

Χάρος αειχώρος

ΚΕΙΜΕΝΑ ΠΟΛΕΟΔΟΜΙΑΣ, ΧΩΡΟΤΑΞΙΑΣ ΚΑΙ ΑΝΑΠΤΥΞΗΣ

ΤΟΜΟΣ 3
VOLUME 3

ΤΕΥΧΟΣ 2
ISSUE 2

ΝΟΕΜΒΡΙΟΣ 2004
NOVEMBER 2004



ΣΥΝΤΑΚΤΙΚΗ ΕΠΙΤΡΟΠΗ - Πανεπιστήμιο Θεσσαλίας
*Τμήμα Μηχανικών Χωροταξίας, Πολεοδομίας
και Περιφερειακής Ανάπτυξης*

ΚΟΚΚΩΣΗΣ ΧΑΡΗΣ
ΜΠΕΡΙΑΤΟΣ ΗΛΙΑΣ
ΟΙΚΟΝΟΜΟΥ ΔΗΜΗΤΡΗΣ
ΠΕΤΡΑΚΟΣ ΓΙΩΡΓΟΣ
ΓΟΥΣΙΟΣ ΔΗΜΗΤΡΗΣ
ΔΕΦΝΕΡ ΑΛΕΞΗΣ
ΨΥΧΑΡΗΣ ΓΙΑΝΝΗΣ

ΣΥΜΒΟΥΛΟΙ ΣΥΝΤΑΞΗΣ

Αραβαντινός Αθανάσιος	- ΕΜΠ
Ανδρικόπουλος Ανδρέας	- Οικονομικό Πανεπιστήμιο Αθηνών
Βασενχόβεν Λουδοβίκος	- ΕΜΠ
Γιαννακούρου Τζίνα	- Πανεπιστήμιο Θεσσαλίας
Γιαννιάς Δημήτρης	- Πανεπιστήμιο Θεσσαλίας
Δελλαδέτσιμας Παύλος	- Χαροκόπειο Πανεπιστήμιο
Δεμαθιάς Ζαχαρίας	- Πάντειο Πανεπιστήμιο
Ιωαννίδης Γιάννης	- Tufts University, USA
Καλογήρου Νίκος	- ΑΠΘ
Καρυίδης Δημήτρης	- ΕΜΠ
Κοσμόπουλος Πάνος	- ΔΠΘ
Κουκλέλη Ελένη	- University of California, USA
Λαμπριανίδης Λόης	- Πανεπιστήμιο Μακεδονίας
Λουκάκης Παύλος	- Πάντειο Πανεπιστήμιο
Λουρή Ελένη	- Οικονομικό Πανεπιστήμιο Αθηνών
Μαντουβάλου Μαρία	- ΕΜΠ
Μελαχροινός Κώστας	- University of London, Queen Mary, UK
Μοδινός Μιχάλης	- Εθν. Κέντρο Περιβ. και Δειφ. Ανάπτυξης (ΕΚΠΙΑΑ)
Μπριασούλη Ελένη	- Πανεπιστήμιο Αιγαίου
Παπαθεοδώρου Ανδρέας	- University of Surrey, UK
Παπτάς Βασίλης	- Πανεπιστήμιο Πατρών
Πρεβελάκης Γεώργιος-Στυλ.	- Universite de Paris I, France
Φωτόπουλος Γιώργος	- Πανεπιστήμιο Πατρών
Χαστάογλου Βίλμα	- ΑΠΘ

Διεύθυνση:

Πανεπιστήμιο Θεσσαλίας
Τμήμα Μηχανικών Χωροταξίας, Πολεοδομίας
και Περιφερειακής Ανάπτυξης

Περιοδικό ΔΕΙΧΩΡΟΣ

Πεδίον Άρεως, 38334 ΒΟΛΟΣ

<http://www.prd.uth.gr/aeihoros> e-mail: aeihoros@prd.uth.gr

τηλ.: 24210 – 74456 fax: 24210 – 74380



ΠΑΝΕΠΙΣΤΗΜΙΑΚΕΣ ΕΚΔΟΣΕΙΣ ΘΕΣΣΑΛΙΑΣ

Ειδικό τεύχος - Αφιέρωμα
Γεωπληροφορική

Επιμέλεια

Μαρίνος Κάβουρας
Ελευθερία Καρνάβου

Επιστημονικό Περιοδικό

αειχώρος



Επιμέλεια έκδοσης : Άννα Σαμαρίνα — Παναγιώτης Πανταζής
Λαγούτ : Παναγιώτης Πανταζής
Σχεδιασμός εξωφύλλου : Γιώργος Παρασκευάς
Εκτύπωση : Αλέκος Ξουράφας
Κεντρική διάθεση : Πανεπιστημιακές Εκδόσεις Θεσσαλίας

Περιεχόμενα

Εισαγωγή	4
Κόκλα Μ., Κάβουρας Μ.	8
Προσδιορισμός σημασιολογικών ιδιοτήτων και σχέσεων για την επίλυση οντολογικής ετερογένειας	
Τομαή Ε., Κάβουρας Μ.	24
Απαραίτητα συστατικά για την ανάπτυξη Θεματικών Γεωγραφικών Οντολογιών	
Γραϊκούσης Γ., Φώτης Γ. και Κουτσόπουλος Κ.	40
Χωροχρονική πρόβλεψη σημειακών προτύπων ζήτησης στοχαστικών χωροθετικών προβλημάτων με χρήση Νευρωνικών Δικτύων	
Παντελέλης Μ., Σουλακέλλης Ν.	62
Διαχείριση και ανάλυση χωροχρονικών δεδομένων αέριας ρύπανσης αστικών περιοχών: Η περίπτωση της Αττικής	
Μανέτος Π., Φώτης Γ.	76
Εφαρμογή μεθόδων εξαγωγής χωρικής γνώσης για τον προσδιορισμό και την απόδοση Χωρικών Προτύπων	
Μελιδόνη Μ., Χατζηχρήστος Θ.	94
Σχεδιασμός και ανάπτυξη περιβάλλοντος διεπαφής με το χρηστή για την ανάλυση χωρικών σημειακών προτύπων με την τεχνική "Ανάλυση Καννάβου" σε περιβάλλον ΓΣΠ	
Σαραφίδης Δ., Παρασχάκης Ι.	112
Ένα χρηστικό περιβάλλον επικοινωνίας για την διάθεση των κτηματολογικών δεδομένων στο διαδίκτυο	
Ντόκου Αικ.	134
Χωρικές διαδικασίες τεκμηρίωσης τρισδιάστατης πληροφορίας ιδιοκτησιακών αντικειμένων	
Καμαριανάκης Γ., Κοντός Δ.	154
Ταξινόμηση των δήμων της Ελλάδας σύμφωνα με τα κοινωνικά και οικονομικά χαρακτηριστικά τους όπως προκύπτουν από την Απογραφή του 2001	

