Semantic Web

Introductory Course

Maria-Iuliana Dascalu, PhD

mariaiuliana.dascalu@gmail.com

University POLITEHNICA of Bucharest
Department of Engineering in Foreign Languages

Agenda

- Course objectives
- Course logistics
- References

- What is World Wide Web (WWW)?
- What is Semantic Web?

Course Objectives

- Main features of the Semantic Web
- XML - Extensible Markup Language & related
- RDF - Resource Description Framework
- OWL - Web Ontology Language
- SPARQL
- Triplestores, graph databases
- JSON-LD
- Microformats
- Semantic search - understanding a searcher's intent through contextual meaning
- Discussions about social web, software agents and semantic web services
- …

Course Logistics

- 5 ECTS
- 2 hours of lectures per week and 2 hour of application per week

- **Grading**: E= E1+E2+E3+E4+E5
  - E1 = 1 written exam in the last week (16th of January): 30%
  - E2 = 1 presentation made by 2 students of 10-15 minutes on a given subject during course : 10%
    https://docs.google.com/spreadsheets/d/1m8JMxegUsHy_ZVxe31WrG0e46gIBFcXq_LiltChpdbs/edit#gid=0 (during 28th Nov, 5th Dec, 12th Dec or 9th Jan)
  - E3 = presence, activity and small homeworks at the lab: 30%
  - E4 = 1 project related to sem. web technologies submitted and presented in 2 phases (10%+20%), in teams of 2 students
  - The projects will be proposed on: 21st of October
    - deadline first phase 23rd of Nov on Moodle and 25th of Nov presentation during lab;
    - deadline second phase 11th of Jan on Moodle and 13th of Jan presentation during lab (the course prof will be also present)
  - E5 =Extra points can be obtained – up to 2 points for exceptional academic activities (e.g. 0.5 points for the best presentation/best project; 2 points for writing an article to be submitted at a conference)
- Conditions for passing:
  - Minimum 6 presences at the course
  - E1 >= 4.5
  - E2 > =4.5
  - E3 >= 4.5 and minimum 6 presences at the lab
  - E4 >= 4.5
  - E >= 4.5

- If problems, please send an e-mail to mariaiuliana.dascalu@gmail.com , with the following subject: [SeWeb] [NAME] [PROBLEM]
- Lab assistant: Iuliana Marin (marin_iulliana@yahoo.com), Ioan Bratosin (ioan.bratosin@gmail.com)

References

- Course Materials
  - First lecture notes: http://mariaiulianadascalu.com/teaching/ (all materials are protected via a password: **ondina**)
  - Moodle
  - http://www.w3schools.com/
  - https://jena.apache.org/
  - http://protege.stanford.edu/
  - Maria-Iuliana Dascalu, Iuliana Marin, Semantic Web: Theory and Applications, Editura Printech (COD CNCSIS: 54), 212 pagini, 2016, ISBN 978-606-23-0705-9
  - Semantic Web in 1 hour, Salih Ismail, 2016
  - Grigoris Antoniou, Frank Van Harmelen - Semantic Web Primer, MIT Press, 2004
  - Dean Allemang, James Hendler - Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL, 2nd edition, Elsevier, 2011
  - Gaševic, Dragan, Djuric, Dragan, Devedžic, Vladan - Model Driven Engineering and Ontology Development, 2nd edition, Springer, 2009
  - Michael C. Daconta, Leo J. Obrst, Kevin T. Smith - The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management, Wiley, 2003

The World Wide Web

- Everything started with Sir Tim Berners-Lee. (CERN, 1989)
- He wanted a common program which all the researchers could use to share their work.
- He built the HTML and URL.
- The ARPANET community expanded the WWW.
- Invention of MOSAIC (Navigator) contributed in the world wide adoption of WWW.

WWW (1)

- The **World Wide Web** is a system of interlinked hypertext documents accessed via the Internet
- WWW <> Internet (global system of interconnected computer networks)
- Based on the client/server model and hypertext
- CERN – 1989, Sir Tim Berners-Lee *et al.*, "*Information Management: A Proposal*", **http://www.w3.org/History/1989/proposal.html**
- Works according to the standards of w3 Consortium (www.w3.org)

WWW (2)

- Idea: integration of disparate information systems in a unified way, with no differences between data sources

