

Συστήματα Οργάνωσης Γνώσης

Χρήστος Παπαθεοδώρου (paratheodor@ionio.gr)

Αναπληρωτής Καθηγητής

Ομάδα Βάσεων Δεδομένων και Πληροφοριακών Συστημάτων,
Τμήμα Αρχιονομίας – Βιβλιοθηκονομίας, Ιόνιο Πανεπιστήμιο
και

Μονάδα Ψηφιακής Επιμέλειας,

Ινστιτούτο Πληροφοριακών Συστημάτων και Προσομοίωσης
Ερευνητικό Κέντρο «Αθηνά»



DBIS

database & information systems group
ionian university

Εισαγωγικά

- Οργάνωση Γνώσης:
 - περικλείει κάθε τύπο και μέθοδο ευρετηρίασης, περιγραφής, ταξινόμησης, διαχείρισης τεκμηρίων, που βοηθούν στην ανάκτηση της πληροφορίας.
 - διανοητική διαδικασία που ενισχύει την πρόσληψη του κόσμου μέσω της ανάπτυξης νοητικών δομών
 - οι νοητικές δομές αυξάνουν την αξία της πληροφορίας αφού καθιστούν ευχερέστερη την αναζήτηση και τη διάχυσή της στις κοινότητες χρηστών
 - παρέχει μια βαθύτερη αντίληψη για το θεματικό περιεχόμενο του τομέα ενδιαφέροντος των χρηστών των υπηρεσιών πληροφόρησης
- Συστήματα Οργάνωσης Γνώσης (ΣΟΓ)
 - εξελιγμένα εργαλεία αναζήτησης ψηφιακής πληροφορίας
 - ταξινομικά συστήματα, καταλόγους θεματικών επικεφαλίδων, αρχεία καθιερωμένων ονομάτων, σημασιολογικά δίκτυα και οντολογίες

Προσεγγίσεις για την Οργάνωση

Γνώσης (1/3)

- Παραδοσιακή: Τα συστήματα ταξινόμησης που χρησιμοποιούν οι Βιβλιοθήκες (π.χ. DDC, LCC, UDC) για να οργανώσουν την επιστημονική βιβλιογραφία. Βασική αρχή οργάνωσης από τα γενικότερα στα ειδικότερα, ενώ τα αντικείμενα ταξινομούνται σε μια μοναδική κατηγορία
- Αναλυτικο-συνθετική (πολυεδρική, facet-analytical): Θεσπίσθηκε από τον Ranganathan, και βασίζεται στην εφαρμογή λογικών αρχών. Η θεματική περιγραφή ενός αντικειμένου μπορεί να αναφέρεται σε περισσότερες από μια έννοιες. Οι έννοιες αναλύονται σε επιμέρους κατηγορίες (ανάλυση εννοιών) οι οποίες συνθέτουν την περιγραφή του θέματος (σύνθεση). Έτσι το θέμα δεν ταξινομείται σε μια μοναδική κατηγορία. Βασικές κατηγορίες για την ανάλυση των θεμάτων είναι οι:
 - Προσωπικότητα (Personality): το βασικό διακριτικό γνώρισμα που χαρακτηρίζει ένα θέμα (π.χ. όνομα)
 - Υλικό (Matter): το φυσικό υλικό από το οποίο αποτελείται μια έννοια
 - Ενέργεια (Energy): κάθε ενέργεια στην οποία ένα θέμα μπορεί να συμμετέχει (π.χ. το ξύλο στην επίπλωση)
 - Χώρος (Space): ο γεωγραφικός τόπος που μπορεί αναφέρεται το θέμα
 - Χρόνος (Time): is the period associated with a subject.

Προσεγγίσεις για την Οργάνωση

Γνώσης (2/3)

- Ανάκτηση Πληροφοριών: Προέρχονται από το χώρο του full text retrieval και της αυτόματης κατηγοριοποίησης κειμένων, με ικανοποιητικά αποτελέσματα. Οι σημασιολογικές σχέσεις βασίζονται σε στατιστικές συσχετίσεις μεταξύ των αντικειμένων με βάση τις συχνότητες εμφάνισης των λέξεών τους. Ακολουθείται μια μεθοδολογία από τα ειδικότερα προς τα γενικότερα (bottom-up)
- Χρηστοκεντρική: Οι σημασιολογικές σχέσεις βασίζονται σε εμπειρικές έρευνες χρηστών και όχι στην επιστημονική βιβλιογραφία
- Βιβλιομετρική: Δύο αντικείμενα (τεκμήρια) συσχετίζονται εννοιολογικά όταν το ένα αναφέρει το άλλο, ή όταν κάποιο τρίτο αντικείμενο αναφέρει και τα δύο (co-citation). Είναι μια μορφή κοινωνικής/συνεργατικής οργάνωσης γνώσης