Προσδιορισμός σημασιολογικών ιδιοτήτων και σχέσεων για την επίλυση οντολογικής ετερογένειας

Μαργαρίτα Κόκλα

Υποψήφια Διδάκτορας, Εθνικό Μετσόβιο Πολυτεχνείο

Μαρίνος Κάβουρας

Αναπληρωτής Καθηγητής, Εθνικό Μετσόβιο Πολυτεχνείο

Περίληψη

Ένα από τα βασικότερα προβλήματα που προκύπτει κατά τη διαδικασία της σημασιολογικής ολοκλήρωσης οντολογιών είναι ο εντοπισμός ομοιοτήτων και ετερογενειών. Ιδιαίτερα στις γεωγραφικές οντολογίες, όπου η σημασιολογία είναι πολύπλοκη, γίνεται επιτακτική η ανάγκη ανάπτυξης μεθόδων για την εξαγωγή και τυποποίηση της σημασιολογίας, αλλά και για την επίλυση της ετερογένειας μεταξύ όμοιων γεωγραφικών εννοιών. Η παρούσα εργασία αξιοποιεί μεθόδους από την Επιστήμη Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing-NLP) για την εξαγωγή σημασιολογικής πληροφορίας από ορισμούς γεωγραφικών εννοιών. Βασίζεται στην παραδοχή ότι οι ορισμοί περιλαμβάνουν πληθώρα σημασιολογικών πληροφοριών οι οποίες είναι δυνατόν να εξαχθούν σύμφωνα με συγκεκριμένους κανόνες και να τυποποιηθούν σε ένα σύνολο σημασιολογικών ιδιοτήτων και σχέσεων. Αυτός ο τρόπος τυποποίησης της σημασιολογίας αξιοποιείται περαιτέρω, προκειμένου ο εντοπισμός και η επίλυση των ετερογενειών μεταξύ όμοιων εννοιών να γίνεται με σαφή και αντικειμενικό τρόπο.

Λέξεις-κλειδιά

σημασιολογική ολοκλήρωση, οντολογίες, σημασιολογικές ιδιότητες και σχέσεις, ετερογένειες.

Determination of semantic information for resolving ontological heterogeneity

The identification and resolution of heterogeneities between different ontologies is a critical process during semantic integration. The need to develop methods for the extraction and formalization of semantics, as well as for the resolution of heterogeneities between similar concepts is a requisite for the complex semantics of geographic ontologies.

The present research relies on the assumption that definitions of geographic categories include a wealth of semantic information which can be extracted and formalized into a set of semantic properties and relations. To this end, an approach is developed which is based on Natural Language Processing (NLP) in order to provide a clear and objective way for: (a) the extraction of semantic information from definitions of geographic categories and (b) the identification and resolution of heterogeneities between similar geographic categories.

Keywords

Semantic integration, ontologies, semantic properties, semantic relations, heterogeneities.

ΕΙΣΑΓΩΓΗ

Η σημασιολογική ολοκλήρωση αποτελεί μια πολύπλοκη διαδικασία, ιδιαίτερα στην περίπτωση της σύνθετης σημασιολογίας των γεωγραφικών οντολογιών. Ένα από τα βασικότερα προβλήματα που προκύπτουν και πρέπει να αντιμετωπιστούν είναι ο εντοπισμός και η επίλυση ετερογενειών, προκειμένου στην τελική ολοκληρωμένη οντολογία να αναδεικνύονται οι ομοιότητες, οι διαφορές και οι επικαλύψεις μεταξύ των αρχικών γεωγραφικών κατηγοριών.

Η συσχέτιση των εννοιών δύο ή περισσότερων οντολογιών γίνεται συνήθως με εργαλεία που βοηθούν έναν ειδικό χρησιμοποιώντας μεθόδους που εξετάζουν γλωσσική και δομική ομοιότητα (McGuinness κ.ά., 2000· Mitra και Wiederhold, 2002· Noy και Musen, 2000), δηλαδή ομοιότητα στους όρους και στην ιεραρχική δομή.

Η γλωσσική και δομική ομοιότητα είναι επαρκείς σε περιπτώσεις οντολογιών με απλή σημασιολογία που ορίζονται κυρίως από όρους που αναπαριστούν τις έννοιες και ιεραρχικές σχέσεις μεταξύ τους. Σε αυτές τις οντολογίες, οι έννοιες γίνονται εύκολα αντιληπτές από τους όρους που τις περιγράφουν και η ομοιότητα στους όρους επιδεικνύει ομοιότητα στη σημασιολογία των εννοιών.

Παρ' όλ' αυτά, οι όροι και οι ιεραρχικές σχέσεις δεν είναι ικανά στοιχεία για να περιγράψουν πλήρως τη σημασιολογία των εννοιών και κυρίως των γεωγραφικών εννοιών. Σε πολλές περιπτώσεις ομοιότητα στους όρους και στις ιεραρχικές σχέσεις δεν συνεπάγεται απαραίτητα και σημασιολογικά ισοδύναμες έννοιες. Όμως, η ανάγκη ακριβούς και σαφούς προσδιορισμού των συσχετίσεων μεταξύ των εννοιών της ολοκληρωμένης οντολογίας κάνει επιτακτική την ανάδειξη έστω και μικρών διαφορών στη σημασιολογία όμοιων εννοιών.

Σκοπός της παρούσας εργασίας είναι η ανάπτυξη μιας μεθόδου για την εξαγωγή και τυποποίηση της σημασιολογίας των γεωγραφικών εννοιών όχι μόνο σύμφωνα με τους όρους και τις ιεραρχικές τους σχέσεις, αλλά και με βάση το νόημά τους. Αυτή η μεθοδολογία ενσωματώνεται σε ένα πλαίσιο εντοπισμού ομοιοτήτων και ετερογενειών μεταξύ των αρχικών οντολογιών προκειμένου η διαδικασία της σημασιολογικής ολοκλήρωσης να γίνεται με σαφή και αντικειμενικό τρόπο. Από τη στιγμή που έχει εξαχθεί και τυποποιηθεί η σημασιολογία των γεωγραφικών εννοιών, και έχουν επιλυθεί οι τυχόν ετερογένειες, μπορούν στη συνέχεια να ακολουθηθούν συστηματικές διαδικασίες ολοκλήρωσής τους, όπως αναφέρεται από τους Kanouira και Kokla (2002) και Κάβουρα (2002).

