

ΑΝΟΙΚΤΑ ΔΙΑΣΥΝΔΕΔΕΜΕΝΑ ΔΕΔΟΜΕΝΑ ΥΠΗΡΕΣΙΩΝ ΚΑΙ ΠΡΟΪΟΝΤΩΝ ΑΠΟ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ, ΣΕ ΠΕΡΙΒΑΛΛΟΝΤΑ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΜΠΟΡΙΟΥ

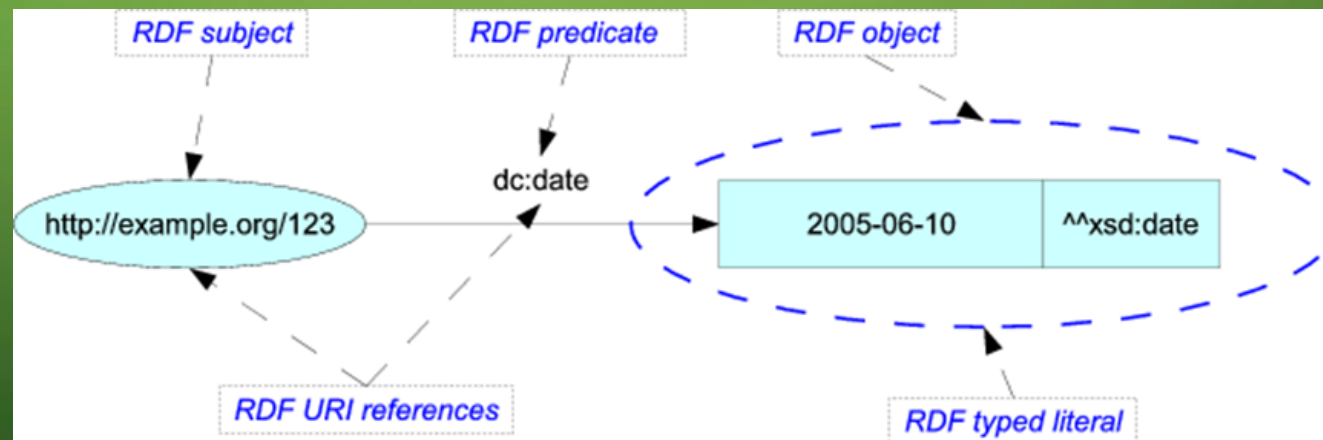
ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΤΗΣ ΚΑΡΑΜΠΕΛΑ ΑΝΑΣΤΑΣΙΑΣ

ΒΑΣΙΚΗ ΟΡΟΛΟΓΙΑ

- **Big Data:** μεγάλα σύνολα δεδομένων που ξεφεύγουν από τις δυνατότητες καταγραφής, αποθήκευσης και ανάλυσης των παραδοσιακών τεχνικών επεξεργασίας δεδομένων
- **Linked Open Data:** σύνολο αρχών σχεδιασμού για την κοινή χρήση διασυνδεδεμένων δεδομένων που διαβάζονται από μηχανές στον Ιστό σε συνδιασμό με τα Ανοικτά Δεδομένα
- **Linked Open Datasets (LODs):** ανοιχτές βάσεις δεδομένων σε μορφή RDF (Resource Description Framework) και συνδέσμους RDF για την σύνδεση διαφορετικών πηγών δεδομένων
- **Semantic Annotation:** προσθήκη ετικετών - προσάρτηση σε έγγραφο κειμένου ή άλλου μη δομημένου περιεχομένου, μεταδεδομένων σχετικά με έννοιες που σχετίζονται με αυτό

ΤΕΧΝΟΛΟΓΙΕΣ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ (1)

- Τα LODs αξιοποιούν το πρωτόκολλο HTTP και την τεχνολογία RDF για τη διασύνδεση δομημένων δεδομένων, οι συνδέσεις διέπονται από κανόνες του σημασιολογικού ιστού.
- Τα RDF περιέχουν περιγραφές αντικειμένων (λειτουργικά και μη-λειτουργικά χαρακτηριστικά)
- Τα RDF σχηματίζουν γράφους μοντελοποιώντας τις μεταξύ τους σχέσεις
- Γλώσσα ερωτημάτων SPARQL (μορφές XML, JSON, CSV, TSV)



ΤΕΧΝΟΛΟΓΙΕΣ ΤΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ (2)

