



Semantic Web

Εισαγωγή

- ▶ Στην ερευνητική κοινότητα έχουν προταθεί τεχνικές που βασίζονται στη σημασιολογική (semantic) αναπαράσταση της ανάκτησης πληροφοριών
- ▶ Παράδειγμα: η σημασιολογική περιγραφή μιας σελίδας με αναγνωριστικά των χρηστών, οργανισμών, τοποθεσιών, κ.λπ.
- ▶ Αυτά τα αναγνωριστικά θεωρούνται ως σημασιολογικά στοιχεία (semantic features)
- ▶ Συνήθως, έχουμε ένα μικρό αριθμό στοιχείων – πολλαπλές τιμές αναπαριστούν την ίδια έννοια π.χ. GR, Hellas, Greece
- ▶ Ένα άλλο πλεονέκτημα είναι πως έχουμε ένα δομημένο τρόπο αναπαράστασης



Εισαγωγή

- ▶ Γενικά, όταν οι χρήστες θέτουν ένα ερώτημα, βάζουν ένα μικρό αριθμό όρων
- ▶ Συνήθως, αυτές είναι οι λέξεις που αναμένουν οι χρήστες να δουν στα αποτελέσματα
- ▶ Αυτό οδηγεί σε ένα πρόβλημα: *index term synonymy*: δεν θα χρησιμοποιούν όλα τα έγγραφα τις ίδιες λέξεις
- ▶ Συνεπώς, δεν θα εξαχθούν όλα τα έγγραφα που περιγράφουν την ίδια έννοια
- ▶ Επίσης, οι όροι στα ερωτήματα μπορεί να έχουν πολλές σημασίες: *query term polysemy*: αφού οι μηχανές αναζήτησης δεν μπορούν να καταλάβουν τις διαφορετικές έννοιες των λέξεων, είναι προφανές ότι θα επιστρέψουν και άσχετα αποτελέσματα



Εισαγωγή

- ▶ Επιπρόσθετα, οι μηχανές αναζήτησης δεν γνωρίζουν τους σημασιολογικούς σύνδεσμούς ανάμεσα στους όρους
- ▶ Ένας περιορισμός των κλασικών προσεγγίσεων είναι το ότι δεν μπορούμε να χειριστούμε ερωτήματα που απαιτούν γνώση και δεδομένα που δεν είναι διαθέσιμα στα έγγραφα
- ▶ Άλλος περιορισμός είναι το γεγονός ότι οι μηχανές αναζήτησης δεν αποδίδουν καλά όταν απαιτείται κάποιου είδους συλλογιστική πάνω στα δεδομένα



Οντολογίες

- ▶ Όταν προσπαθούμε να χτίσουμε με φορμαλισμό τη γνώση που επιτρέπει την αυτόματη μετάφραση και συλλογιστική τότε πρέπει να κατασκευάσουμε κάποια συγκεκριμένα μοντέλα
- ▶ Τέτοια μοντέλα καλούνται οντολογίες (ontologies) στο τομέα της σημασιολογικής αναπαράστασης της πληροφορίας
- ▶ Οι οντολογίες χρησιμοποιούνται για να προσφέρουν τα σχετικά semantics στο μοντέλο μας
- ▶ Τα semantics αυτά υιοθετούνται ώστε να γίνει κατανοητή η πληροφορία από τις μηχανές



Οντολογίες

- ▶ Η συσχέτιση της πληροφορίας με οντολογίες τις κάνει πιο εύκολα επεξεργάσιμες από τις μηχανές
- ▶ Οι οντολογίες μπορούν να χρησιμοποιηθούν για indexing, querying, referencing πάνω σε μη-σημασιολογικά μοντέλα π.χ. βάσεις δεδομένων
- ▶ Μια οντολογία μπορεί να χαρακτηριστεί πλειάδα:
 $O = \langle C, R, I, A \rangle$
- ▶ C: ένα σύνολο κλάσεων (classes) που αναπαριστούν έννοιες
- ▶ R: ένα σύνολο σχέσεων/συσχετίσεων (αλλιώς αναφέρονται ως ιδιότητες – properties, κατηγορήματα - predicates)
- ▶ I: ένα σύνολο από στιγμιότυπα (instances) για τις διαθέσιμες κλάσεις που μπορούν να συνδεθούν με άλλα στιγμιότυπα ή συσχετίσεις
- ▶ A: ένα σύνολο από αξιώματα



