

Σημασιολογική επισήμανση τεκμηρίων

Θοδωρής Κοτσιλιέρης

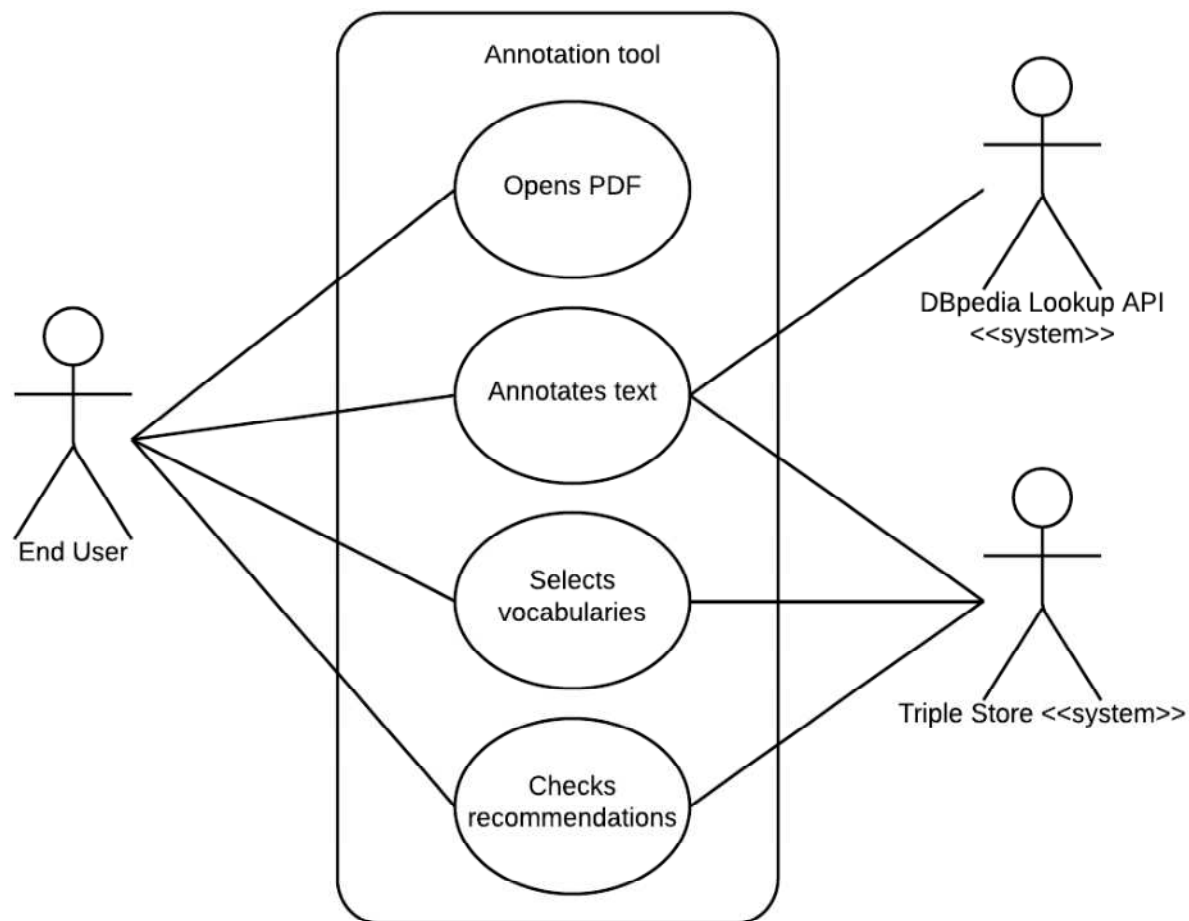
Επίκουρος Καθηγητής ΤΕΙ Πελοποννήσου

Το κίνητρο

Ένα εργαλείο σημασιολογικής επισήμανσης με:

- Δυνατότητα σημασιολογικού σχολιασμού κειμένου σε έγγραφα PDF
- Υποστήριξη πολλαπλών οντολογιών
- Υποστήριξη σύνθετων λεξιλογίων με ιδιότητες
- Εφαρμογή μιας λειτουργικότητα πρότασης παρόμοιων κειμένων
- Υποστήριξη πολλαπλών χρηστών
- Συμμόρφωση με τις απαιτήσεις των Linked Data

Τα βασικά σενάρια λειτουργίας



Άνοιγμα PDF αρχείου



JavaScript Instrumentation for Browser Security

Dachuan Yu Ajay Chander Nayeem Islam Igor Serikov

DoCoMo Communications Laboratories USA, Inc.

