

Τάσεις και τεχνολογίες Σημασιολογικού Ιστού

Ηρακλής Θεοφανίδης

Τμήμα Πληροφορικής Και
Ηλεκτρονικών Συστημάτων
Θεσσαλονίκη, Κεντρική
Μακεδονία, Ελλάδα

iraklistheofanidis@gmail.com

Αχιλλέας Γκέκας

Τμήμα Πληροφορικής Και
Ηλεκτρονικών Συστημάτων
Θεσσαλονίκη, Κεντρική
Μακεδονία, Ελλάδα

gkekas57@gmail.com

Ερρίκος Καλτσόπουλος

Τμήμα Πληροφορικής Και
Ηλεκτρονικών Συστημάτων
Θεσσαλονίκη, Κεντρική
Μακεδονία, Ελλάδα

ekal638@gmail.com

1.Περίληψη

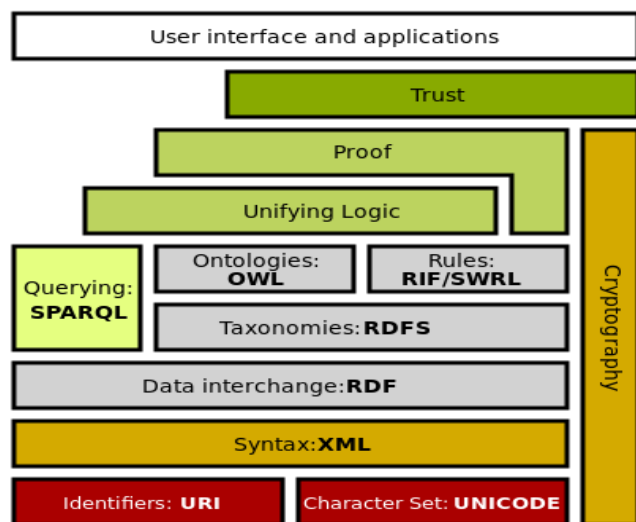
Σε αυτό το άρθρο,θα μιλήσουμε για κάποιες βασικές τεχνολογίες του σημασιολογικού Ιστού.Έπειτα θα επεκτείνουμε το ταξίδι μας σε πιο πρόσφατες τεχνολογίες.Επιπλέον θα αναφέρουμε κάποιες χρήσεις αυτών σε εφαρμογές .Τέλος θα μιλήσουμε για κάποιες τελευταίες τάσεις και που εφαρμόζονται.

2.Εισαγωγή

Μέχρι πρότινος στον παγκόσμιο Ιστό υπήρχε πάρα πολύ πληροφορία την οποία δεν μπορούσε ο υπολογιστής να καταλάβει και να επεξεργαστεί.Εξαιτίας αυτής της ανάγκης δημιουργήθηκε ο Σημασιολογικός Ιστός.

Με τον όρο Σημασιολογικό Ιστό αναφερόμαστε σε μία καινούργια μορφή Παγκόσμιου Ιστού (World Wide Web), που έρχεται να επεκτείνει τον ήδη υπάρχον.Αυτό το επιτυγχάνει χρησιμοποιώντας μεταδεδομένα τα οποία θα είναι κατανοητά από υπολογιστικά συστήματα , τα οποία θα μπορούν να επεξεργαστούν και να τα συλλέξουν.

Επιπλέον ο Σημασιολογικός ιστός έχει μία αρχιτεκτονική η οποία έχει δημιουργηθεί απο το 2001 και πάνω σε αυτήν την αρχιτεκτονική βασίζονται οι πιο πολλές νέες τεχνολογίες .



Εικόνα 1.Αρχιτεκτονική Σημασιολογικού Ιστού

Ο Σημασιολογικός Ιστός βασίζεται σε τεχνολογίες που ήδη υπάρχουν (URI και XML) αλλά και σε νέες τεχνολογίες

(RDF, RDFS, OWL, κα.), οι οποίες αναπτύσσονται με την βοήθεια της κοινότητας. Δεδομένου ότι ο νέος Ιστός σκοπεύει να είναι μια μεγάλη βάση όπου δεδομένα από διαφορετικά πεδία θα συνδέονται μεταξύ τους, αναμένεται να παίξει μεγάλο ρόλο στη ζωή μας.

3.Γλώσσες Εναιολογικής Αναπαράστασης

3.1 XML

Η XML είναι μία γλώσσα σήμανσης (markup language) η οποία περιέχει ένα σύνολο κανόνων για την ηλεκτρονική κωδικοποίηση κειμένων . Έχει αρκετά κοινά με την HTML, αφού είναι και οι δύο, γλώσσες σήμανσης. Αποτελεί ένα δομημένο έγγραφο με βάση της ετικέτες και έχει ιεραρχική δομή. Η XML δημιουργήθηκε με σκοπό να ξεπεραστούν κάποιες αδυναμίες της HTML, όπως η μη ύπαρξη κατάλληλης δομής στις πληροφορίες, αναγνώρισης από τις μηχανές και περιγράφει με σαφήνεια τις πληροφορίες και κάποιες συσχετίσεις μεταξύ τους και όχι απλά την εμφάνισή τους. Επίσης μπορεί να γίνει επεξεργασία των πληροφοριών, έτσι ώστε να χρησιμοποιηθούν από διάφορες εφαρμογές. Μπορούμε να την χαρακτηρίσουμε ως μεταγλώσσα που επιτρέπει την αναπαράσταση δομημένων πληροφοριών στον Ιστό, την ανταλλαγή τους μεταξύ διαφόρων εφαρμογών και την χρήση τους από αυτές, καθώς και την προσπέλασή τους από πράκτορες λογισμικού.

Για να γίνει καλύτερα κατανοητή η σημασία της XML και η διαφορά της σε σχέση με την HTML, παρακάτω παραθέτουμε ένα παράδειγμα περιγραφής ενός βιβλίου, πρώτα χρησιμοποιώντας HTML:

```
<h2>A Semantic Web Primer</h2>
<i>by <b>Grigoris Antoniou</b> and <b>Frank van
Harmelen</b></i><br>
MIT Press 2004<br>
ISBN 0-262-01210-3
```

Εικόνα 2. παράδειγμα χρήσης Html

Ένα παράδειγμα αναπαράστασης της παραπάνω πληροφορίας σε XML θα μπορούσε να είναι το εξής:

```
<book>
<title>A Semantic Web Primer</title>
<author>Grigoris Antoniou</author>
<author> Frank van Harmelen </author>
<publisher>MIT Press</publisher>
<year>2004</year>
<isbn>0-262-01210-3</isbn>
</book>
```

Εικόνα 3. Παράδειγμα XML

Όπως παρατηρούμε, τόσο η HTML όσο και η XML, όντας γλώσσες σήμανσης και οι δύο βασίζονται στα λεγόμενα tags (π.χ. και). Μία από τις διαφορές στην XML είναι ότι όλα τα tags θα πρέπει να είναι κλειστά (π.χ. το book), ενώ στην HTML κάποια tags όπως π.χ. το
, μπορούν να μένουν ανοιχτά. Το περιεχόμενο, μαζί με τα tags από τα οποία περιβάλλεται, ονομάζεται στοιχείο (element).

Κάτι άλλο που παρατηρούμε, είναι ότι κάποιος μπορεί να διαβάσει τα παραπάνω παραδείγματα αρκετά εύκολα, μιας και οι δύο γλώσσες σχεδιάστηκαν για να είναι εύκολα κατανοητές και εύχρηστες από τους ανθρώπους

Η XML πέρα από γλώσσα για τη σήμανση εγγράφων, αποτελεί και ένα ευρέως διαδεδομένο format για την ανταλλαγή δεδομένων ανάμεσα σε εφαρμογές. Μάλιστα, η χρήση της XML ως μέσο ανταλλαγής δεδομένων είναι πολύ περισσότερο διαδεδομένη σε σχέση με τον πραγματικό σκοπό για τον οποίο προοριζόταν κατά τη δημιουργία της, δηλαδή ως γλώσσα σήμανσης εγγράφων. Οι επιχειρήσεις πολύ συχνά ανταλλάσσουν έγγραφα με τους συνεργάτες τους και ανανεώνουν ανάλογα τις βάσεις δεδομένων τους. Εάν δεν υπάρχει ένα προ-συμφωνημένο κοινό πρότυπο όπως η XML, ο κάθε συνεργάτης ξεχωριστά θα πρέπει να αναπτύσσει το δικό του λογισμικό για την επεξεργασία και ανάκτηση τέτοιων εγγράφων, με αποτέλεσμα να αυξάνονται τα κόστη.

3.2 XML SCHEMA

Το XML Schema αποτελεί μία σημαντικά πιο πλούσια γλώσσα σε σχέση με το Document Type Definition, για τον προσδιορισμό της δομής XML εγγράφων. Ένα από τα κύρια χαρακτηριστικά του είναι ότι για τη δημιουργία του χρησιμοποιείται η XML αυτή καθ' αυτή σε αντίθεση με τα DTDs. Μία τέτοια σχεδιαστική απόφαση παρέχει δύο σημαντικά πλεονεκτήματα: αυξάνεται κατά πολύ η ευκολία ανάγνωσης ενός XML Schema, ενώ επιτυγχάνεται σημαντική επαναχρησιμοποίηση της τεχνολογίας. Δεν είναι πλέον απαραίτητη η δημιουργία δύο διαφορετικών ειδών λογισμικού, όπως στην περίπτωση του DTD, δηλαδή ένα για την επεξεργασία των DTDs και ένα για την

επεξεργασία των XML εγγράφων, μιας και χρησιμοποιείται XML και στις δύο περιπτώσεις.

Μία ακόμα βασική βελτίωση σε σχέση με τα DTDs, είναι η δυνατότητα που παρέχεται για την επαναχρησιμοποίηση και βελτίωση των ίδιων των XML σχημάτων. Ένα XML Schema επιτρέπει τη δημιουργία νέων τύπων με την επέκταση ή τον περιορισμό υφιστάμενων σχημάτων. Σε συνδυασμό με τη βασισμένη σε XML σύνταξη, αυτό το χαρακτηριστικό καθιστά δυνατή τη δημιουργία XML σχημάτων από άλλα σχήματα που έχουν δημιουργηθεί ήδη, με αποτέλεσμα τη μείωση του συνολικού χρόνου και κόστους. Τέλος, το XML Schema παρέχει ένα πλούσιο σετ από τύπους δεδομένων που μπορούν να χρησιμοποιηθούν σε έγγραφα XML, ενώ τα DTDs επιτρέπουν μόνο τη χρήση στοιχειοσειρών (strings).

3.3 RDF

Το RDF (Resource Description Framework) είναι μια δομή για να περιγράψεις και να συσχετίζεις πόρους στο web. Είναι σχεδιασμένο για να το διαβάζει και να το καταλαβαίνει ο υπολογιστής και να το χρησιμοποιούν οι εφαρμογές ώστε να κάνει πιο έξυπνες αναζητήσεις και συσχετίσεις. Δεν είναι σχεδιασμένο για να το χρησιμοποιεί ο άνθρωπος άλλωστε ένας από τους λόγους που δημιουργήθηκε ήταν επειδή ο υπολογιστής δεν μπορούσε να κατανοήσει την πληροφορία που υπήρχε στο web. Επίσης χρησιμοποιείται για να δημιουργήσουμε knowledge graphs. Το RDF βασίζεται στα υπάρχοντα πρότυπα Web: XML και URL (URI). Με την χρήση RDF επίσης μπορεί να γίνει η χρήση μοντέλων δεδομένων σε προτάσεις. Μια τριπλετα RDF αποτελείται από Υποκείμενο - κατηγορήμα - αντικείμενο όπου το υποκείμενο έχει μια ιδιότητα η οποία έχει μια τιμή. Κάθε τριπλετα μπορεί να είναι μέρος άλλης τριπλετας. Ακόμη κάθε κομμάτι δεν είναι υποχρεωτικό να έχει κάποια τιμή. Παρακάτω θα ακολουθήσει παράδειγμα.

Που χρησιμοποιείται όμως το RDF?

