|  |  |
| --- | --- |
| Εικόνα που περιέχει εμπρόσθια όψη, κέρμα  Περιγραφή που δημιουργήθηκε αυτόματα | Εθνικό Μετσόβιο Πολυτεχνείο  Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  Τομεας Τεχνολογιασ Πληροφορικησ και Υπολογιστων |

**Οι Συζητήσεις Του Κοινοβουλίου Ως Διασυνδεδεμένα Ανοιχτά Δεδομένα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**ΠΑΠΑΝΙΚΟΛΑΟΥ ΙΩΑΝΝΗ**

**Επιβλέπων :** Παναγιώτης Τσανάκας

Καθηγητής Ε.Μ.Π.

**Συνεπιβλέπων :** Μάριος Κόνιαρης

Ε.ΔΙ.Π. ΕΜΠ

Αθήνα, Σεπτέμβριος 2023

Η σελίδα αυτή είναι σκόπιμα λευκή.

|  |  |
| --- | --- |
| Εικόνα που περιέχει εμπρόσθια όψη, κέρμα  Περιγραφή που δημιουργήθηκε αυτόματα | Εθνικό Μετσόβιο Πολυτεχνείο  Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  Τομεας Τεχνολογιασ Πληροφορικησ και Υπολογιστων |

**Οι Συζητήσεις Του Κοινοβουλίου Ως Διασυνδεδεμένα Ανοιχτά Δεδομένα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**ΙΩΑΝΝΗ ΠΑΠΑΝΙΚΟΛΑΟΥ**

**Επιβλέπων :** Παναγιώτης Τσανάκας

Καθηγητής Ε.Μ.Π.

**Συνεπιβλέπων :** Μάριος Κόνιαρης

Ε.ΔΙ.Π. ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 1η Σεπτεμβρίου 2023.

*(Υπογραφή) (Υπογραφή) (Υπογραφή)*

................................... ................................... ...................................

Παναγιώτης Τσανάκας Βασιλική Καντερέ Ευγενία Τζαννίνη

Καθηγητής Ε.Μ.Π. Επίκουρη Καθηγήτρια Ε.Μ.Π. Καθηγήτρια Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2023

*(Υπογραφή)*

...................................

**ΙΩΑΝΝΗΣ ΠΑΠΑΝΙΚΟΛΑΟΥ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © ΙΩΑΝΝΗΣ, ΠΑΠΑΝΙΚΟΛΑΟΥ, 2023.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

**Περίληψη**

Η αυξανόμενη διαθεσιμότητα ψηφιακών δεδομένων και οι εξελίξεις στις τεχνολογίες σημασιολογικού ιστού έχουν ανοίξει νέες δυνατότητες για την ανάλυση και την εξαγωγή γνώσης από διάφορες πηγές. Η παρούσα διπλωματική εργασία διερευνά την εφαρμογή αυτών των τεχνολογιών για τη μετατροπή των συζητήσεων του Ελληνικού κοινοβουλίου σε συνδεδεμένα δεδομένα, επιτρέποντας πλουσιότερη και πιο ουσιαστική ανάλυση του νομοθετικού λόγου.

Η εργασία ξεκινά με την παρουσία ενός εκτενούς θεωρητικού υποβάθρου. Έπειτα εστιαζόμενο σε πιο πρακτικά ζητήματα, αναλύεται ένα σύστημα το οποίο ξενικά με την μετατροπή των ακατέργαστων κειμενικών δεδομένων των συζητήσεων του Ελληνικού Κοινοβουλίου σε δομημένα έγγραφα XML, τα οποία χρησιμεύουν ως ενδιάμεση αναπαράσταση, επιτρέποντας την περαιτέρω επεξεργασία και ανάλυση. Στη συνέχεια, τα έγγραφα XML μετατρέπονται σε μορφή σχήματος Resource Description Framework Schema (RDFS), επιτρέποντας την αναπαράσταση εννοιών, σχέσεων και μεταδεδομένων που σχετίζονται με τις συζητήσεις.

Εν συνεχεία, στόχος είναι η ανάδειξη της δύναμης των σημασιολογικών ερωτημάτων με τη χρήση της SPARQL, μιας γλώσσας ερωτημάτων για την αναζήτηση δεδομένων RDF. Με τη μετατροπή των συζητήσεων του Ελληνικού Κοινοβουλίου σε διασυνδεδεμένα δεδομένα και την παροχή ενός σχήματος RDFS, καθίσταται δυνατή η εκτέλεση σύνθετων ερωτημάτων, η ανάκτηση ουσιαστικών πληροφοριών και η δημιουργία συνδέσεων μεταξύ διαφόρων πτυχών των συζητήσεων. Η προσέγγιση αυτή επιδεικνύει τις δυνατότητες των τεχνολογιών του σημασιολογικού ιστού στην ενίσχυση της προσβασιμότητας, της ευρεσιμότητας και της χρηστικότητας των κοινοβουλευτικών διαδικασιών, συμβάλλοντας στον ευρύτερο τομέα της ηλεκτρονικής δημοκρατίας και της διαφανούς διακυβέρνησης.

**Λέξεις Κλειδιά:** « Σημασιολογικός Ιστός, Συνδεδεμένα Δεδομένα, XML, RDF, SPARQL, Οι Συζητήσεις Του Ελληνικού Κοινοβουλίου »

Η σελίδα αυτή είναι σκόπιμα λευκή.

**Abstract**

The increasing availability of digital data and developments in semantic web technologies have opened up new possibilities for analysing and extracting knowledge from various sources. This thesis explores the application of these technologies to transform the debates of the Greek parliament into linked data, allowing for richer and more meaningful analysis of legislative discourse.

The paper begins with the presentation of an extensive theoretical background. Then, focusing on more practical issues, a system is analysed which is capable of converting the raw textual data of the Greek Parliament debates into structured XML documents, which serve as an intermediate representation, allowing further processing and analysis. The XML documents are then converted to a Resource Description Framework Schema (RDFS) format, allowing the representation of concepts, relationships and metadata associated with the discussions.

Next, the aim is to demonstrate the power of semantic queries using SPARQL, a query language for querying RDF data. By converting the debates of the Greek Parliament into interlinked data and providing an RDFS schema, it is possible to perform complex queries, retrieve meaningful information and create links between different aspects of the debates. This approach demonstrates the potential of semantic web technologies in enhancing the accessibility, findability and usability of parliamentary procedures, contributing to the broader field of e-democracy and transparent governance.

**Keywords: «**Semantic Web, Linked Data, XML, RDF, SPARQL, Greek Parliament Debates»

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας περιεχομένων

1 Εισαγωγή 1

1.1 Γενικότερη Ανάλυση Προβλήματος της Εργασίας 2

1.2 Αντικείμενο διπλωματικής 2

1.2.1 Συνεισφορά 3

1.3 Σχετικά Παραδείγματα του Εξωτερικού 3

1.4 Οργάνωση εργασίας 3

[2 Θεωρητικό υπόβαθρο 10](#_Θεωρητικό_υπόβαθρο)

2.1 Σημασιολογικός Ιστός - Semantic Web 10**.**

2.2 Ανοιχτά Διασυνδεδεμένα Δεδομένα 10

2.3 Τεχνολογίες Υποστήριξης Διασυνδεδεμένων Δεδομένων 10

2.3.1 XML 11

2.3.2 RDF 11

2.3.3 Λεξιλόγια Σημασιολογικού Ιστού 11

2.3.4 SPARQL 11

2.4 Ανοικτά Κυβερνητικά Δεδομένα – Akoma Ntoso 12

3 Σύστημα Διαχείρισης Πρακτικών 13

3.1 Αρχιτεκτονική Συστήματος 14

3.2 Μέθοδος Αντλήσης Αρχείων - Βάση Δεδομένων 14

3.3 Δομή Πρακτικών Βουλής 14

3.4 Εξαγωγή και Ανάλυση Κειμένων 14

3.4.1 ANLTR4 - REGEΧ 14

3.5 Διαχείριση Δεδομένων 14

3.5.1 Μετατροπή Αρχείων Κειμένου σε XML αρχεία 14

3.5.2 Μετατροπή Αρχείων XML σε RDF αρχεία 14

3.5.3 Επεξεργασία RDF - SPARQL 14

4 Έλεγχος Συστήματος 15

4.1 Μεθοδολογία ελέγχου 15

4.2 Αναλυτική παρουσίαση ελέγχου/παραδείγματος 15

5 Επίλογος 15

5.1 Σύνοψη και συμπεράσματα 15

5.2 Μελλοντικές επεκτάσεις 15

6 Βιβλιογραφία 15

# Εισαγωγή

## Περιγραφή Προβλήματος

Η παρούσα διπλωματική εργασία επικεντρώνεται στην αξιοποίηση των αρχών των συνδεδεμένων δεδομένων και των τεχνολογιών του σημασιολογικού ιστού για τη βελτίωση της προσβασιμότητας και της ανάλυσης των συζητήσεων του Ελληνικού Κοινοβουλίου. Το Ελληνικό Κοινοβούλιο, ως το ανώτατο νομοθετικό όργανο στην Ελλάδα, παράγει έναν τεράστιο όγκο πολύτιμων κειμενικών δεδομένων κατά τη διάρκεια των συζητήσεών του. Ωστόσο, οι υπάρχουσες μέθοδοι πρόσβασης και αξιοποίησης αυτών των δεδομένων παρουσιάζουν αρκετές προκλήσεις και περιορισμούς.