{yu,chander,nayeem,iserikov}@docomolabs-usa.com

Abstract

It is well recognized that JavaScript can be exploited to launch browser-based security attacks. We propose to battle such attacks using program instrumentation. Untrusted JavaScript code goes

Aiming to thwart various attacks launched through JavaScript, we propose to regulate the behavior of untrusted JavaScript code using program instrumentation. Incoming script goes through a rewriting process which identifies and modifies questionable operations. At runtime, the modified script will perform necessary

Επισήμανση κειμένου

- Η διαδικασία του σημασιολογικού σχολιασμού είναι η ακόλουθη:
 1. Ο χρήστης επιλέγει κάποιο κείμενο από το έγγραφο PDF
 2. Ο χρήστης προσθέτει σημασιολογικό εμπλουτισμό στο σχολιασμό
 3. Ο χρήστης κάνει κλικ στο "Προσθήκη επισήμανσης"
- Παραλλαγές στο βήμα σημασιολογικού εμπλουτισμού εξαρτώνται από:
 - Το είδος των σχολιασμών ο χρήστης θέλει να εισάγει.
 - i. Απλή επισήμανση, δηλαδή σχολιασμοί που αποτελούν στιγμιότυπα κάποιας κλάσης. Αυτή είναι η βασική περίπτωση χρήσης σχολιασμού για τους λόγους ταξινόμησης.
 - ii. Σύνθετες επισημάνσεις, δηλαδή σχολιασμός που αποτελεί μέρος μιας σχέσης.
 - Τι είδους σημασιολογικής σήμανσης ο χρήστης εφαρμόζει στο σχολιασμό:
 - i. πόρους της DBpedia
 - ii. κατηγορίες

Παράδειγμα επισήμανσης κειμένου

The image shows a two-part screenshot of a text annotation interface. The left part shows a search for a subject, with 'Resource Description Framework' highlighted in a red box. The right part shows the same interface applied to a document snippet, with a red callout box explaining the annotation process.

Left Panel (Search):

- Text: "s of the extractio... cessing, and find... cts. This allows u... papers in XML."
- Text: "F, RDF graphs"
- Search input: "Subject"
- Question: "Is this the same as any of the below?"
- Result: **Resource Description Framework** (highlighted in red)
- Result description: "The Resource Description Framework (RDF) is a family of World Wide W... has come to be used as a general method for conceptual description or... syntax formats."
- Other result: "Zodiak Media Group [agent, organisation, company]"

Right Panel (Annotation):

- Text: "the Semantic Proc... tific Papers"
- Text: "n and Iana Atanassova"
- Text: "sité du Québec à Montréal... 8, Succ. Centre-ville... ntreal (Quebec) Canada... m.ca, iana.atanassova@nlp-"
- Text: "rid method for the extraction :... tific papers using machine lea... es. Our protocol consists of the"
- Annotation form: "author" (selected), "Property", "Object"
- Value: "http://eis.lai.uni-bonn.de/semann/0.2/sdeo#Author" (highlighted in red)
- Button: "Add annotation"
- Visual representation: "sdeo#Author" in a grey oval connected to the text "Iana Atanassova" in the document.

Red Callout Box:

Specifying an annotation "Iana Atanassova" to be an instance of http://eis.lai.uni-bonn.de/semann/0.2/sdeo#Author

Επιλογή Λεξιλογίου

- Η SemAnn Discourse Elements Ontology φορτώνεται στο εργαλείο
- Όμως ο χρήστης μπορεί να καθορίσει οποιοδήποτε άλλο λεξιλόγιο



Text Annotations Table Annotations Vocabulary Selection

Semann Discourse Elements Ontology

http:// Add vocabulary



Text Annotations Table Annotations Vocabulary Selection

Semann Discourse Elements Ontology
 http://xmlns.com/foaf/0.1

http:// Add vocabulary

Ontology successfully loaded with 631 triples!

