentiWordNet τη βαθμολογία της λέξης «meet», εμφανίζονται 13 αποτελέσματα για τη χρήση του ως ρήμα. Όταν χρησιμοποιείται με την έννοια της «συνάντησης» πρόκειται για ουδέτερο ρήμα με βαθμολογία 0. Κάτι τέτοιο όμως δεν εκφράζει την έννοια που εντοπίσαμε. Επιλέγεται η βαθμολογία 0.625, καθώς είναι εκείνη που αντιστοιχεί στο «meet» όταν χρησιμοποιείται με την έννοια «satisfy» ή «fulfill» (π.χ. «meet a need»).

Στο σημείο αυτό, θα πρέπει να σημειωθεί πως ορισμένοι από τους παράγοντες δεν εμπεριείχαν στοιχείο υποκειμενικότητας. Για παράδειγμα, ο παράγοντας «the project activities are planned to be expanded» δεν περιέχει κάποιο στοιχείο διαμόρφωσης της γνώμης, και απλά παραθέτει κάποιο γεγονός. Σε αυτές τις περιπτώσεις δίνεται η βαθμολογία +1 ή -1 στον παράγοντα, ανάλογα με το αν πρόκειται για παρουσία ή απουσία κάποιου γεγονότος. Επίσης, σε παράγοντες που δεν περιέχουν επιθετικό – επιρρηματικό προσδιορισμό, αλλά περιλαμβάνουν κάποιο υποκειμενικό επίθετο, ρήμα ή επίρρημα, η επιλογή του οποίου υπονοεί τη γνώμη του ειδικού, ανατίθεται η βαθμολογία του υποκειμενικού στοιχείου της πρότασης. Για παράδειγμα, στον παράγοντα «The wide-based and long-term support basis is encouraging for future sustainability» ανατίθεται η βαθμολογία του θετικού ρήματος «encouraging» (0.625).

Η στήλη των σχετικών διδαγμάτων περιλαμβάνει τα διδάγματα που εντοπίστηκαν στις αναφορές που μελετήθηκαν. Αυτά μπορεί να αναφέρονταν ξεκάθαρα στο σημείο της αναφοράς που προορίζεται για τα διδάγματα, ή να βρίσκονται στο ελεύθερο κείμενο των απαντήσεων. Διδάγματα που αναφέρονται σε παραπάνω από μία έννοιες - παράγοντες, καταχωρούνται σε κάθε έννοια στην οποία αναφέρονται.

Σύμφωνα με την εφαρμογή όλων των παραπάνω, δημιουργήθηκε ο Πίνακας του Παραρτήματος Α1.

ΚΕΦΑΛΑΙΟ 6^ο: ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΕΣ ΕΝΕΡΓΕΙΕΣ

6.1 Συμπεράσματα από την εφαρμογή του συστήματος

Σχεδιάστηκε ένα σύστημα αναπαράστασης της γνώσης που περιλαμβάνεται σε αναφορές αξιολόγησης έργων κατά το σύστημα ROM. Κατά την εφαρμογή του συστήματος προέκυψε ένα πλήθος συμπερασμάτων που αφορούν στο σχεδιασμό του και στην αποτελεσματικότητά του. Έχοντας ως δείγμα ένα σύνολο αναφορών αξιολόγησης του συστήματος ROM, στις οποίες εφαρμόστηκε η διαδικασία εντοπισμού παραγόντων και διδαγμάτων, έγινε δυνατή η αξιολόγηση του σχεδιασμού του συστήματος και ο εντοπισμός στοιχείων που μπορούν να εξελιχθούν, διορθωθούν.

Αρχικά, το βασικό στοιχείο που παρατηρήθηκε είναι πως ο διαχωρισμός των κατηγοριών που επιλέχθηκαν και χρησιμοποιήθηκαν, είναι επαρκής και κατάλληλος για τις ανάγκες του συστήματος. Δεδομένου ότι όλοι οι παράγοντες που εντοπίστηκαν μπορούσαν να καταταχθούν σε κάποια κατηγορία με ικανοποιητική ακρίβεια, δεν απαιτείται η δημιουργία νέων κατηγοριών. Αυτό επιβεβαιώνεται και από το γεγονός πως παράγοντες που ανήκουν στην ίδια κατηγορία έχουν κοινά στοιχεία και βρίσκονται στο ίδιο επίπεδο γενικότητας. Επιπλέον, όλες οι κατηγορίες χρειάστηκε να χρησιμοποιηθούν, γεγονός που αποδεικνύει πως δεν υπάρχει περιττή κατηγοριοποίηση, και το επίπεδο λεπτομέρειας των κατηγοριών ήταν το κατάλληλο καθώς καθιστούσε εύκολη και σαφή την επιλογή κατηγορίας για κάθε παράγοντα που εντοπιζόταν.