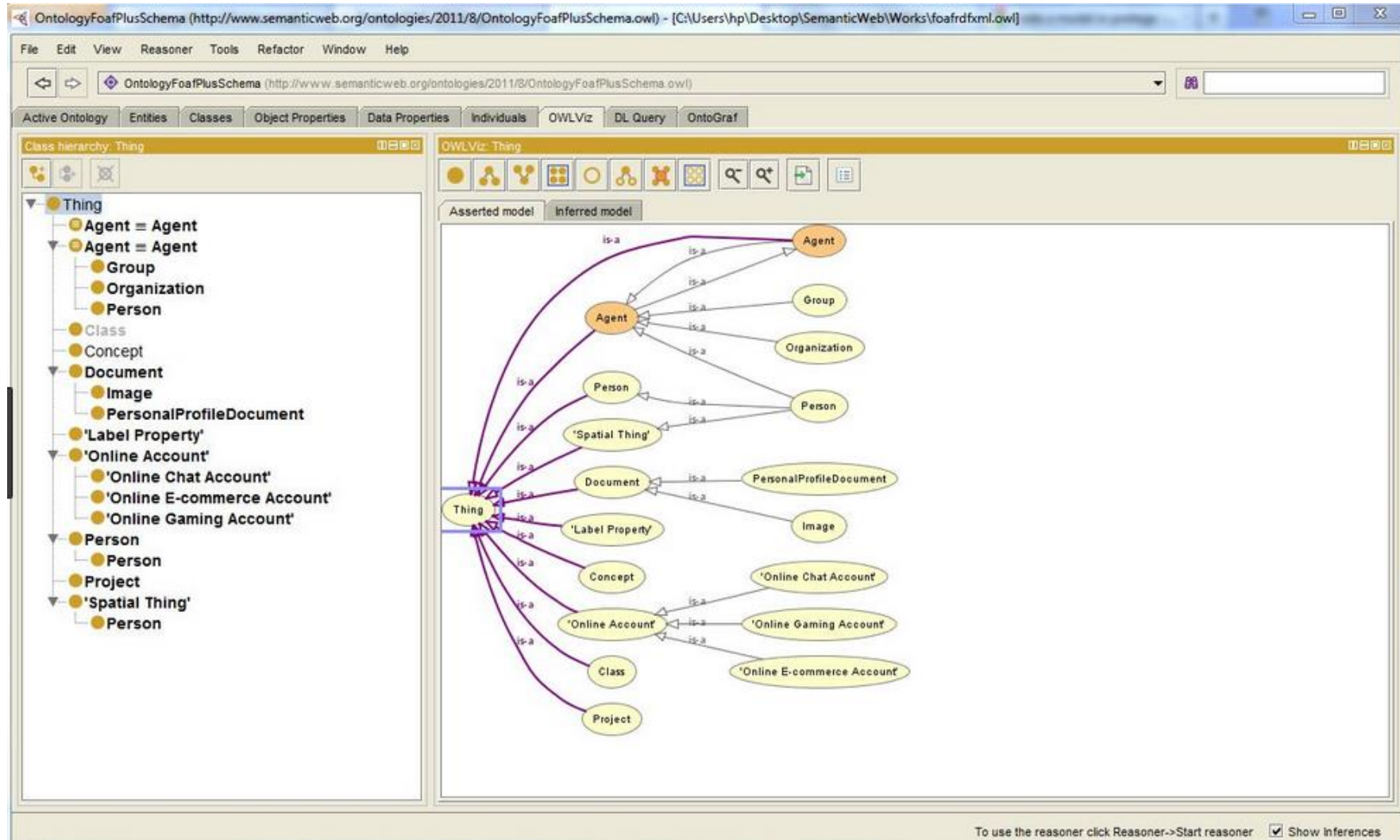
Οντολογίες

▶ Παραδείγματα:

- ▶ C: Product, Vehicle
- ▶ R: Product hasPrice Price, Vehicle hasHeight Height
- ▶ I: product_2 compatibleWith product_3, product_2 hasPrice 170
- ▶ A: ένα προϊόν έχει τιμή πάνω από 150 €, τότε δεν έχει έξοδα αποστολής



Οντολογίες



Οντολογίες

- ▶ Οι οντολογίες μπορούν να κατηγοριοποιηθούν σε *lightweight* και *heavyweight* ανάλογα με την πολυπλοκότητα της γλώσσας που θα χρησιμοποιηθεί
- ▶ Ανάλογα με το είδος της σημασιολογίας που θα περιγραφεί κατατάσσονται σε:
 - ▶ *Schema ontologies*: έχουν τον ίδιο προσανατολισμό όπως τα σχήματα των βάσεων δεδομένων – κάθε κλάση αντιστοιχεί σε ένα πίνακα σε μια βάση
 - ▶ *Topic ontologies*: πρόκειται για τις ταξινομίες που καθορίζουν ιεραρχίες από *topics*, *subjects*, *categories* – υιοθετούνται για κατηγοριοποίηση διαφόρων αντικειμένων
 - ▶ *Lexical ontologies*: πρόκειται για σχήματα που αποτυπώνουν διάφορες λεξοκογραφικές έννοιες



Οντολογίες

- ▶ Μια βάση γνώσης (knowledge base) είναι μια πιο γενική έννοια σε σχέση με μια οντολογία
- ▶ Η βάση γνώσης επιτρέπει τη συλλογιστική (reasoning)
- ▶ Μπορεί να περιλαμβάνει αξιώματα, ορισμούς, κανόνες, γεγονότα, εντολές
- ▶ Όμως, μια βάση γνώσης δεν στοχεύει στο να αναπαραστήσει κάποιο πεδίο
- ▶ Τα σημασιολογικά αποθετήρια (semantic repositories) επιτρέπουν την αποθήκευση, επερώτηση και τη διαχείριση δομημένων δεδομένων σε σχέση με τη σημασιολογία τους
- ▶ Επιτρέπουν τη διαχείριση μιας υποδομής μιας βάσης γνώσης