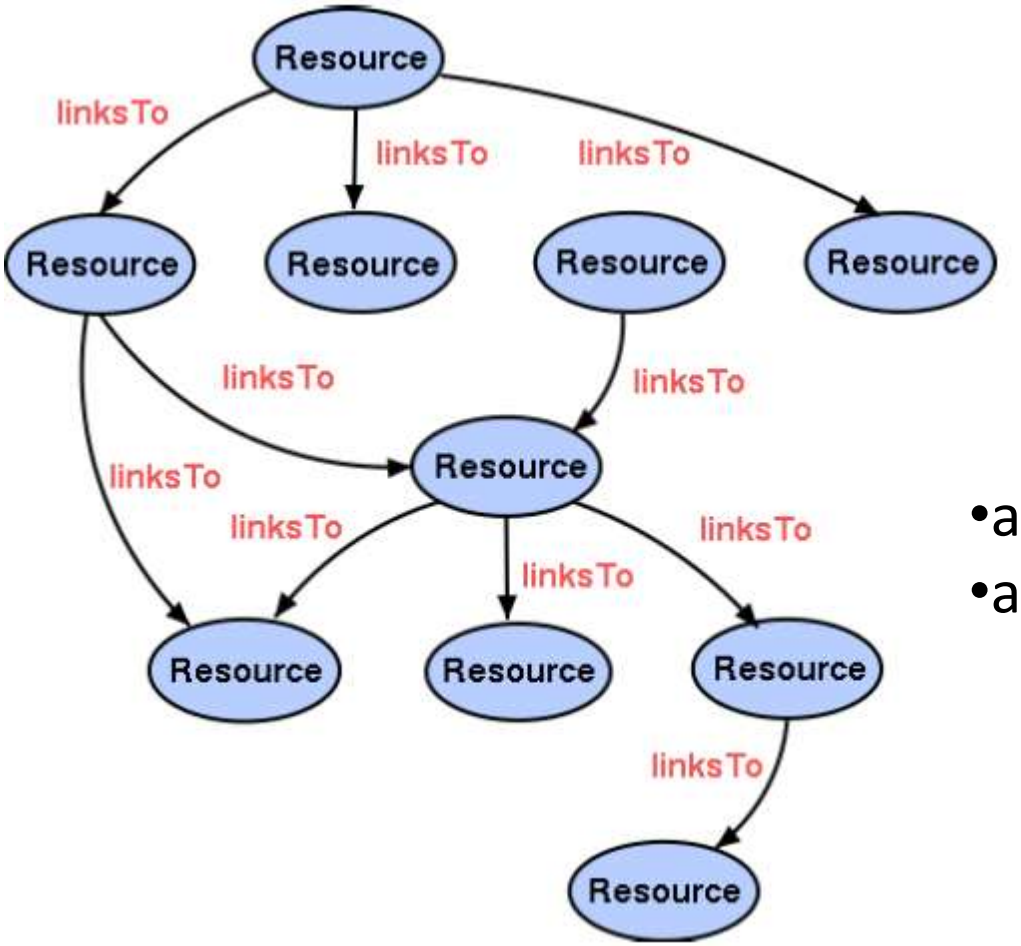
- Main objectives (Tim Berners-Lee):
  - Device independence
  - Software independence
  - Scalability
  - Multimedia

- *"Anything can link to anything"*

WWW (3)

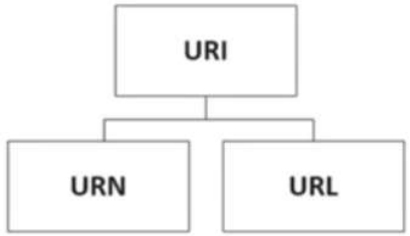
- Resources are identified by their address – uniform resource identifier (URI) (RFC 2396, 3986)
  - allows a uniform semantic interpretation of common syntactic conventions
- There is a protocol (e.g. HTTP) to access the content of resources
- Resources are seen as documents (web pages) and include mark-ups (annotations)
- Mark-ups contain URIs
- See for details **www.w3.org/TR/webarch/**

Web Architecture



- an URI identifies a resource
- a resource is represented from
  - data
  - metadata

URI, URL and URN



- URI – Uniform Resource Identifier: URI is a text which is used to identify any resource or name on Internet.
- URL – Uniform Resource Locator: URL includes location as well as the protocol to retrieve the resource (For example :HTTP Protocol is used to retrieve resource)

  http://example.com/mypage.html
  ftp://example.com/download.zip
  mailto:user@example.com
  file:///home/user/file.txt
- URN – Uniform Resource Name: Identifies a resource by a unique and persistent name, but doesn't necessarily tell you how to locate it on the internet. It usually starts with the prefix urn: For example:

  urn:isbn:9781921987564 to identify a book by its ISBN number.

Classic Web

- www=space for stocking & representing data, via HTML, CSS…
- Problem:
  - I send: <grade student="33">10</grade>
  - You want: <points id="33">A</points>
- Conclusion:
  - computers have to be capable of processing data in an intelligent way
  - need for new models of knowledge representation

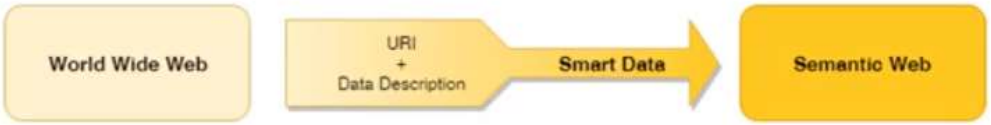
Growth and Criticism of WWW

- Network Effect – Exponential Growth
- 50 billion web pages / 1.12 billion websites.
- Data owned by individuals and corporations.
- Need for Open data and Linked Data.
- Criticism of WWW:
  - Content creators
  - Content users
  - Heterogeneity – AAA slogan

AAA slogan: Anyone can say Anything about Any topic

Rise of Artificial Intelligence

- Dumb Data – It needs to be linked and needs description on it own.
- Creation of Smart Data.



- Either create data that is machine understandable or transform existing data.
- Need for semantic web at this point is non-negotiable.

Semantic Web

- Links between resources have semantics and can be extended
- Resources can be extended and classified using conceptual specifications
- At the programmatic level, there are entities that are able to process information and to reason intelligently, giving complex services to users / machines
- Users can share knowledge regardless of their storage / representation