Παραδείγματα χρήσεις του rdf είναι για να περιγράψει :

- Για προϊόντα προς αγορά όπως η τιμή, διαθεσιμότητα
- Πληροφορίες για Ιστοσελίδες όπως το περιεχόμενο, τον συγγραφέα, την ημερομηνία δημιουργίας ή τροποποίησης
- Πληροφορίες για το περιεχόμενο εικόνων στο διαδίκτυο
- Περιεχόμενο για μηχανές αναζήτησης
- Ηλεκτρονικές Βιβλιοθήκες

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:edu="http://www.example.org/">
  <rdf:Description rdf:about="http://www.princeton.edu">
    <geo:lat>40.35</geo:lat>
    <geo:long>-74.66</geo:long>
    <edu:hasDept rdf:resource="http://www.cs.princeton.edu"
      dc:title="Department of Computer Science"/>
  </rdf:Description>
</rdf:RDF>

```

Εικόνα 4. Παράδειγμα RDF

Σε ένα έγγραφο RDF / XML υπάρχουν δύο τύποι κόμβων: κόμβοι πόρων και κόμβοι ιδιότητας. Οι κόμβοι πόρων είναι τα αντικείμενα των δηλώσεων και συνήθως έχουν μια ιδιότητα `rdf: about` σε αυτά δίνοντας το URI του πόρου που αντιπροσωπεύουν. Σε αυτό το παράδειγμα, ο κόμβος `rdf: Description` είναι ο μοναδικός κόμβος πόρου.

Οι κόμβοι πόρων περιέχουν (μόνο) κόμβους ιδιότητας, οι οποίοι αντιπροσωπεύουν δηλώσεις. Υπάρχουν τρεις δηλώσεις σε αυτό το παράδειγμα, όλες με το θέμα `<http://www.princeton.edu>`, και με τα προκαθορισμένα `geo: lat`, `geo: long` και `edu: hasDept`.

Οι κόμβοι των ιδιοτήτων, με τη σειρά τους, περιέχουν κυριολεκτικές τιμές, όπως "40.35" και "-74.66", ή μια αναφορά σε έναν πόρο αντικειμένου χρησιμοποιώντας το χαρακτηριστικό `rdf: resource` ή μπορεί να περιέχουν ως αντικείμενό τους έναν πλήρη κόμβο πόρων.

Subject	Predicate	Object

<code><http://www.princeton.edu></code>	<code>edu:hasDept</code>	<code><http://www.cs.princeton.edu></code>
<code><http://www.princeton.edu></code>	<code>geo:lat</code>	"40.35"
<code><http://www.princeton.edu></code>	<code>geo:long</code>	"-74.66"
<code><http://www.cs.princeton.edu></code>	<code>dc:title</code>	"Department of Computer Science"

Εικόνα 5.

RDF Schema(RDFS)

Το RDF παρέχει έναν τρόπο για την περιγραφή πηγών πληροφοριών, με τη χρήση απλών προτάσεων. Παρόλα αυτά, το RDF από μόνο του δεν είναι ικανό να ορίζει την έννοια των όρων που χρησιμοποιούνται σε αυτές τις προτάσεις και συγκεκριμένα να υποδηλώνει ότι αυτές οι προτάσεις περιγράφουν συγκεκριμένα είδη ή κλάσεις πόρων χρησιμοποιώντας συγκεκριμένες ιδιότητες για την περιγραφή τους. Για παράδειγμα, εάν μας ενδιέφερε η περιγραφή βιβλιογραφικών πόρων, θα θέλαμε να περιγράψουμε κατηγορίες όπως `ex:Book` ή `ex:MagazineArticle` και ιδιότητες όπως `ex:author`, `ex:title` και `ex:subject`. Άλλες εφαρμογές μπορεί να χρειάζεται να περιγράψουν κλάσεις όπως `ex2:Person` και `ex2:Company` και ιδιότητες όπως `ex2:age`, `ex2:jobTitle`, `ex2:stockSymbol` και `ex2:numberOfEmployees`. Το RDF αυτό καθ' αυτό δεν παρέχει κανένα μέσο για τον προσδιορισμό κλάσεων και ιδιοτήτων. Τέτοιου είδους κλάσεις και ιδιότητες περιγράφονται ως ένα RDF λεξιλόγιο, μέσω ενός μηχανισμού που ονομάζεται RDF Schema (RDFS).

Σκόπιμα τονίσαμε τη λέξη μέσω στην προηγούμενη παράγραφο, μιας και το RDF Schema δεν παρέχει ένα συγκεκριμένο λεξιλόγιο που να αναφέρεται σε συγκεκριμένες εφαρμογές (όπως για παράδειγμα οι κλάσεις και οι ιδιότητες που περιγράψαμε παραπάνω). Αντί αυτού, το RDFS παρέχει τις προϋποθέσεις που απαιτούνται για να περιγράψει κανείς τέτοιες κλάσεις και ιδιότητες και για να υποδηλώσει ποιες κλάσεις και ιδιότητες αναμένεται να χρησιμοποιηθούν μαζί (για παράδειγμα να καθορίσει ότι η ιδιότητα `ex:jobTitle` θα χρησιμοποιηθεί για την περιγραφή του `ex:Person`). Με άλλα λόγια, το RDF Schema παρέχει ένα σύστημα τύπων για το RDF. Το σύστημα αυτό του RDFS είναι παρόμοιο, έως ένα βαθμό, με τα συστήματα τύπων των αντικειμενοστραφών γλωσσών προγραμματισμού όπως η Java. Για παράδειγμα, το RDF Schema επιτρέπει τον ορισμό πόρων ως στιγμιότυπα (instances) μίας ή περισσότερων κλάσεων. Επιπλέον, επιτρέπει την οργάνωση των κλάσεων σε μία ιεραρχική δομή. Παραδείγματος χάριν, μία κλάση `ex:Dog` μπορεί να οριστεί ως υποκλάση του `ex:Mammal`, το οποίο με τη σειρά του αποτελεί υποκλάση του `ex:Animal`, εννοώντας ότι οποιοσδήποτε πόρος που είναι στιγμιότυπο της κλάσης `ex:Dog` είναι επίσης έμμεσα στιγμιότυπο της κλάσης `ex:Animal`.

Οι ιδιότητες του RDF Schema, παρέχονται και οι ίδιες υπό την μορφή ενός RDF λεξιλογίου, δηλαδή ως ένα εξειδικευμένο σετ από προκαθορισμένες πηγές RDF με το

δικό τους ειδικό νόημα. Οι πηγές του λεξιλογίου του RDF Schema προσδιορίζονται με το πρόθεμα `rdfs:` το οποίο παραπέμπει στο URI `http://www.w3.org/2000/01/rdf-schema#`. Οι περιγραφές του λεξιλογίου που είναι γραμμένες στη γλώσσα του RDF Schema, αποτελούν έγκυρους RDF γράφους. Συνεπώς, ένα πρόγραμμα λογισμικού το οποίο δεν έχει αναπτυχθεί συγκεκριμένα για να επεξεργάζεται και το λεξιλόγιο του RDF Schema, από τη μία μεριά έχει τη δυνατότητα να ερμηνεύσει ένα σχήμα ως έναν έγκυρο RDF γράφο που αποτελείται από διάφορες πηγές και ιδιότητες, αλλά από την άλλη μεριά δεν έχει τη δυνατότητα να «κατανοήσει» το νόημα των όρων που χρησιμοποιεί το RDF Schema. Για την κατανόηση αυτών των νοημάτων, το λογισμικό θα πρέπει να υποστηρίζει και μία εκτεταμένη γλώσσα που περιλαμβάνει όχι μόνο το λεξιλόγιο `rdf:`, αλλά και το λεξιλόγιο `rdfs:` μαζί με τα ενσωματωμένα νοήματά τους.

3.4 OWL

Κάπου στο 2000 η Darpa ξεκίνησε μια νέα τεχνολογία με όνομα DAML και σε συνδυασμό με την OIL δημιούργησαν την DAML +OIL η οποία ήταν επέκταση της RDFS. Όπου και η DAML +OIL υποστηρίζει τις κλάσεις, ένωση, τομή και να θέτει περιορισμούς σε αριθμούς συνόλων. Η owl (Ontology Web Language) ξεκίνησε σαν ερευνά για την αναθεώρηση των DAML και OIL ώστε να χρησιμοποιηθούν στον σημασιολογικό ιστό.

Η Owl είναι ένας συνδυασμός από Markup language η οποία είναι σχεδιασμένη για χρήση από εφαρμογές οι οποίες χρειάζονται να επεξεργαστούν δεδομένα. Με την γλώσσα οντολογιών OWL προσπαθούμε να πετύχουμε ότι και με το RDFS, να ορίσουμε, δηλαδή, οντολογίες για την περιγραφή ενός πεδίου γνώσης, οι οποίες περιέχουν κλάσεις, ιδιότητες και τις σχέσεις μεταξύ τους, με την διαφορά ότι με την OWL μας δίνεται η δυνατότητα να εκφράσουμε πιο περίπλοκες συσχετίσεις, δημιουργώντας, με τον τρόπο αυτό, εφαρμογές που περιέχουν συστήματα και κανόνες συλλογιστικής. Εν ολίγοις, OWL είναι η επέκταση του RDFS, περιέχει το λεξιλόγιο του, αλλά προσφέρει μεγαλύτερη εκφραστικότητα στην οποία ο κόσμος είναι ένα σύνολο από κλάσεις, ιδιότητες και άτομα.

Η γλωσσά OWL χωρίζεται σε τρεις γλώσσες με διαφορετική εκφραστική ισχύ. Οι τρεις αυτές γλώσσες είναι η Owl Lite με την μικρότερη εκφραστική ισχύ, OWL DL, OWL Full με την μεγαλύτερη εκφραστική ισχύ όπου και οι τρεις αυτές γλώσσες ακολουθούν την σύνταξη `rdf`.

3.4.1 OWL Lite

Η Owl Lite έχει σχεδιαστεί ώστε να υποστηρίζει την ιεραρχία της ταξινόμησης και να έχει απλούς περιορισμούς ιδιοτήτων. Επίσης περιέχει ένα υποσύνολο του λεξιλογίου της Owl. Έτσι η owl lite υποστηρίζει βασικές έννοιες περιγραφικής Λογικής, Μεταβατικές και αντίστροφες ιδιότητες. Ακόμη ικανοποιεί περιορισμούς όπως είναι η μη χρήση ιδιοτήτων `px`:

`owl: hasValue`, `owl: unionOf`, όπως `owl: oneOf` αλλά δεν επιτρέπονται οι δηλώσεις `owl: equivalentClass` μεταξύ ανωνύμων κλάσεων. Τέλος, είναι εύκολη στην εκμάθηση και λόγω της απλότητας στην έκφραση την προτιμούμε όπου υπάρχει ανάγκη για ταχύτητα.

3.4.2 OWL DL

Η OWL DL δεν είναι συμβατή με την `rdf` αλλά υποστηρίζει τους συλλογισμούς και διατηρεί ως έναν βαθμό την υπολογιστική πληρότητα επίσης παρέχει μεγάλη εκφραστικότητα. Ένα έγγραφο RDF για να γίνει συμβατό με την owl dl απαιτεί αρκετή τροποποίηση και επέκταση.

Τα αρχικά DL προέρχονται από τα συστήματα.

Περιγραφικής Λογικής (Description Logic), για την περιγραφή τάξεων αντικειμένων και την πραγματοποίηση ελέγχων υπαγωγής και ταξινόμησης ανάμεσα σε τάξεις και οντότητες. Η OWL DL έχει ιδιότητες επισημείωσης και είναι περιγραφική, αφού μπορούμε να χρησιμοποιήσουμε συστήματα συλλογιστικής. Το λεξικό που χρησιμοποιείται σε αυτή είναι υποσύνολο της OWL Full και ικανοποιεί κάποιους περιορισμούς όπως:

- Ένας πόρος μπορεί να είναι κάτι συγκεκριμένο και μόνο αυτό. Για παράδειγμα, μια κλάση μπορεί να είναι μόνο μια κλάση και όχι παράλληλα και τύπος δεδομένων ή αντικείμενο, μια ιδιότητα μπορεί να έχει συγκεκριμένες τιμές και όχι μερικές από έναν τύπο δεδομένων.
- Οι ιδιότητες τύπου δεδομένων και οι ιδιότητες αντικειμένου είναι αμοιβαία αποκλειόμενα σύνολα.
- Οι περιορισμοί πληθικότητας δεν είναι μεταβατικές ιδιότητες. Δε μπορούν να τεθούν περιορισμοί από μεταβατικές ιδιότητες στα στοιχεία συνόλων.

3.4.3 OWL Full

Η OWL Full χρησιμοποιεί όλες τις θεμελιώδεις αρχές των άλλων υπογλωσσών και σε συνδυασμό με την RDF και το RDF Schema, επιτρέπει μέγιστη εκφραστικότητα χωρίς όμως εγγυήσεις επιλυσιμότητας, αφού είναι τόσο ισχυρή που δεν μπορεί να παρέχει αποδοτική υποστήριξη συλλογισμών. Η OWL Full είναι αναδρομικά συμβατή με

RDF, RDFS, OWL Lite και OWL DL, περιέχει ολόκληρη τη γλώσσα OWL και χρησιμοποιεί όλα τα δομικά συστατικά της. Οποιοδήποτε έγγραφο είναι έγκυρο στις παραπάνω γλώσσες είναι έγκυρο και στην OWL Full. Το ίδιο ισχύει και για τα συμπεράσματα. Εξαιτίας όμως αυτής της εκφραστικότητας δεν είναι σίγουρο ότι θα λήγουν οι διαδικασίες συλλογιστικής σε πεπερασμένο χρόνο .

Παράδειγμα σύνταξης της Owl
Η βάση της σύνταξης της Owl είναι η xml.

Επικεφαλίδα OWL

Αν και δεν είμαστε υποχρεωμένοι να βάζουμε επικεφαλίδα, γενικά είναι μια καλή πρακτική να τοποθετούμε ώστε να βοηθάμε τους άλλους να κατανοούν τι περιέχει η οντολογία.