Ο ιστότοπος του Ελληνικού Κοινοβουλίου χρησιμεύει ως η κύρια πηγή πληροφοριών σχετικά με τις κοινοβουλευτικές διαδικασίες. Παρέχει πρόσβαση σε απομαγνητοφωνημένα κείμενα και αρχεία των συζητήσεων, επιτρέποντας στους πολίτες, τους ερευνητές και τους πρωταγωνιστές της πολιτικής σκηνής να ενημερώνονται για τις νομοθετικές δραστηριότητες. Ωστόσο, η τρέχουσα δομή του ιστοτόπου υποστηρίζει κυρίως την παραδοσιακή περιήγηση και αναζήτηση με βάση κυρίως την ημερομηνία, χωρίς προηγμένα χαρακτηριστικά για την εξερεύνηση δεδομένων, τη διασύνδεση και τη σημασιολογική αναζήτηση.

Ένα από τα κύρια προβλήματα που αντιμετωπίζει ο υφιστάμενος δικτυακός τόπος είναι η έλλειψη δομημένης αναπαράστασης δεδομένων. Οι συζητήσεις παρουσιάζονται κυρίως ως αδόμητο κείμενο, σε μορφή pdf, word ή txt, γεγονός που καθιστά δύσκολη την εξαγωγή συγκεκριμένων πληροφοριών ή την εκτέλεση ουσιαστικής ανάλυσης σε λεπτομερές επίπεδο. Αυτός ο περιορισμός εμποδίζει την ολοκληρωμένη έρευνα, τον εντοπισμό τάσεων και την ικανότητα να γίνουν συνδέσεις μεταξύ διαφορετικών συζητήσεων, θεμάτων ή συμμετεχόντων/ομιλητών.

Η διπλωματική αυτή αποσκοπεί στην αντιμετώπιση αυτών των προκλήσεων προτείνοντας ένα πλαίσιο που μετατρέπει τις συζητήσεις του Ελληνικού Κοινοβουλίου σε συνδεδεμένα δεδομένα χρησιμοποιώντας τεχνολογίες σημασιολογικού ιστού. Με την αξιοποίηση της γλώσσας προγραμματισμού Python και του ANTLR-4, τα μη δομημένα κειμένα θα αναλυθούν και θα μετατραπούν σε δομημένα έγγραφα XML. Στη συνέχεια, τα έγγραφα XML θα μετατραπούν σε μορφή σχήματος RDFS (Resource Description Framework Schema), επιτρέποντας την αναπαράσταση εννοιών, σχέσεων και μεταδεδομένων που σχετίζονται με τις συζητήσεις.

Ο πρωταρχικός στόχος είναι να δημιουργηθεί μια αναπαράσταση συνδεδεμένων δεδομένων των συζητήσεων του Ελληνικού Κοινοβουλίου, η οποία θα επιτρέπει τη σημασιολογική αναζήτηση μέσω SPARQL. Αυτή η προσέγγιση θα διευκολύνει τις προηγμένες αναζητήσεις, τα σύνθετα ερωτήματα και την εξαγωγή ουσιαστικών πληροφοριών από το σύνολο δεδομένων. Επιπλέον, η εργασία θα διερευνήσει τη δημιουργία ενός σχήματος RDFS ειδικά για τον τομέα των κοινοβουλευτικών συζητήσεων της Ελλάδος, αποτυπώνοντας την απαραίτητη γνώση για ολοκληρωμένη ανάλυση.

Συνολικά, η παρούσα διπλωματική επιδιώκει να συμβάλει στον τομέα της ηλεκτρονικής δημοκρατίας και της διαφανούς διακυβέρνησης αναδεικνύοντας τις δυνατότητες των τεχνολογιών του σημασιολογικού ιστού στις κοινοβουλευτικές συζητήσεις.

## Αντικείμενο διπλωματικής

Εδώ αναφερόμαστε συγκεκριμένα στο τί θα κάνει η διπλωματική. Αναφέρουμε λεπτομερώς α) τα προβλήματα που θα λύσει (και που ήδη έχουν περιγραφεί γενικά στην προηγούμενη ενότητα), και β) πώς σκοπεύει να τα λύσει.

Είναι σημαντικό κάποιος που θα διαβάσει την ενότητα αυτή να καταλάβει σε σημαντικό βαθμό τον σκοπό της διπλωματικής σας και τις τεχνικές δυσκολίες της, χωρίς να είναι αναγκαίο να δει όλα τα άλλα κεφάλαια. **Η ενότητα αυτή θέλει πολύ προσοχή και καλύτερα να τη γράψετε αφού έχετε γράψει όλα τα υπόλοιπα κεφάλαια.**

### Συνεισφορά

Εδώ παραθέτουμε αριθμητικά συγκεκριμένες ενέργειες/λύσεις/μεθοδολογίες που παρουσιάζει η διπλωματική και λύνουν τα προβλήματα που υποσχεθήκαμε στην προηγούμενη ενότητα ότι θα λύσει η διπλωματική. Συνήθως η υποενότητα αυτή έχει την παρακάτω μορφή:

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Μελετήσαμε συστήματα κ.λ.π.
2. Υλοποιήσαμε 3 αλγορίθμους υπολογισμού κ.λ.π.
3. Αξιολογήσαμε την επίδοση των αλγορίθμων και βρήκαμε ότι κ.λ.π.
4. Ενσωματώσαμε τους αλγορίθμους σε πρότυπο σύστημα κ.λ.π.
5. ...

## Σχετικά Παραδείγματα του Εξωτερικού

Είναι ζωτικής σημασίας να εξετάσουμε σχετικά διεθνή παραδείγματα προκειμένου να κατανοήσουμε την εφαρμογή των συνδεδεμένων δεδομένων στις κοινοβουλευτικές διαδικασίες από παγκόσμια άποψη. Τα Πρακτικά του Ευρωπαϊκού Κοινοβουλίου και τα Πρακτικά του Κοινοβουλίου του Καναδά είναι δύο γνωστά παραδείγματα που εξετάζονται στην παρούσα ενότητα.

Τα Πρακτικά του Ευρωπαϊκού Κοινοβουλίου είναι ένα αξιοσημείωτο έργο που χρησιμοποιεί τις αρχές των συνδεδεμένων δεδομένων για να παρέχει πρόσβαση στις κοινοβουλευτικές συζητήσεις και τις σχετικές πληροφορίες (<https://linkedpolitics.ops.few.vu.nl/web/html/home.html>). Οι χρήστες μπορούν να περιηγηθούν και να εξετάσουν συζητήσεις, ομιλίες, αρχεία ψηφοφορίας και άλλες νομοθετικές δραστηριότητες στον ιστότοπο, ο οποίος παρέχει ένα πλήρες αρχείο των κοινοβουλευτικών διαδικασιών. Το σύνολο δεδομένων περιλαμβάνει κάθε συζήτηση στην ολομέλεια του Ευρωπαϊκού Κοινοβουλίου (ΕΚ) από τον Ιούλιο του 1999 έως τον Ιανουάριο του 2014, καθώς και προσωπικά δεδομένα για κάθε μέλος του κοινοβουλίου. Περιέχει δεδομένα σχετικά με τις ημερήσιες συνεδριάσεις του Ευρωπαϊκού Κοινοβουλίου, το πρόγραμμα των συζητήσεων, τους λόγους και τις μεταφράσεις τους σε 21 διαφορετικές γλώσσες. Ακόμα περιλαμβάνει πληροφορίες για τους ρόλους των ομιλητών και τα έθνη που εκπροσωπούν, καθώς και τη συμμετοχή των εθνικών κομμάτων, των ευρωπαϊκών κομμάτων και των επιτροπών. Τα δεδομένα είναι προσβάσιμα μέσω ενός SPARQL API, στο οποίο δύναται να γίνονται πιο στοχευμένα ερωτήματα, με σκοπό προηγμένες αναλύσεις και λήψη συγκεκριμένων πληροφοριών.

Τα Πρακτικά του Κοινοβουλίου του Καναδά (<https://hansard.opennwt.ca/debates/>) είναι μια άλλη σημαντική περίπτωση. Ο ιστότοπος παρέχει πρόσβαση στα αρχεία των συζητήσεων που έχουν διεξαχθεί στο Κοινοβούλιο του Καναδά από τις αρχές τις δεκαετίας του 1990 μέχρι και σήμερα. Οι συζητήσεις αυτές καλύπτουν ένα ευρύ φάσμα θεμάτων, επιτρέποντας στους χρήστες να αποκτήσουν εικόνα των νομοθετικών συζητήσεων και των διαδικασιών λήψης αποφάσεων. Φυσικά, μπορούν να αντληθούν πληροφορίες σχετικά με τους ομιλητές που συμμετέχουν στις συζητήσεις, συμπεριλαμβανομένων των ονομάτων τους, των ψήφων τους και των θέσεων τους στο κοινοβούλιο. Οι χρήστες μπορούν να έχουν πρόσβαση στα δεδομένα μέσα από ένα πλήρως φιλικό προς τον χρήστη UI, στο οποίο οι ομιλίες καταγράφονται ανάλογα με το θέμα, την ημερομηνία και τον ομιλητή.