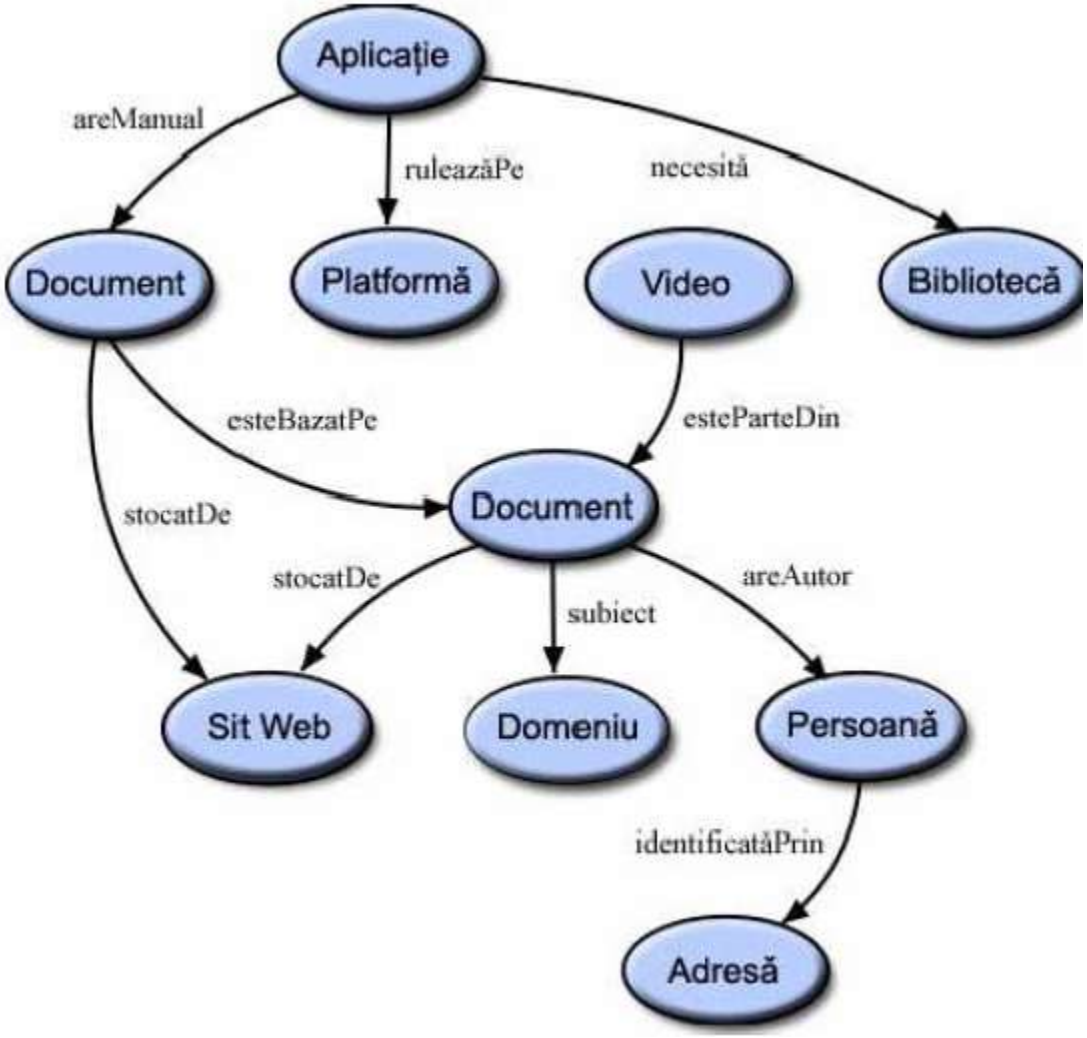
Definition

- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

(Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001)

- SW meets AI
- SW meets personalization.
- SW meets Computer vision.
- SW meets …

Semantic Web Architecture



- Nodes and links have semantic meanings

"The Semantic Web will enable machines to comprehend semantic documents and data, not human speech and writings." (Tim Berners-Lee, 2001)

https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

Characteristics of Semantic Web

- Main interest area: knowledge management
- Objectives:
  - Information retrieval
  - Data mining & relation mining
  - Maintenance of knowledge repositories
  - Automated generation of documents
- Has several levels
- For each level, the Web Consortium proposes dedicated languages, based on XML