```
01. <rdf:RDF
02.   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
03.   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
04.   xmlns:owl="http://www.w3.org/2002/07/owl#"
05.   xmlns:dc="http://purl.org/dc/elements/1.1/"
06.
07.   <!-- OWL Header Example -->
08.   <owl:Ontology rdf:about="http://www.linkeddatatools.com/plants">
09.     <dc:title>The LinkedDataTools.com Example Plant Ontology</dc:title>
10.     <dc:description>An example ontology written for the LinkedDataTools.com RDFS &
11.       OWL introduction tutorial</dc:description>
12.   </owl:Ontology>
13.
14.   <!-- Remainder Of Document Omitted For Brevity... -->
15. </rdf:RDF>
```

Εικόνα 6. Παράδειγμα OWL

στο παραπάνω παράδειγμα υπάρχει ο τίτλος και η περιγραφή της οντολογίας επίσης σε αυτό το μέρος μπορούμε να βάλουμε και versioning ώστε να παρακολουθούμε τα updates

Κλάσεις OWL

Στην Owl έχουμε επίσης κλάσεις (owl:Class) και υποκλάσεις (rdfs:subClass) για να δηλώσουμε ιεραρχίες

```
<owl:Class rdf:ID="associateProfessor">
<rdfs:subClassOf
rdf:resource="#academicStaffMember"/>
</owl:Class>
```

Εικόνα 7. Παράδειγμα Class OWL

Ιδιότητες OWL

Στην owl έχουμε δύο τύπους ιδιοτήτων, ιδιότητες αντικειμένων (owl: ObjectProperty), σύνδεση αντικειμένων με άλλα αντικείμενα
ιδιότητες τύπων δεδομένων (owl: DatatypeProperty)
Σύνδεση αντικειμένων με κάποιες τιμές