ΟΝΤΟΛΟΓΙΑ

- συλλογή από URIs με συγκεκριμένη έννοια και περιεχόμενο (έγγραφο RDFs)
- ρόλος η ταξινόμηση των αντικειμένων με βάση την σημασία τους - περιγραφή των αντικειμένων (Individuals), των κλάσεων (Classes- οι συλλογές των διαφόρων θεμάτων), των χαρακτηριστικών τους (Attributes- οι ιδιότητες των αντικειμένων) και των μεταξύ τους σχέσεων (Relations- οι τρόποι με τους οποίους σχετίζονται τα αντικείμενα με τις κλάσεις)

ΚΟΙΝΌΤΗΤΑ SCHEMA.ORG

- δημιουργεί, συντηρεί και προωθεί σχήματα για τα δομημένα δεδομένα, όπως αυτά που υπάρχουν σε ιστοσελίδες, emails κ.α.
- Χρησιμοποιείται σε πάνω από 10 εκατομμύρια ιστοτόπους και σε εφαρμογές της Google και της Microsoft
- Με βάση το λεξικό Scgema.org οι ιστοσελίδες “μαρκάρονται” με microdata, όπως τα γνωστά tags της γλώσσας HTML5 βελτιώνοντας την παρουσίαση τους στα αποτελέσματα των μηχανών αναζήτησης

ΟΝΤΟΛΟΓΙΑ GOODRELATIONS ΣΤΟ ECOMMERCE

- Η ισχυρότερη οντολογία για την αναπαράσταση και δημοσιοποίηση πληροφοριών (προϊόντα, προσφορές, εκπτώσεις, τιμές ή όρους χρήσης των αγορών στο διαδίκτυο) στο κλάδο του ecommerce
- Μπορεί να χρησιμοποιηθεί σε όλες της γλώσσες που υποστηρίζουν σύνταξη προτάσεων RDF (όπως RDF/XML, Turtle, RDFa, JSON-LD κ.α.), Microdata κα
- Το Νοέμβριο του 2012 ενσωματώθηκε πλήρως στο λεξικό schema.org και αποτελεί πλέον το επίσημο μοντέλο αναπαράστασης δεδομένων ηλεκτρονικού εμπορίου
- η χρήση του μοντέλου GoodRelations βελτιώνει την απόδοση των μηχανών αναζήτησης των Google, Yahoo, Bing και Yandex, και αντίστοιχα υποστηρίζεται και χρησιμοποιείται από τις παραπάνω μέσω του schema.org

ΓΛΩΣΣΑ ΑΝΑΚΤΗΣΗΣ ΠΛΗΡΟΦΟΡΙΩΝ ΑΠΌ LODS: SPARQL

- Τα ερωτήματα SPARQL εκτελούνται σε σύνολα δεδομένων RDF αποτελούμενα από RDF γράφους. Βάσεις ανοικτών διασυνδεδεμένων δεδομένων DBpedia, Yago
- Ένα SPARQL endpoint δέχεται ερωτήματα και επιστρέφει αποτελέσματα μέσω του πρωτοκόλλου HTTP (XML, JSON, RDF, HTML)
- Περιορισμός αποτελεσμάτων: FILTER, LIMIT, SELECT DISTINCT, ORDER BY DESC

- PREFIX foaf: <<http://xmlns.com/foaf/0.1/>>

```
SELECT ?name
```

```
WHERE {?person foaf:name ?name .}
```