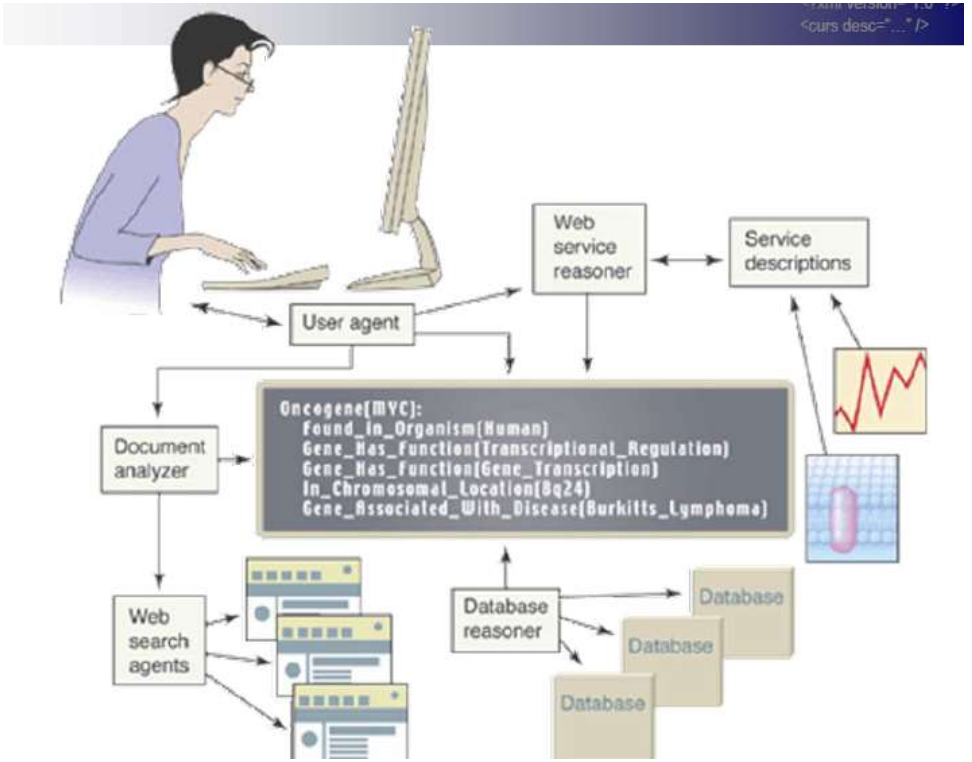
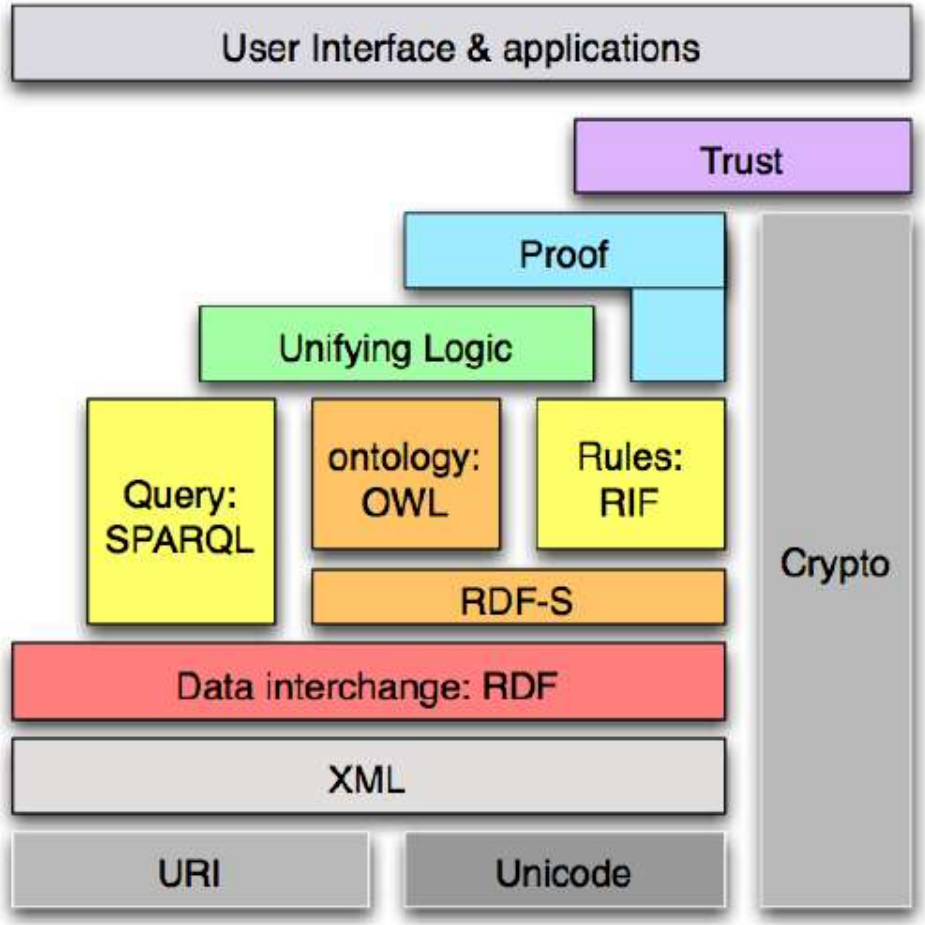


Image by Jim Hendler

Semantic Web Levels (Berners-Lee, 2006)



Opportunities for…

- Ambient intelligence
- Human-computer interaction, computer – computer interaction
- Bioinformatics & Computational Biology
- Grid computing & semantic Grid computing
- E-business
- E-learning
- Simulations
- GIS – Geographical Information Systems

…..

Challenges of Semantic Web (1)

- Providing a semantic infrastructure
  - Ontologies=> rules=> inferences=>reliability
  - Architectures for service integration (agents)
  - Broker architecture (semantic)
  - Semantic Web Support in OS
  - Semantic desktop
- Offering a “Web of trust”
  - Security technologies
  - Dedicated infrastructure models for confidence (trust) and reputation (reputation)
  - e-government
  - Interaction protocols

Challenges of Semantic Web (2)

- Increasing the quality of software in industry
  - Methodologies, tools and development environments which are agent-oriented, service-oriented (SOA), grid-oriented….
  - Integration and interoperability with existing software (legacy)
- Standardization
- Interaction with the user
  - Tools for real user profiling - understanding the user needs
  - Support for customizing applications
  - Knowledge acquisition tools
  - Social networking

Semantic Web “Bricks”

- XML – XML- eXtensible Markup Language
- RDF – Resource Description Framework
- OWL – Web Ontology Language