```
11. <!-- Define the family property -->
12. <owl:DatatypeProperty rdf:about="http://www.linkeddatatools.com/plants#family"/>
13.
14. <!-- Define the similarlyPopularTo property -->
15. <owl:ObjectProperty
16.   rdf:about="http://www.linkeddatatools.com/plants#similarlyPopularTo"/>
17.
18. <!-- Define the Orchid class instance -->
19. <rdf:Description rdf:about="http://www.linkeddatatools.com/plants#orchid">
20.
21.   <!-- Orchid is an individual (instance) of the flowers class -->
22.   <rdf:type rdf:resource="http://www.linkeddatatools.com/plants#flowers"/>
23.
24.   <!-- The orchid is part of the 'Orchidaceae' family -->
25.   <plants:family>Orchidaceae</plants:family>
26.
27.   <!-- The orchid is similarly popular to the magnolia -->
28.   <plants:similarlyPopularTo
29.     rdf:resource="http://www.linkeddatatools.com/plants#magnolia"/>
30. </rdf:Description>
```

Εικόνα 8. Παράδειγμα Ιδιοτήτων OWL

Κάποιες ακόμη ιδιότητες που είναι καλό να γνωρίζει κάποιος

- owl:versionInfo : πληροφορίες για την έκδοση της οντολογίας (με τη μορφή σχολίων φυσικής γλώσσας)
- rdfs:label : περιγραφή σε φυσική γλώσσα για κλάσεις, οντότητες και ιδιότητες
- rdfs:comment : εισαγωγή σχολίου
- rdfs:seeAlso : περισσότερες σχετικές πληροφορίες
- rdfs:isDefinedBy : καθορισμός των σχετικών διευθύνσεων που παρέχουν πληροφορίες για την οντολογία.

4. Linked Data

Τα Linked Data είναι μια μέθοδος δημοσίευσης δομημένων δεδομένων χρησιμοποιώντας λεξιλόγια όπως το schema.org που μπορούν να συνδεθούν μεταξύ τους και να ερμηνευτούν από μηχανές. Χρησιμοποιώντας συνδεδεμένα δεδομένα, statements encoded in triples μπορούν να διαδοθούν σε διαφορετικούς ιστότοπους.

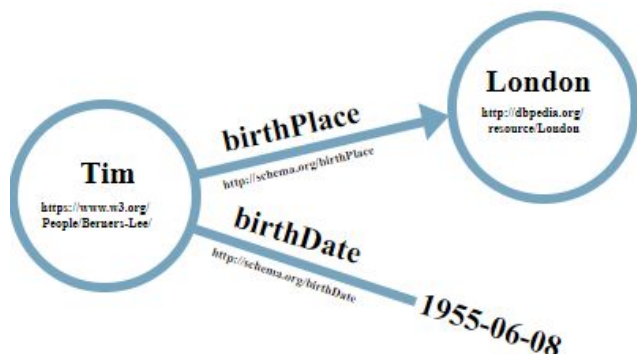
Ένα triple είναι το πιο θεμελιώδες μέρος του τρόπου με τον οποίο η πληροφορία κωδικοποιείται σε ένα γράφημα γνώσης χρησιμοποιώντας τεχνολογίες σημασιολογικού ιστού. Ένα triple περιλαμβάνει μια έκφραση subject–predicate–object .

Κάθε οντότητα συνδέεται με την επόμενη μέσω σχέσεων που εκφράζονται με απλά triple. Όπως και στην πραγματικότητα, όπου από απλούς τοπικούς κανόνες, προκύπτουν πολυπλοκότητα. Με την ανάμιξη δισεκατομμυρίων triple, παίρνετε τον Σημασιολογικό Ιστό.

π.χ. triple

Το αντικείμενο του πρώτου triple, για το birthPlace, περιέχει έναν σύνδεσμο (ένα IRI) σε κάποια άλλη πηγή (Λονδίνο). Το αντικείμενο του δεύτερου triple (η γέννηση) δεν είναι ένας σύνδεσμος, αλλά μια λεγόμενη κυριολεκτική

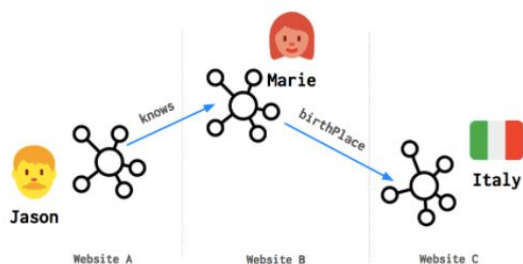
αξία. Η κυριολεκτική τιμή δεν μπορεί να έχει ιδιότητες, αφού δεν είναι πηγή.



Εικόνα 9. .Linked Data

Πως Λειτουργει το Linked Data (Ενα παράδειγμα):

Ας πούμε ότι στην Ιστοσελίδα Α μπορούμε να παρουσιάσουμε την οντότητα Jason και το γεγονός ότι γνωρίζει τη Μαρία. Στην ιστοσελίδα Β μπορούμε να παράξουμε όλες τις πληροφορίες σχετικά με τη Marie και στην ιστοσελίδα C μπορούμε να βρούμε πληροφορίες για τη γενέτειρα της Μαρίας.



Εικόνα 10 .Παράδειγμα Linked data

Κάθε σελίδα περιέχει τα δομημένα δεδομένα για να περιγράψει μια οντότητα (Jason, Marie και Ιταλία) και τη σύνδεση με την οντότητα που θα μπορούσε να περιγραφεί σε διαφορετική σελίδα ή ακόμα και σε διαφορετικό ιστότοπο.

Το 2006, ως εφευρέτης του World Wide Web, ο δημιουργός και υποστηρικτής του Σημασιολογικού Ιστού και ο Διευθυντής του W3C ο Tim Berners-Lee περιέγραψε τα συνδεδεμένα δεδομένα ως εξής:

“Ο Σημασιολογικός Ιστός δεν αφορά μόνο την τοποθέτηση δεδομένων στο διαδίκτυο. Πρόκειται για τη δημιουργία δεσμών, έτσι ώστε ένα άτομο ή μια μηχανή να μπορεί να εξερευνήσει τον ιστό των δεδομένων. Με τα συνδεδεμένα

δεδομένα, όταν έχετε κάποια από αυτά, μπορείτε να βρείτε άλλα σχετικά δεδομένα.”

5. Query Languages

5.1 Sparql

Τα ερωτήματα σχετικά με τα Linked Data γίνονται με τη χρήση μιας σημασιολογικής γλώσσας ερωτήματος που ονομάζεται SPARQL, η οποία επιτρέπει την ανάκτηση και τον χειρισμό δεδομένων που είναι αποθηκευμένα στη μορφή του RDF (Resource Description Framework).

Το SPARQL, (Πρωτόκολλο SPARQL και RDF Language Query), επιτρέπει στους χρήστες να αναζητούν πληροφορίες από βάσεις δεδομένων ή οποιαδήποτε άλλη πηγή δεδομένων που μπορεί να αντιστοιχιστεί στο RDF. Το πρότυπο SPARQL έχει σχεδιαστεί και εγκριθεί από το W3C και βοηθά τους χρήστες και τους προγραμματιστές να επικεντρωθούν σε αυτό που θα ήθελαν να μάθουν αντί για το πώς οργανώνεται μια βάση δεδομένων.

Ο Sir Tim Berners-Lee, είπε χαρακτηριστικά:

"Η προσπάθεια χρήσης του Σημασιολογικού Ιστού χωρίς SPARQL είναι σαν να προσπαθείτε να χρησιμοποιήσετε μια σχεσιακή βάση δεδομένων χωρίς SQL. Το SPARQL καθιστά δυνατή την αναζήτηση πληροφοριών από βάσεις δεδομένων και άλλες ποικίλες πηγές σε άγρια φύση, σε ολόκληρο τον ιστό. "

5.2 RQL

Η RQL (RDF Query Language) αναπτύχθηκε αρχικά από το Ινστιτούτο Πληροφορικής στο Ηράκλειο, στο πλαίσιο του Ευρωπαϊκού Έργου IST MESMUSES.3. Η RQL υιοθετεί τη σύνταξη της OQL (ένα πρότυπο γλώσσας ερωτήματος για αντικειμενοστρεφείς βάσεις δεδομένων) και, όπως το OQL, καθορίζεται μέσω ενός συνόλου ερωτημάτων πυρήνα, ενός συνόλου βασικών φίλτρων και ενός τρόπου δημιουργίας νέων ερωτημάτων μέσω λειτουργικής σύνθεσης και τους iterators.

Τα βασικά ερωτήματα είναι τα βασικά δομικά στοιχεία της RQL, τα οποία δίνουν πρόσβαση στα ειδικά για RDFS περιεχόμενα ενός τριπλού RDF. Η RQL επιτρέπει ερωτήματα όπως Class (ανάκτηση όλων των κλάσεων), Property (ανάκτηση όλων των ιδιοτήτων) ή Employee (επιστρέφοντας όλες τις εμφανίσεις της κλάσης με όνομα Employee).

5.3 RDQL

Η RDQL (RDF Data Query Language) είναι μια γλώσσα ερωτήματος για το RDF που αναπτύχθηκε αρχικά για τα μοντέλα Jena. Η RDQL είναι μια εφαρμογή της γλώσσας ερωτήματος RDF SquishQL, η οποία προέρχεται από το rdflib. Αυτή η κλάση των γλωσσών επερωτήσεων αφορά το RDF ως τριπλά δεδομένα, χωρίς πληροφορίες σχεδίων ή οντολογιών, εκτός εάν περιλαμβάνονται ρητά στην πηγή RDF.

Εκτός από το RDF4J, τα ακόλουθα συστήματα παρέχουν RDQL (όλες οι υλοποιήσεις αυτές είναι γνωστό ότι προέρχονται από την αρχική γραμματική): Jena, RDFStore, PHP XML Classes, 3Store και RAP (API RDF για PHP).

5.4 SeRQL

Το SeRQL (Γλώσσα ερωτημάτων RDF Sesame, προφέρεται "κύκλος") είναι μια γλώσσα ερωτήματος RDF / RDFS που αναπτύχθηκε από τον προγραμματιστή του Sesame Aduna ως μέρος του Sesame (τώρα RDF4J). Συνδυάζει επιλεκτικά τα καλύτερα χαρακτηριστικά (που θεωρούνται από τους δημιουργούς του) άλλων γλωσσών επερωτήσεων (RQL, RDQL, N-Triple, N3) και προσθέτει μερικά χαρακτηριστικά γνωρίσματα. Από αυτή τη γραφή, το SeRQL παρέχει προηγμένες λειτουργίες που δεν είναι ακόμα διαθέσιμες στο SPARQL. Μερικά από τα σημαντικότερα χαρακτηριστικά της SeRQL είναι:

- Μετασχηματισμός γραφήματος.
- Υποστήριξη σχήματος RDF.
- Υποστήριξη τύπου δεδομένων XML Schema.
- Σύνταξη έκφρασης έκφρασης διαδρομής.
- Προαιρετική αντιστοίχιση διαδρομής.

5.5 N3

Το N3 παρέχει μια σύνταξη με βάση το κείμενο για το RDF. Επομένως, το μοντέλο δεδομένων του N3 συμμορφώνεται με το μοντέλο δεδομένων RDF. Επιπρόσθετα, το N3 επιτρέπει να ορίσετε κανόνες, οι οποίοι σημειώνονται χρησιμοποιώντας μια ειδική σύνταξη, για παράδειγμα: `Y rdfs: label "foo" =>? Ya: QueryResult`. Τέτοιοι κανόνες, αν και δεν αποτελούν γλώσσα ερώτησης *per se*, μπορούν να χρησιμοποιηθούν για τους σκοπούς ερώτησης. Για το σκοπό αυτό, τα ερωτήματα πρέπει να αποθηκεύονται με κανόνες σε ένα ειδικό αρχείο, το οποίο χρησιμοποιείται σε συνδυασμό με τα δεδομένα. Η εντολή φίλτρου CWM επιτρέπει την αυτόματη επιλογή των

δεδομένων που δημιουργούνται από τους κανόνες. Παρόλο που το N3 πληροί την ορθογωνικότητα, το κλείσιμο και την ιδιότητα ασφαλείας, το N3 ως γλώσσα επερωτήσεων είναι δυσκίνητο. Το N3 υποστηρίζεται από δύο ελεύθερα διαθέσιμα συστήματα, δηλαδή Euler και CWM Σημασιολογία RDF. Η θεματολογία πρέπει να παρέχεται από τους συνήθεις κανόνες. Για τη σύγκρισή μας, συνεργαστήκαμε με την CWM στην έκδοση της 21ης Μαρτίου 2004

5.6 N-Triples

Το N-Triples είναι μια μορφή γραμμής, απλού κειμένου για την κωδικοποίηση ενός γράφου RDF. Έχει σχεδιαστεί να είναι ένα σταθερό υποσύνολο του N3 [N3] [N3-Primer] και επομένως μπορούν να χρησιμοποιηθούν εργαλεία N3 όπως *cwm* [CWM], *n-triples2kif* [N-TRIPLES2KIF] και *Euler* [EULER]. Το *cwm* μπορεί να εκτυπώσει αυτή τη μορφή όταν καλείται ως `"cwm -ntriples"`. Συνιστάται, αλλά δεν απαιτείται, ότι το περιεχόμενο N-Triples αποθηκεύεται σε αρχεία με επίθημα `'.nt'` για να τα διακρίνει κανείς από το N3. Ο τύπος μέσου εκτύπωσης στο Internet / τύπος MIME του N-Triples είναι κείμενο / απλό και η κωδικοποίηση χαρακτήρων είναι US-ASCII 7-bit.

5.7 Versa

Το Versa είναι μια γλώσσα επερωτήσεων σχεδιασμένη για τον συγκεκριμένο σκοπό της εξαγωγής πληροφοριών από ένα γράφημα RDF με πολύ αρθρωτό τρόπο. Ένα ερώτημα Versa διευκολύνει την απομόνωση των πόρων και των συναφών τιμών ιδιοτήτων τους μέσω συγκεκριμένων προτύπων και περιορισμών όπως καθορίζεται από μια έκφραση Versa (κάπως που θυμίζει τις εκφράσεις XPath και τις κανονικές εκφράσεις). Ένα ερώτημα Versa εκτελείται με την υποβολή μιας έκφρασης Versa σε έναν επεξεργαστή ερωτήματος Versa που σχετίζεται με ένα γράφημα RDF από το οποίο ο χρήστης επιθυμεί να εξάγει πληροφορίες. Μια έκφραση Versa είναι μία από τις ακόλουθες:

- Nested (parenthesized) expressions
- Literals (instances of Versa data types)
- Variable References
- Particles (syntactic sugar): `"*" or "`
- Traversal Expressions
- Function Calls

Το Versa υιοθετεί μια ενδιαφέρουσα προσέγγιση στο ότι το κύριο δομικό στοιχείο του lan-guage είναι μια λίστα με τους πόρους του RDF. Τα τρισδιάστατα RDF παίζουν ένα ρόλο στις λεγόμενες traversaloperations, οι οποίες έχουν τη μορφή ListExpr - ListExpr -> BoolExpr. Αυτές οι εκφράσεις επιστρέφουν μια λίστα με όλα τα αντικείμενα που ταιριάζουν τριπλάσια.Το Versa χρησιμοποιεί τον διαχειριστή διανομής για να επιλύσει αυτόν τον περιορισμό. Δημιουργεί μια λίστα με λίστες, η οποία επιτρέπει την επιλογή πολλών ιδιοτήτων μιας δεδομένης λίστας πόρων. Παράλληλα, υποστηρίζει ορισμένους κανόνες, καθώς επιτρέπει την διαδοχική διέλευση των χαρακτήρων. Προσαρμοσμένες ενσωματώσεις, προβολές, πολλαπλά μοντέλα και χειρισμός δεδομένων δεν εφαρμόζονται. Ωστόσο, το Versa πληροί τα κριτήρια ορθογωνικότητας και ασφάλειας. Η γλώσσα Versa υποστηρίζεται από το 4Suite, το οποίο είναι ένα σύνολο XML και RDFtools9.

5.8 Turtle

Η turtle είναι μια επέκταση των N-τριπλών (N-TRIPLES) λαμβάνοντας προσεκτικά τα πιο χρήσιμα και κατάλληλα πράγματα που προστέθηκαν από την N3 διατηρώντας το στο μοντέλο RDF.

Η συνιστώμενη σύνταξη XML για RDF, RDF / XML ([RDF-XML]) έχει ορισμένους περιορισμούς που επιβάλλονται από την XML και τη χρήση χώρων ονομάτων XML που εμποδίζουν την κωδικοποίηση όλων των γραφημάτων RDF (ορισμένες απαγορεύσεις URI είναι απαγορευμένες και το XML 1.0 απαγορεύει την κωδικοποίηση ορισμένων σημείων κωδικοποίησης Unicode). Αυτοί οι περιορισμοί δεν ισχύουν για την turtle.

Η turtle προορίζεται να είναι συμβατή με ένα υποσύνολο της Notation 3 και είναι γενικά χρησιμοποιήσιμο σε συστήματα που υποστηρίζουν το N3.

Όλα τα RDF γραμμένα σε Turtle θα πρέπει να μπορούν να χρησιμοποιηθούν στο τμήμα της γλώσσας επερώτησης της SPARQL για τα triple μοτίβα και για τα triple RDF στη ρήτρα CONSTRUCT. Αυτό επιτρέπει τη χρήση RDF γραμμένο σε turtle για να επιτρέπεται η διαμόρφωση "ερωτημάτων μέσω παραδείγματος", χρησιμοποιώντας τα δεδομένα για να δημιουργήσετε ένα αρχικό ερώτημα το

οποίο στη συνέχεια μπορεί να επεξεργαστεί για να χρησιμοποιήσει μεταβλητές όπου αναζητούνται δεσμεύσεις

6. Rules

6.1 SWRL

Οι γλώσσες και τα συστήματα κανόνων χρησιμοποιούνται ευρέως σε εφαρμογές επιχειρήσεων, αλλά δεν υπάρχει αρκετή διαλειτουργικότητα μεταξύ των τωρινών συστημάτων. Η SWRL είναι μια νέα τεχνολογία Σημαιολογικού Ιστού, η οποία επεκτείνει την OWL-DL και επιτρέπει τη χρήση περιγραφικής λογικής με ορισμένα είδη κανόνων , ενισχύοντας με αυτόν τον τρόπο στην αφηρημένη σύνταξη της OWL και την διαλειτουργικότητα μεταξύ των εφαρμογών. Άρα, μπορούμε να πούμε ότι αυτό που προσφέρει η SWRL είναι η δήλωση κανόνων εντός μιας οντολογίας. Με την SWRL μπορούν να γραφούν κανόνες τύπου Horn, για να εξάγουμε συμπεράσματα από υπάρχουσες βάσεις γνώσεις OWL.

Οι κανόνες έχουν τη μορφή μιας έννοιας μεταξύ ενός προηγούμενου (σώματος) και ενός επακολουθού (κεφαλή). Η επιδιωκόμενη έννοια μπορεί να αναγνωσθεί ως: κάθε φορά που πληρούνται οι προϋποθέσεις που καθορίζονται στην προηγούμενη αναμονή, τότε πρέπει να διατηρηθούν και οι συνθήκες που καθορίζονται στην επακόλουθη.

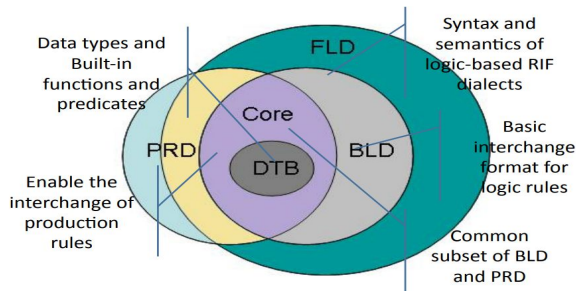
6.2 RIF

Η μορφή ανταλλαγής κανόνων (RIF) είναι μια σύσταση του W3C.Παρόλο που αρχικά οραματίστηκε από πολλούς ως ένα «στρώμα κανόνων» για τον σηματολογικό ιστό, στην πραγματικότητα ο σχεδιασμός του RIF βασίζεται στην παρατήρηση ότι υπάρχουν πολλές "γλώσσες κανόνων" και αυτό που χρειάζεται είναι η ανταλλαγή κανόνων μεταξύ τους.

Το RIF περιλαμβάνει 5 βασικές διαλέκτους, μια διάλεκτο Core που επεκτείνεται σε έναν βασικό λογικό διάλογο (BLD) και έναν διαλεκτικό κανόνα παραγωγής (PRD),την διάλεκτο FLD και την διάλεκτο DTB.

Ο στόχος του RIF είναι να καθορίσει ένα πρότυπο για την ανταλλαγή κανόνων μεταξύ των συστημάτων κανόνων, ιδίως μεταξύ των μηχανών κανόνων Web. Το RIF επικεντρώνεται στην ανταλλαγή και όχι στην οριοθέτηση μιας ενιαίας γλώσσας ενός κανόνα, διότι, σε αντίθεση με άλλα πρότυπα του Σημαιολογικού Ιστού, όπως το RDF, το OWL και το SPARQL, ήταν αμέσως σαφές ότι μία

γλώσσα δεν καλύπτει όλα τα δημοφιλή παραδείγματα χρησιμοποιώντας κανόνες για την εκπροσώπηση της γνώσης και την επιχειρηματική μοντελοποίηση. Ακόμη και η ανταλλαγή κανόνα από μόνη της αναγνωρίστηκε γρήγορα ως αποθαρρυντικό έργο.



Εικόνα 11.

7. Reasoners

7.1 RacerPro

Τα εργαλεία λογικής (Reasoners) χρησιμοποιούν την Περιγραφική Λογική για εξάγουν λογικά συμπεράσματα από τις οντολογίες. Το RacerPro μπορεί να επεξεργαστεί οντολογίες γραμμένες σε OWL Lite και OWL DL. Επίσης συμπεριλαμβάνει σε δοκιμαστικό στάδιο μία έκδοση της γλώσσας SWRL.

Το RacerPro έχει τις παρακάτω δυνατότητες:

- Ελέγχει τη συνέπεια της OWL οντολογίας στην περιγραφή των δεδομένων. Ανακαλύπτει υποκλάσεις ήδη ορισμένων κλάσεων σύμφωνα με τον ορισμό της οντολογίας.
- Ανακαλύπτει συνώνυμα για τους πόρους της οντολογίας, είτε αυτά αφορούν κλάσεις, είτε άτομα.
- Προσφέρει ένα open-source API (application programming interface) για τη δημιουργία επερωτήσεων μέσω ενός συστήματος OWL-QL επεξεργασίας ερωτημάτων για την επεξεργασία πληροφορίας που υπάρχει στα OWL κείμενα.
- Προσφέρει ένα HTTP Interface για την ανάκτηση πόρων που εισάγονται από τον Παγκόσμιο Ιστό (πρωτόκολλο βασισμένο σε XML).
- Προσφέρει την δυνατότητα να δίνει απαντήσεις σε διαδικασίες ανάκτησης πληροφορίας (information retrieval), για παράδειγμα να ανακτά τα επόμενα αποτελέσματα μίας επερώτησης. Επίσης, το RacerPro υποστηρίζει μία έξυπνη μηχανή απάντησης ερωτημάτων, η οποία τα κατηγοριοποιεί με βάση την πολυπλοκότητά τους, εξυπηρετεί τα απλά ενώ στη συνέχεια δίνει στην

εφαρμογή του χρήστη τη δυνατότητα να αποφασίσει αν θα υπολογιστούν και τα επόμενα.

7.2 Pellet

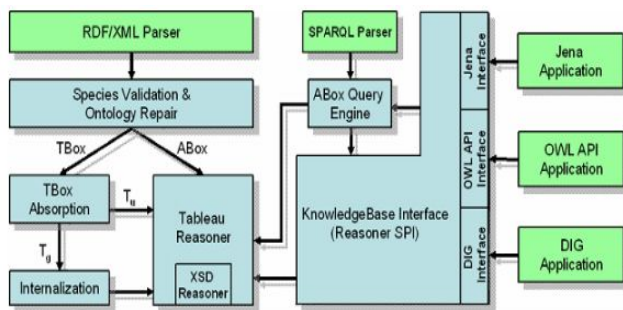
Αν και ο RacerPro reasoner παρέχει πολλές δυνατότητες, έχει δύο βασικά μειονεκτήματα. Α) Είναι εμπορικό προϊόν (παρέχεται όμως ελεύθερα η άδεια για ακαδημαϊκή χρήση, η οποία όμως και πρέπει να ανανεώνεται κάθε εξάμηνο). Β) Η επικοινωνία με το σημασιολογικό API (Jena) που χρησιμοποιήσαμε παρέχεται μόνο μέσω ενός DIG Interface το οποίο θέτει κάποιους περιορισμούς στα ερωτήματα που μπορούν να διαμορφωθούν. Ο Pellet Reasoner από την άλλη παρέχει εκτός από το DIG interface και ένα interface επικοινωνίας το οποίο δεν θέτει τους περιορισμούς του DIG και είναι ο reasoner που συνίσταται από τους δημιουργούς του Jena, ενός από τα πιο διαδεδομένα σημασιολογικά APIs.

Οι δυνατότητες του Pellet περιγράφονται παρακάτω:

- Επιτρέπει την ανάλυση και επιδιόρθωση της οντολογίας. Η OWL-DL επιτρέπει την ύπαρξη κάποιων κανόνων που έχουν σκοπό να εξασφαλίσουν τη λογική συνοχή της οντολογίας. Το Pellet μπορεί να ελέγξει την εγκυρότητα των κανόνων αυτών και έτσι να «διορθώσει» την οντολογία.
- Υποστηρίζει την εξαγωγή συμπερασμάτων με βάση την συνεπαγωγή.
- Υποστηρίζει αλγόριθμους που βελτιώνουν την ταχύτητα των ερωτημάτων. Οι αλγόριθμοι έχουν ως σκοπό να μειώσουν τον αριθμό των ελέγχων εγκυρότητας που απαιτούνται, οδηγώντας έτσι σε ταχύτερα αποτελέσματα.
- Επιτρέπει reasoning σε δεδομένα τύπου datatype. Η γλώσσα XML, περιέχει ένα πλούσιο λεξιλόγιο που συμπεριλαμβάνει αλφαριθμητικά, ακραίους και αριθμούς κινητής υποδιαστολής και τύπους ημερομηνίας και ώρας. Το Pellet μπορεί να θέσει περιορισμούς στα παραπάνω δεδομένα και να ελέγξει την εγκυρότητά τους.
- Επιτρέπει την εξαγωγή συμπερασμάτων από πολλαπλές οντολογίες με τη βοήθεια της γλώσσας αναπαράστασης γνώσης E-Connections
- Υποστηρίζει την εύρεση, όχι μόνο των μη έγκυρων κλάσεων, αλλά και των εσφαλμένων αξιωμάτων ή υπερκλάσεων στην οντολογία που προκαλούν την ασάφεια.

Η αρχιτεκτονική του Pellet φαίνεται στο παρακάτω. Ο πυρήνας του είναι το ταμπλό εξαγωγής συμπερασμάτων που ελέγχει τη σαφήνεια του γνωσιολογικού μοντέλου,

δηλαδή το ζεύγος A-Box και T-Box. Ο reasoner έχει τη δυνατότητα να ελέγχει τη σαφήνεια των δεδομένων. Οι οντολογίες OWL φτάνουν στον reasoner μετά από μια φάση επικύρωσης του τύπου τους και επιδιόρθωσης αυτών. Όσο ο reasoner φορτώνει την οντολογία τα αξιώματα για τις κλάσεις τοποθετούνται στο T-Box, ενώ η πληροφορία για τα instances αποθηκεύεται στο A-Box. Τα αξιώματα του T-Box οδηγούνται στον DL προ-επεξεργαστή πριν οδηγηθούν στον βασικό επεξεργαστή. Ο θεμελιώδης επεξεργαστής έχει μόνο την δυνατότητα να ελέγχει την εγκυρότητα του ABox με βάση ένα T-Box. Όλες οι υπόλοιπες διεργασίες εξαγωγής συμπερασμάτων μπορούν να μετασχηματιστούν με την βοήθεια του SPI (System Programming Interface) σε ελέγχους εγκυρότητας. Υπάρχουν τρεις τύποι διασύνδεσης με άλλες εφαρμογές: α) Υπάρχει το DIG Interface, β) το Jena Interface και το γ) OWL-API Applications interface.

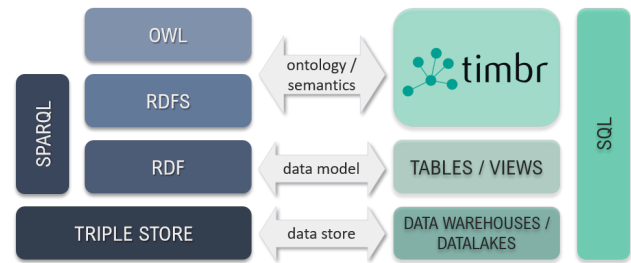


Εικόνα 12. Αρχιτεκτονική Pellet

8. Timbr

Το timbr είναι μια πλατφόρμα γνώσεων που χρησιμοποιεί SQL - τη πιο γνωστή γλώσσα της βάσης δεδομένων, για να εξαλείψει τα τεχνολογικά εμπόδια στην είσοδο για τη χρήση Knowledge graph και να εφαρμόσει τις αρχές του Σημασιολογικού Ιστού για να παρέχει λειτουργικότητα Knowledge graph. που έχουν σημασία σε ολόκληρο τον οργανισμό και καθιστούν τα δεδομένα κατανοητά τόσο για τους επιχειρηματικούς χρήστες όσο και για τις μηχανές.