Αξίζει να σημειωθεί πως, κατά τη μελέτη των έργων, πολλοί από τους παράγοντες που επηρεάζουν τη βαθμολογία τους εντοπιζόνταν παραπάνω από μία φορές. Έργα που αφορούν σε πολύ διαφορετικούς επιστημονικούς τομείς, μπορεί να είχαν τις ίδιες αδυναμίες σε επίπεδο οργάνωσης ή διαχείρισης. Το γεγονός αυτό επιβεβαιώνει την αρχική παραδοχή επάρκειας των κοινών κατηγοριών για όλα τα έργα, ανεξάρτητα από τη φύση τους. Όπως αναμενόταν, πολλοί από τους λόγους που επηρεάζουν τη βαθμολογία ενός έργου δεν σχετίζονται με το είδος του έργου, αλλά με τις διαδικασίες διαχείρισής του, και έτσι παράγοντες και διδάγματα που εμφανίζονται, για παράδειγμα, σε έργα ύδρευσης, μπορεί να εντοπιστούν και σε έργα εκσυγχρονισμού κάποιας περιοχής. Επομένως, η κατηγοριοποίηση με βάση τα κριτήρια αξιολόγησης των έργων και όχι με βάση τη φύση τους, ήταν επιτυχής.

Σχετικά με τη μεθοδολογία ανάθεσης βαρών στις εκφράσεις των παραγόντων, παρατηρήθηκε πως τα αριθμητικά αποτελέσματα που προέκυπταν μέσα από αυτή ήταν αντιπροσωπευτικά. Με άλλα λόγια, μέσω της χρήσης της συγκεκριμένης μεθοδολογίας και του SentiWordNet 3.0 έγινε δυνατή η διαφοροποίηση ανάμεσα σε έννοιες με έντονο χαρακτήρα και με περισσότερο ή λιγότερο έντονο. Τα αριθμητικά αποτελέσματα που προέκυψαν, τόσο για τους συνδυασμούς επιρρήματος – επιθέτου, όσο και για τους συνδυασμούς επιθέτου – ουσιαστικού, έχουν λογική βάση και μπορούν να θεωρηθούν ενδεικτικά της άποψης των εμπειρογνομόνων σχετικά με τις ενέργειες που περιγράφουν.

Τέλος, αξίζει να αναφερθεί η χρησιμότητα του συστήματος αναπαράστασης γνώσης που διαφάνηκε από την εφαρμογή του σε αναφορών των έργων. Γνώση που συμπυκνωνόταν σε ελεύθερο κείμενο, μπόρεσε να κωδικοποιηθεί και να κατηγοριοποιηθεί, δίνοντας έτσι τη δυνατότητα εύκολης ανάκτησής της. Επιπλέον, η εξαγωγή συμπερασμάτων

διευκολύνθηκε καθώς πλέον δεν είναι απαραίτητη η μελέτη ολόκληρης της αναφοράς για τον εντοπισμό διδαγμάτων, αλλά αρκεί μια αναζήτηση στο σύστημα. Με άλλα λόγια, η ύπαρξη ενός συστήματος που περιλαμβάνει μειονεκτήματα ή πλεονεκτήματα τακτικών αλλά και τρόπους για τη βελτίωση των βαθμολογιών των έργων, μπορεί να χρησιμεύσει σε μελλοντικά έργα εάν υιοθετηθεί η γνώση του συστήματος.

6.2 Μελλοντικές ενέργειες

Όπως όλα τα συστήματα αναπαράστασης γνώσης, το σύστημα που σχεδιάστηκε δεν είναι στατικό αλλά διαθέτει περιθώρια εξέλιξης και βελτίωσης, έτσι ώστε να εξυπηρετεί τους σκοπούς του.