Προσεγγίσεις για την Οργάνωση Γνώσης (3/3)

- Ανάλυση πεδίου: Βασίζεται στην όσο πιο πλούσια γνώση του πεδίου που επιτυγχάνεται από την ευρετηρίαση, την ανάπτυξη θησαυρών, τις μελέτες χρηστών, τις βιβλιομετρικές έρευνες ενός επιστημονικού πεδίου
- Επιστημολογικές προσεγγίσεις που βασίζονται στην ερμηνευτική και σημειωτική ανάλυση των αντικειμένων
- Σύγχρονες προσεγγίσεις που βασίζονται στην τεχνολογία της ηλεκτρονικής δημοσίευσης, όπως στις γλώσσες σημειοθέτησης, την αρχιτεκτονική των δικτυακών τόπων, την αναπαράσταση των αντικειμένων και την τυπολογία τους κ.λπ.

Χαρακτηριστικά Συστημάτων Οργάνωσης Γνώσης

- Χρησιμοποιούνται για την οργάνωση του υλικού μιας συλλογής τεκμηρίων σε θεματικές κατηγορίες
 - Θεωρούν από μια συγκεκριμένη οπτική γωνία μια συλλογή και τα αντικείμενά της
 - Μια οντότητα μπορεί να χαρακτηριστεί με διαφορετικούς τρόπους ανάλογα με το ΣΟΓ που χρησιμοποιείται
- Λειτουργούν ως γέφυρα μεταξύ του υλικού και των αναγκών του χρήστη και καθοδηγούν / διευκολύνουν την αναζήτηση του χρήστη
 - Τα ΣΟΓ χρησιμοποιούνται για τη συσχέτιση εννοιών με αντικείμενα μιας συλλογής από εξειδικευμένα πρόσωπα. Μια έννοια αποδίδεται σε ένα αντικείμενο
 - Οι χρήστες που αναζητούν αντικείμενα χρησιμοποιώντας ΣΟΓ, συνδέουν τις έννοιες που αντιλαμβάνονται τον κόσμο με τις αναπαραστάσεις τους στο ΣΟΓ

Συστήματα Οργάνωσης Γνώσης (1/5)

- Κατάλογοι όρων :
 - Καταλόγοι καθιερωμένων ονομάτων (*Authority Files*): Ελεγχόμενα λεξιλόγια των πολλαπλών μορφών ονομάτων ή οντοτήτων ενός τομέα ή πεδίου (Library of Congress Name Authority File, Getty Geographic Authority File)
 - Γλωσσάρια (*Glossaries*): Κατάλογοι ονομάτων συνήθως ενός συγκεκριμένου γνωστικού τομέα ή έργου με τους ορισμούς τους
 - Λεξικά (*Dictionaries*): Αλφαβητικοί κατάλογοι λέξεων, των ορισμών τους, της προέλευσης τους, των διαφορετικών μορφών τους, των συνωνύμων, κ.λπ.
 - Γεωγραφικά λεξικά (*Gazetters*): Κατάλογοι ονομάτων τοποθεσιών που συνοδεύονται από τις γεωγραφικές συντεταγμένες τους

Συστήματα Οργάνωσης Γνώσης (2/5)

- Ταξινομικά συστήματα και κατηγορίες :
 - Καταλόγοι θεματικών επικεφαλίδων (Subject Headings): Ελεγχόμενα λεξιλόγια που αντιπροσωπεύουν τα θέματα των αντικειμένων μιας συλλογής με περιορισμένη ιεραρχική διάρθρωση (LCSH, MeSH)
 - Ταξινομικά συστήματα, ταξονομίες και σχήματα κατηγοριοποιήσεων (Classification Schemes, Taxonomies, Categorization Schemes): Κατατάσσουν ή ομαδοποιούν θέματα σε μεγάλες κατηγορίες είτε με ιεραρχική διάταξη είτε με χρήση αλφαριθμητικής σημειογραφίας χωρίς ρητά εκφρασμένες τις σχέσεις ανάμεσα στους όρους (DDC, UDC)