Έλεγχος συστάσεων

- Οι προτάσεις για παρόμοια έγγραφα εμφανίζονται ως αποτέλεσμα της σημασιολογικής σύγκρισης σχολιασμών του εγγράφου με εκείνους όλων των άλλων εγγράφων στη βάση δεδομένων.
 - i. Ο χρήστης κάνει κλικ στο "Find Similar" κουμπί
 - ii. Οι συστάσεις για άλλα παρόμοια έγγραφα που εμφανίζονται με επεξηγήσεις

Recent developments in vacuum arc deposition - Plasma Science, IEEE Transactions on - no.157, Recent Developments in Vacuum Arc Deposition.pdf

Shares same category [Coatings](http://dbpedia.org/resource/Physical_vapor_deposition_paper) (here: http://dbpedia.org/resource/Physical_vapor_deposition_paper; <http://dbpedia.org/resource/Coating>). **Abstract** common context of the match

Shares same category [Coatings](http://dbpedia.org/resource/Coating) (here: <http://dbpedia.org/resource/Coating>, paper: http://dbpedia.org/resource/Cathodic_arc_deposition).

Shares same category [Coatings](http://dbpedia.org/resource/Coating) (here: <http://dbpedia.org/resource/Coating>, paper: http://dbpedia.org/resource/Diamond-like_carbon).

Shares same category [Semiconductor device fabrication](http://dbpedia.org/resource/Semiconductor_device_fabrication) (here: http://dbpedia.org/resource/Semiconductor_device_fabrication, paper: http://dbpedia.org/resource/Physical_vapor_deposition_paper; [http://dbpedia.org/resource/Evaporation_\(deposition\)](http://dbpedia.org/resource/Evaporation_(deposition))).

matches based on common supercategories of dbpedia resurces used in the current paper and suggested paper

Shares same category [Thin film deposition](http://dbpedia.org/resource/Thin_film_deposition) (here: http://dbpedia.org/resource/Thin_film_deposition, paper: http://dbpedia.org/resource/Physical_vapor_deposition_paper; http://dbpedia.org/resource/Vacuum_deposition).

Shares same category [Carbides](http://dbpedia.org/resource/Carbides) (here: <http://dbpedia.org/resource/Carbides>, paper: http://dbpedia.org/resource/Zirconium_carbide, paper: <http://dbpedia.org/resource/Carbide>).

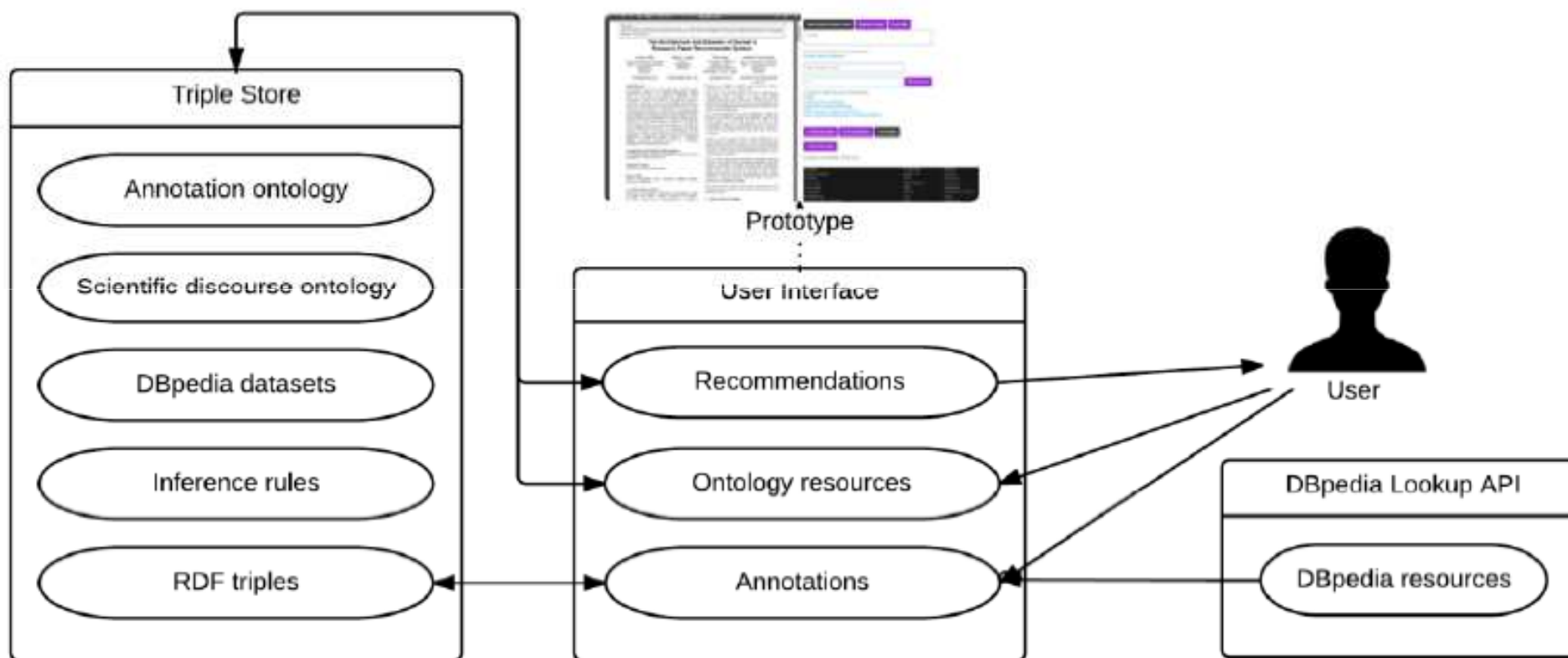
The Architecture and Datasets of Docear's Research Paper Recommender System - The Architecture and Datasets of Docear's Research Paper Recommender System.pdf