Ένα πρωταρχικό σημείο βελτίωσης του συστήματος αφορά στις βαθμολογίες επιθέτων και επιρρημάτων που χρησιμοποιήθηκαν για τον υπολογισμό των βαρών των εννοιών (κεφ. 4.3.6). Τα αποτελέσματα που δίνουν οι συγκεκριμένες βαθμολογίες είναι αντιπροσωπευτικά, ωστόσο η ακρίβειά τους θα μπορούσε να βελτιωθεί εάν τα επίπεδα θετικότητας ή αρνητικότητάς τους αναφέρονταν στα ίδια τα έργα. Πιο συγκεκριμένα, εκφράσεις όπως «organized», ή «well-planned», μπορεί να ενέχουν διαφορετική σημαντικότητα όταν χρησιμοποιούνται για να περιγράψουν την πορεία ενός έργου από ό,τι στην καθημερινή ζωή. Επιπλέον, δεδομένου ότι τα άτομα τα οποία αξιολογούν και αποδίδουν τις βαθμολογίες σε ένα έργο είναι γνωστά και συγκεκριμένα, θα είχε ενδιαφέρον ο εντοπισμός της σημαντικότητας τέτοιων εννοιών για τους ίδιους τους εμπειρογνώμονες. Αυτό θα μπορούσε να γίνει με τη χρήση ερωτηματολογίων ή/και συνεντεύξεων, όπου οι ειδικοί θα καλούνται να βαθμολογήσουν την αρνητικότητα ή θετικότητα των εκφράσεων που οι ίδιοι χρησιμοποιούν στις αναφορές τους. Η στατιστική ανάλυση αυτών των αποτελεσμάτων και η χρήση ενός μέσου όρου βαθμολογίας για κάθε έκφραση, θα έδινε τη δυνατότητα για ένα αποτέλεσμα περισσότερο ακριβές και αντιπροσωπευτικό της γνώμης τους.

Αρχικά, το σύστημα προορίζεται για μια μελλοντική υλοποίησή του σε βάση δεδομένων. Με αναζήτηση μέσω της βάσης αυτής, ο χρήστης θα μπορεί να επιλέξει τους τομείς του έργου που τον ενδιαφέρουν, και σε αυτούς να εντοπίσει παράγοντες που έχουν επηρεάσει την πορεία έργων και σχετικά συμπεράσματα από την αξιολόγησή τους. Η μελλοντική εξέλιξη του συστήματος σχετίζεται με την αυτοματοποίησή του, που ήταν άλλωστε και ένας από τους καθοριστικούς παράγοντες σχεδιασμού του συστήματος. Ένα χειροκίνητο (μη αυτόματο) σύστημα αναπαράστασης γνώσης, απαιτεί την παρουσία ειδικών για τον καθορισμό κανόνων και την επιλογή χωρίων του κειμένου που πρέπει να καταχωρηθούν. Αντίθετα, τα αυτόματα συστήματα που βασίζονται στην εκμάθηση μηχανής (learning-based systems), μπορούν να προχωρούν στη διαδικασία αυτή χωρίς τη παρουσία ειδικού, και έχουν τη δυνατότητα χειρισμού του μεγάλου όγκου αναφορών του συστήματος ROM. Χρειάζεται, ωστόσο, η ύπαρξη χειροκίνητων παραδειγμάτων για να «εκπαιδεύσουν» τα μοντέλα εκμάθησης μηχανής στην εξαγωγή γνώσης. Ως ένα τέτοιο παράδειγμα, εάν εμπλουτιστεί με στοιχεία από επιπλέον αναφορές έργων, μπορεί να χρησιμεύσει το σύστημα που σχεδιάστηκε, με μελλοντική εξέλιξη την αυτοματοποίηση του εντοπισμού παραγόντων και διδαγμάτων σε ελεύθερο κείμενο των αναφορών του συστήματος ROM. Αξίζει να σημειωθεί πως στόχος της διπλωματικής δεν ήταν η συμπλήρωση του συστήματος με όλους τους πιθανούς παράγοντες που μπορεί να εντοπιστούν σε έργα, καθώς κάτι τέτοιο θα ήταν πρακτικά αδύνατο να συμβεί χειροκίνητα. Ωστόσο, ο σχεδιασμός του συστήματος είναι τέτοιος ώστε να μπορεί να χρησιμοποιηθεί μελλοντικά, είτε χειροκίνητα, είτε αυτόματα, με την απαιτούμενη σαφήνεια και ευκολία. Σε αυτή την εξέλιξη, το σύστημα που σχεδιάστηκε θα είναι ιδιαίτερα χρήσιμο καθώς περιλαμβάνει τα απαραίτητα στοιχεία για την κατανοητή αναπαράσταση γνώσης, όπως και αντιπροσωπευτικά παραδείγματα που μπορούν να χρησιμεύσουν στη σύγκριση αποτελεσμάτων.