## Οργάνωση κειμένου

Εδώ περιγράφουμε τα κεφάλαια της διπλωματικής: 1 πρόταση για το τι θα έχει κάθε κεφάλαιο. Συνήθως η ενότητα αυτή έχει την παρακάτω μορφή (δεν θα σας πάρει πάνω από 1 μεγάλη παράγραφο):

Εργασίες σχετικές με το αντικείμενο της διπλωματικής παρουσιάζονται στο Κεφάλαιο 2 . Το Κεφάλαιο 3 συζητά θέματα μοντελοποίησης. Στο Κεφάλαιο 4 αναπτύσσουμε κ.λ.π.

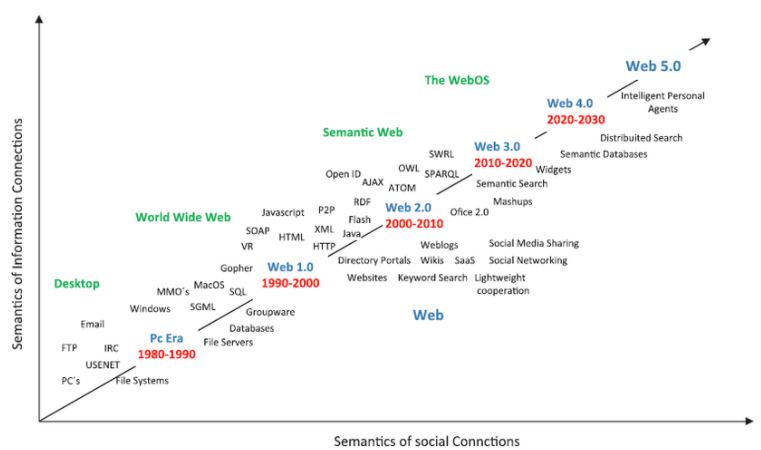
# Θεωρητικό υπόβαθρο

Σύντομη περιγραφή του θεωρητικού υποβάθρου

## Σημασιολογικός Ιστός – Semantic Web

Με την πάροδο των χρόνων, η ανάπτυξη του διαδικτύου έχει φέρει επανάσταση στον τρόπο με τον οποίο επικοινωνούμε, έχουμε πρόσβαση σε πληροφορίες και διεξάγουμε επιχειρηματικές δραστηριότητες. Από τα πρώτα βήματα του, ως ένα απλό δίκτυο διασυνδεδεμένων εγγράφων, το διαδίκτυο έχει υποστεί αξιοσημείωτες αλλαγές με την πάροδο των ετών. Μια τέτοια «τελευταία» αλλαγή είναι η μετάβαση στον σημασιολογικό ιστό, γεγονός που αποσκοπεί στη βαθύτερη κατανόηση και νοηματοδότηση του περιεχομένου του ιστού. Αυτή η μετατόπιση έχει ανοίξει νέες δυνατότητες για την αναπαράσταση γνώσης, την ολοκλήρωση δεδομένων και την ευφυή αυτοματοποίηση.

Η μετάβαση από τον παραδοσιακό ιστό στον σημασιολογικό ιστό αποτελεί σημαντικό ορόσημο στην εξέλιξη του διαδικτύου. Στα πρώτα στάδια του ιστού, οι πληροφορίες παρουσιάζονταν κυρίως σε αδόμητες μορφές, γεγονός που καθιστούσε δύσκολη έως και αδύνατη την αποτελεσματική επεξεργασία και ερμηνεία του περιεχομένου από τις μηχανές. Ωστόσο, καθώς ο ιστός μεγάλωνε σε μέγεθος και πολυπλοκότητα, προέκυψε η ανάγκη για έναν πιο έξυπνο και αποτελεσματικό τρόπο οργάνωσης και κατανόησης του τεράστιου όγκου των διαθέσιμων πληροφοριών. Έτσι γεννιέται η έννοια του σημασιολογικού ιστού, η οποία αποσκοπούσε στο να προσδώσει στο περιεχόμενο του ιστού σαφές νόημα και να επιτρέψει στις μηχανές να κατανοήσουν τα δεδομένα.

Σχήμα 2-1: Η εξέλιξη του Διαδικτύου (Osorio, Ortiz 2013 ‐ <https://myeltcafe.com/wp-content/uploads/2020/12/123123123.jpg> )

Στον πυρήνα του σημασιολογικού ιστού βρίσκεται η χρήση οντολογιών, οι οποίες παρέχουν μια τυπική αναπαράσταση της γνώσης σε συγκεκριμένους τομείς. Οι οντολογίες ορίζουν τις έννοιες, τις σχέσεις και τις ιδιότητες ενός συγκεκριμένου τομέα, διευκολύνοντας την κοινή κατανόηση μεταξύ ανθρώπων και μηχανών. Με την αξιοποίηση των οντολογιών, ο σημασιολογικός ιστός επιτρέπει τον σχολιασμό και τη δόμηση του περιεχομένου του ιστού, καθιστώντας το αναγνώσιμο από μηχανές.

Τα συνδεδεμένα δεδομένα είναι μια άλλη βασική πτυχή του σημασιολογικού ιστού. Περιλαμβάνει τη δημιουργία συνδέσεων μεταξύ διαφορετικών πηγών δεδομένων στον ιστό, χρησιμοποιώντας τυποποιημένα πρωτόκολλα και τεχνολογίες όπως το RDF (Resource Description Framework) και τα URIs (Uniform Resource Identifiers). Συνδέοντας δεδομένα σε διαφορετικούς τομείς και πηγές, ο σημασιολογικός ιστός προωθεί την ολοκλήρωση, τη διασύνδεση και την επαναχρησιμοποίηση δεδομένων. Αυτός ο διασυνδεδεμένος ιστός δεδομένων επιτρέπει μια πιο ολοκληρωμένη και πλαισιωμένη κατανόηση των πληροφοριών, ξεπερνώντας τα όρια των μεμονωμένων εγγράφων ή ιστοτόπων. Φυσικά, οι παραπάνω έννοιες θα αναλυθούν εκτενώς στην συνέχεια.

Προφανώς, η μετάβαση στον σημασιολογικό ιστό υπόσχεται πολλά για διάφορους τομείς. Ενδεικτικά, στον τομέα της ανάκτησης πληροφοριών και των μηχανών αναζήτησης, ο σημασιολογικός ιστός μπορεί να βελτιώσει την ακρίβεια και τη συνάφεια των αποτελεσμάτων αναζήτησης, κατανοώντας το νόημα και το πλαίσιο πίσω από τα ερωτήματα των χρηστών. Σε έναν άλλο τομέα, όπως αυτού του ηλεκτρονικό εμπόριο, ο σημασιολογικός ιστός επιτρέπει πιο έξυπνες συστάσεις προϊόντων και εξατομικευμένες εμπειρίες αγορών με βάση τη βαθύτερη κατανόηση των προτιμήσεων και των αναγκών των πελατών. Τέλος, στην υγειονομική περίθαλψη, ο σημασιολογικός ιστός διευκολύνει τη διαλειτουργικότητα και την ενοποίηση των ιατρικών δεδομένων, οδηγώντας σε βελτιωμένη περίθαλψη των ασθενών, ερευνητική συνεργασία και λήψη αποφάσεων βάσει δεδομένων.

Επιπλέον, ο σημασιολογικός ιστός έχει επιπτώσεις στις εφαρμογές τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης (ML). Παρέχοντας ένα τυποποιημένο πλαίσιο για την αναπαράσταση δεδομένων και την ολοκλήρωση της γνώσης, ο σημασιολογικός ιστός μπορεί να βελτιώσει την εκπαίδευση και την απόδοση των μοντέλων τεχνητής νοημοσύνης, επιτρέποντάς τους να κάνουν πιο τεκμηριωμένες προβλέψεις και συστάσεις. Ανοίγει επίσης το δρόμο για την ανάπτυξη ευφυών πρακτόρων και chatbots που μπορούν να κατανοούν και να απαντούν σε ερωτήματα χρηστών με πιο φυσικό και συνειδητό τρόπο.  
 Πλέον, μοναδικός στόχος της τεχνολογίας είναι η πρόοδος της, επομένως όπως μας μαρτυρά και το Σχήμα 2-1, είναι ότι πλησιάζει η εδραίωση του Semantic Web ή του Web 3.0 και η σταδιακή μας μετάβαση στο νεότερο διαδίκτυο με αρίθμηση 4.0, αν βέβαια επιβεβαιωθούν αυτές οι προβλέψεις, μέχρι το τέλος αυτής της δεκαετίας.

## Ανοιχτά Διασυνδεδεμένα Δεδομένα

Η ιδέα των Ανοικτών Συνδεδεμένων Δεδομένων είναι ένα θεμελιώδες στοιχείο που, εκτός από τον Σημασιολογικό Ιστό, βελτιώνει την αποτελεσματικότητα και τη διαλειτουργικότητα των δεδομένων. Ενθαρρύνοντας την τυποποιημένη και ανοικτή δημοσίευση και σύνδεση δομημένων δεδομένων, τα Ανοικτά Συνδεδεμένα Δεδομένα προωθούν τα ιδανικά του Σημασιολογικού Ιστού.