Η παρούσα εργασία αξιοποιεί μεθόδους από την Επιστήμη Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing-NLP) για την εξαγωγή σημασιολογικής πληροφορίας από ορισμούς γεωγραφικών κατηγοριών. Η σημασιολογική πληροφορία τυποποιείται με τη μορφή σημασιολογικών στοιχείων (ιδιότητες και σχέσεις) που λαμβάνουν συγκεκριμένες τιμές. Ο εντοπισμός ομοιοτήτων και ετερογενειών γίνεται μέσω της σύγκρισης των σημασιολογικών στοιχείων και των τιμών τους. Τέλος, προτείνεται και ένας τρόπος επίλυσης των ετερογενειών που εντοπίστηκαν προκειμένου να προχωρήσει η διαδικασία σημασιολογικής ολοκλήρωσης στο τελικό βήμα δημιουργίας της ενοποιημένης οντολογίας.

ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΟΡΙΣΜΩΝ

Στα πλαίσια της Επεξεργασίας Φυσικής Γλώσσας αναπτύσσονται μεθοδολογίες για την αυτόματη εξαγωγή σημασιολογικής πληροφορίας από ορισμούς. Σύμφωνα με τους Jensen και Binot (1987), οι ορισμοί περιλαμβάνουν αφθονία γνώσης που εκφράζεται σε φυσική γλώσσα και η οποία μπορεί να αξιοποιηθεί από τα Συστήματα Επεξεργασίας Φυσικής Γλώσσας.

Οι ορισμοί θεωρούνται ως ένα είδος κειμένου το οποίο έχει ιδιαίτερη δομή και περιεχόμενο. Αποτελούν σημαντικές πηγές πληροφοριών και εκφράζουν την επιστημονική γνώση κάθε γνωστικού πεδίου. Η έρευνα γύρω από τους ορισμούς αναζητά μεθόδους αξιοποίησης του πλούτου πληροφορίας που ενυπάρχει σε αυτήν την ειδική κατηγορία

κειμένου. Στις υπάρχουσες γεωγραφικές οντολογίες οι ορισμοί αποτελούν τις κυριότερες και συνήθως τις μοναδικές περιγραφές των κατηγοριών, αφού άλλα στοιχεία όπως ιδιότητες, λειτουργίες, αξιώματα, κ.λπ., είτε απουσιάζουν, είτε ορίζονται επιφανειακά.

Οι ορισμοί διακρίνονται σε δύο τμήματα: στο γένος (genus) και στα χαρακτηριστικά γνωρίσματα του γένους (differentiae). Το γένος είναι ο ανώτερος όρος του λήμματος (της λέξης που ορίζεται), δηλαδή η ορισμένη λέξη είναι "ένα είδος" (a kind of) του γένους. Για παράδειγμα, στον ορισμό: "*river: large natural stream of water*", το "υδατόρευμα" (stream) είναι το γένος της έννοιας "ποτάμι" (river).

Τα χαρακτηριστικά γνωρίσματα (differentiae) είναι όλα τα υπόλοιπα στοιχεία του ορισμού εκτός από το γένος και τα οποία διαφοροποιούν τις λέξεις που έχουν το ίδιο γένος. Έτσι, στον ορισμό: "*brook: natural stream of water smaller than a river (and often the tributary of a river)*", η λέξη "ρυάκι" (brook) έχει το ίδιο γένος με τη λέξη "ποτάμι" (river), αλλά διαφοροποιούνται λόγω των χαρακτηριστικών γνωρισμάτων (π.χ., "large" ή "smaller than a river" και "often the tributary of a river").

ΕΞΑΓΩΓΗ ΣΗΜΑΣΙΟΛΟΓΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

Η μεθοδολογία εξαγωγής σημασιολογικής πληροφορίας που υιοθετήθηκε στην παρούσα έρευνα, αναπτύχθηκε από τους Jensen και Binot (1987) και στη συνέχεια επεκτάθηκε από τη Vanderwende (1995). Η προσέγγιση αυτή, συνίσταται:

- α) στην συντακτική ανάλυση των ορισμών και
- β) στην εφαρμογή κανόνων που εξετάζουν την ύπαρξη συγκεκριμένων συντακτικών και λεξιλογικών προτύπων.

Σκοπός της συντακτικής ανάλυσης είναι ο προσδιορισμός της δομής του ορισμού, δηλαδή της λειτουργίας κάθε λεκτικής μονάδας, καθώς και της σχέσης μεταξύ των λεκτικών μονάδων. Για τη συντακτική ανάλυση χρησιμοποιείται ένα κατάλληλο εργαλείο που ονομάζεται συντακτικός αναλυτής (parser). Το αποτέλεσμα της συντακτικής ανάλυσης αναπαρίσταται με τη μορφή ενός αρχείου. Το αρχείο αυτό περιλαμβάνει συντακτική πληροφορία όπως τμήμα του λόγου, λήμμα, μορφολογική πληροφορία και κάποια πληροφορία υποκατηγοριοποίησης.

Οι δομές που προκύπτουν από τη συντακτική ανάλυση υποβάλλονται σε ένα σύνολο ευριστικών κανόνων (Dolan κ.ά., 1993). Οι κανόνες αυτοί εξετάζουν την ύπαρξη συντακτικών και λεξιλογικών προτύπων (patterns), δηλαδή συγκεκριμένων λέξεων ή φράσεων που χρησιμοποιούνται συστηματικά στους ορισμούς για να εκφράσουν συγκεκριμένες σημασιολογικές ιδιότητες και σχέσεις. Τα πρότυπα που εφαρμόζονται στο γένος του ορισμού εξάγουν τη σχέση εξειδίκευσης (IS-A). Τα πρότυπα που εφαρμόζονται στα χαρακτηριστικά γνωρίσματα εξάγουν άλλες ιδιότητες και σχέσεις όπως: θέση, χρόνος σκοπός, αιτία, κ.λπ.

- Για παράδειγμα, η σημασιολογική ιδιότητα ΣΚΟΠΟΣ (PURPOSE) προσδιορίζεται στους ορισμούς της αγγλικής γλώσσας από συγκεκριμένες φράσεις που περιλαμβάνουν την πρόθεση "for" (π.χ., *for (the) purpose(s) of, for, used for, intended for*) και ακολουθούνται από μια φράση εμπριέχουσα κύριο όνομα, μια μετοχή ενεστώτα ή μια πρόταση εμπριέχουσα απαρέμφατο. Έτσι, ο κανόνας για την εξαγωγή αυτής της ιδιότητας από ορισμούς μπορεί να εκφραστεί ως εξής:
 - "Αν το ρήμα *used* ακολουθείται από μια φράση εισαγόμενη με την πρόθεση *for*, τότε ορίζεται μια σχέση ΣΚΟΠΟΥ (PURPOSE) με τιμή την κυρίως λέξη της φράσης".
 - Για παράδειγμα, η σημασιολογική ιδιότητα ΣΚΟΠΟΣ εξάγεται από τον ορισμό: "canal: a manmade or improved natural waterway used for transportation", με τιμή "transportation".
- Η σημασιολογική ιδιότητα ΘΕΣΗ (LOCATION) προσδιορίζεται στους ορισμούς της αγγλικής γλώσσας από αναφορικές προτάσεις που εισάγονται με το αναφορικό επίρρημα "όπου" (where) ή από φράσεις που εισάγονται με τις προθέσεις "in", "on", "from". Έτσι, ο κανόνας για την εξαγωγή αυτής της ιδιότητας από ορισμούς μπορεί να εκφραστεί ως εξής:
 - Αν το γένος ακολουθείται από αναφορική πρόταση που εισάγεται με το επίρρημα "where" ή από μια φράση εισαγόμενη με την πρόθεση "in" ή "on" ή "above" κ.λπ., τότε ορίζεται μια σχέση ΘΕΣΗΣ (LOCATION) με τιμή την κυρίως λέξη της φράσης.
 - Στον παρακάτω ορισμό εξάγεται η σημασιολογική ιδιότητα ΘΕΣΗ με τιμή "desert region": "saltpan: a shallow basin in a desert region".