ΠΑΡΑΡΤΗΜΑ

Α1. Πίνακας Καταχωρήσεων Συστήματος Αναπαράστασης Γνώσης

Παράγοντας	Κατηγορίες	BCS ID	Αριθμός Ερώτησης	Βάρος	Σχετικό Δίδαγμα
The project promotes support to decentralization and local governance	M11iii, M11iv	5048	1.1	1	
The project team showed considerable capacity to adapt to the overall timeframe	M32	5146	3.2	1	
The project aims are in line with the Government's priorities	M11i	5146	1.1	1	
Meeting expected socio-economic results by the end of the covered period	M31i	5146	1.1	0.875	
The OO, the PP and the planned results remain highly appropriate to meet the beneficiaries' needs	M11ii, M12iii	5146	1.2	1	
The LF is clear	M12i	5146	1.2	0.625	
The LF is coherent	M12i	5146	1.2	0.75	
The LF contains suitable OVIs	M12i, M12iia, M12iib	5146	1.2	0.875	
The LF reflects a realistic approach to the project implementation	M12iii	5146	1.2	0.5	
timely provision of all inputs	M21ia	5146	2.1	1	
The project design can lead to a positive impact in the sector both during project implementation and after completion	M4, M42iii	5146	1.2	0.95	
all activities have been implemented strictly as scheduled	M22iii	5146	2.2	1	
results planned for the previous period have been achieved	M31	5146	3.1	1	
there were no negative impacts on target groups	M42iv	5146	3.2	0.77	

it is likely that the PP will be achieved during the timeframe of the project	M23i	5146	3.2	0.75	Clear and concrete nature of PP
There are no external factors which may jeopardize the project's direct impact.	M41ii	5146	4.1	1	
the consistency of the governmental support for the initiatives and activities is evident	M52i, M13	5146	5.3	0.625	
It is very likely that the region will continue to make use of the relevant results	M42iii, M52ii	5146	5.2	0.77	
partner avoided possible delay in implementation of activities from its side	M24i, M31i	5146	Key lessons	0.75	Partner financed its participation in the programme with its own budget for 6 months, prior to the signing of this grant contract and adjusted signature date of grant contracts of partners, to be in accordance with other partners
the ongoing project directly contributes to the Government Strategy	M11i, M11iii	4795	1.1	0.625	
project fully relevant to the present needs of the main project beneficiary	M11ii	4795	1.1	1	
the project design includes a relatively well developed Logframe	M12i	4795	1.2	0.8	

the correctly associated targets are in most cases hardly measurable	M12ii, M12iia	4795	1.2	0,53	MS and BC Project Partners: To review the Logframe Matrix agreed, in terms of the need to identify more measurable and time bound OVIs
the correctly associated targets are not time bound	M12iia	4795	1.2	-1	MS and BC Project Partners: To review the Logframe Matrix agreed, in terms of the need to identify more measurable and time bound OVIs
planned results are appropriate for achieving the PP within the given timeframe of the project	M23iii, M12iii	4795	1.2	1	
the original project design does not formulate any particular risk management arrangements	M12iv	4795	1.2	-0.875	
partner has provided all necessary material resources in line with requirements of the contract	M21i	4795	2.1	1	
The project inputs are regularly monitored	M21iii	4795	2.1	1	
The project implementation of activities is regularly monitored	M22	4795	2.1	1	
adherence to the original time schedule	M22i	4795	2.2	1	
the project has delivered major outputs	M23	4795	2.3	0.875	