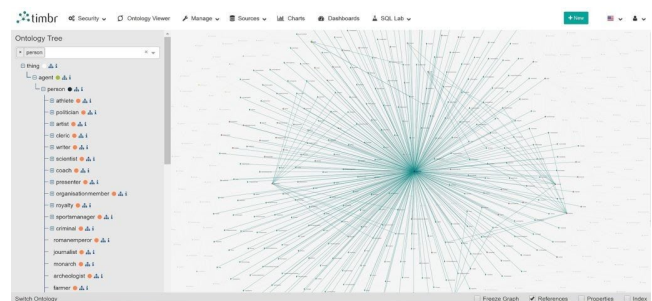
Εφαρμογή Σημασιολογικού σε SQL



Εικόνα 13. Αρχιτεκτονική timbr

Το timbr χρησιμοποιεί τυπικό SQL για την εφαρμογή των αρχών του Σημασιολογικού Ιστού, επιτρέποντας την οντολογική μοντελοποίηση των δεδομένων ως συνδεδεμένων, εμπλουτισμένων με το περιβάλλον εννοιών με δυνατότητες συμπερασμάτων και διαγράμματα γραφημάτων.

ΔΥΝΑΤΟΤΗΤΕΣ ΜΟΝΤΕΛΙΣΜΟΥ ΙΣΧΥΡΗΣ ΟΝΟΛΟΓΙΑΣ



Εικόνα 14. εφαρμογής timbr

Ο μοντελοποιητής οντολογιών επιτρέπει **στους** χρήστες του timbr το να δημιουργήσουν εννοιολογικά σχήματα, τα οποία προέρχονται από συνδεδεμένες εμπλουτισμένες με το περιβάλλον έννοιες που αντιπροσωπεύουν αφηρημένους ή κοινούς όρους επιχείρησης, όπως προϊόν, πελάτη ή έσοδα, για να προσφέρουν ενοποιημένη και ενοποιημένη προβολή δεδομένων σε ολόκληρο τον οργανισμό.

Εύκολη χαρτογράφηση εννοιών στην οντ

Το timbr επιτρέπει τη μοντελοποίηση δεδομένων ως συνδεδεμένων, εμπλουτισμένων με το περιβάλλον ιδεών με δυνατότητες αντιστοίχισης και διαγράμματος γραφημάτων, ενώ παράλληλα γίνεται ερώτημα σε τυπική SQL, μετασχηματίζοντας δεδομένα σε έξυπνα συνδεδεμένα δεδομένα, απρόσκοπτα προσβάσιμα σε λύσεις

επιχειρηματικής ευφυΐας, μηχανική μάθηση και επεξεργασία φυσικής γλώσσας.

Η μοντελοποίηση των εννοιών επεκτείνει τους ορισμούς του σχήματος και του πίνακα με τις αφαιρέσεις που διευκολύνουν την τυποποίηση των δεδομένων οντολογία σε φυσικούς πίνακες σε Data lakes και Data warehouses. Όλες οι έννοιες στην οντολογία είναι προσβάσιμες σε οποιαδήποτε πλατφόρμα ερωτήσεων και εμφανίζονται ως πίνακες στα εργαλεία Business Intelligence.

9. Τασεις Σημασιολογικου Ιστου

9.1 Machine Learning

Η μάθηση μέσω μηχανής (ML) είναι ένας κλάδος της τεχνολογίας αιχμής που παρέχει στις μηχανές τη δυνατότητα αυτόματης εκμάθησης, βελτίωσης και προσαρμογής από την εμπειρία, χωρίς ρητό προγραμματισμό για την εκτέλεση συγκεκριμένων εργασιών.

Τα πρότυπα του Σημασιολογικού Ιστού χρησιμοποιούνται για την περιγραφή μεταδεδομένων, αλλά έχουν επίσης μεγάλο δυναμικό ως γενικό format για την επικοινωνία και την ενσωμάτωση δεδομένων. Σε ένα ευρύ φάσμα πιθανών εφαρμογών, η μηχανική μάθηση θα παίζει έναν όλο και πιο σημαντικό ρόλο: Οι λύσεις μηχανικής μάθησης έχουν αναπτυχθεί για να υποστηρίξουν τη διαχείριση οντολογιών, και να ενσωματώσουν σημασιολογικές πληροφορίες στον web mining. Η μηχανική μάθηση θα χρησιμοποιείται όλο και περισσότερο για την ανάλυση των καταναμημένων πηγών δεδομένων που περιγράφονται σε format του Σημασιολογικού Ιστού.

Η μηχανική μάθηση βοηθά τον σημασιολογικό ιστό μέσω της οντολογικής αξιολόγησης, της κατασκευής, της διαχείρισης, της εξέλιξης και της βελτίωσης, και επιπλέον της συγχώνευσης της χαρτογράφησης. Συγκρίνει πολλά μέτρα ομοιότητας και αλγόριθμους που χρησιμοποιούνται για τη χαρτογράφηση ή τη συγχώνευση δύο οντολογιών με αλγορίθμους μηχανικής μάθησης.

Ο Σημασιολογικός ιστός ολοκληρώνει διάφορα καθήκοντα για την εκμάθηση μηχανών, όπως η χρήση διαφόρων εργαλείων για την εξήγηση και την ανταλλαγή δεδομένων για μεταγενέστερη χρήση με μεθόδους μηχανικής μάθησης σε ένα κανονικό μοντέλο, χρησιμοποιώντας οντολογικές δομές για την ενίσχυση της εργασίας εκμάθησης μηχανών μάθησης. Για παράδειγμα, ένα άτομο θα έχει την ικανότητα να δίνει στον υπολογιστή μια εντολή, "να αναζητά όλες τις τρομοκρατικές επιθέσεις οι οποίες έχουν στοχευθεί από τις

ομάδες των δραστών της Αλ Κάιντα στην Ινδία. "Μη αξιόπιστα, ένας υπολογιστής μπορεί να έχει την ικανότητα να κάνει έρευνα για τις ομάδες των δραστών της Αλ Κάιντα και να επιστρέψει τις συνέπειες για τον πελάτη. Αυτή η αναζήτηση θα γίνεται με την αντιστοίχιση λέξεων χωρίς την κατανόηση της μηχανής. Με τη βοήθεια των SWT, ο υπολογιστής θα έχει την ικανότητα πρόσβασης στις οντολογίες, ώστε να αναγνωρίζει (ή φαίνεται να καταλαβαίνει) τη σχέση μεταξύ των τρομοκρατικών επιθέσεων, της Αλ Κάιντα και της Ινδίας.

9.2 Deep Learning

Σε αυτήν την κατηγορία, οι τεχνολογίες του Σημασιολογικού Ιστού χρησιμοποιούνται για τη βελτίωση των μεθόδων βαθιάς εκμάθησης, καθιστώντας τους αλγόριθμους ταξινόμησης πιο ενημερωμένους για τη σημασιολογία των δεδομένων από τα οποία μαθαίνουν. Μια χρήση του Deep learning είναι σε ένα πείραμα που έγινε σε μια οντολογία για τα καρδιαγγειακά νοσήματα (CVDO). Άλλη μια συνεργασία Σημασιολογικού ιστού και Deep learning ήταν για τη μείωση του θορύβου για το RDF.

Η βαθιά εκμάθηση είναι ένας αλγόριθμος για πολυεπίπεδες (πολύπλοκες) εκφράσεις που μαθαίνουν αυτόματα τη δυνητική (σιωπηρή) κατανομή των δεδομένων ώστε να έχουν μοντέλο. Με άλλα λόγια, ο αλγόριθμος βαθιάς μάθησης αποκτά αυτόματα χαρακτηριστικά χαμηλού ή υψηλού επιπέδου που απαιτούνται για ταξινόμηση. Η βαθιά εκμάθηση σύντηξης μπορεί να λάβει καλύτερη εικόνα των χαρακτηριστικών των δεδομένων. Ταυτόχρονα, επειδή το μοντέλο έχει πολλά επίπεδα, παραμέτρους και επαρκή χωρητικότητα, έχει τη δυνατότητα να αντιπροσωπεύει δεδομένα μεγάλης κλίμακας. Ως εκ τούτου, το πρόβλημα των σημασιολογικών συναισθημάτων με τέτοια χαρακτηριστικά δεν είναι προφανές. Επιπλέον, από την άποψη των χαρακτηριστικών αναγνώρισης προτύπων και ταξινομητές, το πλαίσιο βάσης της βαθιάς εκμάθησης συνδυάζει τα χαρακτηριστικά και τους ταξινομητές σε ένα πλαίσιο και χρησιμοποιεί δεδομένα για να μάθει χαρακτηριστικά, γεγονός που μειώνει το τεράστιο φόρτο εργασίας των χειροκίνητων χαρακτηριστικών σχεδιασμού κατά τη χρήση. Η ανάλυση του συναισθήματος είναι πολύ σημαντική στο μελλοντικό μας δίκτυο, όπως το Twitter. Επομένως, όχι μόνο η επίδραση της ανάλυσης αισθήσεων μπορεί να είναι καλύτερη και υπάρχουν πολλές ευκολίες στη χρήση της.

Ο πυρήνας της σημασιολογικής ανάλυσης συναισθημάτων είναι να προσθέσετε τα μεταδεδομένα που μπορούν να

κατανοηθούν από τον υπολογιστή στο έγγραφο δικτύου για βαθιά εκμάθηση, ώστε ολόκληρο το Διαδίκτυο να γίνει παγκόσμιο μέσο ανταλλαγής πληροφοριών.

Ο απώτερος στόχος της μελέτης του συναισθηματικού σημασιολογικού αναλυτικού υποδείγματος είναι να επιτευχθεί μια ολοκληρωμένη προσομοίωση της ανθρώπινης νοημοσύνης. Ωστόσο, μελέτες στην κοινότητα ψυχολογίας έχουν διαπιστώσει ότι, αν οι άνθρωποι χάνουν ορισμένα συναισθηματικά συστατικά, η λήψη αποφάσεων είναι επίσης δύσκολη. Επομένως, η ικανότητα της συναισθηματικής σημασιολογικής ανάλυσης είναι απαραίτητη για την κανονική ανθρώπινη συμπεριφορά. Επομένως, η προσομοίωση της ανθρώπινης νοημοσύνης πρέπει να περιλαμβάνει την προσομοίωση συναισθηματικών δραστηριοτήτων σημασιολογικής ανάλυσης. Σκοπός της τεχνητής συναισθηματικής σημασιολογικής ανάλυσης είναι η διερεύνηση μερικών από τους ρόλους που διαδραματίζει η συναισθηματική σημασιολογική ανάλυση στη βιολογία και η ανάπτυξη τεχνικών και μεθόδων για την ενίσχυση της αυτονομίας, της προσαρμοστικότητας και των δυνατοτήτων κοινωνικής αλληλεπίδρασης των υπολογιστών ή των ρομπότ. Η συναισθηματική σημασιολογική ανάλυση και υπολογισμός αποτελούν τη βάση για τη δημιουργία ενός αρμονικού περιβάλλοντος ανθρώπου-μηχανής. Τα ερευνητικά περιεχόμενα περιλαμβάνουν κυρίως: την απόκτηση σημάτων, την ανάλυση και αναγνώριση πληροφοριών, την κατανόηση και την έκφραση σημασιολογικής ανάλυσης συναισθημάτων. Προς το παρόν, οι κύριοι τομείς εφαρμογής της συναισθηματικής σημασιολογικής ανάλυσης είναι οι εξής: επεξεργασία έκφρασης αναγνώρισης και σύνθεσης χειρονομίας, κατανόηση γλωσσών, συναισθηματική ανάλυση και υπολογισμός συναισθηματικής σημασίας, ανάλυση συναισθημάτων και εξατομικευμένη υπηρεσία.

9.3 Neural Networks

Τα Νευρωνικά Δίκτυα μπορούν να οριστούν ως μεγάλα σύνολα διασυνδεδεμένων απλών μονάδων που εκτελούνται παράλληλα για να εκτελεσουν μια κοινή παγκόσμια διεργασία. Αυτές οι μονάδες συνήθως υποβάλλονται σε μια διαδικασία μάθησης η οποία ενημερώνει αυτόματα τους παραμέτρους δικτύου για ένα πιθανώς εξελισσόμενο περιβάλλον εισόδου. Είναι συστήματα επεξεργασίας πληροφοριών εμπνευσμένα από την ικανότητα του ανθρώπινου εγκεφάλου να μάθει από τις παρατηρήσεις και να γενικεύσει με την αφαίρεση, σύμφωνα με τα ακόλουθα χαρακτηριστικά:

- Η γνώση αποκτάται από το δίκτυο μέσω μιας μαθησιακής διαδικασίας.
 - Χρησιμοποιούνται ισχυρά σημεία σύνδεσης μεταξύ νευρώνων, γνωστά ως συναπτικά βάρη που χρησιμοποιούνται για την αποθήκευση της γνώσης.
- Τα Νευρωνικά Δίκτυα χρησιμοποιούνται για να αναγνωρίζουν αυτόματα τη σύνδεση μεταξύ ιστοσελίδων με παρόμοιο περιεχόμενο και να βελτιώνουν από κοινού τη συλλογή σημασιολογικών πληροφοριών με θησαυρούς συνωνύμων ή βάσει εγγράφων κειμένου. Ο SEMantic INTEgrator (SEMINT) είναι ένα εργαλείο που βασίζεται σε νευρωνικά δίκτυα για να βοηθήσει στην αναγνώριση των χαρακτηριστικών αντιστοιχίας σε ετερογενείς βάσεις δεδομένων. Υποστηρίζει πρόσβαση σε μια ποικιλία από συστήματα βάσεων δεδομένων και χρησιμοποιεί πληροφορίες τόσο σε επίπεδο schema όσο και σε επίπεδο instance ώστε να παράγουν κανόνες για την αυτόματη αντιστοίχιση των αντίστοιχων χαρακτηριστικών. Κατά την αντιστοίχιση, διαφορετικές σημασιολογικές πτυχές, όπως τα ονόματα ιδεών, η έννοια ιδιότητες και σχέσεις ιδεών, συμβάλλουν σε διαφορετικούς βαθμούς στην αντιστοίχιση των αποτελεσμάτων. Επομένως, πρέπει να αποδοθεί ένας φορέας βαρών σε αυτές τις πτυχές, καθώς δεν είναι μια ασήμαντη εργασία και η τρέχουσα ερευνητική εργασία εξαρτάται από την ανθρώπινη ευρετική. Ένα μοντέλο νευρωνικού δικτύου μαθαίνει και προσαρμόζει αυτά τα βάρη, με σκοπό την αποφυγή κάποιων μειονεκτημάτων τόσο στις προσεγγίσεις αντιστοίχισης οντολογιών που βασίζονται σε κανόνες όσο και στη μάθηση, οι οποίες αγνοούν τις πληροφορίες που μπορεί να παρέχουν τα δεδομένα των στιγμιότυπων.