Τα πρότυπα που προσδιορίζονται στους ορισμούς δεν μπορούν να χρησιμοποιηθούν το ίδιο αξιόπιστα για την εξαγωγή σημασιολογικών πληροφοριών από ελεύθερο κείμενο. Επιπλέον, στη βιβλιογραφία δεν υπάρχει πλήρης λίστα σημασιολογικών ιδιοτήτων και σχέσεων που είναι δυνατό να εξαχθούν από ορισμούς. Γι' αυτό το λόγο, στα πλαίσια της παρούσας έρευνας αναλύθηκαν διαφορετικές οντολογίες, πρότυπα, κατηγοριοποιήσεις (π.χ., CYC Upper Level Ontology, WordNet, CORINE Land Cover, DIGEST, SDTS, κ.λπ.) για την εύρεση προτύπων που χρησιμοποιούνται συστηματικά στους ορισμούς γεωγραφικών κατηγοριών για να εκφράσουν συγκεκριμένα σημασιολογικά στοιχεία (ιδιότητες και σχέσεις). Τα κυριότερα από αυτά παρουσιάζονται στον πίνακα 1 (Κανούρας κ.ά., 2003).

Η εφαρμογή της παραπάνω μεθοδολογίας οδηγεί στην ανάλυση των ορισμών των γεωγραφικών κατηγοριών και στην τυποποίησή τους σύμφωνα με τα σημασιολογικά τους στοιχεία. Έτσι, για κάθε γεωγραφική έννοια προσδιορίζεται ένα σύνολο σημασιολογικών

ιδιοτήτων και σχέσεων και οι αντίστοιχες τιμές που περιγράφουν τη σημασιολογία της.

Στον πίνακα 2 παρουσιάζονται τα σημασιολογικά στοιχεία και οι τιμές για ένα σύνολο γεωγραφικών κατηγοριών σχετικών με την υδρογραφία από τις οντολογίες CYC και WordNet και το πρότυπο ανταλλαγής γεωγραφικών πληροφοριών DIGEST. Η σημασιολογική σύγκριση μεταξύ των γεωγραφικών κατηγοριών θα βασιστεί σε αυτό το σύνολο των σημασιολογικών στοιχείων και τιμών. Προκειμένου να διεξαχθεί αποτελεσματικά το επόμενο βήμα της μεθοδολογίας που είναι ο εντοπισμός και η επίλυση των ετερογενειών απαιτείται μια επεξεργασία των τιμών των σημασιολογικών στοιχείων του Πίνακα 2 κυρίως για τον προσδιορισμό συνωνύμων. Για παράδειγμα, οι τιμές "conduit" και "channel" της σημασιολογικής σχέσης "IS-A" είναι συνώνυμες. Με εξέταση ενός online λεξικού (π.χ., WordNet) διαπιστώνεται ότι οι λέξεις αυτές μπορούν να αντικατασταθούν από τη συνώνυμη ή υπερώνυμη λέξη "way".

Πίνακας 1: Σημασιολογικές ιδιότητες και σχέσεις γεωγραφικών κατηγοριών

Σημασιολογικές Ιδιότητες	Σημασιολογικές σχέσεις
ΣΚΟΠΟΣ	IS-A
ΑΙΤΙΑ	ΕΙΝΑΙ ΤΜΗΜΑ ΤΟΥ
ΘΕΣΗ	ΑΠΟΤΕΛΕΙΤΑΙ ΑΠΟ
ΧΡΟΝΟΣ	ΓΕΙΤΟΝΕΥΕΙ ΜΕ
ΚΑΛΥΨΗ	ΠΕΡΙΒΑΛΛΕΤΑΙ ΑΠΟ
ΜΕΓΕΘΟΣ	ΣΥΝΔΕΕΤΑΙ ΜΕ

ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΕΠΙΛΥΣΗ ΕΤΕΡΟΓΕΝΕΙΩΝ

Η σύγκριση εννοιών συνίσταται στον προσδιορισμό των μεταξύ τους ομοιοτήτων και ετερογενειών. Η διαδικασία αυτή βασίζεται στα διαθέσιμα στοιχεία που περιγράφουν τη σημασιολογία των εννοιών, όπως είναι οι όροι και οι ορισμοί τους. Σύμφωνα όμως με τα παραπάνω, οι ορισμοί μπορούν να αναλυθούν περαιτέρω με βάση τις σημασιολογικές τους ιδιότητες και σχέσεις. Έτσι, αν υποθέσουμε ότι ο ορισμός μιας έννοιας αναλύεται σε ένα σύνολο σημασιολογικών ιδιοτήτων και σχέσεων και στις αντίστοιχες τιμές αυτών, τότε μια έννοια E_i μπορεί να εκφραστεί από την τριάδα $\langle O_{E_i}, \Sigma_{E_i}, \Gamma_{\sigma_i E_i} \rangle$, όπου O_{E_i} είναι ο όρος, Σ_{E_i} είναι το σύνολο των σημασιολογικών ιδιοτήτων και σχέσεων και $\Gamma_{\sigma_i E_i}$ το σύνολο των αντίστοιχων τιμών, δηλαδή:

$$\Sigma_{E_i} = \{ \sigma_{1E_i}, \sigma_{2E_i}, \dots, \sigma_{nE_i} \}$$

$$\Gamma_{\sigma_i E_i} = \{ \tau_{\sigma_1 E_i}, \tau_{\sigma_2 E_i}, \dots, \tau_{\sigma_n E_i} \}$$