Σημασιολογικός Ιστός

- ▶ Ο Σημασιολογικός Ιστός (Semantic Web) στοχεύει στο να προσθέσει σημασιολογία στο WWW
- ▶ Η τρέχουσα μορφή του Παγκόσμιου Ιστού περιλαμβάνει πληροφορία που αποτυπώνεται στη γλώσσα html
- ▶ Ο Σημασιολογικός Ιστός προσφέρει:
 - ▶ Μια κοινή μορφή για την ανταλλαγή των δεδομένων
 - ▶ Μια γλώσσα για την καταγραφή του πως τα δεδομένα σχετίζονται με τα αντικείμενα του πραγματικού κόσμου



Σημασιολογικός Ιστός

- ▶ Η πληροφορία που προσφέρεται στο Σημασιολογικό Ιστό μπορεί να προσπελαστεί από σημασιολογικές μηχανές αναζήτησης
- ▶ Τα πρότυπα και οι τεχνολογίες του Σημασιολογικού Ιστού μπορούν να υιοθετηθούν στη σημασιολογική αναζήτηση
- ▶ Ένα ανοιχτό ζήτημα στο Σημασιολογικό Ιστό είναι ο καθορισμός αποδοτικών μεθόδων πρόσβασης στη σημασιολογική πληροφορία



RDF(S) & OWL

- ▶ Έχει υλοποιηθεί μια οικογένεια προτύπων ως η βάση του Σημασιολογικού Ιστού
- ▶ Η γλώσσα RDF (Resource Description Framework) είναι μια γλώσσα μετα-δεδομένων που αποτελεί το βασικό μοντέλο δεδομένων του Σημασιολογικού Ιστού
- ▶ Επιτρέπει την περιγραφή των resources και των σχέσεων μεταξύ τους
- ▶ Το κύριο μοντέλο περιγραφής είναι το διάνυσμα <Subject, Predicate, Object>
- ▶ Subject: είναι το resource που θέλουμε να περιγράψουμε
- ▶ Predicate: είναι ένα resource που καθορίζει μια σχέση
- ▶ Object: είναι ένα resource ή literal που αναπαριστά μια τιμή της ιδιότητας