9.4 Natural Language Processing

Χάρη στην ικανότητά του NLP να αναλύει την εμφάνιση των όρων σε μια πρόταση και να αποθηκεύει και να δημιουργεί σχέσεις σε μια βάση δεδομένων γραφημάτων, συμβάλλει στη βελτίωση των βασικών όρων στα αρχεία ιατρικής περίθαλψης και φαρμάκων. Το NLP θα συνεχίσει να είναι μια τάση με τις επιχειρήσεις να αναλύουν τη συμπεριφορά και την εμπειρία των καταναλωτών, καθώς και το συναίσθημα προς τα προϊόντα και τις υπηρεσίες.

Το NLP (Natural language processing) και το SW (Semantic Web) είναι δύο διαφορετικές τεχνολογίες αλλά έχουν συμπληρωματικούς ρόλους στη διαχείριση δεδομένων.

Ο συνδυασμός αυτών των δύο τεχνολογιών καθιστά δυνατή την ομοιόμορφη συγχώνευση δομημένων και αδόμητων δεδομένων, για παράδειγμα, ένας machine

reader είναι ένα εργαλείο που μετατρέπει το κείμενο φυσικής γλώσσας σε σωστή δομημένη γνώση και το τελευταίο, σύμφωνα με την κοινή σημασιολογία, μπορεί να ερμηνευθεί από μηχανές. Οι τεχνολογίες του Σημασιολογικού Ιστού στοχεύουν στη μετατροπή των μη δομημένων δεδομένων σε σημαντικές αναπαραστάσεις, οι οποίες ωφελούν εξαιρετικά τη χρήση τεχνολογιών NLP, επιτρέποντας έτσι εφαρμογές όπως η σύνδεση κειμένου με Linked Open Data, η σύνδεση κειμένων μεταξύ τους, η σημασιολογική αναζήτηση, η οπτικοποίηση πληροφοριών και η μοντελοποίηση της συμπεριφοράς του χρήστη στο web.

Χρησιμοποιήσαμε αυτόν τον συνδυασμό σε διάφορες εφαρμογές, όπου αντιπαραβάλλουμε με ένα τεράστιο όγκο μη δομημένων ή ανοργάνωτων πληροφοριών. Εξάγουμε δομημένα δεδομένα από ένα έγγραφο κειμένου χρησιμοποιώντας NLP και στη συνέχεια συνδέουμε αυτά τα εξαγόμενα δεδομένα μέσω σημασιολογικών τεχνολογιών. Συνδυάζοντας και τις δύο τεχνολογίες, μπορούμε να αντλήσουμε σχετικές πληροφορίες και να τις συνδέσουμε με αυστηρά επιστημονικά δεδομένα (π.χ. με οργανικές ενώσεις, μονοπάτια κλπ.) που έχουν ήδη δηλωθεί και τοποθετούνται σε μια οργανωμένη βάση δεδομένων. Ο συνδυασμός NLP και SW παρέχει ένα περιβάλλον με το οποίο μπορούμε να βάλουμε το πολύπλοκο ερώτημα και να φέρουμε την πραγματική λογική απάντηση. Μπορούμε να σχεδιάσουμε μια κατάλληλη δομή για τα δεδομένα κειμένου (textual data) με τη βοήθεια του RDF και τις τεχνολογίες επεξεργασίας φυσικής γλώσσας.

Ως εκ τούτου, ο σημασιολογικός ιστός μπορεί να ενισχύσει το τεχνικό επίπεδο των τεχνολογιών φυσικής γλώσσας και η NLP μπορεί επίσης να βοηθήσει στην παροχή και χρήση ενός καλύτερου σημασιολογικού ιστού. Οι πραγματικές προκλήσεις για τον ερευνητή είναι να προσελκύσει τις τεχνικές NLP πλησιέστερα στην οντολογική μηχανική και να επεκτείνει τις υπάρχουσες τεχνικές NLP στις εφαρμογές που βασίζονται σε οντολογίες.

9.5 Text Mining and Text Analytics

Οι επιχειρήσεις θα αναζητούν όλο και περισσότερο εργαλεία για να βλέπουν και να ανακαλύπτουν όλες τις πληροφορίες σε μεγάλα κείμενα. Αυτά συνήθως περιέχουν εξαιρετικά πολύτιμες γνώσεις που είναι συχνά δύσκολο να αποκτήσουν πρόσβαση ή να οργανωθούν λόγω της αδόμητης και ετερογενούς φύσης τους.

Η εξόρυξη κειμένου κάνει ευκολότερη την εξαγωγή πληροφοριών από τεράστιους όγκους δεδομένων και τις

διαρθρώνει ως σημαντικά γεγονότα, βασικούς όρους ή πρόσωπα. Μόλις εξάγονται δομημένα δεδομένα από ελεύθερα κείμενα, επιτρέπει στους οργανισμούς να επιτύχουν αποτελεσματική ευρετηρίαση και βελτιωμένη αναζήτηση ή να προσφέρουν εξατομικευμένες συστάσεις στους πελάτες.

Αποσκοπεί στην εξαγωγή χρήσιμης πληροφορίας από πηγές δεδομένων μέσω της αναγνώρισης και της διερεύνησης ενδιαφερόντων προτύπων. Οι πηγές δεδομένων είναι συλλογές κειμένων. Τα ενδιαφέροντα πρότυπα αναζητούνται σε μη δομημένα δεδομένα κειμένων, δηλαδή στα έγγραφα της συλλογής και όχι σε δομημένα δεδομένα Βάσεων δεδομένων.

Η ανάλυση κειμένων περιλαμβάνει την ανάκτηση πληροφοριών, τη λεξική ανάλυση για τη μελέτη των κατανομών συχνότητας λέξεων, την αναγνώριση προτύπων, την επισήμανση / σχολιασμό, την εξαγωγή πληροφοριών, τις τεχνικές εξόρυξης δεδομένων, συμπεριλαμβανομένης της ανάλυσης σύνδεσης και σύνδεσης, της απεικόνισης και των προγνωστικών αναλύσεων. Ο πρωταρχικός στόχος είναι, ουσιαστικά, η μετατροπή του κειμένου σε δεδομένα για ανάλυση, με εφαρμογή της επεξεργασίας φυσικής γλώσσας (NLP) και αναλυτικών μεθόδων. Μια τυπική εφαρμογή είναι η σάρωση ενός συνόλου εγγράφων γραμμένων σε μια φυσική γλώσσα και είτε η μοντελοποίηση του εγγράφου για σκοπούς πρόβλεψης ταξινόμησης είτε η συμπλήρωση μιας βάσης δεδομένων ή ενός ευρετηρίου αναζήτησης με τις πληροφορίες που εξάγονται.

Σήμερα, καθώς είναι διαθέσιμο ένα τεράστιο ποσό πληροφοριών στον ιστό, πάνω από ένα δισεκατομμύριο σελίδες είναι ευρετηριασμένες μέσω μηχανών αναζήτησης και ως εκ τούτου, η αναζήτηση των επιθυμητών πληροφοριών είναι ένα εξαιρετικά δύσκολο έργο. Αυτός ο πλούτος των πόρων έχει ενθαρρύνει την απαίτηση για την ανάπτυξη αυτόματων τεχνικών εξόρυξης στον Παγκόσμιο Ιστό, δημιουργώντας έτσι τον όρο "Εξόρυξη Διαδικτύου". Οι SWT σκοπεύουν να αντιμετωπίσουν το ζήτημα της εξαγωγής πληροφοριών από το διαδίκτυο, προσφέροντας στη μηχανή κατανοητή σημασιολογία καλύτερη υποστήριξη μηχανών για τον πελάτη. Μια ολοκληρωμένη προσέγγιση των SWT και της διαδικτυακής εξόρυξης παρέχει μια καλύτερη μέθοδο για την εξόρυξη συναφών και σημασιολογικών πληροφοριών από τον ιστό, δημιουργώντας έτσι τον όρο "Εξόρυξη Σημασιολογικού Ιστού". Ο συνδυασμός των SWT και της εξόρυξης ιστού ενισχύει τις συνέπειες της εξόρυξης ιστού

χρησιμοποιώντας τη σημασιολογική δομή στον ιστό. Η Οντολογία έχει την ικανότητα να απομακρύνει δεδομένα από μια μεγάλη συλλογή δεδομένων. Χρησιμοποιείται κατά την προεπεξεργασία για τη βελτίωση των αποτελεσμάτων της ομαδοποίησης. Τα SWT καθιστούν την εξόρυξη Web απλούστερη και μπορεί επίσης να ενισχύσει την αποτελεσματικότητα της διαδικασίας. Με τη χρήση τεχνολογιών σημασιολογικού ιστού με Linked Open Data, υποστηρίζουμε την ανακάλυψη γνώσης. Η προσπάθεια πίσω από τον Σημασιολογικό Ιστό πρέπει να είναι μηχανικά ερμηνεύσιμη, να υπάρχει σημασιολογικός σχολιασμός (παρατήρηση) στα έγγραφα του διαδικτύου, προκειμένου να αποκτηθεί η γνώση και να μην υπάρχουν ανοργάνωτα δεδομένα. Έχουν γίνει διάφορα είδη ερευνητικών εργασιών. Ωστόσο, η ολοκλήρωση του σημασιολογικού ιστού με συνδεδεμένα ανοιχτά δεδομένα στον τομέα της εξόρυξης δεν έχει ακόμη ολοκληρωθεί. Στην εξόρυξη δεδομένων, η επιλογή των δεδομένων ανάλογα με τις ανάγκες των χρηστών θεωρείται δύσκολη υπόθεση επειδή δεν υπάρχει σχέση μεταξύ των συνόλων δεδομένων. Τα δεδομένα που υπάρχουν στο διαδίκτυο αλληλοσυνδέονται μεταξύ τους μέσω οντολογίας που παρέχει επιτυχή αναζήτηση. Υπάρχουν δύο βασικές δυσκολίες στη σημασιολογική εξόρυξη ιστού: πρώτον, η φύση των δεδομένων είναι αδόμητη και δεύτερον, η εξαγωγή δεδομένων από τον ιστό. Η δυσκολία στην παρούσα μηχανή αναζήτησης είναι η ασαφής συμπεριφορά των λέξεων. Για το λόγο αυτό, μια μηχανή αναζήτησης είναι ενεργοποιημένη για να διαφοροποιήσει την έννοια και επομένως επιστρέφει όλους τους ιστότοπους που περιέχουν μια δεδομένη παράμετρο αναζήτησης. Μπορούμε να επιλύσουμε αυτό το πρόβλημα χρησιμοποιώντας σημασιολογικές τεχνολογίες που έχουν την ικανότητα να διαφοροποιούν τις έννοιες των λέξεων. Ο Σημασιολογικός ιστός επιτρέπει εξόρυξη γνώσης αντί για εξόρυξη δεδομένων επειδή ενσωματώνει την γνώση πεδίου και την ιδιότητα σημασιολογικής σχολιασμού που παρέχει την έννοια των δεδομένων. Ως εκ τούτου, τα SWT βοηθούν τους ανθρώπους να διερευνήσουν τα γεγονότα από τις λεπτομέρειες για να απαντήσουν συγκεκριμένες ερωτήσεις.

9.6 Internet of things

Η πραγματικότητα του Ίντερνετ των πραγμάτων (IoT), με τον αυξανόμενο αριθμό συσκευών και την ποικιλομορφία τους, προκαλεί σύγχρονες προσεγγίσεις και τεχνολογίες για μια πιο έξυπνη ενσωμάτωση των δεδομένων, εφαρμογών και υπηρεσιών τους. Ενώ ο Ιστός θεωρείται ως μια βολική πλατφόρμα για την ενσωμάτωση των πραγμάτων, ο Σημασιολογικός Ιστός μπορεί να βελτιώσει περαιτέρω την

ικανότητά του να κατανοεί τα δεδομένα των πραγμάτων και να διευκολύνει τη διαλειτουργικότητά τους.

Το IoT έχει απαιτήσει μια σημασιολογική ραχοκοκαλιά για να ευδοκιμήσει. Περισσότερες από 25 δισεκατομμύρια συσκευές εκτιμήθηκαν ότι συνδέονται με το Διαδίκτυο το 2015 και 50 δισεκατομμύρια μέχρι το 2020. Η διαλειτουργικότητα μεταξύ του IoT είναι μία από τις πιο βασικές απαιτήσεις για την υποστήριξη της ανακάλυψης αντικειμένων, την αντιμετώπιση και την παρακολούθηση εκτός από την αποθήκευση πληροφοριών, ασφάλεια, αναπαράσταση και ανταλλαγή. Το IoT θα περιλαμβάνει ουσιαστικά διαφορετικά σύνολα συσκευών και διαφορετικές στρατηγικές επικοινωνίας μεταξύ των συσκευών. Αυτός ο τύπος ετερογενούς συστήματος θα πρέπει να εξελιχθεί σε ένα πρόσθετο οργανωμένο σύνολο λύσεων, όπου τα περιεχόμενα του IoT έχουν την ικανότητα να επικοινωνούν με διαφορετικές οντότητες και περιλαμβάνονται έντονα στην υποδομή και τις υπηρεσίες Διαδικτύου, αντί του ιδιαίτερου τρόπου με τον οποίο είναι που σχετίζονται με το IoT. Η δυναμική, η ποικιλομορφία, τα δίκτυα, τα δεδομένα και η ετερογένεια των συσκευών αποτελούν βασικά ζητήματα των τεχνολογιών του Διαδικτύου. Οι τεχνολογίες του σημασιολογικού ιστού αποδείχθηκαν καρποφόρες σε διάφορους τομείς όσον αφορά την αντιμετώπιση του προβλήματος της ετερογένειας όσον αφορά i) τη διασύνδεση τέτοιων δεδομένων (ii) τη συναγωγή νέων γνώσεων για την ανάπτυξη έξυπνων εφαρμογών, (iii) την παροχή διαλειτουργικότητας στη διαχείριση δεδομένων.