Πίνακας 2: Παράδειγμα σημασιολογικών στοιχείων και τιμών γεωγραφικών κατηγοριών

ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΚΑΙ ΣΧΕΣΕΙΣ

	ΑΡΧΙΚΕΣ ΚΑΤΗΓΟΡΙΕΣ	IS-A	ΚΑΛΥΨΗ (COVER)	ΣΚΟΠΟΣ (PURPOSE)	ΙΔΙΟΣΥΣΤΑΣΙΑ (NATURE)	ΜΕΓΕΘΟΣ (SIZE)	ΘΕΣΗ (LOCATION)	ΕΙΝΑΙ ΤΜΗΜΑ ΤΟΥ (IS PART OF)	ΑΠΟΤΕΛΕΙΤΑΙ ΑΠΟ (HAS PART)	ΠΕΡΙΒΑΛΛΕΤΑΙ ΑΠΟ (SURROUNDED BY)
CYC	BODY OF WATER	body	water		natural or artificial					
	stream	body	fresh water		natural					
	river	stream	water		natural	large volume				
	lake	body	water							land-locked
	sea	body	salt water			as large as or larger than a lake				
	canal	waterway			boats or irrigation	artificial				
WordNet	BODY OF WATER	part	water					earth's surface		
	stream	body	running water		natural		on or under the earth			
	river	stream	water		natural	large stream (larger than a creek)				
	brook	stream	water		natural	smaller than river				
	lake	body	water (usually fresh)							land
	sea	body	salt water			large body				land
	WAY	artifact			affording passage from one place to another				a road or path	
	watercourse, waterway	conduit						through which water flows		
	canal	strip	water		boats or irrigation		long and narrow			
	headrace	waterway			feeds water to a mill or water wheel or turbine					
DIGEST	HYDROGRAPHY - INLAND WATER	waterway body					inland			
	stream	watercourse			natural					

Διαφορετικοί συνδυασμοί των στοιχείων O_{E_i} , Σ_{E_i} και $T_{\sigma_i E_i}$ οδηγούν σε ένα σύνολο πιθανών αποτελεσμάτων σύγκρισης μεταξύ όμοιων εννοιών όπως φαίνεται στον Πίνακα 3. Οι πιο σαφείς περιπτώσεις είναι η "ισοδυναμία" (ίδιοι όροι, σημασιολογικά στοιχεία και τιμές) και η "διαφορά" (διαφορετικοί όροι και σημασιολογικά στοιχεία).

Πίνακας 3: Διαφορετικές περιπτώσεις συνδυασμών των O_{E_i} , Σ_{E_i} , $T_{\sigma_i E_i}$.

Όροι	Σημασιολογικά στοιχεία	Τιμές	Αποτέλεσμα σύγκρισης
$O_{E1}=O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	ισοδυναμία
$O_{E1}=O_2$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	ομωνυμία-διαφορά
$O_{E1}=O_2$	$\Sigma_{E1}\neq\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	σύγκρουση
$O_{E1}=O_2$	$\Sigma_{E1}\neq\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	ομωνυμία-διαφορά
$O_{E1}\neq O_2$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	συνωνυμία - ισοδυναμία
$O_{E1}\neq O_2$	$\Sigma_{E1}\neq\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	σύγκρουση
$O_{E1}\neq O_2$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	σύγκρουση
$O_{E1}\neq O_2$	$\Sigma_{E1}\neq\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	διαφορά

Ο παραπάνω πίνακας βασίζεται στην υπόθεση ότι οι σημασιολογικές ιδιότητες και σχέσεις και οι αντίστοιχες τιμές τους μπορεί να είναι ίδιες ή διαφορετικές, οπότε η σύγκριση δύο εννοιών καταλήγει είτε στην ταύτιση (ισοδυναμία) είτε στη διαφορά. Όμως, η ισοδυναμία και η διαφορά δεν καλύπτουν όλες τις δυνατές περιπτώσεις σύγκρισης δύο εννοιών, οι οποίες είναι:

- ταύτιση (ισοδυναμία), όταν οι δύο έννοιες είναι ισοδύναμες
- διαφορά, όταν οι δύο έννοιες είναι διαφορετικές
- συμπερίληψη, όταν η μία έννοια είναι γενικότερη της άλλης
- επικάλυψη, όταν οι δύο έννοιες έχουν κοινά, χωρίς όμως η μία να είναι σαφώς γενικότερη της άλλης.

Επομένως, προκειμένου να καλύψουμε όλες τις δυνατές περιπτώσεις, υποθέτουμε ότι οι σημασιολογικές ιδιότητες και σχέσεις μπορεί να είναι ίδιες ($\Sigma_{E1}=\Sigma_{E2}$), διαφορετικές ($\Sigma_{E1}\neq\Sigma_{E2}$), υποσύνολο η μια της άλλης ($\Sigma_{E1}<\Sigma_{E2}$) ή επικαλυπτόμενες ($\Sigma_{E1}\otimes\Sigma_{E2}$) και οι τιμές τους μπορεί αντίστοιχα να είναι ίδιες ($T_{\sigma E1}=T_{\sigma E2}$), διαφορετικές ($T_{\sigma E1}\neq T_{\sigma E2}$), ειδικότερες ή γενικότερες ($T_{\sigma E1}<T_{\sigma E2}$) ή επικαλυπτόμενες ($T_{\sigma E1}\otimes T_{\sigma E2}$). Προφανώς, η σύγκριση των τιμών $T_{\sigma_i E_i}$ αναφέρεται μόνο στις κοινές σημασιολογικές ιδιότητες και σχέσεις. Για παράδειγμα, η περίπτωση $\Sigma_{E1}>\Sigma_{E2}$ και $T_{\sigma E1}=T_{\sigma E2}$, υπονοεί ότι οι κοινές σημασιολογικές ιδιότητες των εννοιών E_1 και E_2 , έχουν ίδιες τιμές. Το σύμβολο $T_{\sigma E1}\otimes T_{\sigma E2}$ αναφέρεται στην περίπτωση που δύο έννοιες έχουν επικαλυπτόμενες τιμές για κάποια σημασιολογική ιδιότητα ή σχέση ή στην περίπτωση που ισχύει $T_{\sigma_1 E1}>T_{\sigma_1 E2}$ και $T_{\sigma_2 E1}<T_{\sigma_2 E2}$, δηλαδή όταν η τιμή μιας σημασιολογικής ιδιότητας σ_1 είναι γενικότερη για την έννοια E_1 , ενώ η τιμή μιας άλλης σημασιολογικής ιδιότητας σ_2 είναι γενικότερη για την έννοια E_2 .

Επομένως, καταλήγουμε στον ακόλουθο Πίνακα όλων των δυνατών συνδυασμών μεταξύ των τριών στοιχείων O_{E_i} , Σ_{E_i} , $T_{O_iE_i}$, εκτός από την περίπτωση που οι σημασιολογικές ιδιότητες και σχέσεις είναι διαφορετικές (σαφής περίπτωση διαφορετικών εννοιών), οπότε δεν εξετάζονται περαιτέρω οι τιμές τους. Εκτός από τους δυνατούς συνδυασμούς των στοιχείων O_{E_i} , Σ_{E_i} , $T_{O_iE_i}$, ο πίνακας 4 περιλαμβάνει το αποτέλεσμα της σύγκρισης των στοιχείων O_{E_i} , Σ_{E_i} , $T_{O_iE_i}$ και την ενέργεια επίλυσης.