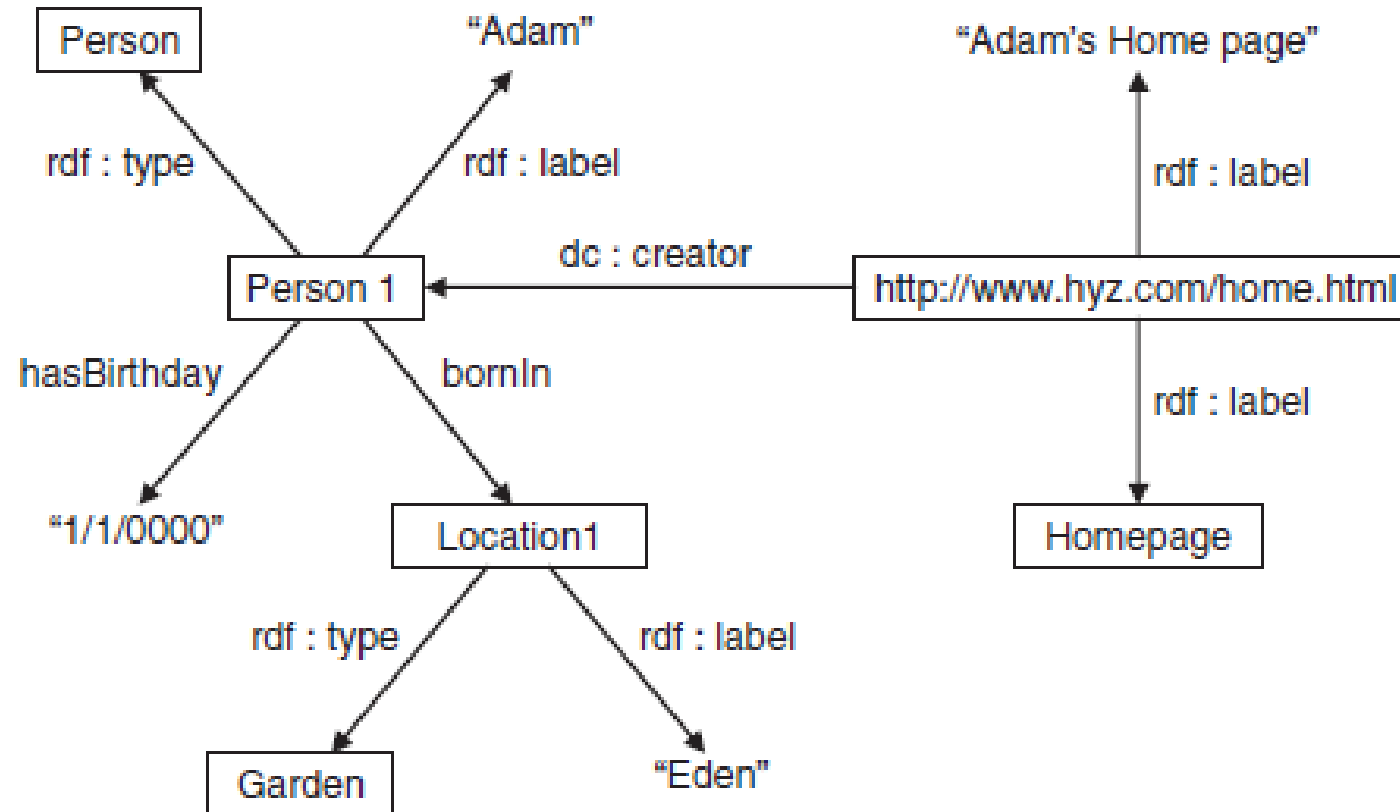


RDF(S) & OWL

- ▶ Ένα σύνολο από τριπλέτες RDF μπορούμε να τις δούμε σαν ένα γράφο
- ▶ Οι κόμβοι είναι τα resources ή τα literals
- ▶ Κάθε statement είναι μια ακμή πάνω στο γράφο
- ▶ Οι ακμές συνδέουν Subjects με Objects
- ▶ Τα resources είναι στιγμιότυπα κάποιων κλάσεων
- ▶ Αυτό καθορίζεται με το `rdf:type` (<resource, rdf:type, class)
- ▶ Δύο είναι οι κλάσεις του συστήματος: `rdf:Class`, `rdf:Property`
- ▶ Η RDFS (RDF Schema) είναι μια γλώσσα σχήματος που επιτρέπει τον ορισμό νέων κλάσεων και ιδιοτήτων