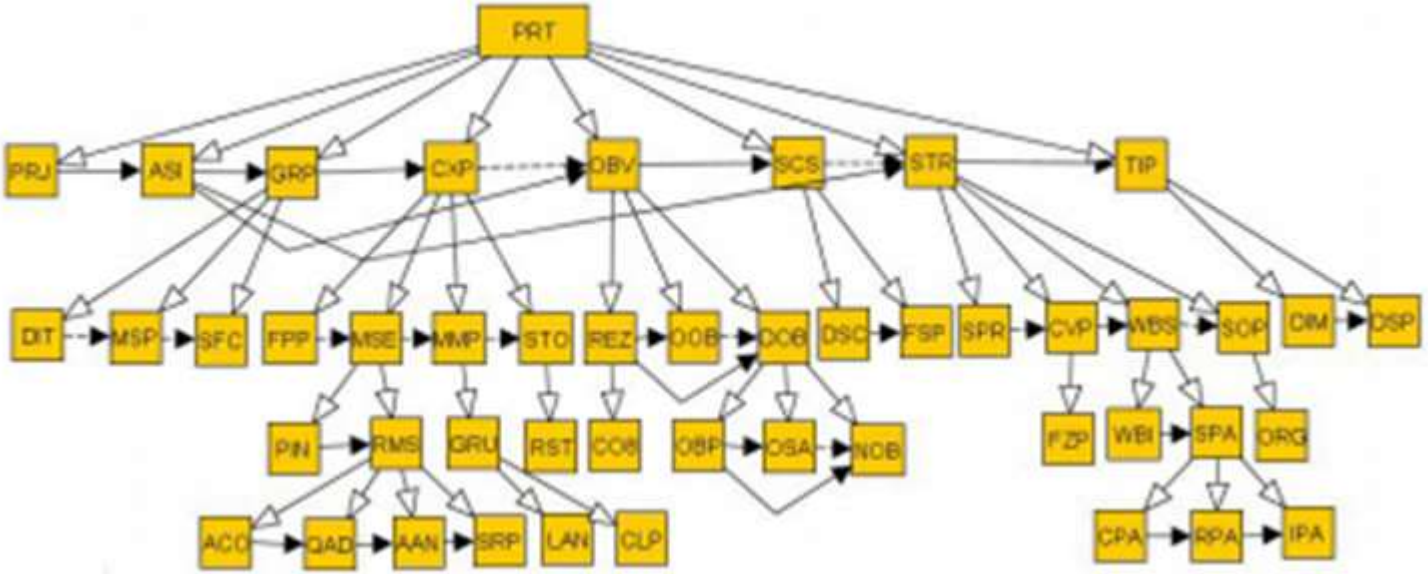
- way of expressing and storing metadata
- way to "structure" and describe terms
- way to "explain" resources to allow **automated reasoning**

A Quick Insight on Ontologies

- In computer science and information science, an **ontology** formally represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts.

Application 1

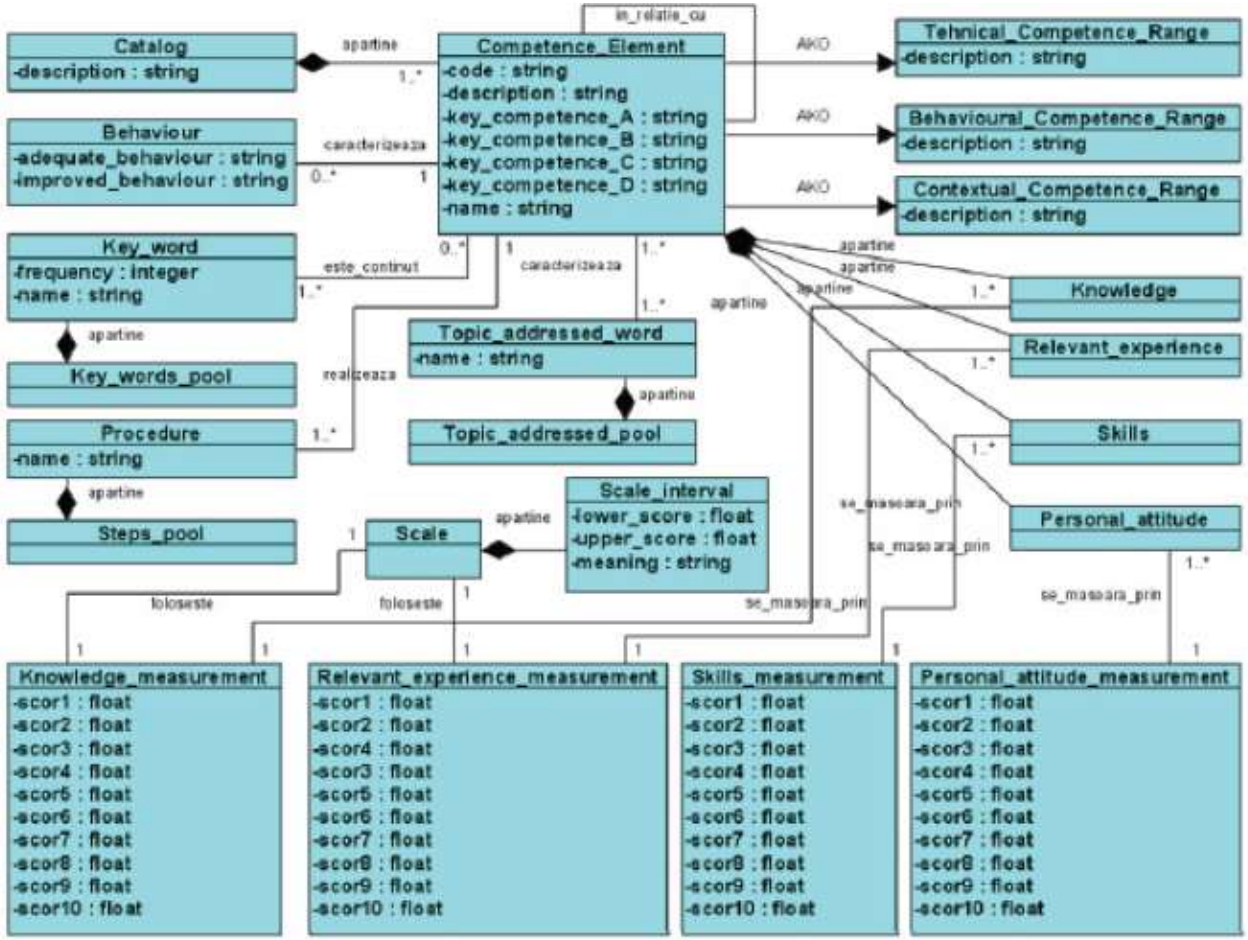
Project Management Ontology



An Ontological-based Model for Competences in Sustainable Development Projects: a Case Study for Project's Commercial Activities