Ορισμένοι συνδυασμοί του πίνακα 4 δεν καταλήγουν σε μονοσήμαντο, αλλά σε δισημάντο αποτέλεσμα σύγκρισης. Επομένως, απαιτείται περαιτέρω διερεύνηση προκειμένου να διαπιστωθεί ποιο αποτέλεσμα επιφέρει η σύγκριση των συγκεκριμένων εννοιών. Για παράδειγμα, ο συνδυασμός $O_{E_1}=O_{E_2}$, $\Sigma_{E_1}>\Sigma_{E_2}$, $T_{O_{E_1}}=T_{O_{E_2}}$ μπορεί να αντιστοιχεί σε δύο περιπτώσεις. Στην πρώτη περίπτωση, η έννοια E_1 περιγράφεται με περισσότερες σημασιολογικές ιδιότητες και σχέσεις και επομένως είναι ειδικότερη έννοια από την E_2 . Στη δεύτερη περίπτωση, οι επιπρόσθετες σημασιολογικές ιδιότητες της έννοιας E_1 δεν επηρεάζουν τη σημασιολογία της, αλλά σχετίζονται με το σκοπό και τις ιδιαιτερότητες της οντολογίας στην οποία ανήκει. Για παράδειγμα, οι οντολογίες πεδίου περιλαμβάνουν συνήθως εξειδικευμένη γνώση (domain knowledge) η οποία δεν περιλαμβάνεται σε γενικές οντολογίες. Χαρακτηριστικό παράδειγμα αυτής της περίπτωσης αποτελούν οι κατηγορίες του CORINE Land Cover, αφού οι ορισμοί τους περιλαμβάνουν εξειδικευμένη γνώση για την καλύτερη αναγνώριση των κατηγοριών από τις δορυφορικές εικόνες.

Στη συνέχεια κάθε μία από τις περιπτώσεις σύγκρισης κατηγοριών πρέπει να αντιμετωπιστεί με τον κατάλληλο τρόπο. Οι τρεις πρώτες περιπτώσεις (ισοδυναμία, διαφορά, συμπερίληψη) επιλύονται εύκολα. Στην ισοδυναμία ορίζεται αντιστοιχία (ισότητα) μεταξύ των εννοιών των δύο οντολογιών, οπότε στην τελική οντολογία εμφανίζονται ως μία έννοια. Στην αντίθετη περίπτωση, όταν δηλαδή οι δύο έννοιες είναι διαφορετικές, τότε δεν ορίζεται καμιά αντιστοιχία και στην τελική οντολογία συμπεριλαμβάνονται και οι δύο έννοιες. Στην περίπτωση που η μια έννοια είναι ευρύτερη της άλλης, τότε ορίζεται μια σχέση εξειδίκευσης (is-a) στην τελική οντολογία. Η περίπτωση της εννοιολογικής επικάλυψης είναι δυσκολότερη να επιλυθεί. Γι' αυτήν την περίπτωση είναι απαραίτητη η εφαρμογή της διαδικασίας σημασιολογικής παραγοντοποίησης, προκειμένου να διαχωριστούν τα κοινά από τα διακριτά τμήματα των επικαλυπτόμενων εννοιών.

Πίνακας 4: Περιπτώσεις σύγκρισης των στοιχείων O_{Ei} , Σ_{Ei} , T_{oiEi}

Όροι	Σημσιολογικά στοιχεία	Τιμές	Αποτέλεσμα σύγκρισης	Ενέργεια επίλυσης
$O_{E1}=O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	ισοδυναμία	$E_1=E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	ομωνυμία-διαφορά	$E_1\neq E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη	$E_1>E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη	$E_1<E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}\neq\Sigma_{E2}$...	ομωνυμία-σύγκρουση	$E_1\neq E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	πιο λεπτομερής ορισμός ή συμπερίληψη	$E_1=E_2$ ή $E_1>E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	ομωνυμία-σύγκρουση	$E_1\neq E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη	$E_1>E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1<E_2$ ή $E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	πιο λεπτομερής ορισμός ή συμπερίληψη	$E_1=E_2$ ή $E_1<E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	ομωνυμία-σύγκρουση	$E_1\neq E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1>E_2$ ή $E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη	$E_1<E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	ισοδυναμία ή επικάλυψη	$E_1=E_2$ ή $E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	ομωνυμία-σύγκρουση	$E_1\neq E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1>E_2$ ή $E_1\emptyset E_2$
$O_{E1}=O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1<E_2$ ή $E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	συνωνυμία	$E_1=E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	διαφορά - σύγκρουση	$E_1\neq E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη	$E_1>E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}=\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη	$E_1<E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}\neq\Sigma_{E2}$...	διαφορά	$E_1\neq E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	συμπερίληψη ή ισοδυναμία	$E_1>E_2$ ή $E_1=E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	διαφορά	$E_1\neq E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1>E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}<\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1<E_2$ ή $E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	συμπερίληψη ή ισοδυναμία	$E_1<E_2$ ή $E_1=E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	διαφορά	$E_1\neq E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	επικάλυψη	$E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη ή επικάλυψη	$E_1>E_2$ ή $E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}>\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη	$E_1<E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}=T_{\sigma E2}$	συνωνυμία ή συμπερίληψη	$E_1=E_2$ ή $E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}\neq T_{\sigma E2}$	διαφορά	$E_1\neq E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}\emptyset T_{\sigma E2}$	διαφορά	$E_1\neq E_2$ ή $E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}>T_{\sigma E2}$	συμπερίληψη	$E_1>E_2$ ή $E_1\emptyset E_2$
$O_{E1}\neq O_{E2}$	$\Sigma_{E1}\emptyset\Sigma_{E2}$	$T_{\sigma E1}<T_{\sigma E2}$	συμπερίληψη	$E_1<E_2$ ή $E_1\emptyset E_2$

ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗ

Η Σημασιολογική Παραγοντοποίηση είναι η διαδικασία ανάλυσης-διάσπασης των αρχικών εννοιών στα συστατικά τους ή σε θεμελιώδεις υπο-έννοιες που ονομάζονται *σημασιολογικοί παράγοντες*. Οι υπο-έννοιες αυτές είναι σαφείς, χωρίς σημασιολογική επικάλυψη μεταξύ τους και αποτελούν τους οικοδομικούς λίθους της ολοκληρωμένης οντολογίας.