Ωστόσο, μία από τις προκλήσεις με τις υπάρχουσες εφαρμογές IoT είναι ότι οι συσκευές δεν είναι (ή ελάχιστα) συμβατές μεταξύ τους, επειδή τα δεδομένα τους εξαρτώνται από ιδιόκτητες μορφές και δεν χρησιμοποιούν κοινό λεξιλόγιο για να εξηγήσουν συμβατά δεδομένα IoT. Τα SWT έχουν ήδη αποδείξει τα οφέλη τους σε τομείς εκτός από το Διαδίκτυο. Οι τεχνολογίες σημασιολογικού ιστού χρησιμοποιούνται σε διαδικτυακό τόπο για να μειωθεί η πρόκληση όσον αφορά τη διαλειτουργικότητα των δεδομένων που παράγονται από συσκευές που χρησιμοποιούνται ήδη στην πραγματική ζωή. Ο Σημασιολογικός Ιστός ασχολείται με το IoT και το WoT. Με την υλοποίηση του πραγματικού δυναμικού των σημασιολογικών τεχνολογιών, έχουν προταθεί διάφορα πλαίσια IoT που αντιμετωπίζουν τα ζητήματα διαλειτουργικότητας των δεδομένων χρησιμοποιώντας SWT και πρότυπα. Για την ανάπτυξη του IoT, πολλές σημασιολογικές τεχνολογίες όπως οντολογίες, , RDF, Linked data, κ.λπ. μπορεί να χρησιμοποιηθούν ως κύρια

λύση. Οι σημασιολογικές τεχνολογίες έχουν αποδείξει τη χρησιμότητά τους σε διάφορους τομείς και μερικά από τα διάφορα προβλήματα που αντιμετωπίζουν τα SWT είναι (i) η ελάττωση της ετερογένειας με τη χρήση σημασιολογικής διαλειτουργικότητας, (ii) η απλή ενσωμάτωση της εφαρμογής των δεδομένων, (iii) νέες γνώσεις για την ανάπτυξη εφαρμογών και την παροχή έξυπνων λύσεων και iv) παροχή διαλειτουργικότητας μεταξύ διαφορετικών διαδικασιών δεδομένων με εκπροσώπηση, διαχείριση και αποθήκευση δεδομένων.

9.7 Big Data

Το «Big Data» είναι ένας όρος που έχει χρησιμοποιηθεί για να εξηγήσει ένα μεγάλο όγκο δεδομένων που έχουν δημιουργηθεί τα τελευταία 20 χρόνια. Το Facebook, το Twitter και άλλα κοινωνικά μέσα περα από το να δημιουργούν πολλά δεδομένα επιπλέον καθιστούν εφικτή την πρόσβαση των προγραμματιστών. Υπάρχουν πολλά προβλήματα στην οργάνωση των δεδομένων που προέρχονται από διάφορες πηγές και σε διάφορες μορφές. Μερικές από τις προκλητικές μελλοντικές κατευθυντήριες γραμμές έρευνας για τα μεγάλα δεδομένα είναι οι κλιμακωτές αρχιτεκτονικές για μαζική παράλληλη επεξεργασία δεδομένων και επεξεργασία δεδομένων ροής, κλιμακούμενα συστήματα αποθήκευσης, θέματα ασφάλειας και ιδιωτικότητας σε μεγάλα δεδομένα, ανάλυση μεγάλων κλιμάκων δεδομένων για κοινωνικά δίκτυα, προστασία προσωπικών δεδομένων, Big Data Analytics για την Επιχειρηματική Ευφυΐα και την έξυπνη υγειονομική περίθαλψη, σημασιολογία. Ανάμεσά τους, το «Big Data Analytics» έχει γίνει το κεντρικό σημείο της έρευνας. Ένα από τα πραγματικά ζητήματα για την ανάπτυξη του Big Data Analytics είναι η ποικιλομορφία των πληροφοριών που βασίζονται στο διαδίκτυο, επειδή τα δεδομένα που συλλέγονται είναι μη δομημένα στη φύση τους. Η διεθνής κοινότητα W3C ενθάρρυνε μια κοινή μορφή δεδομένων για να καταστήσει τα δεδομένα στον ιστό πιο αξιόπιστα και πιο εύκολα να διαβαστούν. Για τους οργανισμούς με βάση το Big Data Analytics, οι SWT τους βοηθούν στις επιχειρήσεις, επιτρέποντας στη συνέχεια να κάνουν ακόμα καλύτερη κρίση σε πραγματικό χρόνο. Για τις οργανώσεις των επιχειρήσεων προς τους καταναλωτές (με επικεφαλής την Google), ο σημασιολογικός ιστός τους επιτρέπει να προσφέρουν στους καταναλωτές άμεσες απαντήσεις. Η προσέγγιση για την αντιμετώπιση του προβλήματος που συνδέεται με μεγάλα δεδομένα είναι η αποθήκευση δεδομένων μια δομημένη μορφή και

χαρακτηρίζει τα σύνολα δεδομένων ως γραφήματα. Αυτό επιτρέπει στους πράκτορες λογισμικού να υποβάλλουν ερωτήματα στις βάσεις δεδομένων. Η επεξεργασία των συνδεδεμένων δεδομένων καθιστά εφικτή την εύρεση πληροφοριών. Ένα Knowledge graph είναι ένα καλό παράδειγμα μεγάλων δεδομένων για τον σημασιολογικό ιστό. Το Knowledge graph προστέθηκε στο Google το 2012-2013 και παρέχει έναν ενημερωμένο αλγόριθμο που λέγεται 'Hummingbird'. Η σημασιολογική γνώση χρησιμοποιείται από το γράφημα γνώσεων Google, το οποίο αυξάνει τις συμβατικές σελίδες αποτελεσμάτων μηχανών αναζήτησης. Ένα άλλο καλό παράδειγμα είναι τα κοινωνικά μέσα ενημέρωσης. Το πρωτόκολλο Facebook Open Graph δίνει τη δυνατότητα σε οποιαδήποτε ιστοσελίδα να γίνει ένα πλούσιο αντικείμενο σε ένα κοινωνικό γράφημα. Η Αναζήτηση Graph Facebook είναι σημασιολογική μηχανή αναζήτησης που εισήχθη το 2013 για να παρέχει απαντήσεις σε ερωτήματα φυσικής γλώσσας των χρηστών αντί για μια λίστα συνδέσμων.

10. REFERENCES

- [1] <https://www.slideshare.net/onlyjiny/the-semantic-web-3-uri>
- [2] <https://computer.howstuffworks.com/semantic-web3.htm>
- [3] <https://quintupledev.wordpress.com/2016/02/29/difference-between-uri-url-and-urn/>
- [4] <https://stackoverflow.com/questions/4913343/what-is-the-difference-between-uri-url-and-urn>
- [5] https://www.w3schools.com/xml/xml_rdf.asp
- [6] <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf/>
- [7] <https://www.obitko.com/tutorials/ontologies-semantic-web/rdf-elements.html>
- [8] <http://ikee.lib.auth.gr/record/135325/files/GRI-2014-13311.pdf>
- [9] <https://www.w3.org/TR/sparql11-protocol/>
- [10] https://www.webopedia.com/TERM/O/Ontology_Web_Language.html
- [11] http://dione.lib.unipi.gr/xmlui/bitstream/handle/unipi/9227/Tsartsali_Aikaterini.pdf?sequence=1&isAllowed=y
- [12] <https://www.w3.org/standards/semanticweb/data>
- [13] <http://repository.library.teimes.gr/xmlui/bitstream/handle/123456789/4610/%CE%95%CE%99%CE%A3%CE%91%CE%93%CE%A9%CE%93%CE%97%20%CE%A3%CE%A4%CE%9F%20%CE%A3%CE%97%CE%9C%CE%91%CE%A3%CE%99%CE%9F%CE>

- %9B%CE%9F%CE%93%CE%99%CE%9A%CE%9F%20%CE%99%CE%A3%CE%A4%CE%9F%20%CE%9A%CE%91%CE%99%20%CE%A3%CE%A4%CE%97%20%CE%93%CE%9B%CE%A9%CE%A3%CE%A3%CE%91%20%CE%91%CE%9D%CE%91%CE%A0%CE%91%CE%A1%CE%91%CE%A3%CE%A4%CE%91%CE%A3%CE%97%CE%A3%20%CE%9F%CE%9D%CE%A4%CE%9F%CE%9B%CE%9F%CE%93%CE%99%CE%A9%CE%9D%20%28OWL%29..pdf?sequence=1&isAllowed=y
- [14] <https://wordlift.io/blog/en/entity/triple/>
- [15] <https://wordlift.io/blog/en/entity/linked-data/>
- [16] <http://www.linkeddatatools.com/introducing-rdfs-owl>
- [17] <https://www.xml.com/pub/a/2001/01/24/rdf.html>
- [18] <http://ikee.lib.auth.gr/record/298020/files/GRI-2018-21666.pdf>
- [19] https://nemertes.lis.upatras.gr/jspui/bitstream/10889/4225/1/msc_mertis_539.pdf
- [20] <https://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html>
- [21] https://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=UTConvertQ1
- [22] <https://techterms.com/definition/unicode>
- [23] <https://unicodebook.readthedocs.io/unicode.html>
- [24] <http://thinkzarahatke.com/author/amty.html>
- [25] https://www.w3.org/2001/sw/wiki/Semantic_Web_terminology
- [26] <http://graphdb.ontotext.com/documentation/free/introduction-to-semantic-web.html#what-s-in-this-document>
- [27] https://www.researchgate.net/publication/221466538_A_Comparison_of_RDF_Query_Languages
- [28] <https://www.xml.com/pub/a/2005/07/20/versa.html>
- [29] <https://www.w3.org/TeamSubmission/n3/>
- [30] <https://www.w3.org/TR/rdf-testcases/#ntriples>
- [31] <https://www.w3.org/2001/sw/wiki/RIF>
- [32] https://www.sti-innsbruck.at/sites/default/files/courses/fileadmin/documents/semweb09-10/08_SW-RIF.pdf
- [33] <https://ieeexplore.ieee.org/document/5467076>
- [34] <https://www.timbr.ai/timbr/>
- [35] http://tw.rpi.edu/media/2018/08/29/14ad2/makni_bassem_dissertation_dl4rdfs.pdf
- [36] <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>
- [37] <https://allegrograph.com/articles/semantic-web-and-semantic-technology-trends-in-2019/>
- [38] https://www.researchgate.net/publication/221466797_Towards_Machine_Learning_on_the_Semantic_Web
- [39] <https://tandfonline.com/doi/pdf/10.1080/1206212X.2019.1570666?needAccess=true>
- [40] <https://www.slideshare.net/semwebcompany/semantics-as-the-basis-of-advanced-cognitive-computing>
- [41] <https://www.infoworld.com/article/2610447/cognitive-computing-can-take-the-semantic-web-to-the-next-level.html>
- [42] <https://www.ontotext.com/blog/taking-semantic-web-next-level-cognitive-computing/>
- [43] <https://www.forbes.com/sites/bernardmarr/2016/03/23/what-everyone-should-know-about-cognitive-computing/#2a9a82315088>
- [44] <https://searchenterpriseai.techtarget.com/definition/cognitive-computing>
- [45] <https://www.ontotext.com/blog/taking-semantic-web-next-level-cognitive-computing/>
- [46] <http://www.semantic-web-journal.net/blog/call-papers-special-issue-semantic-deep-learning>
- [47] <https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-018-0181-1>
- [48] https://www.researchgate.net/publication/226213344_Semantic_Web_Technologies_and_Artificial_Neural_Networks_for_Intelligent_Web_Knowledge_Source_Discovery
- [49] <https://2019.semantics.cc/building-bridges-between-databases-and-semantic-web>

