

Διασυνδεδεμένα δεδομένα, σχήματα δομημένων περιγραφών και λεξιλόγια

Εθνικό Μετσόβιο Πολυτεχνείο

8 Ιουλίου 2015

- Το μοντέλο δεδομένων RDF παρέχει έναν τρόπο διατύπωσης προτάσεων σχετικά με διάφορους πόρους.
example:alice example:knows example:bob
example:alice example:name "Alice"
- Δεν κάνει καμία παραδοχή ως προς το τι αντιπροσωπεύουν τα διάφορα URI σε μια πρόταση.
- Στην πράξη, το μοντέλο RDF χρησιμοποιείται σε συνδυασμό με λεξιλόγια που παρέχουν σημασιολογικές πληροφορίες για τους περιγραφόμενους πόρους.
 - Το example:alice αντιπροσωπεύει κάποια γυναίκα.
 - Το example:bob αντιπροσωπεύει κάποιον άνδρα.
 - Οι γυναίκες και οι άνδρες είναι άνθρωποι.
 - Η ιδιότητα example:knows σχετίζει μεταξύ τους ανθρώπους.

- Ο ορισμός λεξιλογίων για δεδομένα RDF γίνεται με χρήση της γλώσσας του RDF Schema.
- Ουσιαστικά, το RDF Schema είναι το ίδιο ένα λεξιλόγιο που επιτρέπει τη σημασιολογική μοντελοποίηση δεδομένων RDF.
- Παρέχει μηχανισμούς για την περιγραφή ομάδων ομοειδών πόρων και των μεταξύ τους σχέσεων.
 - Π.χ. μπορεί να οριστεί η ομάδα πόρων `example:Person` που αντιπροσωπεύουν ανθρώπους.
 - Π.χ. μπορεί να οριστεί ότι το υποκείμενο και το αντικείμενο της ιδιότητας `example:knows` είναι πόροι της κλάσης `example:Person`.

- Οι πόροι που αντιπροσωπεύουν κατηγορίες ομοειδών αντικειμένων καλούνται κλάσεις.
example:Person, example:Woman, example:Institute, example:City
- Οι πόροι που ανήκουν σε μια κλάση καλούνται στιγμιότυπα της κλάσης.
example:alice, example:NTUA, example:Athens
- Για τη δήλωση ενός στιγμιοτύπου ως μέλους μιας κλάσης χρησιμοποιείται η ιδιότητα `rdf:type`.
example:alice `rdf:type` example:Woman
example:NTUA `rdf:type` example:Institution

Ορίζονται οι εξής βασικές κλάσεις:

- **rdfs:Resource**: Η κλάση όλων των πόρων. Ό,τι περιγράφεται από ένα μοντέλο RDF είναι πόρος.
- **rdfs:Class**: Η κλάση των πόρων που αντιπροσωπεύουν κλάσεις.
`example:Person` `rdf:type` `rdfs:Class`
`example:Institution` `rdf:type` `rdfs:Class`
- **rdf:Property**: Η κλάση των πόρων που αντιπροσωπεύουν ιδιότητες.
`example:knows` `rdf:type` `rdf:Property`
`example:studies` `rdf:type` `rdf:Property`
- **rdfs:Literal**: Η κλάση των (λεκτικών, αριθμητικών, κλπ) τιμών.

Ορίζονται οι εξής βασικές ιδιότητες:

- **rdf:type**: Συμμετοχή στιγμιοτύπου σε κλάση.
example:alice **rdf:type** example:Person
example:alice **a** example:Person
- **rdfs:subClassOf**: Για τη δήλωση υποκλάσεων.
example:Male **rdfs:subClassOf** example:Person
- **rdfs:subPropertyOf**: Για τη δήλωση υποϊδιοτήτων.
example:isFriendOf **rdfs:subClassOf** example:knows
- **rdfs:domain**: Για τη δήλωση πεδίου ορισμού.
example:knows **rdfs:domain** example:Person
- **rdfs:range**: Για τη δήλωση πεδίου τιμών.
example:knows **rdfs:range** example:Person
example:name **rdfs:range** rdfs:Literal