Συστήματα Οργάνωσης Γνώσης (3/5)

- Κατάλογοι σχετιζόμενων όρων :
 - Θησαυροί (Thesauri): Υποδεικνύουν τις σχέσεις που υπάρχουν ανάμεσα σε όρους και κύρια τις σχέσεις ιεραρχίας, ισοδυναμίας και τις σχέσεις συσχέτισης (AAT)
 - Σημασιολογικά δίκτυα (Semantic networks): Δομούν τις έννοιες και τους όρους όχι σε ιεραρχίες αλλά με τη μορφή δικτύου (με κόμβους τις έννοιες και ακμές τις μεταξύ τους σχέσεις) ενώ οι σχέσεις ανάμεσα στους όρους είναι ευρύτερες σε σχέση με του θησαυρού (WorldNet)
 - Οντολογίες (Ontologies): Είναι από τα πιο πρόσφατα συστήματα και αναπτύσσονται κυρίως στο πλαίσιο του Σημασιολογικού Ιστού ως ειδικά μοντέλα εννοιών (specific concept models). Αναπαριστούν πολύπλοκες σχέσεις ανάμεσα σε αντικείμενα και περιλαμβάνουν κανόνες και αξιώματα ενώ περιγράφουν τη γνώση ενός συγκεκριμένου τομέα (SUMO)

Συστήματα Οργάνωσης Γνώσης (4/5)

- Άλλοι τύποι συστημάτων οργάνωσης γνώσης είναι:
 - Θεματικοί χάρτες (Topic maps): Βασίζονται στην αρχή ότι τα αντικείμενα (occurrences) περιστρέφονται γύρω από δομές γνώσης που είναι τα θέματα (topics) με τις μεταξύ τους συσχετίσεις. Περιγράφουν δομές γνώσης και η πλοήγηση σε αυτές οδηγεί στον εντοπισμό αντικειμένων
 - Εννοιολογικοί χάρτες (Concept maps): Οπτική αναπαράσταση των εννοιών ενός και των σχέσεών τους. Οι έννοιες αποτελούν κόμβους σε ένα γράφο και συνδέονται με ακμές που έχουν ετικέτες της μορφής «εγείρει το» , «έχει σαν αποτέλεσμα», «απαιτείται από», «συνεισφέρει σε»
 - Βιβλιομετρικοί χάρτες (Bibliometric maps): Με βάση τις βιβλιογραφικές αναφορές που συνυπάρχουν σε επιστημονικά τεκμήρια (co-citation analysis) μπορούν να κατασκευαστούν χάρτες που αποτυπώνουν την τρέχουσα γνώση (Atlas of Science)
 - Δακτύλιοι συνωνύμων (synonym ring): Συνώνυμοι όροι συνδέονται μεταξύ τους και σχηματίζουν «κλάσεις ισοδυναμίας». Οι κλάσεις ισοδυναμίας έχουν προσδιορισθεί αναλυτικά (π.χ. Syntactic variants, Phrase variants, Quasi-synonyms, Synonyms, Absolute synonyms, Cognitive synonyms, Contextual synonyms)

Συστήματα Οργάνωσης Γνώσης (5/5)

- Υπερκείμενο
- Τυπολογίες (Typologies): Το υλικό ταξινομείται με βάση τον τύπο του (π.χ. οι τύποι των ΣΟΓ)
- Βιβλιογραφικοί κατάλογοι (OPAC, WebOPAC)
- Κοινωνικές ταξονομίες (folksonomies, folk+taxonomy): Επισημειώσεις περιεχομένου (αντικειμένων) με σκοπό τη θεματική κατηγοριοποίησή του που εισάγονται από κοινότητες χρηστών και έχουν τη μορφή ετικετών/λέξεων κλειδιών (tags, tag cloud) χωρίς μεταξύ τους συσχετίσεις (Delicious)
- Βιβλιογραφίες, εγκυκλοπαίδειες, γλώσσες και συμβολικά συστήματα
- Εννοιολογικά συστήματα και θεωρίες (για συγκεκριμένα επιστημονικά πεδία)
- Κανόνες, μοντέλα δρώντων και διεργασιών, τομείς κοινωνικής δραστηριότητας