- PREFIX foaf: <<http://xmlns.com/foaf/0.1/>>

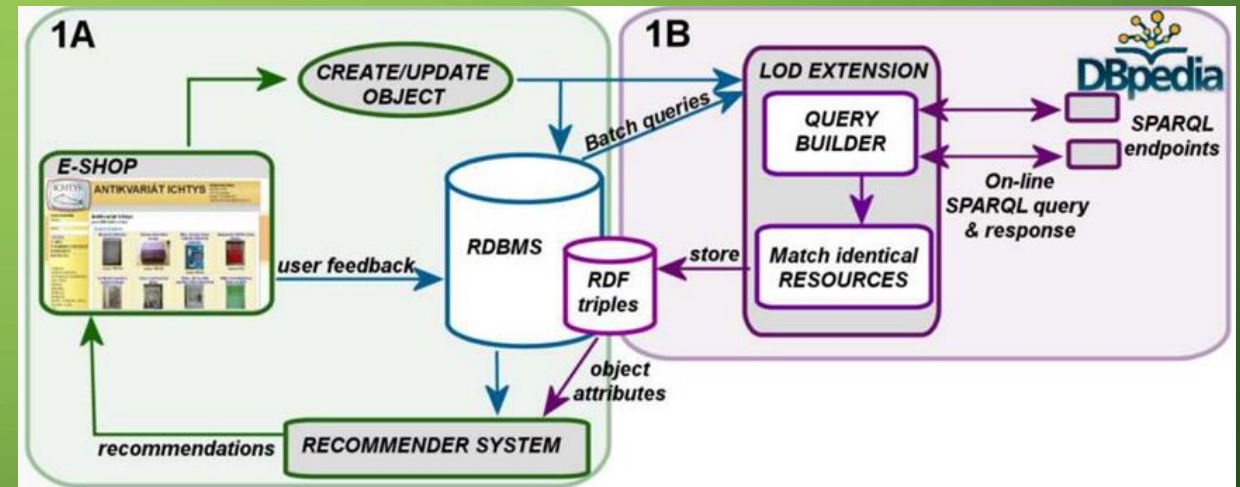
```
SELECT *
```

```
WHERE {?person foaf:name ?name .
```

```
        ?person foaf:mbox ?email .}
```

ΑΛΓΟΡΙΘΜΟΙ ΑΝΑΖΗΤΗΣΗΣ ΣΕ LINKED OPEN DATA

- Μηχανές αναζήτησης Open Data που ανιχνεύουν διασυνδεδεμένα δεδομένα: Swoogle Semantic Web Search Engine (SWSE), Falcons, Sindica και Watson
- Αλγόριθμοι αναζήτησης: Pop-Rank, ObjectRank (Swoogle)
- Αναζήτηση RDF Τριπλετών προς τα εμπρός και προς τα πίσω



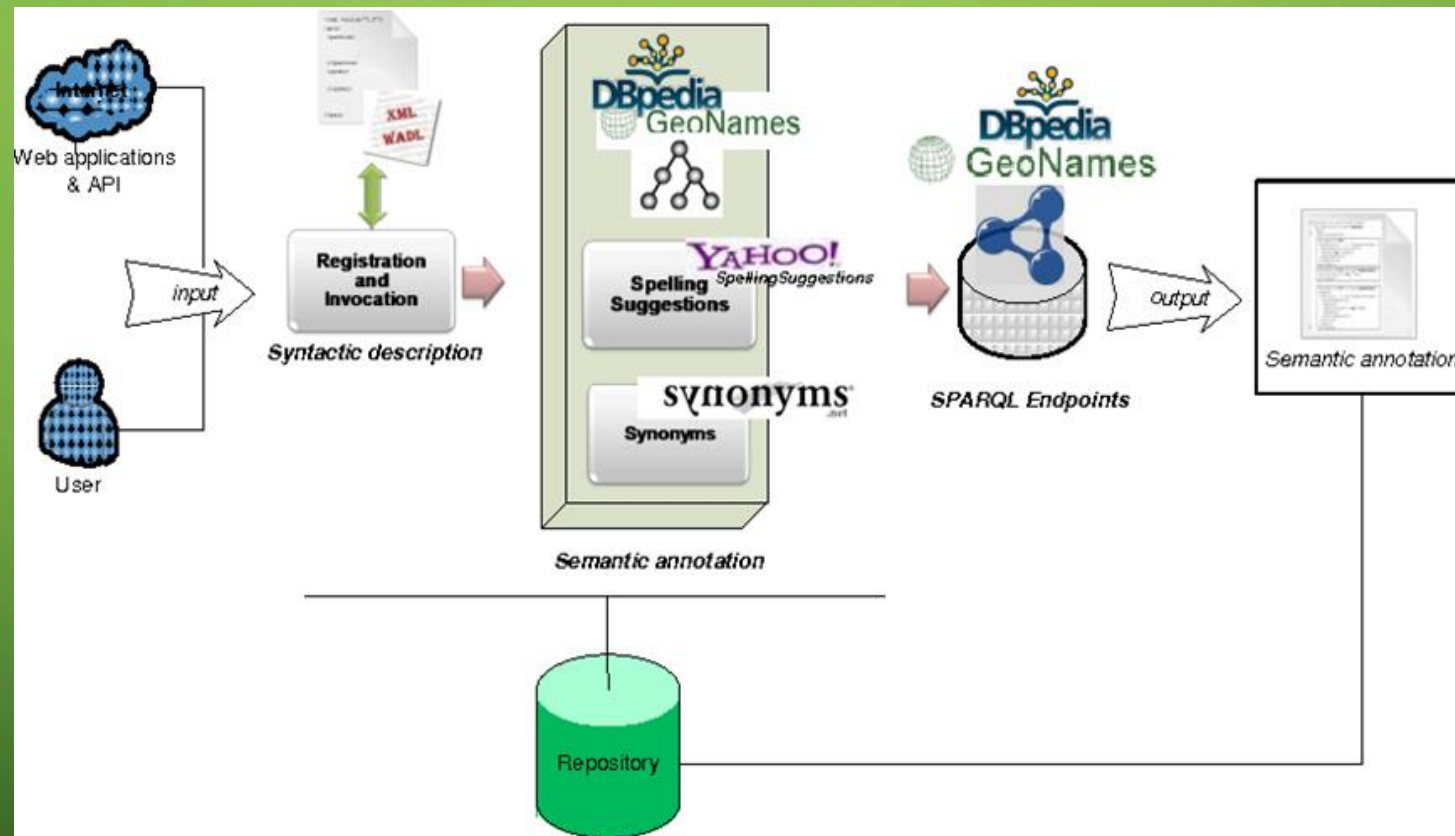
Recommendation System ενός eshop που συνδέεται με βάση ανοικτών διασυνδεδεμένων δεδομένων