The cooperation between partners is believed to be one of the major accomplishments of the project	M24i	4795	2.4	0.97	
Partners have expressed extreme satisfaction in regard to the smooth cooperation with the staff and their full commitment to the project objectives	M24i, M41iii	4795	2.4	0.94	
considerable impacts are visible	M4	4795	4.1	1	Broad definition of the OO
the project is considered to be embedded in local structures to almost the maximum possible degree	M52ii	4795	5.2	0.75	
project target groups will indisputably continue to make use of the relevant results to be achieved	M42iii, M52ii	4795	5.2	Cannot find the term «indisputably» in SentiWordNet 3.0	
Good governance of the project is adequately reflected	M54ii	4795	7.3	0.75	
The Project Purpose (PP) is in line with the actual development policy of the Government	M53, M11i	4548	1.1	1	
The Overall Objectives (OO) are in line with the actual development policy of the Government	M53, M11i	4548	1.1	1	
The project remains relevant to the needs of the target group	M11ii	4548	1.1	1	
There is a Logframe (LF) in which the PP is supported logically by the expected results	M12i, M12iii, M12v	4548	1.2	1	
The objectives are in line with the strategies of the EC	M11iii	4548	1.1	1	
The LF contains relevant Objectively Verifiable Indicators (OVI) on the level of the activities	M12i, M12ii, M22iv	4548	1.2	1	

The Objectively Verifiable Indicators (OVI) on the level of the activities are, not accompanied by baseline data	M22ii	4548	1.2	-1	The design of the project, could have been more precise and could have incorporated clearer indicators to measure the long term impact of the intervention .
the indicators are qualitative rather than quantitative and do not allow to measure progress achieved	M12ii	4548	1.2	-1	The design of the project, could have been more precise and could have incorporated clearer indicators to measure the long term impact of the intervention .
the set of OVI does not allow close follow up on the outcomes / results	M12iib	4548	1.2	-1	The design of the project, could have been more precise and could have incorporated clearer indicators to measure the long term impact of the intervention .
No risk management strategy is in place	M12iv	4548	1.2	-1	

The key stakeholders have been involved closely in the design and management process of the project	M13	4548	1.2	0.625	
the timescale and range of activities is realistic	M21ii	4548	1.3	1	
The inputs were provided timely by all partners	M21ia	4548	1.3	1	
the Work Plan is not suitable as a monitoring instrument	M21iii	4548	3.1	-0.875	
The quality of the reporting to the EC is poor	M21iv	4548	Lessons learned	-0.625	Request partners to prepare a more detailed work programme and report against it, introducing clear aggregated figures over the total project period
unclear formulation in the project proposal, lacking OVI	M12iv, M12iib	4548	3.1	-0.5	
The quality of the outputs is high	M23	4548	3.1	0.5	
presenting anecdotic information in the reports	M21iii	4548	3.2	-0.5	Concentrate , in the next progress report for the EC, on result achievement and introduce systemic instruments to present the typology of initiatives, lessons learnt etc.
The project is constantly examining the use of innovative means of communication (Facebook, e-	M14, M42	4548	4.2	0.625	

mail, blogs etc.) to bring the issues closer to the daily reality of the target group, which is successful					
The networking with the relevant parts of the public administration the NGO and the private sector is efficient	M41iii, M24i	4548	4.2	0.5	
The present policies of the Government are very supportive of the activities undertaken by the project	M53	4548	5.4	0.94	
The activities, presently undertaken by the project, are planned to be expanded	M42iii	4548	5.4	1	
The wide-based and long-term support basis is encouraging for future sustainability.	M5	4548	5.1	0.625	