- Άλλες περιγραφικές ιδιότητες:
 - **rdf:label**: Όνομα πόρου αναγνώσιμο από τους ανθρώπους.
example:alice **rdfs:label** "Alice Smith"
 - **rdfs:comment**: Περιγραφή πόρου αναγνώσιμη από τους ανθρώπους.
example:Person **rdfs:comment** "Το σύνολο όλων των ανθρώπων"

- Ορισμός λεξιλογίου

```
example:Person rdf:type rdfs:Class
```

```
example:Person rdfs:comment "Το σύνολο όλων των ανθρώπων"
```

```
example:Woman rdf:type rdfs:Class
```

```
example:Woman rdfs:subClassOf example:Person
```

```
example:Man rdf:type rdfs:Class
```

```
example:Man rdfs:subClassOf example:Person
```

```
example:knows rdf:type rdf:Property
```

```
example:knows rdfs:domain example:Person
```

```
example:knows rdfs:range example:Person
```

- Δεδομένα

```
example:alice rdf:type example:Woman
```

```
example:alice rdfs:label "Alice Smith"
```

```
example:bob rdf:type example:Man
```

```
example:bob rdfs:label "Bob Ryan"
```

```
example:bob rdfs:knows example:alice
```


- Το RDF Schema αποτελεί σημασιολογική επέκταση του RDF.
 - Ορίζει μηχανισμούς εξαγωγής συμπερασμάτων.
 - Π.χ. αφού ο πόρος `example:alice` αντιπροσωπεύει μια γυναίκα και όλες οι γυναίκες είναι άνθρωποι, ο `example:alice` αντιπροσωπεύει έναν άνθρωπο.

```
example:alice rdf:type example:Woman
example:Woman rdfs:subClassOf example:Person
example:alice rdf:type example:Person
```
 - Μια σημασιολογική επέκταση μπορεί να θέτει και περιορισμούς ως προς τις επιτρεπτές τριάδες.

- Οι τριάδες υποκειμένου-κατηγορήματος-αντικειμένου μιας αποθήκης RDF συνιστούν έναν κατευθυντικό γράφο.

```
:alice rdf:type :Woman
```

```
:bob :knows :alice
```

```
:alice :livesIn :Athens
```

```
:NTUA :isLocatedIn :Athens
```

```
:Athens :capitalOf :Greece
```

- Τα `:alice`, `:bob`, `:Woman`, `:NTUA`, `:Athens`, `:Greece` είναι οι κόμβοι του γράφου.
- Οι ιδιότητες `rdf:type`, `:knows`, `:livesIn`, `:isLocatedIn`, `:capitalOf` είναι οι ακμές του γράφου.
- Η SPARQL είναι μια γλώσσα ερωτημάτων επί αποθηκών τριάδων RDF.

- Οι κόμβοι αναπαριστώνται με τα URI τους ή με μεταβλητές.
- Η βασική εντολή για την υποβολή ερωτημάτων είναι η **SELECT**. Δηλώνονται τα ονόματα των μεταβλητών που αποτελούν απάντηση στο ερώτημα.
- Στο σκέλος **WHERE** δηλώνονται οι συνθήκες που πρέπει να ικανοποιούν οι μεταβλητές.
- Οι συνθήκες είναι τριάδες με παρουσία μεταβλητών.
- Οι συνθήκες διαχωρίζονται μεταξύ τους με μια τελεία.
PREFIX ex: <http://example.org/>
SELECT ?x ?y ?z
WHERE
 { ?x ?y ?z . ?z rdf:type ex:Person }
- Η τελευταία έκδοση της γλώσσας επιτρέπει και την ενημέρωση (εισαγωγή, διαγραφή δεδομένων) μιας αποθήκης τριάδων.