Mentions [Marketing](http://dbpedia.org/resource/Marketing). exact match of a dbpedia resource

Περιορισμοί

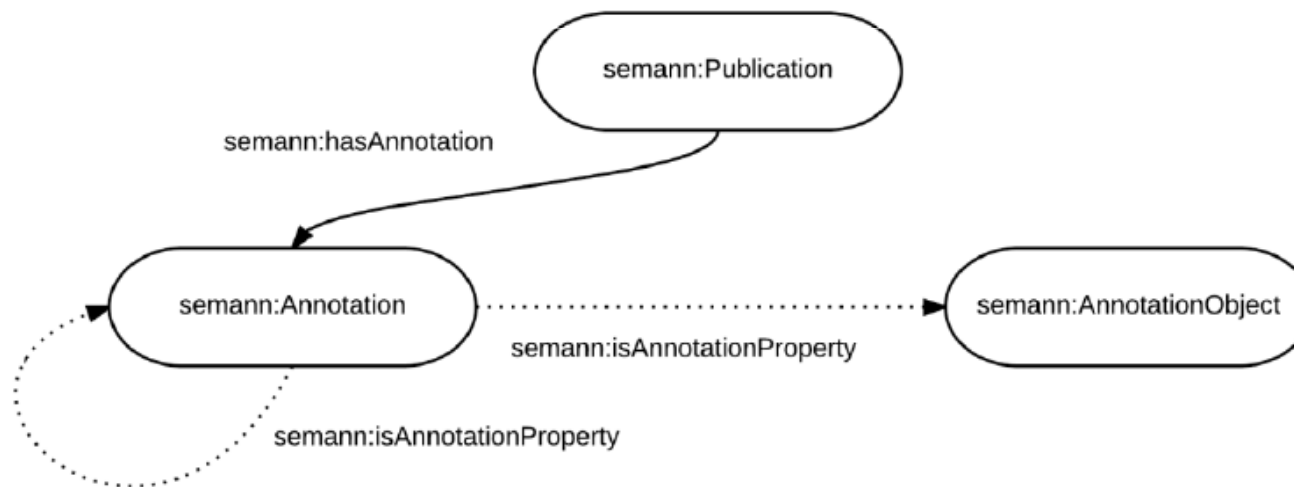
- Χρησιμοποιεί τοπικό `Virtuoso server` κι όχι κάποιον κεντρικό
- Δεν επιτρέπει τη μεταφόρτωση PDF αρχείων από άλλο `server`
- Η ανάλυση και εμφάνιση (`rendering`) των PDF αρχείων δεν είναι πλήρης λόγω περιορισμών του `PDF.js` – Ορισμένα χαρακτηριστικά του `PDF format` δεν υποστηρίζονται

Αρχιτεκτονική



Χρησιμοποιούμενες Οντολογίες


- SemAnn Annotation Ontology –
 - Μοντελοποιεί την πληροφορία που σχετίζεται με τους σχολιασμούς



Χρησιμοποιούμενες Οντολογίες

- SemAnn Annotation Ontology

<http://eis.iai.uni-bonn.de/semann/pdf/example.pdf#page=1?char=74,87&id=0/20/1/1:0,0/20/1/1:13>