Πίνακας : Καταχωρήσεις Συστήματος Αναπαράστασης Γνώσης

A.2 Λεξιλόγιο Επιλεγμένων Όρων που χρησιμοποιούνται στο σύστημα ROM

Όρος	Ερμηνεία
Activity Schedule	Διάγραμμα Gantt που καθορίζει το συγχρονισμό, την σειρά και τη διάρκεια των δραστηριοτήτων του έργου. Μπορεί επίσης να χρησιμοποιηθεί στον εντοπισμό οροσήμων στη διαδικασία επίβλεψης και την ανάθεση ευθυνών
Beneficiaries	Το σύνολο που επωφελείται με οποιονδήποτε τρόπο από την εφαρμογή του έργου. Μια διαχώριση μπορεί να γίνει ανάμεσα σε: <ul style="list-style-type: none"> • Target Groups: Η ομάδα που θα επηρεαστεί άμεσα από το έργο σε επίπεδο σκοπού του έργου • Final Beneficiaries: αυτοί που θα επωφεληθούν μακροπρόθεσμα από το έργο σε κοινωνικό επίπεδο ή σε γενικότερο επίπεδο τομέα, π.χ. οι καταναλωτές στη περίπτωση βελτίωσης της αγροτικής παραγωγής
Contractor	Ο δημόσιος ή ιδιωτικός οργανισμός, κοινοπραξία ή ιδιώτης με τον οποίο συνάπτεται το συμβόλαιο.
Decentralisation	Πέρασμα ευθυνών από την Κομισιόν στην χώρα που επωφελείται από το έργο
Effectiveness	Η συμβολή των αποτελεσμάτων του έργου στην επίτευξη του σκοπού του έργου
Efficiency	Η επίτευξη αποτελεσμάτων σε λογικό κόστος, το πόσο καλά εξελίσσονται οι δραστηριότητες σε αποτελέσματα, και η ποιότητα των επιτευγμένων αποτελεσμάτων
Impact	Η επίδραση του έργου στο ευρύτερο περιβάλλον και η συμβολή του στους στόχους του τομέα, οι οποίοι συνοψίζονται στο ΟΟ του έργου, και στην επίτευξη των πρωταρχικών στόχων της πολιτικής της ΕΕ
Impact Indicators	Δείκτες μέτρησης των μακροπρόθεσμων συνεπειών των αποτελεσμάτων. Μετρούν τους γενικούς σκοπούς από την άποψη της εθνικής ανάπτυξης και της μείωσης της φτώχειας. Π.χ.: ποσοστά αναλφαβητισμού
Intervention Logic	Η στρατηγική πίσω από το έργο. Πρόκειται για την περιγραφή του έργου σε καθένα από τα τέσσερα στάδια που χρησιμοποιούνται στο Logframe
Input Indicators	Δείκτες μέτρησης των οικονομικών, διοικητικών και ρυθμιστικών πόρων που παρέχονται από την κυβέρνηση και χορηγούς. Η διασύνδεση των πόρων που χρησιμοποιούνται και των αποτελεσμάτων που

	επιτυγχάνονται είναι απαραίτητη για την εκτίμηση της αποτελεσματικότητας των δράσεων που πραγματοποιούνται. Παραδείγματα: μερίδιο προϋπολογισμού που διατίθεται για δαπάνες στην εκπαίδευση, κατάργηση της υποχρεωτικής στολής στις σχολές
Logframe	Ο πίνακας στον οποίο παρουσιάζονται τα στοιχεία του έργου: Intervention Logic, Assumptions, Objectively Verifiable Indicators και Sources of Verification
Logical Framework Approach	Μεθοδολογία για τον σχεδιασμό, τη διαχείριση και την αξιολόγηση προγραμμάτων και έργων, συμπεριλαμβανομένων των διαδικασιών ανάλυσης συμβαλλόμενων μελών, προβλημάτων, στόχων, στρατηγικών και της προετοιμασίας πίνακα Logframe
Means	Τα μέσα και οι πηγές (φυσικές ή όχι) που είναι απαραίτητα για την πραγματοποίηση των προγραμματισμένων δραστηριοτήτων και τη διαχείριση του έργου. Συχνά αναφέρονται και ως Inputs. Μπορούν να διαχωριστούν σε ανθρώπινους και υλικούς πόρους
Objectively Verifiable Indicators (OVI)	Μετρήσιμοι δείκτες που επισημαίνουν την επίτευξη ή όχι των στόχων, σε καθένα από τα τρία επίπεδα του Logframe. Οι OVIs παρέχουν τη βάση σχεδιασμού ενός κατάλληλου συστήματος επίβλεψης
Overall Objective (συναντάται και ως 'goal')	Ο ΟΟ επεξηγεί τη σημαντικότητα του έργου για την κοινωνία, σε επίπεδο μακροπρόθεσμων οφελών στους final beneficiaries του έργου και ευρύτερων οφελών σε άλλες ομάδες. Είναι επίσης δείγμα του κατά πόσο το έργο ταιριάζει στις πολιτικές των οργανισμών που συμπεριλαμβάνονται ή της ΕΕ για την περιοχή ή τον τομέα, καθώς και στην ευρύτερη πολιτική της ΕΕ. Ο στόχος αυτός δεν θα επιτευχθεί ολοκληρωτικά χάρη στο έργο (το έργο θα έχει μόνο μια συμβολή) και απαιτεί τη συμβολή άλλων προγραμμάτων και έργων
Outcome Indicators	Δείκτες μέτρησης του αποτελέσματος ως προς τα οφέλη που προσδίδει στις ομάδες – στόχους. Π.χ.: ποσοστό κοριτσιών στις εγγραφές στην πρώτη τάξη του δημοτικού σχολείου
Output Indicators	Δείκτες μέτρησης των άμεσων και συγκεκριμένων συνεπειών των μέτρων που λαμβάνονται και των πόρων που χρησιμοποιούνται. Πχ: αριθμός σχολείων που χτίζονται, αριθμός δασκάλων που εκπαιδεύονται
Partner	Οι ιδιώτες ή/και οι οργανισμοί που συνεργάζονται για την επίτευξη αμοιβαία συμφωνημένων στόχων. Η ιδέα