- Ανάκτηση όλων των ονομάτων
- Αποθήκη τριάδων

```
ex:alice ex:name "Alice Smith" ; rdf:type ex:Person
```

```
ex:bob ex:name "Bob Ryan" ; rdf:type ex:Person
```

```
ex:john ex:name "John Taylor" ; rdf:type ex:Person
```

```
ex:alice ex:knows ex:bob
```

```
ex:john ex:knows ex:bob
```

- Ερώτημα

```
PREFIX ex: <http://example.org/>
```

```
SELECT ?x ?name
```

```
WHERE
```

```
  { ?person ex:name ?name }
```

- Αποτέλεσμα:

```
ex:alice  "Alice Smith"
```

```
ex:bob    "Bob Ryan"
```

```
ex:john   "John Taylor"
```

- Ανάκτηση των ονομάτων των φίλων όλων των ανθρώπων:
- Αποθήκη τριάδων

```
ex:alice ex:name "Alice Smith" ; rdf:type ex:Person
```

```
ex:bob ex:name "Bob Ryan" ; rdf:type ex:Person
```

```
ex:john ex:name "John Taylor" ; rdf:type ex:Person
```

```
ex:alice ex:knows ex:bob
```

```
ex:john ex:knows ex:bob
```

- Ερώτημα

```
SELECT ?x ?name
```

```
WHERE
```

```
  { ?x rdf:type ex:Person . }
```

```
  ?x ex:knows ?y .
```

```
  ?y ex:name ?name }
```

- Αποτέλεσμα:

```
ex:alice    "Bob Ryan"
```

```
ex:john     "Bob Ryan"
```

- Ανάκτηση των ανθρώπων με όνομα John Taylor.

- Αποθήκη τριάδων

```
ex:alice ex:name "Alice Smith" ; rdf:type ex:Person
```

```
ex:bob ex:name "Bob Ryan" ; rdf:type ex:Person
```

```
ex:john ex:name "John Taylor" ; rdf:type ex:Person
```

```
ex:alice ex:knows ex:bob
```

```
ex:john ex:knows ex:bob
```

- Ερώτημα

```
SELECT ?x
```

```
WHERE
```

```
  { ?x rdf:type example:Person .
```

```
    ?x example:name "John Taylor" }
```

- Αποτέλεσμα

```
ex:john
```

- Ανάκτηση των ανθρώπων με ηλικία άνω των 25 ετών:

```
SELECT ?x ?age
WHERE
  { ?x rdf:type example:Person .
    ?x example:age ?age .
    FILTER (?age > 25) }
```

- Ανάκτηση όλων των ανθρώπων και των email τους, αν υπάρχει:

```
SELECT ?x ?email
WHERE
  { ?x rdf:type example:Person .
    OPTIONAL {?x example:email ?email } }
```

- Ανάκτηση όλων των ανθρώπων και των τηλεφώνων τους:

```
SELECT ?x ?tel
WHERE
  { ?x rdf:type example:Person .
    {?x example:telephone ?tel }
    UNION
    {?x example:mobile ?tel } }
```

Σε μια αποθήκη τριάδων RDF μπορούν να επιβληθούν διάφορα καθεστώτα συμπερασμού.



- Ερώτημα: `SELECT ?prop WHERE { ?prop rdf:type rdf:Property }`
 - Χωρίς συλλογιστική: καμία απάντηση.
 - Συμπερασμός RDF: `ex:publishes rdf:type rdf:Property`
- Ερώτημα: `SELECT ?pub WHERE { ?pub rdf:type ex:Publication }`
 - Χωρίς συλλογιστική: `ex:book1`.
 - Συμπερασμός RDF: `ex:book1`.
 - Συμπερασμός RDFS: `ex:book2 rdf:type ex:Publication`
`ex:book3 rdf:type ex:Publication`