Η διαδικασία εξελίσσεται από τις πιο ειδικές στις πιο γενικές κατηγορίες (πίνακας 5). Αρχικά δημιουργείται ένας σημασιολογικός παράγοντας για κάθε έννοια που δεν έχει μερική επικάλυψη με κάποια άλλη. Για παράδειγμα, η έννοια "λίμνη" (lake) που ορίζεται ισοδύναμα και στις τρεις αρχικές οντολογίες σύμφωνα με τον πίνακα 2 (πίνακας 6) αναλύεται σε ένα σημασιολογικό παράγοντα g_3 . Εφόσον στις ισοδύναμες έννοιες αντιστοιχούν κοινοί σημασιολογικοί παράγοντες, ο σημασιολογικός παράγοντας g_3 αντιστοιχεί στην έννοια "λίμνη" (lake) και των τριών οντολογιών (CYC, WordNet και DIGEST).

Όταν ολοκληρωθεί η σημασιολογική παραγοντοποίηση των απλών εννοιών που αποτελούνται από ένα σημασιολογικό παράγοντα, η διαδικασία προχωρά και στις πιο γενικές έννοιες, οι οποίες αποτελούνται από περισσότερους σημασιολογικούς παράγοντες. Για παράδειγμα, η πιο σύνθετη έννοια "υδατόρευμα" (stream) του WordNet περιλαμβάνει τις ειδικότερες έννοιες "ποτάμι" και "ρυάκι". Επομένως, η έννοια αυτή αποτελείται από τους σημασιολογικούς παράγοντες g_1 και g_2 που αντιστοιχούν στις ειδικότερες έννοιες "ποτάμι" (river) και "ρυάκι" (brook) αντίστοιχα.

Σε αυτή τη φάση είναι απαραίτητη η γνώση των ιεραρχικών σχέσεων μεταξύ των εννοιών, τόσο της ίδιας οντολογίας όσο και μεταξύ διαφορετικών οντολογιών. Για τον προσδιορισμό των σχέσεων μεταξύ των εννοιών της ίδιας οντολογίας λαμβάνεται υπόψη η ιεραρχία των αρχικών οντολογιών (εφόσον ορίζεται). Για τον προσδιορισμό των ιεραρχικών σχέσεων μεταξύ των εννοιών διαφορετικών οντολογιών χρησιμοποιούνται τα αποτελέσματα σύγκρισης εννοιών που προέκυψαν από την προηγούμενη φάση της μεθοδολογίας. Ουσιαστικά, υλοποιούνται οι αποφάσεις επίλυσης των ετερογενειών που ελήφθησαν στο προηγούμενο στάδιο. Γι' αυτό το λόγο, η σύγκριση των εννοιών βάσει των όρων, των σημασιολογικών ιδιοτήτων και σχέσεων και των τιμών τους βοηθά ώστε η διαδικασία της σημασιολογικής παραγοντοποίησης να γίνεται με σαφή και αντικειμενικό τρόπο.

Επίσης, στον προσδιορισμό των ιεραρχικών σχέσεων μεταξύ των εννοιών τόσο της ίδιας οντολογίας όσο και μεταξύ διαφορετικών οντολογιών συμβάλλει και η γνώση των τιμών της σημασιολογικής σχέσης "IS-A". Η τιμή αυτής της σημασιολογικής σχέσης για κάθε έννοια φανερώνει τη γενικότερή της έννοια και σε σωστά δομημένες οντολογίες πρέπει να συμπίπτει με την ιεραρχία της οντολογίας.

Πίνακας 5: Σημασιολογική Παραγοντοποίηση

Αρχικές Κατηγορίες		Σημασιολογικοί Παράγοντες								
		g1	g2	g3	g4	g5	g6	g7	g8	g9
CYC	BODY OF WATER	x	x	x	x	x	x			
	stream	x	x							
	river	x								
	lake			x						
	sea				x					
	canal					x	x			
WordNet	BODY OF WATER	x	x	x	x					
	stream	x	x							
	river	x								
	brook		x							
	lake			x						
	sea				x					
	WAY					x	x		x	x
	waterway					x	x		x	
	canal					x	x			
headrace								x		
DIGEST	INLAND WATER	x	x	x		x	x	x		
	stream	x	x							
	canal					x				
	ditch						x	x		
	lake			x						

Πίνακας 6: Σημασιολογικά στοιχεία και τιμές της έννοιας "λίμνη".

	IS-A	ΚΑΛΥΨΗ (COVER)	ΠΕΡΙΒΑΛΛΕΤΑΙ ΑΠΟ (SURROUNDED BY)
<i>lake</i> (CYC)	body	water	land-locked
<i>lake</i> (WordNet)	body	water	land
<i>Lake</i> (DIGEST)	body	water	land

Η περίπτωση επικάλυψης μεταξύ δύο εννοιών διαφορετικών οντολογιών προσδιορίζεται με βάση τα αποτελέσματα σύγκρισης της προηγούμενης φάσης. Η επικάλυψη αντιμετωπίζεται με τη διάσπαση των επικαλυπτόμενων εννοιών σε τρεις σημασιολογικούς παράγοντες: ένας αντιστοιχεί στο κοινό τους τμήμα και οι άλλοι δύο στα μη κοινά τμήματα κάθε έννοιας.

Η χρήση των σημασιολογικών παραγόντων ως θεμελιωδών εννοιών για τη δημιουργία της ολοκληρωμένης οντολογίας συμβάλλει στην αντιμετώπιση των σύνθετων περιπτώσεων εννοιολογικής επικάλυψης και στην καλύτερη αναπαράσταση σύνθετων εννοιών. Για παράδειγμα, η έννοια "κανάλι" (canal) της οντολογίας CYC με τις έννοιες "κανάλι" (canal) και "όδευμα" (ditch) του DIGEST επικαλύπτονται σημασιολογικά, όπως φαίνεται και από τους ορισμούς τους:

- canal (CYC): artificial waterways created to be paths for boats, or for irrigation
- canal (DIGEST): a man-made or improved natural waterway used for transportation
- ditch (DIGEST): a channel constructed for the purpose of irrigation or drainage

Από τον πίνακα των σημασιολογικών στοιχείων αυτών των κατηγοριών προκύπτει ότι ενώ για τα υπόλοιπα σημασιολογικά στοιχεία οι τιμές είναι ισοδύναμες, οι τιμές της σημασιολογικής ιδιότητας ΣΚΟΠΟΣ (PURPOSE) είναι επικαλυπτόμενες. Επομένως, οι τιμές αυτής της ιδιότητας θα καθορίσουν τους σημασιολογικούς παράγοντες που θα προκύψουν. Έτσι ορίζονται τρεις σημασιολογικοί παράγοντες: ο σημασιολογικός παράγοντας g_5 αντιστοιχεί στην τιμή "transportation", ο σημασιολογικός παράγοντας g_6 αντιστοιχεί στην τιμή "irrigation" και ο σημασιολογικός παράγοντας g_7 αντιστοιχεί στην τιμή "drainage".