Προκειμένου να εκφραστούν τα δεδομένα με τρόπο που να είναι αναγνώσιμο από μηχανήματα, τα Ανοιχτά Συνδεδεμένα Δεδομένα επιτάσσουν τη χρήση ανοιχτών προτύπων όπως το RDF (Resource Description Framework). Το RDF επιτρέπει την τριπλή δομή υποκειμένου-κατηγορουμένου-αντικειμένου των δεδομένων, επιτρέποντας τη δήλωση σύνθετων σημασιολογιών και συνδέσεων εντός των δεδομένων. Αυτό διευκολύνει τον συνδυασμό και την ενσωμάτωση συνόλων δεδομένων από πολλές πηγές, δημιουργώντας ένα δίκτυο γνώσης που είναι πιο εκτεταμένο και συνδεδεμένο.

Η χρήση ομοιόμορφων αναγνωριστικών πόρων (URI) για τον ειδικό προσδιορισμό πραγμάτων και εννοιών εντός του οικο-συστήματος των συνδεδεμένων δεδομένων επιβάλλεται επίσης από τις αρχές των ανοικτών συνδεδεμένων δεδομένων. Τα URI χρησιμεύουν ως μόνιμα και παγκοσμίως μοναδικά αναγνωριστικά, επιτρέποντας την απρόσκοπτη αναφορά και σύνδεση πόρων σε διαφορετικά σύνολα δεδομένων και τομείς.

Όλα όσα περιγράψαμε παραπάνω έχει συμπυκνώσει ο Τιμ Μπέρνερς Λι, ο εφευρέτης του Παγκόσμιου Ιστού, σε τέσσερις βασικές και θεμελιώδης αρχές. Πιο συγκεκριμένα:

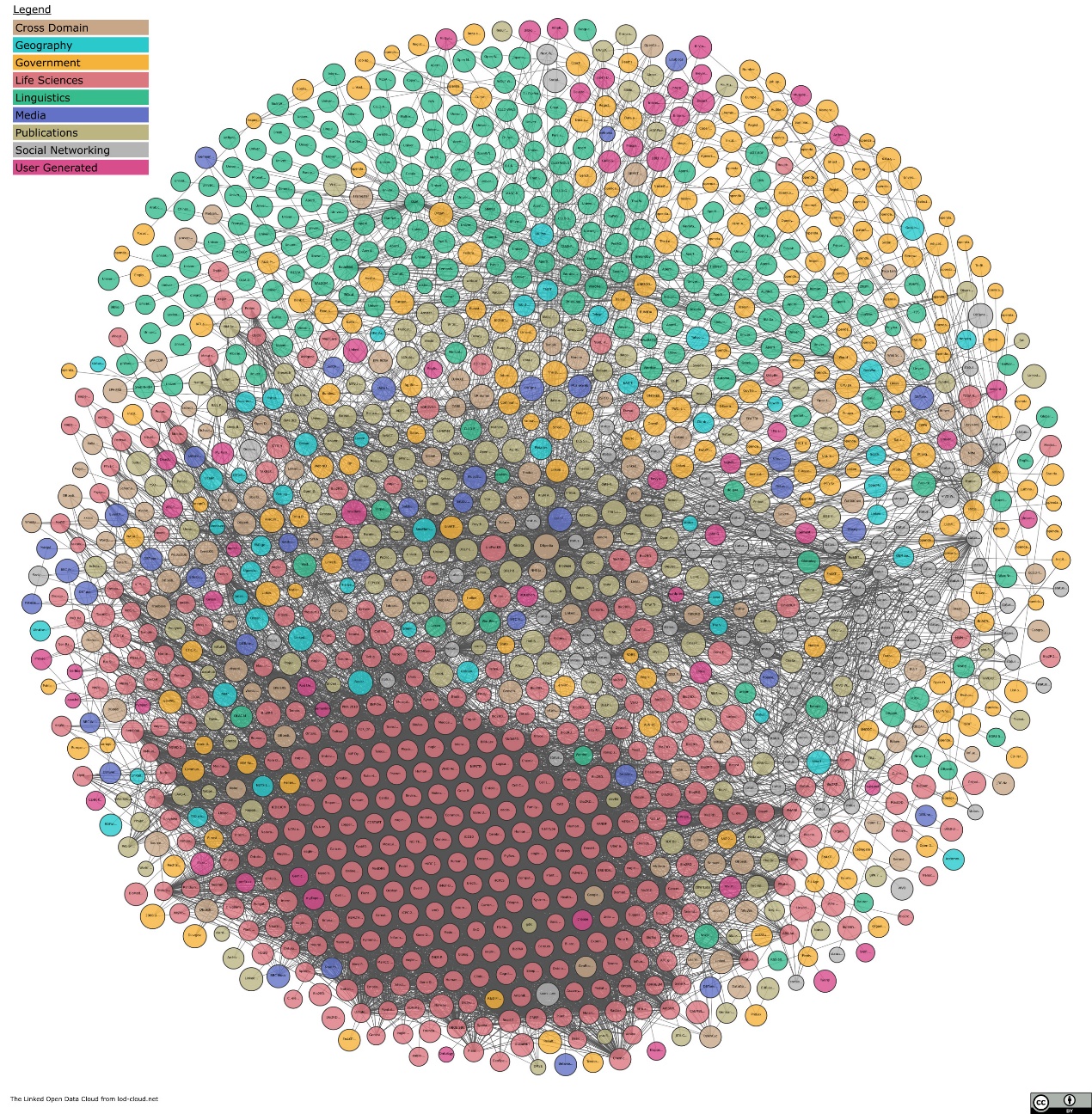
- Για να διασφαλιστεί ότι κάθε οντότητα ή έννοια έχει μια διακριτή ταυτότητα που την ξεχωρίζει μέσα στο τεράστιο διασυνδεδεμένο δίκτυο δεδομένων, ο πρώτος κανόνας απαιτεί τη χρήση URIs ως ονόματα για τα πράγματα.

- Ο δεύτερος κανόνας καθιστά απλή την αναζήτηση αυτών των ονομάτων με τη χρήση HTTP URIs, επιτρέποντας στους χρήστες να έχουν πρόσβαση και να ανακτούν χρήσιμα δεδομένα που σχετίζονται με τους αναγνωρισμένους πόρους.

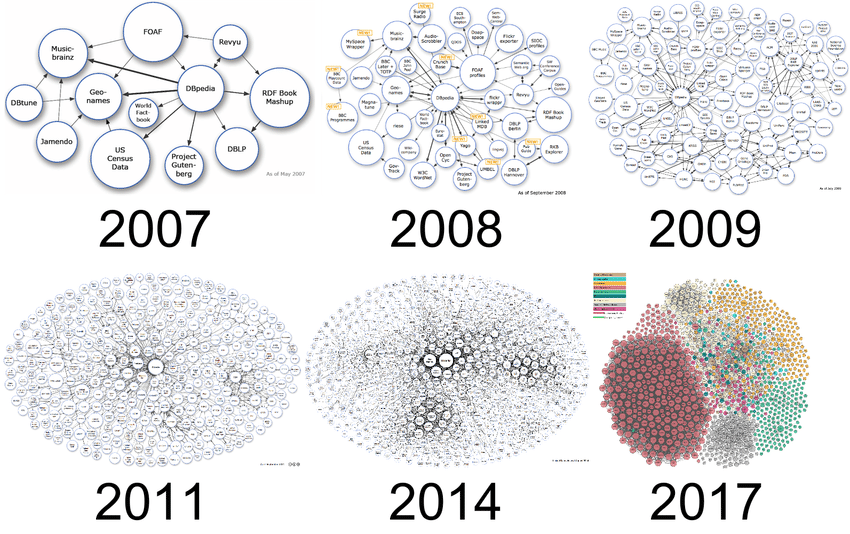
- Για την επιτυχή διευθέτηση και αναζήτηση των δεδομένων, ο τρίτος κανόνας δίνει έμφαση στην παρουσίαση σχετικών πληροφοριών κατά την αναζήτηση ενός URI. Για το σκοπό αυτό χρησιμοποιούνται τυποποιημένες τεχνολογίες όπως το RDF και η SPARQL.

- Η συμπερίληψη συνδέσμων προς άλλα URI τονίζεται επίσης από το τέταρτο κριτήριο, διευκολύνοντας την εύρεση νέων σχετικών πόρων και την ανάπτυξη ενός δικτύου διασυνδεδεμένης γνώσης.