ΤΑ SOCIAL MEDIA ΣΤΟΝ ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΙΣΤΟ (1)

Οντολογίες για αναπαράσταση περιεχομένου Social Media:

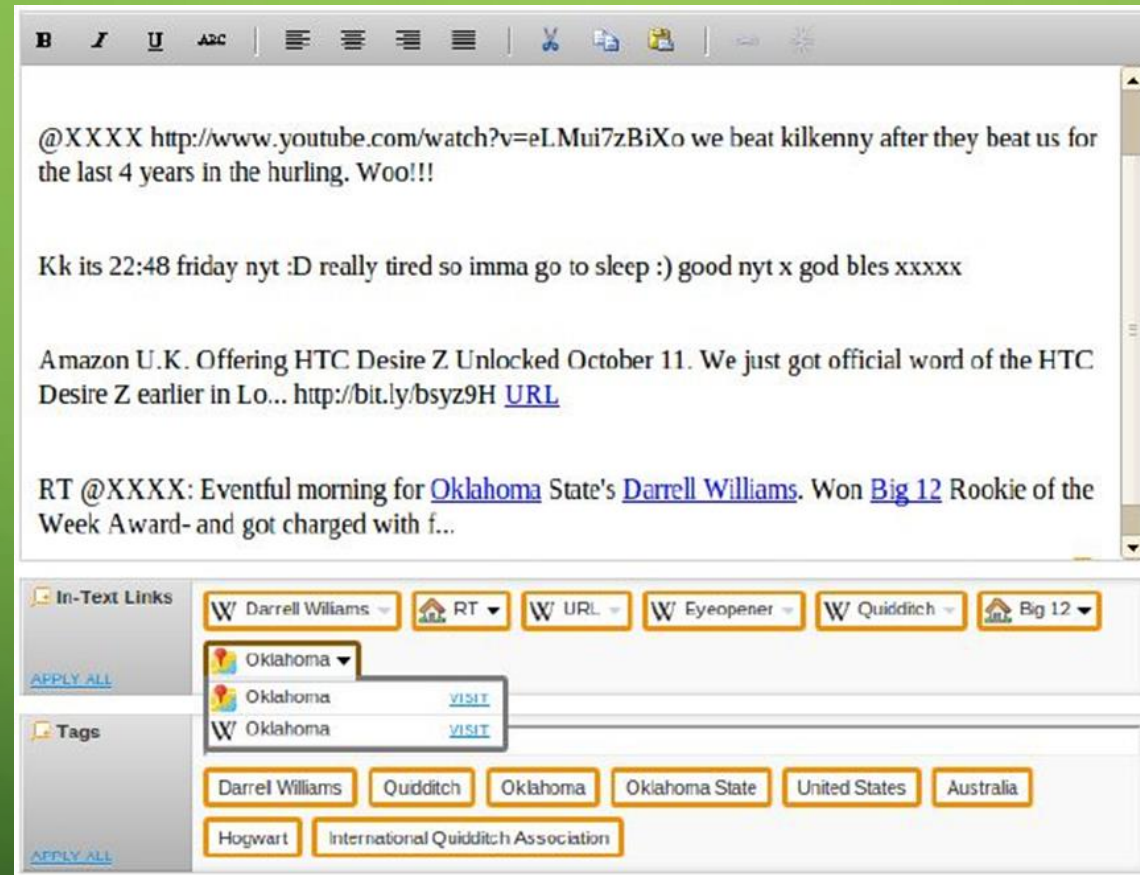
- FOAF (Friend-of-a-Friend): Περιγράφει ανθρώπους και Κοινωνικά Δίκτυα.
- SIOC (Semantically Interlinked Online Communities): Μοντελοποιεί blogs, wikis, online forums, microblogs
- DLPO (Live PostOntology): Χρησιμοποιείται για τη διασύνδεση Κοινωνικών Μέσων, Κοινωνικών Δικτύων και πρακτικών Online Sharing