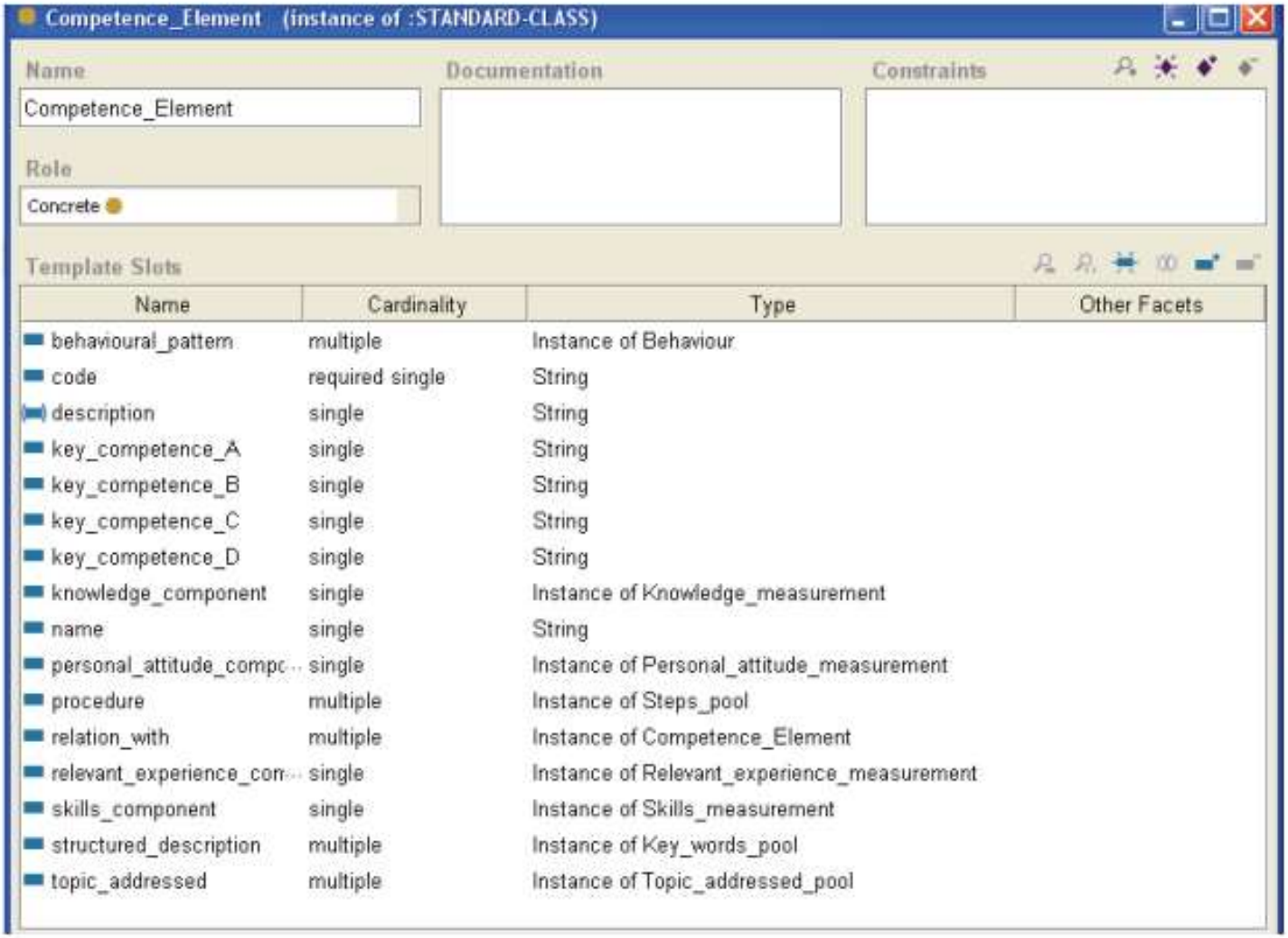
- http://www.amfiteatrueconomic.ro/temp/Article_944.pdf
- PMCatalog Ontology was developed in the framework of the project CONTO, financed by Romanian through the grant 91-037/2007
- PMCatalog Ontology was developed using the Protégé editor.

Class Diagram for a Project Management Ontology (developed in Protégé)



The Catalog Class

The Competence_Element Class



The main use cases of the ontology for commercial activities

- competency gap analysis at individual level
- competency gap analysis at project level, for all project team members
- competency gap analysis at organizational level, for all commercial-related jobs in the organization

- the competency gap analysis represents the process of comparing the requirements profile with the current competency profile, yielding missing competencies; the result is the degree of how well a person fits to requirements.

Managing personnel competences in an IT firm with Jess

- http://www.ecocyb.ase.ro/ABSTRACTS,%202,2010.pdf
- A set of project manager instances and project instances were defined in PMCatalog Ontology