- Το RDF Schema είναι μια γλώσσα για που επιτρέπει την σύνταξη δομημένων λεξιλογίων.
- Ένα λεξιλόγιο μπορεί να είναι π.χ. ένας θησαυρός ή μια ταξινόμια που να αφορά ένα συγκεκριμένο πεδίο γνώσης.
- **Dublin Core:** Λεξιλόγιο για την περιγραφή πόρων του ιστού (εικόνων, video, ιστοσελίδων, κλπ.) και φυσικών πόρων όπως βιβλίων, CD, κλπ.
- **FOAF:** Λεξιλόγιο για την περιγραφή ανθρώπων, των δραστηριοτήτων τους και των μεταξύ τους σχέσεων.
- **SKOS:** Λεξιλόγιο για την σύνταξη ταξονομιών.
- **Schema.org:** Λεξιλόγιο γενικής χρήσης για τον σημασιολογικό χαρακτηρισμό των ιστοσελίδων.

- Βασικό λεξιλόγιο για την περιγραφή πόρων του ιστού (εικόνων, video, ιστοσελίδων, κλπ.) και φυσικών πόρων όπως βιβλίων, CD, έργων τέχνης, κλπ.
- Ονοματοχώρος: <http://purl.org/dc/terms/>
- 15 βασικές ιδιότητες: **contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type.**
- Επιπλέον εξειδικεύσεις: Π.χ. **abstract, dateAccepted, dateCopyrighted, isPartOf, isReferencedBy, extend**, κ.ά.
- Κλάσεις: **Agent, BibliographicResource, LinguisticSystem, Location, PeriodOfTime, PhysicalMedium, PhysicalResource, SizeOrDuration**, κ.ά.

- Παράδειγμα ορισμού της ιδιότητας `dcterms:creator` με χρήση του RDF Schema

```
dcterms:creator rdf:type rdfs:Property
dcterms:creator rdfs:label "Creator@en"
dcterms:creator rdfs:comment
    "An entity primarily responsible for making the resource.@en"
dcterms:creator rdfs:range dcterms:Agent
dcterms:creator rdfs:subPropertyOf dcterms:contributor
```

- Παράδειγμα χρήσης:

```
example:MerchOfVen dcterms:title "Merchant of Venice"@en
example:MerchOfVen dcterms:creator example:Shakespeare
example:MerchOfVen dcterms:created "1689"
example:Shakespeare foaf:familyName "Shakespeare"
example:Shakespeare foaf:givenName "William"
```

- Βασικό λεξιλόγιο για την περιγραφή ανθρώπων, ανθρώπινων δραστηριοτήτων, ανθρώπινων σχέσεων και κοινωνικών δικτύων.
- Ονοματοχώρος: `http://xmlns.com/foaf/0.1/`
- Κλάσεις: `Agent`, `Person`, `Project`, `Organization`, `Group`, `Document`, `Image`.
- Βασικές ιδιότητες: `name`, `title`, `img`, `depicts`, `familyName`, `givenName`, `knows`, `based_near`, `age`, `made`, `primaryTopic`, `member`.
- Επιπλέον ιδιότητες: Π.χ. `nick`, `homepage`, `weblog`, `account`, `schoolHomepage`, `currentProject`, `pastProject` κ.ά.

- Παράδειγμα χρήσης.

```
ex:jim rdf:type foaf:Person
ex:jim foaf:name "Jimmy Wales"
ex:jim foaf:mbox <mailto:jwales@bomis.com>
ex:jim foaf:homepage <http://www.jimmywales.com>
ex:jim foaf:nick "Jimbo"
ex:jim foaf:depiction <http://www.jimmywales.com/aus_img.jpg>
ex:jim foaf:interest <http://www.wikimedia.org>
ex:jim foaf:knows ex:mary
ex:mary rdf:type foaf:Person
ex:mary foaf:name "Mary Beesley"
```