ΤΑ SOCIAL MEDIA ΣΤΟΝ ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΙΣΤΟ (2)



Σύστημα Semantic Annotation με χρήση RESTful API

ΕΜΠΟΡΙΚΕΣ ΔΙΑΔΙΚΤΥΑΚΕΣ ΥΠΗΡΕΣΙΕΣ SEMANTIC ANNOTATION (1)



Προτεινόμενα tags με πιθανούς συνδέσμους στην Wikipedia με το εργαλείο Zemanta

ΕΜΠΟΡΙΚΕΣ ΔΙΑΔΙΚΤΥΑΚΕΣ ΥΠΗΡΕΣΙΕΣ SEMANTIC ANNOTATION (2)

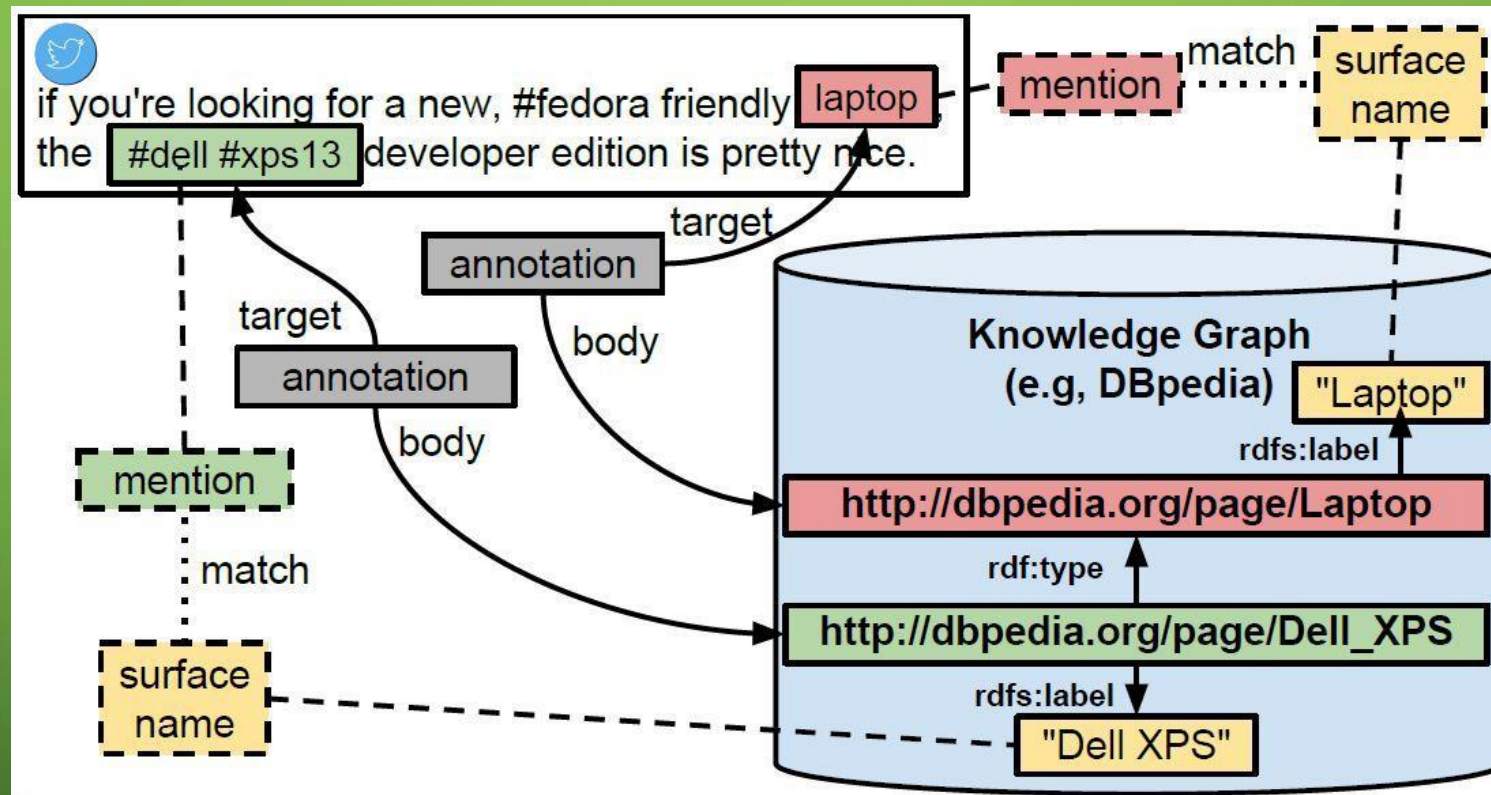
The screenshot displays the Open Calais web interface. On the left, a sidebar contains two main sections: 'Social Tags' and 'Entities'. The 'Social Tags' section shows the tag 'XXXX' with a three-star rating. The 'Entities' section lists various categories with checkboxes and corresponding counts:

- Company**: Amazon U.K. (2), HTC Corporation (2)
- Person**: Darrell Williams (1)
- Position**: (0)
- RT**: (2)
- Product**: HTC Desire Smartphone (2)
- URL**: <http://bit.ly/bsyz9H> (1), http://www.youtube.com/watch?v=eL_Mui7zBiXo (2)

The main content area on the right shows a Twitter feed with several tweets. The first tweet is a retweet (RT) of a tweet from @XXXX about a Quidditch match. The second tweet is from @XXXX linking to a YouTube video. The third tweet is from RT @XXXX mentioning 'Eyeopener vs. Ryerson Quidditch team'. The fourth tweet is from Amazon U.K. offering the HTC Desire Z. The fifth tweet is from RT @XXXX mentioning 'Oklahoma State's Darrell Williams'.

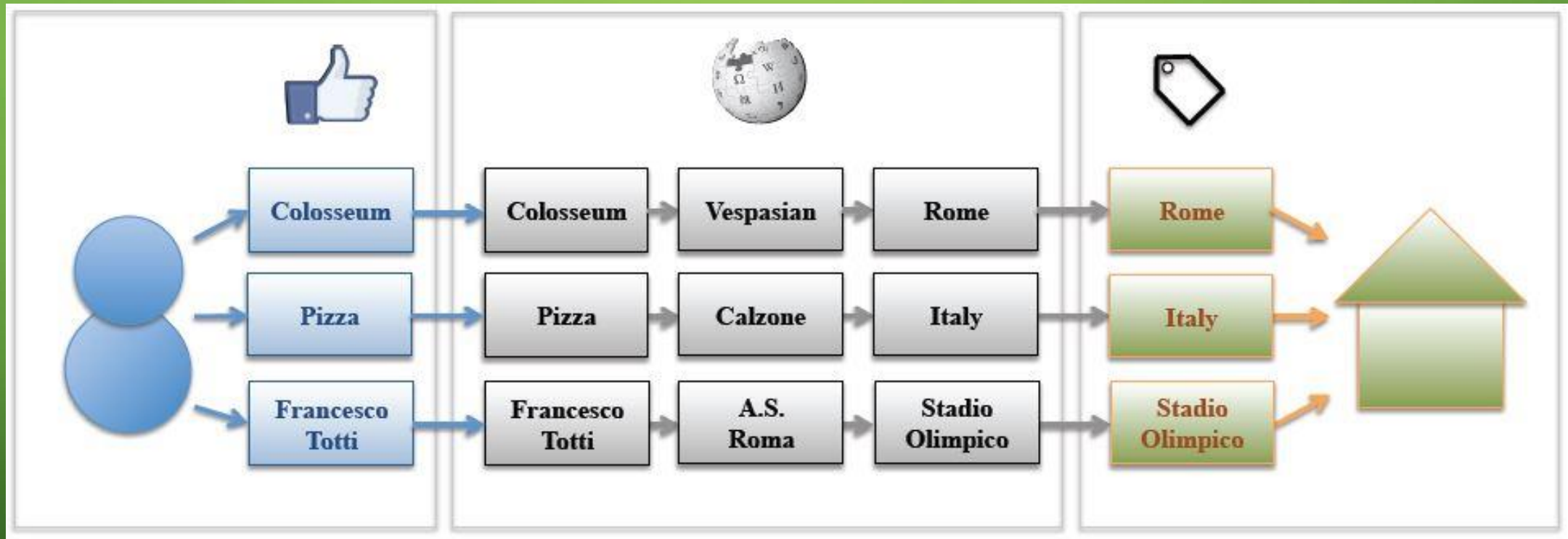
Αντιστοίχιση κειμένου URIs με το εργαλείο Open Calais