Ορισμοί (1/2)

- Έννοια (concept) είναι η νοητική αντιπροσώπευση ενός ιδιαίτερου αντικειμένου ή ενός συνόλου ιδιαίτερων αντικειμένων που έχουν ορισμένες κοινές ιδιότητες
- Ορος (term) είναι μια λέξη ή φράση που αντιπροσωπεύει και προσδιορίζει μια έννοια.
- Πλάτος (έκταση) μίας έννοιας: το σύνολο όλων των αντικειμένων στα οποία αντιστοιχεί μία έννοια
 - π.χ. {ακαδημαϊκές, σχολικές, ειδικές, ...} βιβλιοθήκες
- Βάθος (ένταση) μίας έννοιας: το σύνολο των ουσιωδών χαρακτηριστικών που συνιστούν την έννοια
 - π.χ. στην έννοια {ξύλινο τραπέζι} το χαρακτηριστικό της σύνθεσης (ξύλινο) αποτελεί ουσιώδες χαρακτηριστικό και διακρίνει την έννοια τραπέζι
- Ατομική έννοια: αντιστοιχεί σε ένα μόνο αντικείμενο {νήσος Κέρκυρα}
- Γενική έννοια: αντιστοιχεί σε δύο ή περισσότερα αντικείμενα που αποτελούν ομάδα με κοινές ιδιότητες {νήσος}

Ορισμοί (2/2)

- Χαρακτηριστικό (attribute, characteristic) μιας έννοιας είναι οποιαδήποτε από τις ιδιότητες ή σχέσεις που συνιστούν μία έννοια
 - (π.χ. ένα χαρακτηριστικό της έννοιας δέντρο είναι ο ξυλώδης κορμός)
- Ενδογενή χαρακτηριστικά (σχήμα, μέγεθος, υλικό, χρώμα)
 - π.χ. οδοντωτός (τροχός), κόκκινο (φως), ξύλινο (τραπέζι)
- Εξωγενή χαρακτηριστικά (προέλευση, προορισμός, λειτουργία, θέση, εφευρέτης, τοποθέτηση)
 - π.χ. πεντελικό (μάρμαρο), ευκλείδεια (γεωμετρία), δοκιμαστικός (σωλήνας)

Ορισμοί εννοιών

- Ορισμός μιας έννοιας είναι η λεκτική περιγραφή της έννοιας για διαφοροποίηση από συναφείς έννοιες
- Εντατικός: περιγράφει το βάθος μιας έννοιας
 - π.χ ισόπλευρο τρίγωνο είναι το τρίγωνο που έχει και τις 3 πλευρές ίσες
- Εκτατικός: απαρίθμηση των υποτασσόμενων εννοιών, πλάτος
 - π.χ. τρίγωνο: ισόπλευρο, ισοσκελές, σκαληνό
- Εμμεσος: ορισμός μέσω παραδείγματος

Λεξιλόγιο

- Σύνολο όρων
- Ελεγχόμενο λεξιλόγιο (controlled vocabulary): μόνο συγκεκριμένοι όροι ή είδη όρων επιτρέπεται να συμμετέχουν στις δηλώσεις θεμάτων
 - Ονομάζονται περιγραφείς (descriptors) ή προτιμώμενοι όροι (preferred terms)

Ενότητες μαθήματος

- Εισαγωγή στα ΣΟΓ (Χ. Παπαθεοδώρου).
- Κωδικοποίηση πληροφορίας με τη γλώσσα XML (Χ. Παπαθεοδώρου).
- Εισαγωγή στα μεταδεδομένα. Το πρότυπο μεταδεδομένων Dublin Core – προφίλ εφαρμογών. Τα πρότυπα MODS και EAD (Χ. Παπαθεοδώρου).
- Θεματικοί όροι και Θεματικές γλώσσες, προ-συνδυασμένες, υστερο-συνδυασμένες θεματικές γλώσσες. Σημασιολογικές σχέσεις στις θεματικές γλώσσες. Αρχεία καθιερωμένων όρων - το πρότυπο MADS (Χ. Παπαθεοδώρου).
- Ταξονομίες και ταξινομικά συστήματα, τα πρότυπα DDC και UDC. Πολυεδρική ταξινόμηση (Κ. Κακάλη).
- Θεματικές επικεφαλίδες. Το πρότυπο των θησαυρών, παραδείγματα και εφαρμογές θησαυρών. Πολυεδρικοί θησαυροί (Κ. Κακάλη).
- Εισαγωγή στο σημασιολογικό ιστό. Το πρότυπο RDF (Χ. Παπαθεοδώρου).
- Κωδικοποίηση θησαυρών με το πρότυπο SKOS. Θεματικοί χάρτες και το πρότυπο XTM (προσκεκλημένοι ομιλητές).
- Οντολογίες, γλώσσες οντολογιών RDF Schema και OWL (Χ. Παπαθεοδώρου).
- Επερωτήσεις και συλλογισμοί στο σημασιολογικό ιστό (Χ. Παπαθεοδώρου). Αναπαράσταση γνώσης με κανόνες και πλαίσια και στο σημασιολογικό ιστό – οι γλώσσες RuleML και F-Logic (Χ. Παπαθεοδώρου).
- Κοινωνικές ταξονομίες (folksonomies) στο web2.0 (Κ. Κακάλη).