RDF(S) & OWL



RDF(S) & OWL

```
<?xml version="1.0"?>
```

```
<rdf:RDF
```

```
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:si="https://www.w3schools.com/rdf/">
```

```
<rdf:Description rdf:about="https://www.w3schools.com">
```

```
  <si:title>W3Schools</si:title>
```

```
  <si:author>Jan Egil Refsnes</si:author>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```



RDF(S) & OWL

- ▶ Η OWL (Web Ontology Language) είναι μια γλώσσα που επεκτείνει την RDF(S) ώστε να παραχθούν οντολογίες
- ▶ Έχει τρεις εκδόσεις: ○
 - ▶ WL-Lite: έχει τη 'μικρότερη' εκφραστικότητα
 - ▶ OWL-DL: μπορούμε να τη δούμε σαν μια υπο-γλώσσα που έχει παραχθεί με επιρροές από Descriptions Logic
 - ▶ OWL-Full: έχει τη μεγαλύτερη εκφραστικότητα αλλά είναι μη αποφασίσιμη



RDF(S) & OWL

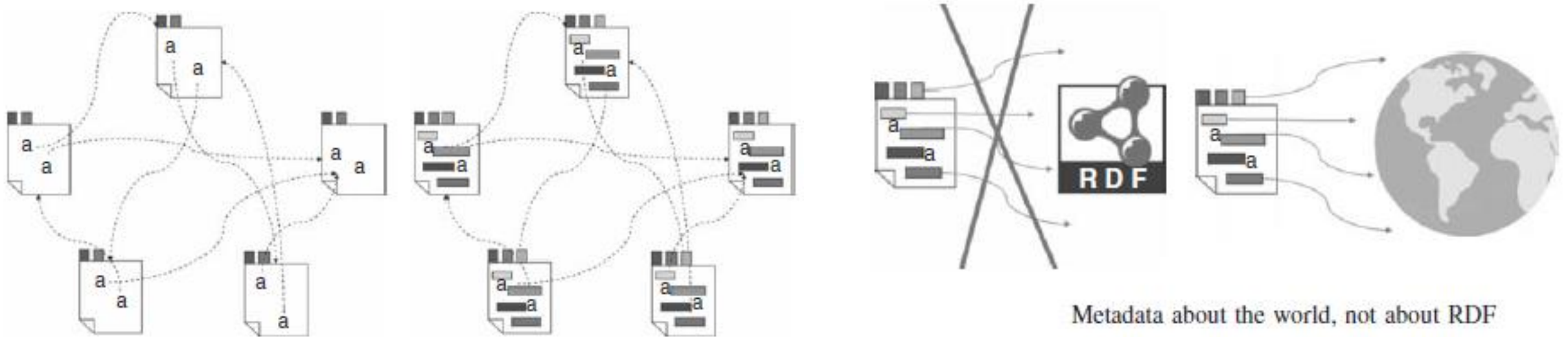
- ▶ Examples

<http://www.linkeddatatools.com/introducing-rdfs-owl>

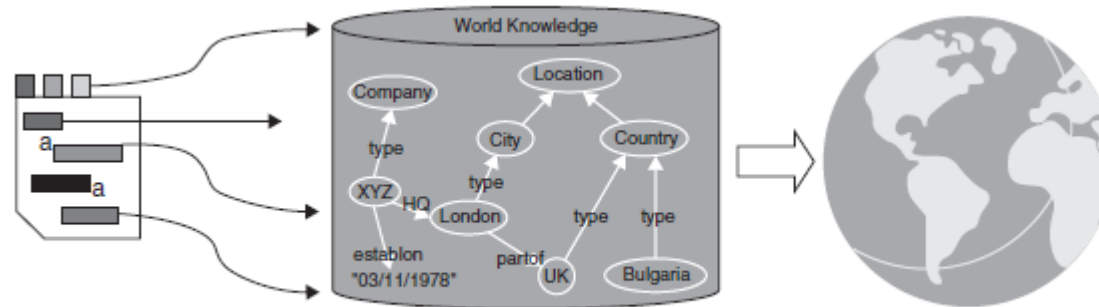


Semantic Annotation

- ▶ Οι επόμενες εικόνες μας δείχνουν τη 'χρήση' του Σημασιολογικού Ιστού



The Current WWW (left) and the Semantic Web (right)



Metadata Referring to World Knowledge

Semantic Search Tools

- ▶ Κάποια εργαλεία για αναζήτηση στο Σημασιολογικό Ιστό έχουν ως ακολούθως:
 - ▶ QuizRDF
 - ▶ TAP
 - ▶ KIM
 - ▶ Squirrel





Multimedia Retrieval

Εισαγωγή

- ▶ Η κατασκευή ευρετηρίων για πολυμεσικά αντικείμενα διαφέρει σε σχέση με τη διαδικασία που ακολουθείται για το κείμενο
- ▶ Συνήθως, οι εικόνες και το βίντεο δεν συνοδεύονται από τα κατάλληλα μετα-δεδομένα
- ▶ Ακόμα, και υπερκεραστούν τα διάφορα προβλήματα δεν σημαίνει πως θα καταλήξουμε σε ένα κατάλληλο annotation



Εισαγωγή

- ▶ Ο καλύτερος μέχρι στιγμής τρόπος για την κατασκευή ευρετηρίων πολυμεσικής πληροφορίας είναι η δημιουργία μιας library card
- ▶ Είναι μια ειδική καταχώρηση στη βάση δεδομένων που αφορά σε μετα-δεδομένα π.χ. συγγραφέας, τίτλος, κ.λπ.
- ▶ Άλλη προσέγγιση είναι να κατηγοριοποιήσουμε την πληροφορία (αυτόματα) προσθέτοντας κάποιες λέξεις κλειδιά που είναι προκαθορισμένα
- ▶ Μπορούμε να έχουμε την αναγνώριση κάποιου τμήματος π.χ. μιας εικόνας ή ενός βίντεο το οποίο να χρησιμοποιηθεί για το annotation
- ▶ Η τεχνική αυτή ονομάζεται feature selection ή automated annotation



Εισαγωγή

- ▶ Συνήθως ο τρόπος αναζήτησης είναι content-based
- ▶ Η βασική ιδέα είναι ότι η πολυμεσική πληροφορία μπορεί να χρησιμοποιηθεί σαν ερώτημα και το σύστημα ανάκτησης ψάχνει να βρει όμοια
- ▶ Σε κάποιες αναζητήσεις, υπάρχει η ανάγκη της επιλογής μόνο ενός αρχείου
- ▶ Αυτή η προσέγγιση ονομάζεται fingerprinting ή known-item search



Piggy-back Text Retrieval

- ▶ Από όλα τα είδη πολυμεσικής πληροφορίας, οι ροές τηλεοπτικών σημάτων προσφέρουν την καλύτερη 'υποδομή' για την εξαγωγή κειμένων
- ▶ Παράδειγμα: υπότιτλοι, αυτόματη μετατροπή ήχου σε κείμενο, teletext, κ.λπ.