SOCIAL CRMS



Παραγωγή RDF από ένα tweet με την χρήση πληροφορίας της βάσης DBpedia-Spotlight

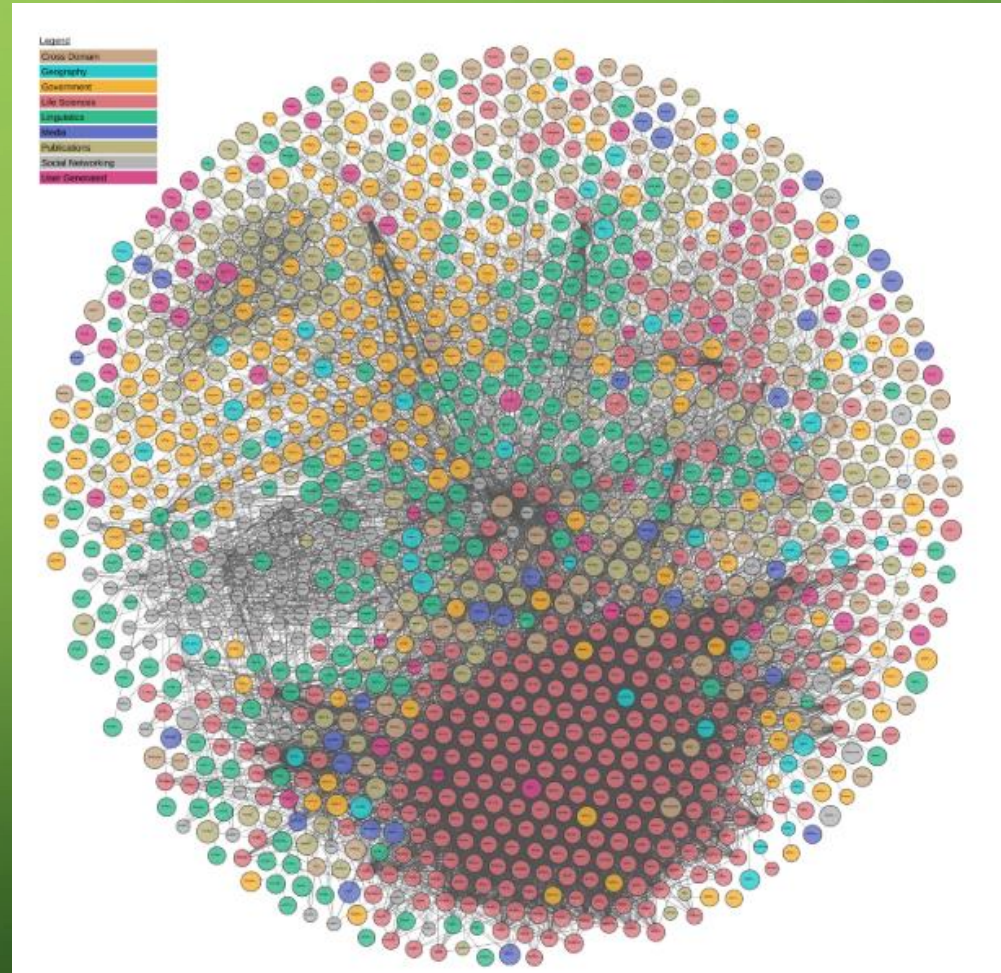
RECOMMENDATION SYSTEMS ΜΕ ΧΡΗΣΗ SOCIAL CONTENT