Πίνακας 7: Σημασιολογικά στοιχεία και τιμές των κατηγοριών "κανάλι" και "όδευμα"

	IS-A	ΚΑΛΥΨΗ (COVER)	ΣΚΟΠΟΣ (PURPOSE)	ΙΔΙΟΣΥΣΤΑΣΙΑ (NATURE)
<i>canal (CYC)</i>	way	water	transportation or irrigation	artificial
<i>canal (DIGEST)</i>	way	water	transportation	artificial or improved natural
<i>ditch (DIGEST)</i>	way		irrigation or drainage	artificial

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η σημασιολογική ολοκλήρωση οντολογιών προϋποθέτει τον εντοπισμό και την επίλυση των ετερογενειών μεταξύ των αρχικών εννοιών. Όμως, η διαδικασία αυτή πραγματοποιείται συνήθως εμπειρικά και βασίζεται σε στοιχεία των κατηγοριών που δεν περιγράφουν πλήρως τη σημασιολογία τους, όπως οι όροι ή κάποια μη σημασιολογικά χαρακτηριστικά, γεγονός που εισάγει σφάλματα στην τελική οντολογία.

Η παρούσα εργασία προτείνει μια μεθοδολογία για τον εντοπισμό και την επίλυση της ετερογένειας μεταξύ των αρχικών εννοιών λαμβάνοντας υπόψη τη σημασιολογία που ενυπάρχει στους ορισμούς. Η μεθοδολογία αυτή αφορά κυρίως υπάρχουσες οντολογίες από τις οποίες απουσιάζουν άλλα στοιχεία περιγραφής της σημασιολογίας όπως ιδιότητες, αξιώματα, λειτουργίες. Βασίζεται στην παραδοχή ότι οι ορισμοί περιλαμβάνουν πληθώρα σημασιολογικών πληροφοριών οι οποίες είναι δυνατόν να εξαχθούν σύμφωνα με συγκεκριμένους κανόνες και να τυποποιηθούν σε ένα σύνολο σημασιολογικών ιδιοτήτων και σχέσεων. Ο εντοπισμός και η επίλυση των ετερογενειών βασίζεται πλέον σε αυτό το τυποποιημένο σύνολο σημασιολογικών στοιχείων και καταλήγει σε μία από τις τέσσερις δυνατές περιπτώσεις σύγκρισης εννοιών (ταύτιση, διαφορά, συμπερίληψη, επικάλυψη).

Δύο είναι τα βασικά πλεονεκτήματα της μεθοδολογίας. Καταρχήν, αξιοποιεί τους ορισμούς που αποτελούν την πλουσιότερη πηγή σημασιολογικών πληροφοριών για κάθε έννοια. Επιπλέον, η διαδικασία τυποποιείται σε τέτοιο βαθμό, ώστε να διεξάγεται με σαφή και αντικειμενικό τρόπο, αποκλείοντας υποκειμενικότητες που προκαλούνται από εμπειρικούς τρόπους αντιμετώπισης του προβλήματος.

Παρ' όλ' αυτά, μια διαδικασία επεξεργασίας της σημασιολογίας εννοιών δεν είναι δυνατόν να αυτοματοποιηθεί πλήρως. Γι' αυτό το λόγο, η πλήρης αυτοματοποίηση δεν αποτελεί έναν εφικτό στόχο της μεθοδολογίας. Στους μελλοντικούς στόχους συγκαταλέγονται ενέργειες που θα ενισχύουν την τυποποίηση της διαδικασίας. Η επεξεργασία και η απλούστευση των τιμών των σημασιολογικών στοιχείων προκειμένου να διευκολύνεται η σύγκρισή τους αποτελεί έναν από αυτούς. Η επεξεργασία αυτή, εκτός από την εύρεση συνωνύμων, είναι δυνατόν να περιλαμβάνει ανάλυση σύνθετων λέξεων στα συστατικά τους (π.χ., "waterway"), μια πολύπλοκη διαδικασία που όμως φανερώνει επιπλέον σημασιολογικές ιδιότητες και σχέσεις (Vanderwende, 1995).

Βιβλιογραφία

- CYCORP, Inc., Upper Cyc Ontology, <http://www.cyc.com/>
- Digital Geographic Information Working Group (DGIWG) (2000) *The Digital Geographic Information Exchange Standard (DIGEST)*, Part 4, Annex A: Feature Codes, Edition 2.1.
- Dolan, W.B., Vanderwende L. και Richardson S.D. (1993) "Automatically Deriving Structured Knowledge Base from On-line Dictionaries", Proceedings of the Pacific Association for Computational Linguistics, Vancouver, British Columbia, April 21-24.
- Jensen K. και Binot J.L. (1987) "Disambiguating prepositional phrase attachments by using on-line dictionary definitions", *Computational Linguistics* 13 (3-4): 251-60.

- Κάβουρας Μ. (2002) "Γεωγραφικές Οντολογίες και Διαλειτουργικότητα", *Αειχώρος*, 1(1): 92-115.
- Kavouras M. και Kokla M. (2002) "A method for the formalization and integration of geographical categorizations", *International Journal of Geographical Information Science*, 16 (5): 439-453.
- Kavouras M., Kokla M. και Tomai E. (2003) "Determination, Visualization and Interpretation of Semantic Similarity among Different Geographic Ontologies", 6th AGILE Conference on Geographic Information Science, Lyon, France.
- McGuinness D. L., Fikes R., Rice J. και Wilder S. (2000) "An environment for merging and testing large ontologies", Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR2000).
- Mitra P. και Wiederhold G. (2002) "Resolving Terminological Heterogeneity In Ontologies", Proceedings of Workshop on Ontologies and Semantic Interoperability at the 15th European Conference on Artificial Intelligence (ECAI), Euzenat J., Gomez-Perez A., Guarino N. και Stuckenschmidt H. (επ.), Lyon, France.
- Noy N. και Musen M. (2000) "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment", Proceedings of the National Conference on Artificial Intelligence (AAAI-2000), Austin, TX, USA.
- Vanderwende L. (1995) "The Analysis of Noun Sequences using Semantic Information Extracted from On-Line Dictionaries", Ph.D. thesis, Faculty of the Graduate School of Arts and Sciences, Georgetown University, Washington, D.C.
- WORDNET - a Lexical Database for English, Cognitive Science Laboratory, Princeton University, <http://www.cogsci.princeton.edu/~wn/>

Μαργαρίτα Κόκλα

*Εθνικό Μετσόβιο Πολυτεχνείο, Σχολή Αγρονόμων και Τοπογράφων Μηχανικών,
Ηρώων Πολυτεχνείου 9, 157 80 Ζωγράφου, Αθήνα, e-mail: mkokla@survey.ntua.gr*

Μαρίνος Κάβουρας

*Εθνικό Μετσόβιο Πολυτεχνείο, Σχολή Αγρονόμων και Τοπογράφων Μηχανικών,
Ηρώων Πολυτεχνείου 9, 157 80 Ζωγράφου, Αθήνα, e-mail: mkav@survey.ntua.gr*