Διδασκαλία και εξέταση

- Διδασκαλία
 - Δευτέρα 12.00 – 15.00
- Ιστοσελίδα μαθήματος
 - <http://eclass.aueb.gr/courses/INF180/>
 - στα Έγγραφα θα βρείτε βιβλιογραφία (άρθρα και υπερ-συνδέσμους) και τις σημειώσεις
 - Μπορείτε να το επιλέξετε από τον κατάλογο μαθημάτων <http://eclass.aueb.gr/modules/auth/opencourses.php?fc=12>
- Εξέταση
 - Παρουσίαση ατομικής εργασίας σε θέματα που θα διανεμηθούν μαζί με αρχική βιβλιογραφία
 - Διάρκεια παρουσίασης 20 - 30 λεπτά ανάλογα με το πλήθος των εργασιών
 - Η εξέταση θα γίνει σε 1 ή 2 ημέρες, ανάλογα με το πλήθος των εργασιών
 - Γραπτή εξέταση

Πρότυπα

- ΕΛΟΤ 402. Αρχές και μέθοδοι ορολογίας
- ΕΛΟΤ 414. Περιλήψεις για τις δημοσιεύσεις και την τεκμηρίωση
- ΕΛΟΤ 561-1. Λεξιλόγιο Ορολογίας
- ΕΛΟΤ 720 :1993. Φύλλα περιλήψεων σε περιοδικές εκδόσεις. (ISO 5122-1979)
- ΕΛΟΤ 743. Μετατροπή του ελληνικού αλφαβήτου με λατινικούς χαρακτήρες
- ΕΛΟΤ 1213:1994. Ευρετήριο ενός δημοσιεύματος. (ISO 999-1996)
- ΕΛΟΤ 1312:1993 (ΣΕΠ). Θεματική ανάλυση τεκμηρίων : Μέθοδοι εξέτασης τεκμηρίων, προσδιορισμού των θεμάτων τους και επιλογής των όρων ευρετηρίασης (ISO 5963-1985)
- ΕΛΟΤ 1321:1993 (ΣΕΠ) Κατευθυντήριες οδηγίες για τη συγκρότηση και ανάπτυξη μονόγλωσσων θησαυρών. (ISO 2788-1986)
- ΕΛΟΤ 1381-1 (ΣΕΠ). Λεξιλόγιο - Μέρος 1 : Βασικές έννοιες. (ISO 5127/1-1983)
- ΕΛΟΤ 1381-2-1997(ΣΕΠ) Λεξιλόγιο - Μέρος 2 : Παραδοσιακά τεκμήρια. (ISO 5127/2-1983)

Βιβλιογραφία

- Gail Hodge, «Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files», <http://www.clir.org/pubs/reports/pub91/contents.html>
- Networked Knowledge Organization, Systems/Services, <http://nkos.slis.kent.edu/>
- Knowledge organization systems (KOS), http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm
- Knowledge Organization, http://www.db.dk/bh/lifeboat_ko/concepts/knowledge_organization.htm
- Birger Hjørland, “What is Knowledge Organization (KO)?” , *Knowledge Organization. International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation* 35(2/3):pp. 86-101 (υπάρχει στο eclass)
- Birger Hjørland, “Semantics and Knowledge Organization”, *Annual Review of Information Science and Technology* 41:367 -405 (υπάρχει στο eclass)