Τα συνδεδεμένα δεδομένα λοιπόν είναι μια έννοια που προωθεί τη διασύνδεση και την ενσωμάτωση διαφορετικών συνόλων δεδομένων στο διαδίκτυο, όπως φαίνεται από το Σχήμα 2-1 (LOD Cloud). Στο διάγραμμα αυτό αποτυπώνεται το μεγαλειώδες δίκτυο των συνδεδεμένων συνόλων δεδομένων, τα οποία εμφανίζονται ως κόμβοι, και τις συνδέσεις τους. Ένα σύνολο δεδομένων, σύμφωνα με τις αρχές που είδαμε παραπάνω, αναπαρίσταται από κάθε κόμβο στο LOD Cloud, το οποίο χρησιμοποιεί μοναδικά αναγνωριστικά (URIs) για τον προσδιορισμό και την αναφορά των πόρων. Οι σύνδεσμοι και οι συνδέσεις μεταξύ των διαφόρων συνόλων δεδομένων αναπαρίστανται από τους συνδέσμους μεταξύ των κόμβων, επιτρέποντας την εύκολη πλοήγηση και την εξερεύνηση πληροφοριών. Το διάγραμμα της εικόνας δείχνει τη δύναμη της διασύνδεσης δεδομένων μεταξύ τους, επιτρέποντας στους χρήστες να περιηγηθούν μεταξύ συνόλων δεδομένων, να βρουν νέα δεδομένα και να αποκτήσουν ολοκληρωμένες γνώσεις, αξιοποιώντας τη συλλογική σοφία που υπάρχει στο οικοσύστημα των συνδεδεμένων δεδομένων.

Σχήμα 2-1 The Linked Open Data Cloud (<http://cas.lod-cloud.net/clouds/lod-cloud-sm.jpg>)

Ακόμα ένα άλλο ενδιαφέρον γράφημα είναι αυτό του Σχήματος 2-2 στο οποίο απεικονίζεται η εκθετική ανάπτυξη των Συνδεδεμένων Ανοικτών Δεδομένων (LOD) από το 2007 έως και τα τελευταία χρόνια, αναδεικνύοντας την αξιοσημείωτη πρόοδο που έχουν σημειώσει τα Διασυνδεδεμένα Δεδομένα με την πάροδο του χρόνου. Το γράφημα καταγράφει το αυξανόμενο δίκτυο πόρων συνδεδεμένων δεδομένων, δείχνοντας την αύξηση του αριθμού των συνόλων δεδομένων που δημοσιεύονται ως LOD. Η επέκταση αυτή αντανακλά την αυξανόμενη αποδοχή και εκτίμηση των αρχών των συνδεδεμένων δεδομένων σε ένα ευρύ φάσμα πεδίων και τομέων. Η ποσότητα και η ποικιλία των διασυνδεδεμένων συνόλων δεδομένων έχουν αυξηθεί τρομερά καθώς όλο και περισσότερες επιχειρήσεις, κοινωνίες και κυβερνήσεις υιοθετούν την ιδέα των συνδεδεμένων δεδομένων. Η ανάπτυξη ενός τεράστιου, διασυνδεδεμένου γράφου γνώσης ως αποτέλεσμα αυτής της ανάπτυξης κατέστησε δυνατή την καλύτερη ενσωμάτωση, ανακάλυψη και χρήση δεδομένων.



Σχήμα 2-2 Ανάπτυξη των συνδεδεμένων ανοικτών δεδομένων από το 2007 (<https://www.researchgate.net/figure/Growth-of-Linked-Open-Data-since-2007-1-The-amount-of-data-sets-published-as-LOD-have_fig2_331748480> )

## Τεχνολογίες Υποστήριξης Δεδομένων

Στον τομέα του Σημασιολογικού Ιστού και των Συνδεδεμένων Δεδομένων περιλαμβάνονται διάφορες τεχνολογίες που είναι ζωτικής σημασίας για την αποτελεσματική αναπαράσταση, ολοκλήρωση, επερώτηση και διαλειτουργικότητα των δεδομένων. Οι τεχνολογίες αυτές προσφέρουν το πλαίσιο για την οργάνωση, τη σύνδεση και τη συλλογή γνώσης από διάφορα σύνολα δεδομένων. Μια ευέλικτη και ευρέως χρησιμοποιούμενη γλώσσα για την κωδικοποίηση δομημένων δεδομένων, η XML (eXtensible Markup Language) προωθεί την ανταλλαγή δεδομένων και τη διαλειτουργικότητα. Προκειμένου να καταστεί δυνατή η ανάπτυξη πλούσιων σημασιολογικών αναπαραστάσεων, το RDF (Resource Description Framework) παρουσιάζει ένα τυποποιημένο μοντέλο για την περιγραφή και τη σύνδεση δεδομένων. Ο πυρήνας των συνδεδεμένων δεδομένων, το RDF προσφέρει έναν αποτελεσματικό τρόπο περιγραφής σχέσεων, ιδιοτήτων και πληροφοριών. Με τη βοήθεια της ισχυρής γλώσσας ερωτημάτων SPARQL (SPARQL Protocol and RDF Query Language), οι χρήστες μπορούν να λάβουν συγκεκριμένα δεδομένα, να πραγματοποιήσουν περίπλοκες συνδέσεις και να αποκτήσουν κατανόηση από τα διασυνδεδεμένα δεδομένα. Αυτές οι τεχνολογίες συνεργάζονται για να δημιουργήσουν μια πλήρη εργαλειοθήκη που επιτρέπει στους ερευνητές, τους προγραμματιστές και τους επαγγελματίες των δεδομένων να αξιοποιούν πλήρως τα έργα του σημασιολογικού ιστού και των συνδεδεμένων δεδομένων.

### Extensible Markup Language (XML)

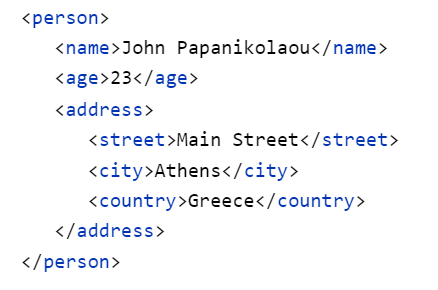
Η eXtensible Markup Language, γνωστή ως XML, είναι μια ευρέως υιοθετημένη τεχνολογία στον τομέα του Σημασιολογικού Ιστού και των συνδεδεμένων δεδομένων, καθώς είναι μια ισχυρή γλώσσα σήμανσης που έχει επηρεάσει σημαντικά την ανάπτυξη της ανταλλαγής πληροφοριών στο διαδίκτυο. Η αφετηρία της εντοπίζεται στις αρχές της δεκαετίας του 1970, όταν κατέστη για πρώτη φορά αναγκαίος ένας δομημένος τρόπος αναπαράστασης και ανταλλαγής δεδομένων.

Στις δεκαετίες του 1980 και 1990 δημιουργήθηκαν διάφορες γλώσσες σήμανσης (όπως η HTML), η καθεμία με μοναδικούς περιορισμούς και προβλεπόμενες χρήσεις. Ωστόσο, για να αντιμετωπιστούν οι αυξανόμενες απαιτήσεις για ανταλλαγή δεδομένων σε διάφορες πλατφόρμες και συστήματα, απαιτήθηκε μια πιο προσαρμόσιμη και επεκτάσιμη λύση, με αποτέλεσμα στα τέλη της δεκαετίας του 1990, να αναπτυχθεί η XML.

Η Κοινοπραξία του Παγκόσμιου Ιστού (W3C) ξεκίνησε τη διαδικασία τυποποίησης της XML το 1996 και καθόρισε τη σύνταξη και τη σημασιολογία της. Ο κύριος στόχος ήταν η ανάπτυξη μιας γλώσσας που θα επέτρεπε στους χρήστες να κατασκευάζουν τις δικές τους μοναδικές γλώσσες σήμανσης, καθιστώντας την αρκετά ευέλικτη ώστε να μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα εφαρμογών και δομών δεδομένων.

Η κύρια καινοτομία της XML είναι το πόσο απλή και κατανοητή είναι. Τα στοιχεία περιέχονται σε αγκύλες και διατάσσονται ιεραρχικά χρησιμοποιώντας μια μέθοδο βασισμένη σε ετικέτες. Τόσο οι άνθρωποι όσο και οι μηχανές μπορούν να ερμηνεύσουν τα δεδομένα, δεδομένου ότι αυτές οι ετικέτες καθορίζουν τη δομή και το νόημά τους.

Παρακάτω παρατίθεται ένα απόσπασμα από μια πιθανή απεικόνιση της XML με χρήση ετικετών για να γίνει κατανοητή η σύνταξη της γλώσσας:



Σχήμα 2.3.1 Δείγμα Κώδικα XML

Στο παράδειγμα του σχήματος 2.3.1, οι ετικέτες XML ενσωματώνουν διαφορετικά κομμάτια πληροφοριών για ένα άτομο. Το βασικό στοιχείο (root) είναι η ετικέτα "person", η οποία περιέχει φωλιασμένες ετικέτες όπως "name", "age" και "address". Κάθε ετικέτα αντιπροσωπεύει ένα μοναδικό κομμάτι δεδομένων. Για παράδειγμα, οι ετικέτες "name" και "age" περιέχουν το όνομα και την ηλικία του ατόμου αντίστοιχα, ενώ η ετικέτα "address" έχει φωλιασμένες τις ετικέτες "street", "city" και "country" για να αντικατοπτρίζει τις πληροφορίες διεύθυνσης.