	του συνεταιρισμού εμπεριέχει κοινούς στόχους, κοινές ευθύνες για τα αποτελέσματα, ξεχωριστές ευθύνες για αναφορά και αμοιβαίες υποχρεώσεις. Στους partners μπορεί να συμπεριλαμβάνονται κυβερνήσεις, μη κυβερνητικές οργανώσεις, η κοινωνία πολιτών, πανεπιστήμια, επαγγελματίες, ιδιωτικές εταιρείες κ.τ.λ.
Project Purpose	Ο κεντρικός στόχος του έργου. Ο στόχος θα πρέπει να αντιμετωπίζει το κύριο πρόβλημα (ή τα κυριότερα προβλήματα) και να καθορίζεται από την άποψη των αιτιόρων οφελών των ομάδων-στόχων. Για μεγαλύτερα/σύνθετα έργα, ενδέχεται να υπάρχουν περισσότεροι από ένας στόχοι
Relevance	Η καταλληλότητα των στόχων του έργου για την επίλυση προβλημάτων και την κάλυψη αναγκών βάσει των προτεραιοτήτων των ομάδων στις οποίες απευθύνεται το έργο, λαμβάνοντας υπόψη και το πολιτικό περιβάλλον στο οποίο εξελίσσεται
Results	Απτά προϊόντα/υπηρεσίες που προκύπτουν ως συνέπειες της εφαρμογής του συνόλου των δραστηριοτήτων του έργου
Stakeholder	Κάθε ιδιώτης, ομάδα ανθρώπων, οργανισμός ή εταιρεία που μπορεί να σχετίζεται με το πρόγραμμα/έργο. Μπορεί να είναι άμεσα ή έμμεσα θετικά ή αρνητικά επηρεαζόμενος από την πορεία και τα αποτελέσματα του έργου
Sustainability	Το ενδεχόμενο να συνεχιστούν οι παροχές που παράγονται από το έργο μετά τη λήξη της περιόδου εξωτερικής στήριξης. Βασικοί παράγοντες που επηρεάζουν την πιθανότητα βιωσιμότητας είναι: ιδιοκτησία δικαιούχων, υποστήριξη πολιτικής, συμβατότητα με την πολιτική, κατάλληλη τεχνολογία, περιβάλλον, κοινωνικά/πολιτισμικά ζητήματα, ισότητα φύλων, θεσμική ικανότητα διαχείρισης και οικονομική και χρηματοοικονομική βιωσιμότητα
Target Group(s)	Η ομάδα που θα επηρεαστεί θετικά από το έργο σε επίπεδο σκοπού έργου

Πίνακας : Λεξιλόγιο Επιλεγμένων Όρων του συστήματος ROM (Co-operation, Handbook for Results-Oriented Monitoring of EC External Assistance (projects and programmes), 2008)