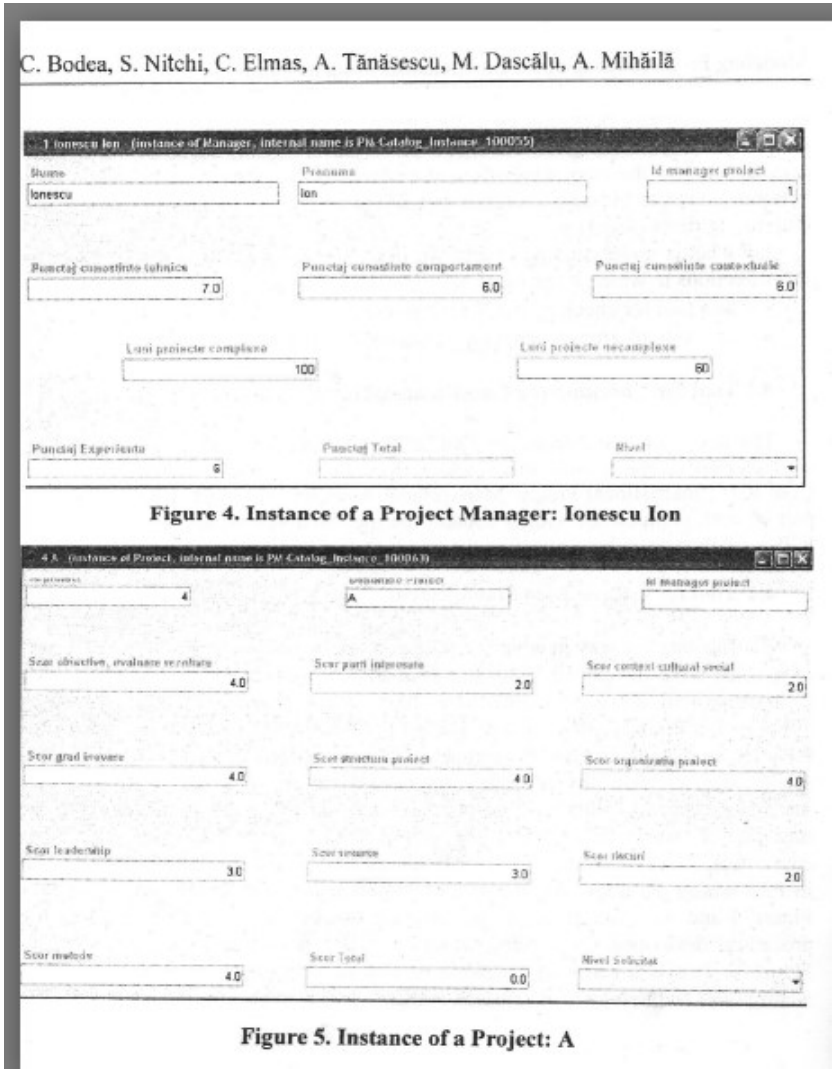


Figure 4. Instance of a Project Manager: Ionescu Ion

Figure 5. Instance of a Project: A

Use Case 1:
Query: Identify the needed competence level of a project manager to successfully conduct project A.
Results: Being a complex project, project A (id4) needs a B-level project


```
(defmodule evalproiect)
;definire reguli pentru evaluarea nivelului de competenta solicitat pentru un proiect
(defrule evalproiect::calculscortotal
?p<-(MAIN::proiect(id_proiect ?idp)(denumire_proiect ?denproiect)
      (scor_ob_evrez ?s1)(scor_partint ?s2)(scor_contcs ?s3)
      (scor_grinovare ?s4)(scor_strproiect ?s5)(scor_orgproiect ?s6)
      (scor_leadership ?s7)(scor_resurse ?s8)(scor_riscuri ?s9)
      (scor_metode ?s10)(nivel_solicitat nil))
=>
   (bind ?stotal (+ ?s1 ?s2 ?s3 ?s4 ?s5 ?s6 ?s7 ?s8 ?s9 ?s10))
   (if (>= ?stotal 25) then
       (bind ?nivelsolicitat B)
       (modify ?p (scor_total ?stotal)(nivel_solicitat ?nivelsolicitat))
       (printout t "Proiectul " ?denproiect " cu numarul " ?idp " solicita nivelul de competenta "
?nivelsolicitat crlf)
   else
        (bind ?nivelsolicitat C)
        (modify ?p (scor_total ?stotal)(nivel_solicitat ?nivelsolicitat))
        (printout t "Proiectul " ?denproiect " cu numarul " ?idp " solicita nivelul de competenta "
?nivelsolicitat crlf)))
```

Figure 7. Jess Code for Ontology Query – Use Case 1

Use Case 2:
Query: Identify all the projects which don't have a compatible project manager and display the managers who have the needed competence level.

```
(defrule evalproiect::linienoua
=>
(printout t crlf crlf "PROIECTELE CARE NU AU ASOCIAT UN MANAGER COMPATIBIL
SUNT URMATOARELE:"))

(defmodule verifcompatib)
;definirea regulii care cauta proiectele la care managerii nu sunt compatibili
(defrule verifcompatib::manageriincompat
(MAIN::proiect(id_proiect ?idp)(denumire_proiect ?denproiect)(id_manager_proiect
?idm)(nivel_solicitat ?n1))
(MAIN::manager(id_manager_proiect ?idm)(nume ?n)(prenume ?p)(nivel ?n2&:(< (str-compare
?n1 ?n2) 0)))
=>
(printout t crlf "Proiectul " ?denproiect " cu id-ul " ?idp " nu are un manager compatibil cu nivelul
solicitat, si anume " ?n1 "." crlf)
(printout t "Managerul asociat acestui proiect este " ?n " " ?p " care este certificat(a) cu nivelul de
competenta " ?n2 "." crlf)
(printout t "Acestui proiect i se pot atribui urmatorii manageri:" crlf)
(assert (MAIN::incompat  ?idp ?denproiect ?n1)))

(defrule verifcompatib::manageripotential
?i<-(MAIN::incompat ?idp ?denproiect ?n1)
    (MAIN::manager(nume ?n)(prenume ?p)(nivel ?n2&:(<= (str-compare ?n2 ?n1) 0)))
=>
(printout t ?n " " ?p " care are nivelul " ?n2 crlf))
```