Στο πλαίσιο των συνδεδεμένων δεδομένων, η XML αποτελεί βασική τεχνική για τη δόμηση και τη δημιουργία εγγράφων αναγνώσιμων από μηχανήματα για μη δομημένα κείμενα. Αυτά τα έγγραφα μπορούν να επεξεργαστούν, να ταξινομηθούν και να επισημανθούν με χρήσιμα στοιχεία και χαρακτηριστικά με τη χρήση XML, με αποτέλεσμα την παραγωγή καλά διαμορφωμένων εγγράφων XML. Τα επόμενα βήματα της μετάφρασης, της ολοκλήρωσης και της αναζήτησης δεδομένων με τη χρήση διαφόρων τεχνολογιών σημασιολογικού ιστού καθίστανται δυνατά με αυτή τη διαδικασία.

Σε καταστάσεις που αφορούν την ενσωμάτωση δεδομένων, όπου διάφορες πηγές πρέπει να εναρμονιστούν σε μια ενιαία αναπαράσταση, η XML είναι επίσης απαραίτητη. Είναι απλούστερη η αντιστοίχιση και ο μετασχηματισμός διαφορετικών συνόλων δεδομένων σε μια ενιαία μορφή λόγω της καθολικής σύνταξης της XML για την αναπαράσταση δομών δεδομένων.

### Resource Description Framework (RDF)

Το Resource Description Framework, γνωστό ως RDF, είναι μια θεμελιώδης τεχνολογία στον Σημασιολογικό Ιστό που χρησιμεύει ως τυποποιημένο μοντέλο για την αναπαράσταση και τη σύνδεση δεδομένων. Το RDF παρέχει μια δομημένη και ευέλικτη προσέγγιση για την περιγραφή των πόρων, των χαρακτηριστικών τους και των μεταξύ τους σχέσεων.

Οι τριπλέτες υποκειμένου-προγνωστικού-αντικειμένου (subject – predicate – object) είναι τα θεμελιώδη δομικά στοιχεία των αναπαραστάσεων πληροφοριών RDF. Οι πόροι που περιγράφονται αντιπροσωπεύονται από το υποκείμενο, το κατηγόρημα και το αντικείμενο, το οποίο αντιπροσωπεύει την τιμή ή έναν άλλο πόρο στον οποίο παραπέμπει η ιδιότητα. Αυτές οι τριπλέτες μπορούν να χρησιμοποιηθούν για τη δημιουργία ενός γραφήματος συνδεδεμένων δεδομένων, το οποίο λειτουργεί ως βάση για έναν knowledge graph.

Εικόνα που περιέχει κύκλος, γραμματοσειρά, κείμενο, γραφικά

Περιγραφή που δημιουργήθηκε αυτόματα

Σχήμα 2.2.2 – Βασική σύνταξη RDF (<https://upload.wikimedia.org/wikipedia/commons/thumb/8/88/Basic_RDF_Graph.svg/800px-Basic_RDF_Graph.svg.png>)

Το RDF ωφελεί τον Σημασιολογικό Ιστό με διάφορους τρόπους. Πρώτον, προσφέροντας ένα συνεπές μοντέλο δεδομένων και λεξιλόγιο για την κωδικοποίηση πληροφοριών, διευκολύνει τη συγχώνευση δεδομένων από πολλές πηγές. Διαφορετικά σύνολα δεδομένων μπορούν να ενσωματωθούν, να συνδεθούν και να αναζητηθούν συλλογικά χάρη στη διαλειτουργικότητα.

Δεύτερον, το RDF διευκολύνει την έκφραση σχέσεων και σημασιολογίας μέσα στα δεδομένα. Το RDF παρέχει παγκοσμίως μοναδικά και μόνιμα αναγνωριστικά για τους πόρους χρησιμοποιώντας τα URI ως αναγνωριστικά. Αυτό επιτρέπει τη σωστή αναφορά και σύνδεση των δεδομένων σε όλα τα σύνολα δεδομένων. Η ικανότητα σύνδεσης των σχετικών πηγών μεταξύ τους βελτιώνει την κατανόηση και την ερμηνεία των δεδομένων.

Χρησιμοποιώντας οντολογίες, το RDF διευκολύνει επίσης την επεκτασιμότητα. Οι έννοιες, οι συνδέσεις και οι περιορισμοί σε έναν τομέα αναπαρίστανται τυπικά και δομικά με οντολογίες. Η γνώση που αφορά συγκεκριμένο τομέα μπορεί να κωδικοποιηθεί χρησιμοποιώντας γλώσσες όπως το RDF Schema (RDFS) ή η Web Ontology Language (OWL).

Τέλος, τα δεδομένα RDF μπορούν να αποθηκευτούν και να αναζητηθούν αποτελεσματικά με τη χρήση του Apache Fuseki, μιας βάσης δεδομένων RDF ανοικτού κώδικα και ενός SPARQL endpoint. Η ικανότητα του Apache Fuseki να αποθηκεύει δεδομένα RDF με ιδιαίτερα κλιμακούμενο και αποτελεσματικό τρόπο είναι ένα από τα χαρακτηριστικά που το διακρίνουν. Χρησιμοποιεί το πλαίσιο Apache Jena, το οποίο παρέχει πλήρεις δυνατότητες διαχείρισης δεδομένων RDF, συμπεριλαμβανομένης της μόνιμης αποθήκευσης δεδομένων και της βελτιστοποίησης ερωτημάτων.

Επιπλέον, το Fuseki προσφέρει ένα SPARQL endpoint που επιτρέπει στους χρήστες να αλληλεπιδρούν με τα αποθηκευμένα δεδομένα RDF χρησιμοποιώντας τη γλώσσα ερωτημάτων SPARQL. Με τη βοήθεια της SPARQL, οι χρήστες μπορούν να εκτελούν εξελιγμένες λειτουργίες σύνδεσης, να λαμβάνουν συγκεκριμένα δεδομένα και να εξετάζουν τις αλληλένδετες σχέσεις που περιλαμβάνονται στα σύνολα δεδομένων. Η εκτέλεση ερωτημάτων SPARQL στην υποκείμενη βάση δεδομένων RDF γίνεται απλή και τυποποιημένη από το τελικό σημείο SPARQL που προσφέρει το Fuseki. Να τονίσουμε πως η λειτουργία και η σύνταξη των ερωτημάτων SPARQL θα αναλυθεί λεπτομερώς σε επόμενη παράγραφο.

### Λεξιλόγια Σημασιολογικού Ιστού

Τα λεξιλόγια του Σημασιολογικού Ιστού, γνωστά και ως οντολογίες, διαδραματίζουν κρίσιμο ρόλο στην αναπαράσταση και οργάνωση των δεδομένων στο οικοσύστημα του Σημασιολογικού Ιστού. Αυτά τα λεξιλόγια παρέχουν ένα κοινό και τυποποιημένο σύνολο όρων, σχέσεων και περιορισμών που επιτρέπουν τη δομημένη περιγραφή και ερμηνεία των δεδομένων.

Προορίζονται για την καταγραφή των ιδεών και των συνδέσεων που είναι μοναδικές σε έναν συγκεκριμένο τομέα. Παρέχουν στη γνώση μια τυπική αναπαράσταση, επιτρέποντας τη μοντελοποίηση και την έκφραση της σημασιολογίας που αφορά έναν συγκεκριμένο τομέα. Τα λεξιλόγια ακόμα επιτρέπουν την ανταλλαγή και την ενσωμάτωση δεδομένων σε διάφορες εφαρμογές και τομείς, περιγράφοντας ιδέες και τις σχέσεις τους.

Η ικανότητα του Σημασιολογικού Ιστού να περιγράφει πολλές μορφές πληροφοριών με τυποποιημένο και διαλειτουργικό τρόπο διευκολύνεται από διάφορα λεξιλόγια, καθένα από τα οποία έχει ξεχωριστό σκοπό και σύνολο ορολογίας. Οι εφαρμογές και τα συστήματα που χρησιμοποιούν αυτά τα λεξιλόγια μπορούν να επωφεληθούν από το σύνολο της γνώσης που αντιπροσωπεύουν συλλογικά, επιτρέποντας βαθύτερη ενσωμάτωση δεδομένων, αποτελεσματικότερη αναζήτηση και βελτιωμένη σημασιολογική διαλειτουργικότητα.

Τα λεξιλόγια του Σημασιολογικού Ιστού έχουν πολλά πλεονεκτήματα. Για αρχή, επιτρέπουν την διασύνδεση δεδομένων παρέχοντας μια κοινή κατανόηση των δεδομένων σε πολλές εφαρμογές και συστήματα. Τα λεξιλόγια καθιστούν δυνατή την ουσιαστική και κατανοητή αναπαράσταση των δεδομένων από τις μηχανές, επιτρέποντας τη βελτιωμένη αναζήτηση, ανάκτηση και ανεύρεση πληροφοριών. Παρέχουν, τέλος, εξελιγμένη εξαγωγή συμπερασμάτων, διευκολύνοντας την αυτόματη παραγωγή νέων πληροφοριών.