Πως οι προτιμήσεις των χρηστών στα social media συνδέονται με δεδομένα από τα ανοιχτά διασυνδεδεμένα δεδομένα

ΟΙ ΠΙΟ ΓΝΩΣΤΕΣ DATASETS ΜΕ SOCIAL MEDIA CONTENT

- **Social Link:** 80.101.425 τριπλέτες , 643.235 συνδέονται με την Dbpedia
- **sears.com:** 75.000.000, 100 συνδέονται με την Dbpedia
- **Linked Crunchbase:** 5.000.000 τριπλέτες
- **VIVO:** 4,514,025, 58 συνδέονται με την Dbpedia
- **Debian Package Tracking System:** 1,500,000 τριπλέτες
- **Social Semantic Web Thesaurus:** 20,000 τριπλέτες, 300 συνδέονται με την Dbpedia



ΟΝΤΟΛΟΓΙΑ ECLASSOWL ΣΤΟΝ ΚΛΑΔΟ ΤΟΥ ECOMMERCE: ΠΑΡΑΔΕΙΓΜΑΤΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΕΡΩΤΗΜΑΤΩΝ

ECLASSOWL ΚΑΙ GOODRELATIONS

- δημοσίευση meta-data για τα προϊόντα ή τις υπηρεσίες ενός ηλεκτρονικού καταστήματος- 2 παράμετροι:
 - Το είδος και τα χαρακτηριστικά του προϊόντος (πχ. “οθόνη τηλεόρασης 20 ιντσών”)
 - Τις λεπτομέρειες της προσφοράς (πχ. Τιμή και τρόποι πληρωμής)
- Η οντολογία eCI@ss είναι πολύ μεγάλη, κάνουμε αναζήτηση στο Web Site της για να βρούμε τις κατάλληλες κλάσεις και ιδιότητες για την περιγραφή ενός αντικειμένου

Παράδειγμα: Το μοντέλο ενός μολυβιού

foo:Pencil_150

```
# Class "Pencil": 24-24-01-01 [ AKF303003 ]  
α eco:C_AKF303003-gen, gr:ProductOrServiceModel ;  
rdfs:label "Heavy-duty pencil, 150 mm long, pointed  
tip"@en ;  
# Length [BAF559001] 150 mm  
eco:P_BAF559001 foo:QuantitativeValueFloat_1 ;  
# Design of tip state [BAG073001] = "pointed"  
[BAC386001]  
eco:P_BAG073001 eco:V_BAC386001 .
```

foo:QuantitativeValueFloat_1

```
α gr:QuantitativeValueFloat ;  
gr:hasUnitOfMeasurement "MMT"^^xsd:string ;  
gr:hasValueFloat "150.0"^^xsd:float .
```


SPARQL QUERIES

Παράδειγμα : Βρίσκει όλα τα μολύβια που έχουν μήκος τουλάχιστον 100 mm.

```
SELECT ?pencil, ?length WHERE
{
  ?pencil a eco:C_AKF303003-gen.
  ?pencil a gr:ActualProductOrServiceInstance.
  ?pencil eco:P_BAF559001 ?value.
  ?value gr:hasMaxValue ?length.
  ?value gr:hasUnitOfMeasurement "MMT"^^xsd:string.
  # Υπάρχει και η παρακάτω εναλλακτική:
  # ?value gr:hasUnitOfMeasurement ?unit.
  # FILTER (regex(?unit, "^MMT"))
  FILTER (?length >=100)}
LIMIT 10
```


ΣΥΜΠΕΡΑΣΜΑΤΑ

- Ελάχιστοι είναι η ερευνητές οι οποίοι γνωρίζουν την εν λόγω τεχνολογία και την αξιοποιούν για την ανάπτυξη των ηλεκτρονικών επιχειρήσεων τους.
- Ένα ακόμη κομμάτι στο οποίο αξιοποιούνται τα ανοιχτά διασυνδεδεμένα δεδομένα είναι για τη βελτίωση των μηχανών αναζήτησης (google) αλλά και στην ανάπτυξη recommendation systems.
- τα social media επηρεάζουν σε μεγάλο βαθμό την online καταναλωτική συμπεριφορά
- εργαλεία για την ανάλυση των δεδομένων των social media, social CRMs καθώς και recommendation systems με χρήση social content.

ΣΑΣ ΕΥΧΑΡΙΣΤΩ