 Crowd-sourced Open Courseware

```
<rdf:Description rdf:about="http://eis.iai.uni-  
bonn.de/semann/pdf/exmple.pdf#page=1?char=74,87&id=0/20/1/1:0,0/20/1/1:13">  
  <rdf:type rdf:resource="http://dbpedia.org/resource/Crowdsourcing" />  
  <rdf:type rdf:resource="http://eis.iai.uni-bonn.de/semann/0.2/owl#Annotation" />  
  <rdfs:label xml:lang="en">Crowd-sourced</rdfs:label>  
</rdf:Description>  
<rdf:Description rdf:about="http://eis.iai.uni-bonn.de/semann/pdf/example.pdf">  
  <rdf:type rdf:resource="http://eis.iai.uni-bonn.de/semann/0.2/owl#Publication" />  
  <semann:hasAnnotation rdf:resource="http://eis.iai.uni-  
bonn.de/semann/pdf/example.pdf#page=1?char=74,87&id=0/20/1/1:0,0/20/1/1:13" />  
  <rdfs:label xml:lang="en">Example Document</rdfs:label>  
</rdf:Description>
```

Χρησιμοποιούμενες Οντολογίες

- SemAnn Annotation Ontology

<A.1> <http://purl.org/spar/cito:disagrees_with> <B.1>

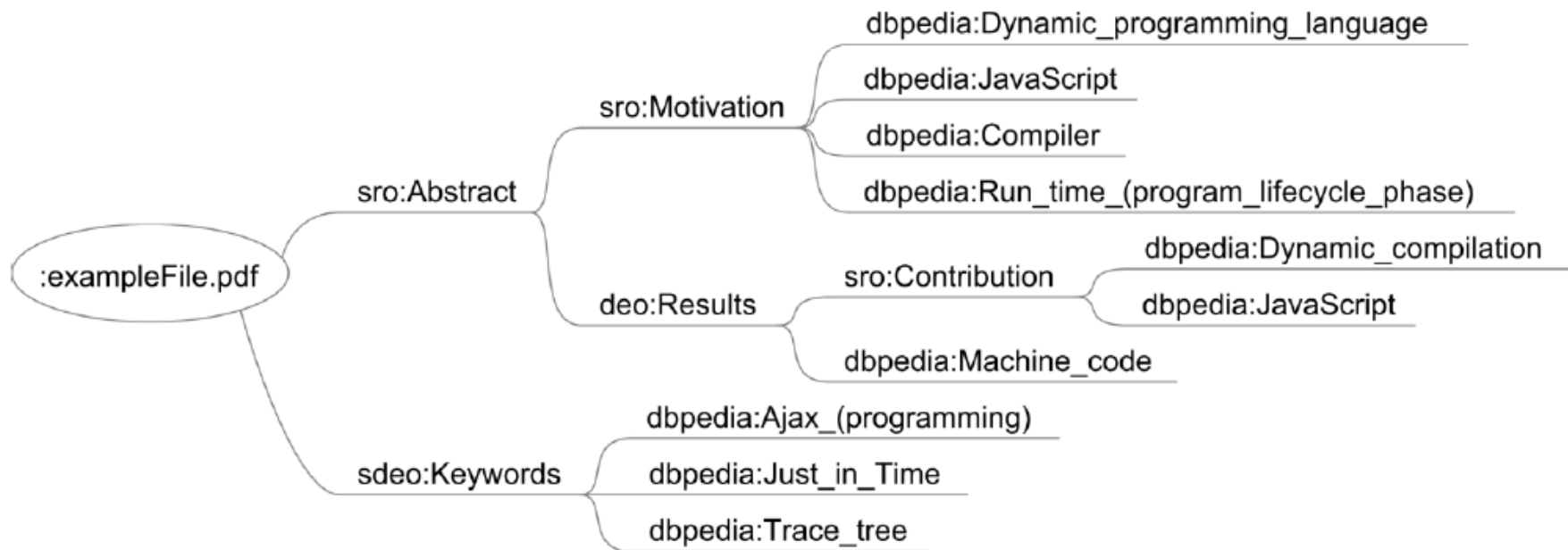
<A.2> <<http://www.essepuntato.it/2011/02/argumentmodel:proves>> <B.2>

<A.3> <<http://purl.org/swan/1.2/swan-commons/citesAsSupportiveEvidence>> <B.3>

<A.4> <<http://purl.org/spar/cito:plagiarizes>> <B.4>

Χρησιμοποιούμενες Οντολογίες

- SemAnn Discourse Elements Ontology (SDEO) –
 - Επεκτείνει την Discourse Elements Ontology (DEO), την οντολογία για την περιγραφή δομικά συστατικά μιας ακαδημαϊκής εργασίας (κι αποτελεί μέρος της SPAR - Semantic Publishing and Referencing Ontologies)



Χρησιμοποιούμενες Οντολογίες

- SemAnn Discourse Elements Ontology (SDEO) –