Πιο συγκεκριμένα, ένα από τα ευρέως χρησιμοποιούμενα λεξιλόγια του Σημασιολογικού Ιστού είναι το RDF Schema (RDFS), το οποίο προσφέρει ένα σύνολο κλάσεων και ιδιοτήτων για τη δημιουργία βασικών οντολογιών. Το RDFS παρέχει τα θεμέλια για τον ορισμό κλάσεων, ιδιοτήτων και σχέσεων μεταξύ τους, επιτρέποντας την ιεραρχική ταξινόμηση, τον προσδιορισμό ιδιοτήτων και τους περιορισμούς τομέα/περιοχής.

Ένα άλλο ισχυρό λεξιλόγιο είναι το Web Ontology Language (OWL), το οποίο παρέχει έναν πιο επίσημο και εκφραστικό τρόπο αναπαράστασης της γνώσης, επιτρέποντας τη διατύπωση περίπλοκων ιδεών, αξιωμάτων και λογικών επιχειρημάτων. Καθιστά δυνατή τη δημιουργία περίπλοκων οντολογιών που υποστηρίζουν πολύπλοκους συλλογισμούς και αυτοματοποιημένη εξαγωγή συμπερασμάτων, γεγονός που προάγει τις δυνατότητες αναπαράστασης γνώσης και εξαγωγής συμπερασμάτων.

Υπάρχουν διάφορα ειδικά λεξιλόγια για διάφορους τομείς και κλάδους, εκτός των RDFS και OWL. Ακολουθούν ορισμένα από τα πιο συνηθισμένα λεξιλόγια του Σημασιολογικού Ιστού:

* FOAF (Friend of a Friend):

Το FOAF είναι ένα λεξιλόγιο που χρησιμοποιείται για τον χαρακτηρισμό των ανθρώπων και των συνδέσεών τους. Δίνει ορισμούς για λέξεις που δηλώνουν προσωπικά δεδομένα όπως ονόματα, διευθύνσεις ηλεκτρονικού ταχυδρομείου και προφίλ στα μέσα κοινωνικής δικτύωσης. Το FOAF επιτρέπει την απεικόνιση των διαπροσωπικών δεσμών, συμπεριλαμβανομένων των φιλικών σχέσεων, των συνεργασιών και της ομαδικής εργασίας.

* SKOS (Simple Knowledge Organization System)

Το SKOS είναι μία οντολογία που επικεντρώνεται στην περιγραφή συστημάτων οργάνωσης γνώσης, όπως ταξινομίες (taxonomies), θησαυροί (thesauri), και συστήματα ταξινόμησης.

* Dublin Core

Για την περιγραφή των στοιχείων μεταδεδομένων των ψηφιακών πόρων, το Dublin Core είναι ένα ευρέως χρησιμοποιούμενο λεξιλόγιο. Παρέχει όρους για τη συλλογή θεμελιωδών λεπτομερειών σχετικά με τους πόρους, συμπεριλαμβανομένων των ονομάτων, των συγγραφέων, των θεμάτων και των ημερομηνιών.

* Schema.org

Σε μια προσπάθεια να παρέχουν ένα κοινό λεξιλόγιο για δομημένα δεδομένα στον ιστό, οι μεγάλες μηχανές αναζήτησης, συμπεριλαμβανομένων των Google, Microsoft και Yahoo!, συνεργάστηκαν για τη δημιουργία του Schema.org. Παρέχει μια μεγάλη ποικιλία ορολογίας για την κατηγοριοποίηση πραγμάτων όπως εταιρείες, αντικείμενα, γεγονότα και σχέσεις σε πολλούς διαφορετικούς τομείς. Η βελτίωση στην ποιότητα και την ποσότητα στα αποτελέσματα των μηχανών αναζήτησης καθίστανται δυνατή λόγω αυτού του λεξιλογίου.

- DBpedia

Μια δημοφιλής και εκτεταμένη οντολογία που αναπαριστά τη γνώση που λαμβάνεται από τη Βικιπαίδεια με οργανωμένο τρόπο. Προσφέρει ένα ευρύ φάσμα κλάσεων, χαρακτηριστικών και συνδέσεων που αντιπροσωπεύουν τα διάφορα πεδία και θέματα που καλύπτονται στα άρθρα της Wikipedia.

### SPARQL

Η SPARQL (SPARQL Protocol and RDF Query Language) είναι μια ισχυρή γλώσσα ερωτημάτων ειδικά σχεδιασμένη για την αναζήτηση δεδομένων RDF (Resource Description Framework) στον Σημασιολογικό Ιστό. Παρέχει ένα τυποποιημένο και εκφραστικό συντακτικό για την ανάκτηση και τον χειρισμό δεδομένων που είναι αποθηκευμένα αυτοτελώς ως RDF είτε προβάλλονται ως RDF μέσω ενδιάμεσου λογισμικού.

Τα ερωτήματα των SPARQL ακολουθούν μια παραπλήσια δομή σύνταξης όπως και τα απλά ερωτήματα SQL, δηλαδή ακολουθούν το μοτίβο SELECT-WHERE-FILTER. Πιο συγκεκριμένα, στη συνθήκη SELECT ορίζουμε μεταβλητές που θα επιστραφούν στα αποτελέσματα του ερωτήματος. Στο WHERE καθορίζονται τα μοτίβα και οι απαιτήσεις για την αντιστοίχιση τριπλών RDF στο σύνολο δεδομένων. Ενώ στην προαιρετική συνθήκη FILTER δίνονται οι απαιτήσεις για το φιλτράρισμα των αποτελεσμάτων βάσει συγκεκριμένων κριτηρίων. Βέβαια, η μόνη βασική διαφορά που αξίζει να αναφερθεί μεταξύ της σύνταξης της SPARQL και της SQL έγκειται στην αντιμετώπιση των προθεμάτων (prefix). Τα προθέματα χρησιμοποιούνται στην SPARQL για τον ορισμό συντομογραφιών των οντολογιών, καθιστώντας το ερώτημα συντομότερο και πιο κατανοητό. Οι χρήστες μπορούν να δώσουν σε ένα URI μιας οντολογίας-λεξικού ένα γρήγορο πρόθεμα χρησιμοποιώντας τη λέξη-κλειδί PREFIX και στη συνέχεια να χρησιμοποιήσουν αυτό το πρόθεμα για να αναφερθούν σε πόρους ή χαρακτηριστικά εντός αυτού του χώρου ονομάτων σε όλη τη διάρκεια του ερωτήματος.

Η SPARQL διαδραματίζει πρωταγωνιστικό ρόλο στο Apache Fuseki, όπως έιδαμε και νωρίτερα. Στο σύνολο δεδομένων RDF που είναι αποθηκευμένα στο Fuseki, οι χρήστες μπορούν να δημιουργήσουν ερωτήματα SPARQL για να αποκτήσουν τα απαραίτητα δεδομένα και να εκτελέσουν περίπλοκες λειτουργίες. Οι χρήστες μπορούν να εξερευνήσουν τον γράφο του RDF, να φιλτράρουν τα αποτελέσματα και να αθροίσουν τα δεδομένα σύμφωνα με τις μοναδικές τους ανάγκες χάρη στην εκφραστική φύση της SPARQL.

## Ανοικτά Κυβερνητικά Δεδομένα – Akoma Ntoso

Τα διασυνδεδεμένα ανοικτά κυβερνητικά δεδομένα, σε συντομογραφία LOGD, είναι μια «ιδέα» που προωθεί την δημοσιοποίηση, τη διαλειτουργικότητα και την ενσωμάτωση των κυβερνητικών δεδομένων στο ανοικτό διαδίκτυο. Εφαρμόζει τις αρχές των Συνδεδεμένων Ανοικτών Δεδομένων στο πεδίο της κυβέρνησης, αποσκοπώντας να μεγιστοποιήσει το πλέγμα των ανοικτών δεδομένων και να προωθήσει την καινοτομία, τη συνεργασία και τη διαφάνεια.

Τα LOGD περιλαμβάνει την κοινή χρήση δημόσιων δεδομένων σε ανοικτούς, τυποποιημένους μορφότυπους, τηρώντας παράλληλα τις αρχές των Συνδεδεμένων Δεδομένων. Προκειμένου να προσδιοριστούν οι πόροι και οι σχέσεις, τα δεδομένα διατίθενται με τη χρήση RDF (Resource Description Framework), το οποίο χρησιμοποιεί μοναδικά αναγνωριστικά (URIs). Τα κυβερνητικά δεδομένα καθίστανται διασυνδεδεμένα όταν ακολουθούνται αυτά τα πρότυπα, καθιστώντας απλή τη σύνδεση και την ανάμειξή τους με άλλα σύνολα δεδομένων για τη δημιουργία ενός πλούσιου ιστού διασυνδεδεμένης γνώσης.

Η υιοθέτηση του LOGD έχει μια σειρά από πλεονεκτήματα. Πρώτα απ' όλα, διευκολύνει την πρόσβαση και τη χρήση κυβερνητικών δεδομένων από πολίτες, ακαδημαϊκούς, εταιρείες και πολιτικούς. Παρέχοντας ανοικτά και οργανωμένα δεδομένα, το LOGD επιτρέπει σε άτομα και οργανισμούς να εξετάζουν, να αξιολογούν και να εξάγουν συμπεράσματα από τα δημόσια δεδομένα.