- Σύσταση της W3C για ένα λεξιλόγιο RDF που επιτρέπει την αναπαράσταση θησαυρών, ταξονομιών και άλλων παρόμοιων μορφών ελεγχόμενου λεξιλογίου.
- Ονοματοχώρος: <http://www.w3.org/2004/02/skos/core#>
- Βασική μονάδα η `skos:Concept`. Πόροι που αντιπροσωπεύουν νοητικές κατηγορίες αντικειμένων ή γεγονότων.
- Οι έννοιες συνοδεύονται από ετικέτες αναφοράς σε φυσική γλώσσα: `skos:prefLabel`, `skos:altLabel`, `skos:hiddenLabel`.

```
example:animals rdf:type skos:Concept
example:animals skos:prefLabel "animals"@en
example:animals skos:altLabel "creatures"@en
example:animals skos:prefLabel "animaux"@fr
example:animals skos:altLabel "bêtes"@fr
example:animals skos:hiddenLabel "betes"@fr
```

- Οι έννοιες οργανώνονται μέσω (`skos:broader`, `skos:narrower`) και συσχετιστικών (`skos:related`) σχέσεων.
`example:animals` `rdf:type` `skos:Concept`
`example:animals` `skos:prefLabel` "animals"@en
`example:mammals` `rdf:type` `skos:Concept`
`example:mammals` `skos:prefLabel` "mammals"@en
`example:zoology` `rdf:type` `skos:Concept`
`example:zoology` `skos:prefLabel` "zoology"@en
`example:animals` `skos:narrower` `example:mammals`
`example:mammals` `skos:broader` `example:animals`
`example:animals` `skos:related` `example:zoology`
- Οι ιδιότητες `skos:broader` και `skos:narrower` είναι ανάστροφες μεταξύ τους, αλλά όχι μεταβατικές.
- Δεν υπάρχει πρόβλεψη για σχέσεις μέρους-όλου.

SKOS: Simple Knowledge Organization System

- Περιλαμβάνονται και ιδιότητες για την τεκμηρίωση των εννοιών: `skos:note`, `skos:definition`, `skos:example`.
- Υποϊδιότητες: `skos:scopeNote`, `skos:historyNote`, `skos:editorialNote`, `skos:changeNote`.
`example:microwaveFrequencies skos:scopeNote`
"Used for frequencies between 1GHz to 300Ghz"@en
`example:organizationsOfScienceAndCulture skos:example`
"academies of science, general museums, world fairs"@en
`example:documentation skos:definition`
"the process of storing and retrieving information in all fields of knowledge"@en.
`example:childAbuse skos:historyNote`
"est. 1975; heading was: Cruelty to children [1952-1975]"@en

SKOS: Simple Knowledge Organization System

- Οι έννοιες μπορούν να οργανωθούν ώστε να κατασκευαστούν λεξικά και ταξονομίες.
- Ένα λεξικό ορίζεται ως στιγμιότυπο της έννοιας `skos:ConceptScheme` και οι διάφορες έννοιες συμμετέχουν στο λεξικό μέσω της ιδιότητας `skos:inScheme`.
- Η κεφαλή μιας ταξονομίας δηλώνεται μέσω της ιδιότητας `skos:hasTopConcept`.

```
example:animalThesaurus rdf:type skos:ConceptScheme
example:animalThesaurus dcterms:title "Simple animal thesaurus"
example:mammals rdf:type skos:Concept
example:mammals skos:inScheme example:animalThesaurus
example:fish rdf:type skos:Concept
example:fish skos:inScheme example:animalThesaurus
example:animalThesaurus skos:hasTopConcept example:mammals
example:animalThesaurus skos:hasTopConcept example:fish
```

- Το Schema.org είναι μια κοινή προσπάθεια των Google, Microsoft, Yahoo και Yandex για την δημιουργία ενός σχήματος για τον σημασιολογικό χαρακτηρισμό των δεδομένων που περιέχουν οι ιστοσελίδες.
- Η εισαγωγή σημασιολογικής πληροφορίας στις σελίδες HTML μπορεί να γίνει μέσω των προτύπων Microdata, RDFa και JSON-LD.
- Η σημασιολογική πληροφορία μπορεί να αναγνωστεί από τις μηχανές αναζήτησης έτσι ώστε να αποκτάται πρόσβαση στη σημασιολογικό περιεχόμενο των σελίδων.