- ▶ Συνεπώς, μπορούμε να κάνουμε αναζήτηση πάνω σε αυτό το κείμενο
- ▶ Παράδειγμα:
 - ▶ https://www.google.com/videohp?gws_rd=ssl
 - ▶ <http://www.blinkx.tv/>
- ▶ Τα βίντεο στα DVD έχουν διαφορετικά κανάλια για τις γλώσσες οπότε εκεί απαιτείται οπτική αναγνώριση χαρακτήρων
- ▶ Οι τεχνικές της εύρεσης της απόστασης (π.χ. edit distance) μπορούν να υιοθετηθούν



Automated Annotations

- ▶ Υιοθετούνται τεχνικές μηχανικής όρασης για συγκεκριμένες λέξεις κλειδιά
- ▶ Οι πιο συχνές τεχνικές είναι οι τεχνικές κατηγοριοποίησης (classification)
- ▶ Όμως, χρειαζόμαστε ένα μεγάλο σύνολο δεδομένων για εκπαίδευση
- ▶ Τα δεδομένα εκπαίδευσης περιλαμβάνουν annotations πάνω σε τμήματα των εικόνων
- ▶ Οι τεχνικές στοχεύουν στο να συνδέσουν περιοχές μια εικόνας με λέξεις
- ▶ Όλες έχουν ως κοινή βάση την ύπαρξη ενός λεξικού που υιοθετείται για το annotation
- ▶ Οι μέθοδοι μηχανικής μάθησης υιοθετούνται και για άλλα είδη πολυμεσικής πληροφορίας



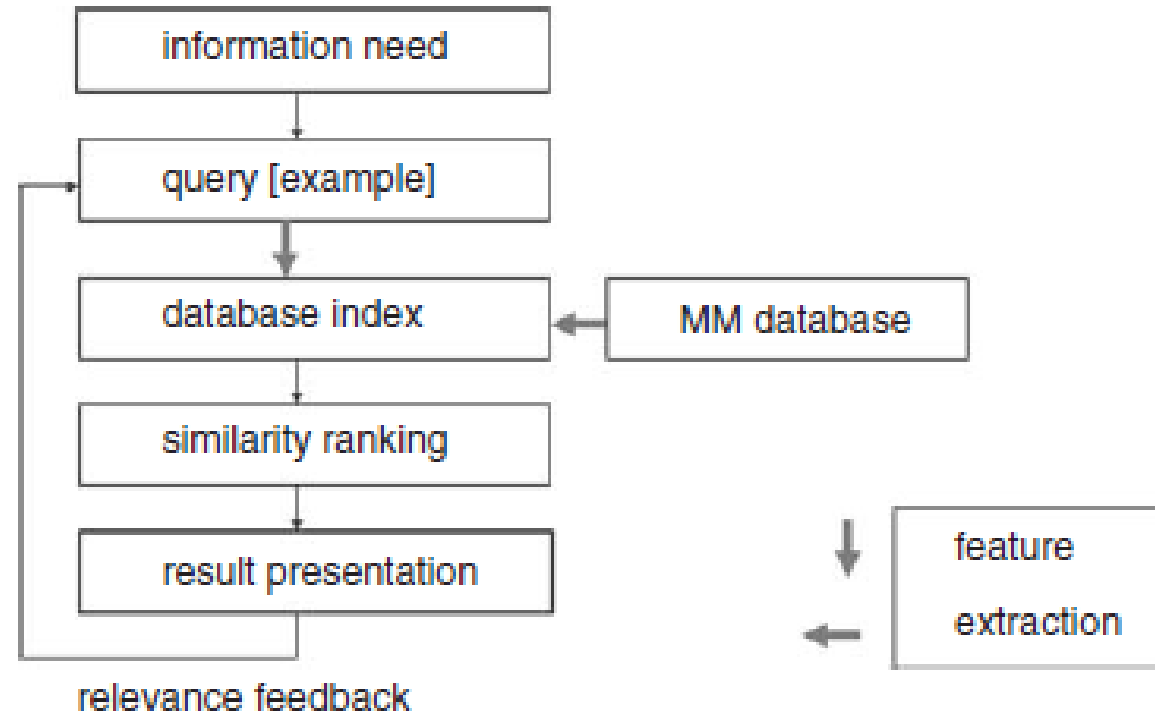
Content-based Retrieval

- ▶ Ο πιο κοινός τρόπος για να δημιουργήσουμε ευρετήρια εικόνων είναι να εξάγουμε χαρακτηριστικά όπως το χρώμα, το σχήμα, τη δομή, κ.λπ.
- ▶ Αυτές οι αναπαραστάσεις είναι διανύσματα τιμών που απεικονίζουν κάποια στατιστικά στοιχεία
- ▶ Οι αποστάσεις των διανυσμάτων αποτελούν μέτρα ομοιότητας
- ▶ Από τη στιγμή που θα καταγραφούν οι παραπάνω τιμές, στη συνέχεια μπορεί να γίνει η σύγκριση των εικόνων με εικόνες που απεικονίζουν τα ερωτήματα
- ▶ Η προσέγγιση αυτή ονομάζεται `query-by-example`



Content-based Retrieval

- ▶ Η γενική αρχιτεκτονική της μεθόδου έχει ως εξής:



Content-based Retrieval

- ▶ Κάθε πολυμεσικό αντικείμενο m οδηγεί στην εξαγωγή k στοιχείων $f_1(m), f_2(m), \dots, f_k(m)$
- ▶ Κάθε στοιχείο είναι ένα διάνυσμα τιμών
- ▶ Η τελική ομοιότητα με ένα ερώτημα q έχει ως εξής:

$$D_w(m, q) = \sum_{i=1}^k w_i d_i(f_i(m), f_i(q))$$

- ▶ Τα w_i είναι τα βάρη που ανατίθενται σε κάθε στοιχείο
 - ▶ Τα βάρη μπορούν να καθοριστούν από το σύστημα ρητά ή να εξαχθούν μέσω της αλληλεπίδρασης με τους χρήστες
-



Fingerprinting

- ▶ Τα αποτυπώματα (fingerprints) είναι μοναδικές αποτιμήσεις σε μια βάση πολυμεσικής πληροφορίας
- ▶ Υπολογίζονται από τα περιεχόμενα των πολυμεσικών αντικειμένων
- ▶ Είναι μικρές αναπαράστασεις και επιτρέπουν τη γρήγορη, αξιόπιστη και μοναδική τοποθεσία μιας πολυμεσικής πληροφορίας
- ▶ Τα fingerprints των ήχων αποτυπώνουν ακόμα και τους διαφορετικούς ήχους του ίδιου τραγουδιού που έχουν ηχογραφηθεί π.χ. σε διαφορετικό στούντιο
- ▶ Ενδιαφέρουσες εφαρμογές περιλαμβάνουν υπηρεσίες που επιτρέπουν τις εταιρείες που χειρίζονται τα πνευματικά δικαιώματα να αναγνωρίσουν το τι έχει παιχτεί





Mobile Search

Εισαγωγή

- ▶ Η επέκταση της χρήσης των κινητών προσφέρει νέους τρόπους επικοινωνίας με το Web