Δεύτερον, το LOGD ενθαρρύνει τη λογοδοσία και την διαφάνεια στις κυβερνητικές δραστηριότητες. Οι κυβερνήσεις μπορούν να καλιεργήσουν την εμπιστοσύνη και να δώσουν τη δυνατότητα στους πολίτες να εξετάζουν και να επικυρώνουν τα γεγονότα καθιστώντας τα δεδομένα άμεσα διαθέσιμα. Το LOGD διευκολύνει την παρακολούθηση των δημόσιων υπηρεσιών, την ενθάρρυνση της συμμετοχής των πολιτών και την καλύτερη κατανόηση των κυβερνητικών διαδικασιών και νόμων.

Επιπλέον, η LOGD προωθεί τη συνεργασία και την ανταλλαγή πληροφοριών μεταξύ των κυβερνήσεων και σε πολλούς τομείς. Τα κυβερνητικά δεδομένα μπορούν να συνδεθούν με άλλα σύνολα δεδομένων για την αποκάλυψη νέων σχέσεων και γνώσεων που επιτρέπουν τη διατομεακή ανάλυση και τη δημιουργία δημιουργικών απαντήσεων στα κοινωνικά προβλήματα.

Κυβερνήσεις σε όλο τον κόσμο έχουν αναλάβει την «ανοικτή» δημοσίευση των τοπικών τους κυβερνητικών αρχείων, παρέχοντας ποικίλες πληροφορίες, συμπεριλαμβανομένων στατιστικών στοιχείων για τον πληθυσμό, τις δαπάνες, τους νόμους και το περιβάλλον. Οι χρήστες μπορούν να έχουν πρόσβαση στα δεδομένα και να τα αναζητούν με τυποποιημένο και διαλειτουργικό τρόπο χάρη στις εξειδικευμένες πύλες, τα API και τα SPARQL endpoints μέσω των οποίων διατίθενται αυτά τα σύνολα δεδομένων.

Η αναπαράσταση και η διαχείριση νομοθετικών και νομικών εγγράφων επιτυγχάνεται χάρη στο Akoma Ntoso, επίσης γνωστό ως LegalDocML, το οποίο είναι ένα πρότυπο βασισμένο στην XML. Το Akoma Ntoso δημιουργεί έναν δομημένο μορφότυπο για την απεικόνιση νομικών εγγράφων που υποστηρίζει τη διαλειτουργικότητα, την προσβασιμότητα και τη σημασιολογία, βασιζόμενο στις ιδέες των Συνδεδεμένων Ανοικτών Δεδομένων.

Πιο συγκεκρίμενα, το Akoma Ntoso παρουσιάζει μια ενδελεχή συλλογή στοιχείων και χαρακτηριστικών που είναι ειδικά σχεδιασμένα για την αποτύπωση της μορφής και του περιεχομένου των νομοθετικών κειμένων. Περιλαμβάνει ένα εύρος χαρακτηριστικών που παρατηρούνται σε νομικά κείμενα, όπως επικεφαλίδες, υποκεφαλίδες, άρθρα, παραγράφους και παραπομπές. Τα νομικά έγγραφα μπορούν να αναπαρασταθούν με ομοιόμορφο και συνεπή τρόπο συμμορφούμενα με το πρότυπο Akoma Ntoso, απλοποιώντας την ανταλλαγή, την ανάλυση και την επαναχρησιμοποίησή τους.

Πρόσθετα οφέλη προκύπτουν από την ενσωμάτωση του Akoma Ntoso με τα Συνδεδεμένα Ανοικτά Κυβερνητικά Δεδομένα. Το Akoma Ntoso καθιστά δυνατή τη διασύνδεση νομικών εγγράφων με άλλα συναφή δεδομένα και σύνολα δεδομένων, αξιοποιώντας τις έννοιες των Συνδεδεμένων Δεδομένων. Με τη χρήση αυτής της δυνατότητας, τα νομικά κείμενα μπορούν να ενσωματωθούν απρόσκοπτα με μεταδεδομένα, δικαστικές αποφάσεις, κανόνες και άλλες σχετικές πηγές δεδομένων, δημιουργώντας ένα πιο λεπτομερές και ολοκληρωμένο γράφημα νομικής γνώσης.

Το Akoma Ntoso χρησιμοποιείται για τη βελτίωση της προσβασιμότητας και της μηχανικής αναγνωσιμότητας των νομικών κειμένων. Η προηγμένη αναζήτηση, η σημασιολογική ανάλυση και η αυτοματοποιημένη επεξεργασία των νομοθετικών δεδομένων καθίστανται δυνατές χάρη στη δομημένη αναπαράσταση των νομικών κειμένων στο Akoma Ntoso. Εκτός από τη δυνατότητα ταχύτερης πρόσβασης σε σχετικές πληροφορίες και τη διευκόλυνση της ανάπτυξης νομικών εφαρμογών και συστημάτων υποστήριξης αποφάσεων, αυτό ενθαρρύνει τη διαφάνεια, την αποτελεσματικότητα και την ομοιομορφία των νομικών διαδικασιών.

Κυβερνήσεις, δικαστήρια και άλλοι οργανισμοί σε όλο τον κόσμο έχουν υιοθετήσει ευρέως το Akoma Ntoso ως πρότυπο για την παρουσίαση νομοθετικών κειμένων. Η χρήση του διασφαλίζει ότι οι νομικές πληροφορίες είναι τυποποιημένες και διαλειτουργικές, προωθώντας τη διεθνή συνεργασία, τη συγκριτική νομική έρευνα και τη δημιουργία νομικών οντολογιών και βάσεων γνώσης.

# Σύστημα Διαχείρισης Πρακτικών

Εδώ λέμε ότι θα ακολουθήσει η περιγραφή της αρχιτεκτονικής του συστήματος και θα γίνει η ανάλυση απαιτήσεων για τις λειτουργίες του.

## Αρχιτεκτονική Συστήματος

Εδώ παρουσιάζουμε τα επιμέρους κομμάτια (υποσυστήματα) από τα οποία θεωρούμε ότι αποτελείται το σύστημά μας. Για κάθε υποσύστημα δίνεται μια σύντομη περιγραφή (η αναλυτική περιγραφή των λειτουργιών του ακολουθεί στην επόμενη ενότητα). Τέλος, δίνουμε ένα γενικό σχήμα που δείχνει τα υποσυστήματα και πώς αυτά επικοινωνούν μεταξύ τους. Το σχήμα αυτό αρκεί να είναι απλό block diagram και όχι Διάγραμμα Ροής Δεδομένων (DFD).

## Μέθοδος Αντλήσης Αρχείων - Βάση Δεδομένων

Εδώ

## Εξαγωγή και Ανάλυση Κειμένων

### ANLTR4 - REGEΧ

## Διαχείριση Δεδομένων

### Μετατροπή Αρχείων Κειμένου σε XML αρχεία

### Μετατροπή Αρχείων XML σε RDF αρχεία

### Επεξεργασία RDF - SPARQL

# Έλεγχος Συστήματος

Εδώ λέμε ότι θα ακολουθήσει ο έλεγχος του συστήματος.

## Μεθοδολογία ελέγχου

….

## Αναλυτική παρουσίαση ελέγχου/παραδείγματος

…

# Επίλογος

Εδώ λέμε ότι θα συνοψίσουμε την παρουσίαση της διπλωματικής.

## Σύνοψη και συμπεράσματα

Εδώ συνοψίζουμε τα αποτελέσματα της διπλωματικής και περιγράφουμε τα συμπεράσματα που προέκυψαν, αρνητικά και θετικά. Επιβεβαιώνουμε τη συνεισφορά της διπλωματικής στα προβλήματα που αναφέραμε στην εισαγωγή.

## Μελλοντικές επεκτάσεις

Εδώ δίνουμε ιδέες για επέκταση της διπλωματικής.

# Βιβλιογραφία

|  |  |
| --- | --- |
| [BBC+99] | P.A. Bernstein, Th. Bergstraesser, J. Carlson, S. Pal, P. Sanders, D. Shutt. Microsoft Repository Version 2 and the Open Information Model. To appear in Information Systems 24(2), 1999. |
| [BCR94] | V. R. Basili, G.Caldiera, H. D. Rombach. The Goal Question Metric Approach. Encyclopedia of Software Engineering - 2 Volume Set, pp. 528-532, John Wiley & Sons, Inc., available at http://www.cs.umd.edu/users/basili/papers.html, 1994 |
| [Dea97] | E. B. Dean, "Quality Functional Deployment from the Perspective of Competitive Advantage", available at http://mijuno.larc.nasa.gov/dfc/qfd.html |
| [JJQV98] | M. Jarke, M.A.Jeusfeld, C. Quix, P. Vassiliadis: Architecture and quality in data warehouses, Proceedings CΑiSE 98, Pisa, Italy, 1998. |
| [JV97] | M. Jarke, Y. Vassiliou. Foundations of data warehouse quality – a review of the DWQ project. In Proc. 2nd Intl. Conference Information Quality (IQ-97), Cambridge, Mass., 1997. |
| [Orr98] | K. Orr. Data quality and systems theory. In Communications of the ACM, 41, 2, pp. 54-57, Feb. 1998. |