- Ονοματώρος: <http://schema.org/>
- Ορίζει μια εκτεταμένη ιεραρχία εννοιών με βασικές έννοιες τις **Action, BroadcastService, CreativeWork, Event, Intangible, MedicalEntity, Organization, Person, Place** και **Product**.
- Απόσπασμα από την ιεραρχία της CreativeWork:

CreativeWork

Article

BlogPosting

NewsArticle

ScholarlyArticle

MedicalScholarlyArticle

Blog

Book

Clip

MovieClip

RadioClip

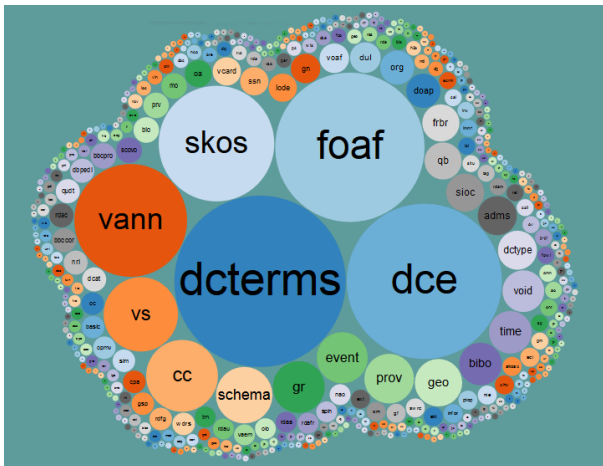
TVClip

- Ορίζεται πλήθος ιδιοτήτων με πεδία ορισμού και τιμών τις έννοιες τις ιεραρχίας.
- Απόσπασμα από τις ιδιότητες με πεδίο ορισμού την έννοια

`schema:CreativeWork`

<code>schema:character</code>	<code>schema:Person</code>
<code>schema:contributor</code>	<code>schema:Person, schema:Organization</code>
<code>schema:comment</code>	<code>schema:Comment</code>
<code>schema:dateCreated</code>	<code>schema>Date</code>
<code>schema:hasPart</code>	<code>schema:CreativeWork</code>
<code>schema:headline</code>	<code>schema:Text</code>
<code>schema:inLanguage</code>	<code>schema:Language schema:Text</code>
<code>schema:keywords</code>	<code>schema:Text</code>
<code>schema:mentions</code>	<code>schema:Thing</code>
<code>schema:translator</code>	<code>schema:Person, schema:Organization</code>

Υπάρχουν πολλά λεξιλόγια που καλύπτουν διαφορετικές ανάγκες:



- Πρότυπο για την εισαγωγή μεταδεδομένων σε ιστοσελίδες.
- Τα εισαγόμενα μεταδεδομένα για να έχουν σημασιολογική αξία πρέπει να αποτελούν όρους κάποιου λεξιλογίου.
- Τα microdata είναι στοιχεία εντός μιας σελίδας που χαρακτηρίζονται από ιδιότητες. Κάθε ιδιότητα είναι ένα ζεύγος ονόματος–τιμής.
- Ένα στοιχείο δηλώνεται μέσω του `itemscope` και μια ιδιότητά του μέσω του `itemprop`.

```
<div itemscope>
  <p>My name is <span itemprop="name">Elizabeth</span>.</p>
  <p>I am <span itemprop="nationality">British</span>.</p>
  <p>I was born on <data itemprop="birthday"
                        value="2009-05-10">May 10th 2009</data>.</p>
</div>
```

- Για την χρήση λεξιλογίων κάθε στοιχείο μπορεί να αντιστοιχιστεί σε έναν τύπο μέσω του `itemtype`.

```
<div itemscope itemtype="schema:Person">
  <p>My name is
    <span itemprop="schema:givenName">Elizabeth</span>.</p>
  <p>I am
    <span itemprop="schema:nationality">British</span>.</p>
  <p>I was born on <data itemprop="schema:birthDate"
    value="2009-05-10">May 10th 2009</data>.</p>
</div>
```