- ▶ Προσφέρονται υπηρεσίες που βασίζονται στη θέση των συσκευών
- ▶ Η θέση του χρήστη είναι ο παράγοντας κλειδί για την αναζήτηση της πληροφορίας
- ▶ Η πλειοψηφία των ερωτημάτων και της πληροφορίας είναι τοπική
- ▶ Αρκετές μηχανές αναζήτησης προσφέρουν location-based εφαρμογές αναζήτησης για κινητές συσκευές (Bing Local, Yahoo! Local, etc)
- ▶ Προφανώς, όπως και στην 'κλασική' αναζήτηση η συνάρτηση που αποτυπώνει τα rankings καθορίζει και την ποιότητα των αποτελεσμάτων



Εισαγωγή

- ▶ Οι διαφορές με την κλασική αναζήτηση είναι:
 - ▶ Η τοπική αναζήτηση συνήθως αφορά σε οντότητες (π.χ. εστιατόρια) και όχι σελίδες – αυτό απαιτεί γνώση του domain και διαχείριση π.χ. των ratings που δίνουν οι χρήστες στις οντότητες
 - ▶ Η τοπική αναζήτηση συνήθως αφορά σε οντότητες που βρίσκονται σε κοντινή απόσταση με τους χρήστες – π.χ. κοντινά εστιατόρια



Ranking Signals

- ▶ Ένα βασικό ζήτημα είναι η αναγνώριση των ranking signals
- ▶ Υπάρχουν αρκετές ερευνητικές προσπάθειες για την επεξεργασία διαφόρων logs από ερωτήματα που τίθενται σε κινητές συσκευές
- ▶ Ο στόχος είναι να αναγνωριστούν διάφορα πρότυπα και να εξαχθούν στατιστικά των αναζητήσεων
- ▶ Κάποια από τα ranking signals είναι:
 - ▶ Η απόσταση
 - ▶ Οι κριτικές των χρηστών
 - ▶ Οι προσωπικές προτιμήσεις
 - ▶ Ο τόπος, ο χρόνος και κοινωνικοί παράγοντες



Απόσταση

- ▶ Οι χρήστες προτιμούν τα αποτελέσματα να είναι σε κοντινή απόσταση
- ▶ Ο υπολογισμός της απόστασης είναι μια εύκολη υπόθεση
- ▶ Όμως, η απόσταση ορισμένες φορές δεν αποτυπώνει την ευκολία πρόσβασης (π.χ. κίνηση)
- ▶ Μελέτες έχουν δείξει πως το 40% των χρηστών ενδιαφέρονται για την απόσταση



Κριτικές των Χρηστών

- ▶ Οι κριτικές των χρηστών μπορούν να ενσωματωθούν στους υπολογισμούς ώστε να βελτιώσουν τα αποτελέσματα
- ▶ Το τελικό σκορ και το πλήθος των κριτικών για κάποια οντότητα μπορεί να επηρεάσει τη συμπεριφορά των υπολοίπων χρηστών
- ▶ Οι χρήστες έχουν την τάση να συμβουλευούνται άλλους όταν πρόκειται να προχωρήσουν σε κάποιες ενέργειες
- ▶ Όμως, κάποιες οντότητες π.χ. επιχειρήσεις μπορεί να μην έχουν κάποια κριτική



Προσωπικές Προτιμήσεις

- ▶ Έρευνες έχουν δείξει πως τα ερωτήματα σε τοπικές αναζητήσεις είναι γενικά και δεν καθορίζονται σαφώς
- ▶ Αυτό σημαίνει πως θα επιστραφεί μια μεγάλη λίστα αποτελεσμάτων πολλά από τα οποία μπορεί να μην ενδιαφέρουν τους χρήστες
- ▶ Για αυτό, το να καταλάβουμε το τις προτιμήσεις των χρηστών είναι πολύ κρίσιμο ζήτημα
- ▶ Όμως, δεν είναι τόσο εύκολο να γίνουν τα προφίλ των χρηστών όπως στις Web αναζητήσεις
- ▶ Λόγω της τοπικότητας των αναζητήσεων δεν είναι εύκολο να εξαχθούν συμπεράσματα για τους χρήστες που μπορεί να αναζητούν πληροφορίες από διαφορετικά σημεία



Τοποθεσία

- ▶ Πρόκειται για μια από τις σημαντικές παραμέτρους