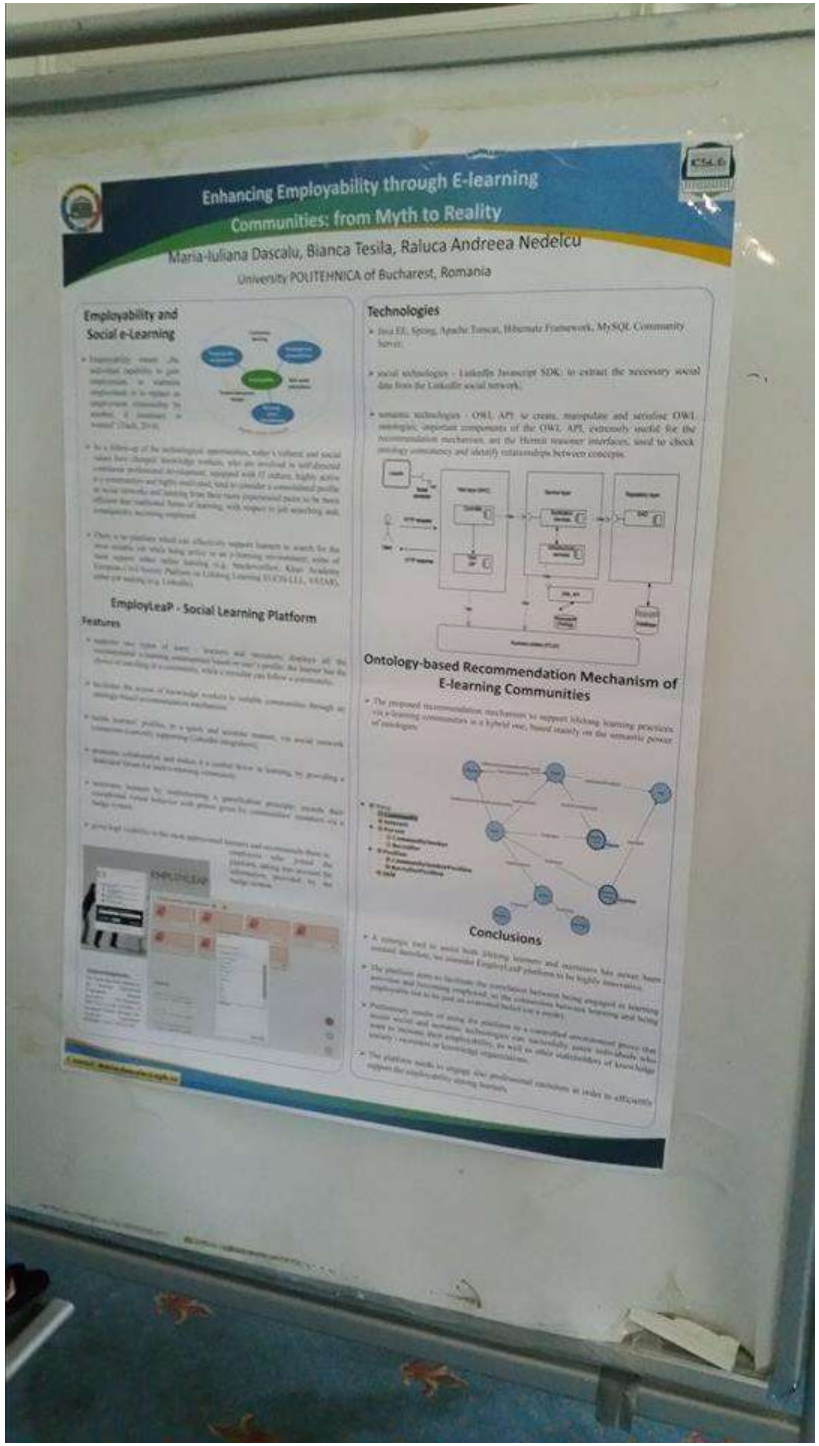
Figure 8. Jess Code for Ontology Query – Use Case 2

In IT companies, the query engine can be a valuable decision tool. The innovative use of technology in enterprises, as a way of manage knowledge, has a tremendous impact on firms' growth (Hitt et al. 2000).

Application 2

Enhancing Employability through E-learning Communities: from Myth to Reality (ICSLE 2015)

- Features of the platform EmployLeaP platoform:
  - facilitates the access of knowledge workers to suitable e-learning communities, taking into account their profiles, through **an ontology-based recommendation mechanism**
  - builds learners' profiles, in a quick and accurate manner, via social network connectors (developed with existent social network APIs)
  - promotes collaboration and makes it a central driver in e-learning, by providing a dedicated forum for each e-learning community
  - motivates learners by implementing a gamification principle: awards their exceptional virtual behavior with points given by communities' members
  - gives high visibility to the most appreciated learners and recommends them to employers who joined the platform



Application 3

Evaluation of Open-Ended Questions

- "Semantic Formative E-Assessment for Project Management Professionals", ECBS-EERC 2015 - The 4th Eastern European Regional Conference on the Engineering of Computer Based Systems, IEEE, by Maria-Iuliana Dascalu, Constanta-Nicoleta Bodea, Iuliana Marin
- Semantic technologies:
  - Text2Onto
    - extracts ontologies from text-based answers
    - based on General Architecture for Text Engineering (GATE)
    - uses Java Annotation Patterns Engine (JAPE) to extract the conceptual relations from text-based answers
  - WordNet
  - Apache OpenNLP: sentence segmentation, tokenization, POS tagging
  - Alignment API: for ontologies' comparison (AlignmentProcess and Evaluator interface)
  - JENA API: for further recommendations and ontological inferences