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Andrew McCallum, D. J. (2003). A Note on the Unification of Information Extraction and Data Mining.
- Assistance, R. -O. (2009). Causes underlying Effectiveness and Impact of EC Development Projects.
- Bouquet, P., Giunchiglia, F., van, H., Serafini, L., & Stuckenschmidt, H. (2003). C-OWL: Contextualizing ontologies. *Proceedings of the Second international Semantic Web Conference, 2870*, 164--179.
- Chklovski, T. (2006). Deriving quantitative overviews of free text assessments on the web. *IUI '06 Proceedings of the 11th international conference on Intelligent user interfaces*.
- D. White, J. F. (2002). Current practice in project management: an empirical study. *International Journal of Project Management*, 20 (1).
- Deming, W. E. (1990). *Sample design in business research*.
- Elma Balic, R. M. (2009). VRBAS, Joint Tourism Development of the Vrbas Valley, BCS. *MR-041105.03*.
- ETUI-REHS. (2004). Project monitoring and evaluation - guidelines.
- Farah Benamara, C. C. (2006). Sentiment Analysis: Adjectives and Adverbs are better.
- Frawley W.J, P. -S. (1991). *Knowledge Discovery in Databases : An Overview, in Knowledge Discovery in Databases*. MIT Press.
- Gruber, T. (1993). A translation approach to portable previous termontologynext term specification. *Knowledge Acquisition 5*, 199–220. .
- Gruber, T. (2008). Ontology. *Encyclopedia of Database Systems, Ling Liu and M. Tamer Özsu (Eds.)* .
- Gruninger, M. U. (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 93-136.
- Jay Liebowitz, I. M. (2003). A set of frameworks to aid the project manager in conceptualizing and implementing knowledge management initiatives. *International Journal of Project Management*, 21 (3), 189-198.
- Jiawei Han, M. K. (2006). *Data mining: concepts and techniques*.
- Kerzner, H. (2009). *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*.
- Krzysztof J. Cios, W. P. (1998). *Data mining methods for knowledge discovery*.
- Leary, D. E. (1998). Enterprise knowledge management. *Computer*, 31 (3), 54-61.
- M. Uschold, M. G. (1996). *Ontologies: Principles, Methods and Applications*.
- Majchrzak A, N. O. (2001). Knowledge reuse for innovation—the missing focus in knowledge management: results of a case analysis at the jet propulsion laboratory. *Academy of Management Proceedings*.

Marr, D. (1982). *Vision*.

McCallum, A. (2008). Information Extraction: Distilling Structured Data from Unstructured Text. *ACM QUEUE* .

Mike Bresnen, L. E. (2003). Social practices and the management of knowledge in project environments. *International Journal of Project Management* , 21 (3), 157-166.

Minsky, M. (1975). *A framework for representing knowledge*.

Natalya F. Noy, D. L. (2000). Ontology Development 101: A Guide to Creating Your First Ontology.

Office, E. C.-o. (2009, May). Causes underlying Effectiveness and Impact of EC Development Projects.

Office, E. C.-o. (2008). *Handbook for Results-Oriented Monitoring of EC External Assistance (projects and programmes)*.

Philipp Cimiano, S. S. (2002). Automatic Acquisition of Taxonomies from Text: FCA meets NLP.

Sarawagi, S. (2008). *Foundations and Trends in Databases, Vol.1, No.3*.

Senigen, S. (2004). Learn the Value of Lessons Learned.

Sinisa Bralic, R. M. (2010). Minicipal Infrastructure Support Programme (MISP) BCS. *MR-116120.02* .

Stefano Baccianella, A. E. (2008). SENTIWORDNET 3.0: An Enhanced Lexical Resource.

Stewart, A. (1997). Under the Hood at Ford. *Webmaster* , 26-34.

Un Yong Nahm, R. J. (2000). A Mutually Beneficial Integration of Data Mining and Information Extraction. *Seventeenth National Conference on Artificial Intelligence(AAI-2000)*, (σσ. 627-632). Austin, Texas.

Usama Fayyad, G. P.-S. (1996). The KDD project for Extracting Useful Knowledge from Volumes of Data. *COMMUNICATIONS OF THE ACM* , 28-34.

Wilks, J. C. (1996). Information Extraction.

Ντάνος, Χ. (2010). Προτεινόμενη Μεθοδολογία Εξόρυξης Κειμένου από Αναφορές Αξιολόγησης Έργων.