- ▶ Οι χρήστες σε διαφορετικές θέσεις τείνουν να λαμβάνουν διαφορετικές αποφάσεις
- ▶ Συνήθως, οι γεωγραφικές και οι κοινωνικές παράμετροι επηρεάζουν τις αποφάσεις
- ▶ Επίσης, οι χρήστες αναζητούν με βάση τη γεωγραφική θέση σε σχέση με τη θέση τους



Χρόνος

- ▶ Επηρεάζει τη διαδικασία λήψης απόφασης από τους χρήστες
- ▶ Η συμπεριφορά των χρηστών αλλάζει ανάλογα με το χρονικό σημείο των αποφάσεων (π.χ. Σαββατοκύριακα)
- ▶ Επίσης, οι ενέργειες των χρηστών τείνουν να γίνονται άμεσα αμέσως μετά τις αναζητήσεις



Κοινωνικοί Παράγοντες

- ▶ Οι χρήστες επηρεάζονται από άλλους όταν κάνουν αναζητήσεις
- ▶ Σε σχετική έρευνα, το 63% των χρηστών δήλωσε πως η πιο πρόσφατη αναζήτησή τους από την κινητή συσκευή έτυχε επιρροής από κάποιον άλλο
- ▶ Μετά τις αναζητήσεις, οι χρήστες τείνουν να μοιράζονται την εμπειρία τους με άλλους



Ranking Heuristics

- ▶ Οι περισσότερες μέθοδοι που υιοθετούνται για τους υπολογισμούς του ranking στηρίζονται στη μηχανική μάθηση
- ▶ Οι μέθοδοι συνδυάζουν διάφορες παραμέτρους ώστε να βγάλουν το τελικό αποτέλεσμα
- ▶ Τα τελικά αποτελέσματα, κριτικές, ομοιότητες, κ.λπ. συνήθως αποτυπώνονται στη διεπαφή των χρηστών
- ▶ συνεπώς, οι χρήστες μπορούν να μελετήσουν τις συγκεκριμένες πληροφορίες



Likelihood of Clicks / Relevance

- ▶ Έχουν προταθεί τεχνικές με τις οποίες υπολογίζεται η πιθανότητα μια οντότητα να δεχθεί την προτίμηση των χρηστών (rating)
- ▶ Υιοθετείται ένα training dataset
- ▶ Οι αναφορές στα δεδομένα εκπαίδευσης ταξινομούνται και χωρίζονται σε ίσου μεγέθους περιοχές (bins)
- ▶ Το μέσο σκορ σε κάθε περιοχή επιλέγεται για την αναπαράσταση της περιοχής
- ▶ Η πιθανότητα για μια οντότητα υπολογίζεται με τη χρήση του λόγου των clicks σε κάθε περιοχή και του μεγέθους της περιοχής



The Mean Normalization Scheme

- ▶ Συσχετίζουμε τις προτιμήσεις των χρηστών με τις κατηγορίες των οντοτήτων
- ▶ Συγκρίνουμε το σκορ μιας οντότητας με το μέσο σκορ της αντίστοιχης κατηγορίας
- ▶ Με αυτό τον τρόπο μπορούμε να διαπιστώσουμε αν η οντότητα είναι ανάμεσα σε αυτές που έχουν λάβει υψηλό σκορ
- ▶ Οι χρήστες δεν ξέρουν το μέσο κάθε κατηγορίας αλλά έχουν μια εκτίμησή του
- ▶ Τείνουν να επιλέγουν οντότητες που εκτιμούν ότι είναι πάνω από το μέσο όρο



Cluster-based Smoothing of Ratings

- ▶ Δημιουργούμε συστάδες με ratings
- ▶ Η ιδέα είναι να πάρουμε το μέσο rating σε κάθε συστάδα
- ▶ Μια οντότητα σε μια συστάδα θα πάρει περίπου το ίδιο rating
- ▶ Δύο ειδών συστάδες μπορούν να υιοθετηθούν:
 - ▶ Κατηγορίας: χρησιμοποιούμε τα ratings των οντοτήτων στην ίδια κατηγορία για να εκτιμήσουμε οντότητες που δεν έχουν ratings – χρειάζεται η εκτίμηση της πρωταρχικής κατηγορίας ιδιαίτερα αν μια οντότητα ανήκει σε πολλές κατηγορίες
 - ▶ Αλυσίδας: εκτιμούμε το rating μέσω των ratings όλων των οντοτήτων που ανήκουν στην ίδια αλυσίδα

