Towards a graduate curriculum for the Semantic Web education

Jan Maluszynski[1], Jörg Diederich[2], Luis Moniz Pereira[3], Norbert Eisinger[4], and Artur Wilk[1]

[1] Linköping University, Department of Computer and Information Science, 581 83 Linköping, Sweden
janma@ida.liu.se, artwi@ida.liu.se
[2] L3S Research Center, Hannover, Germany
diederich@l3s.de
[3] Departmento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
lmp@di.fct.unl.pt
[4] Ludwig-Maximilians-Universität München, Germany
Norbert.Eisinger@ifi.lmu.de

1 Introduction

One of the objectives of REWERSE Education and Training activity is to develop a curriculum recommendation for the graduate Semantic Web education. Understanding of the structure of this rapidly developing field is essential for providing recommendations for higher education curricula on M.Sc. and Ph.D levels, as well as for industrial courses and for supporting Semantic Web education with learning materials. An important aspect of the structuring effort is identification of relations between the Semantic Web and the existing body of knowledge in Computer Science.

This extended abstract summarizes the existing draft proposal for such a structuring (REWERSE deliverable E-D7) with the objective to stimulate further discussion. It also refers to the IEEE/ACM Computer Science Curriculum CC2001 in an attempt to identify the undergraduate learning units which may be required as prerequistes in some options of the graduate Semantic Web education. The long range objective of the work is to develop recommendations for structure and options of graduate Semantic Web education.

The present version of the proposed structure builds upon the previous work:

- REWERSE deliverable E-D1 presented information on already offered university courses relevant for the Semantic Web.
- Analysis of the information in E-D1 resulted in a preliminary structure presented in REWERSE deliverable E-D5, which was the subject of a discussion initiated by Hannover and involving members of Knowledge Web. The result of this discussion is a refined draft structure at https://wiki-sop.inria.fr/wiki/bin/view/Acacia/KnowledgeWeb

The proposed structure is used in the joint Knowledge Web and REWERSE educational infrastructure REASE[5] for classification of learning materials.

2 Specifying Prerequistes

Different options of graduate Semantic Web education may require different background knowledge. This section attempts to identify prerequisites which may be needed in some options but not necessarily in all of them.

2.1 Relevant learning units in CC2001

We take IEEE/ACM CC2001 as standard reference to undergraduate curricula in Computer Science. We refer to the areas and units specified therein, using the terminology and the unit codes of CC2001.

As a general guidance for the identification of prerequisites for graduate Semantic Web education we take the introductory sentence at the main page of W3C Semantic Web Activity (http://www.w3.org/2001):

"The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries."

and quoted therein the statement by Tim Berners-Lee, James Hendler and Ora Lassila:

"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".

Thus the prerequisites for Semantic Web education should include topics in the following areas of the CS Body of Knowledge as defined in CC2001:

Information Management (IM) The following IM topics, as specified in CC2001, seem to be relevant in Semantic Web education

IM1 Information models and systems,
IM2 Database systems,
IM3 Data modeling, including conceptual models (entity-relationship and UML), relational data models, object-oriented models and semistructured data,
IM5 Database query languages.

Generally the field of Information Management is very broad and its structuring in CC2001 may not be fully adequate for purposes of Semantic Web education. However, some topics like conceptual modelling or semistructured data are of direct importance for Semantic Web education.

[5] `http://rease.semanticweb.org`

Intelligent Systems (IS) The following IS topics, as specified in CC2001, are of particular importance

IS3 Knowledge representation and reasoning (including review of propositional and predicate logic, resolution and theorem proving),
IS5 Advanced knowledge representation and reasoning (with focus on description logics, on nonmonotonic reasoning and on reasoning on action and change),
IS6 Agents,
IS7 Natural language processing.

Courses on Knowledge Representation and Reasoning, Agents, and Natural Language processing offered in undergraduate curricula may not be sufficient for graduate Semantic Web education. In particular, we note that, due to an unfortunate decision, CC2001 does not include logic programming thus neglecting its importance in Knowledge Representation and Reasoning. These topics are very relevant for graduate Semantic Web education, among others as a prerequisite for studying rules on the Semantic Web. Thus a graduate Semantic Web program should offer specialized advanced courses on relevant topics not covered by undergraduate curricula.

Net-centric Computing (NC) The following NC topics, as specified in CC2001 are relevant as prerequisites for Semantic Web education:

NC1 Introduction to net-centric computing,
NC2 Communication and networking,
NC3 Network security,
NC5 Building web applications.

2.2 Other Foundational Topics

The above listed CC2001 topics in undergraduate education give a general background for Semantic Web education. Some of them may need additional advanced courses. Also some of advanced foundational topics relevant for the Semantic Web are not covered by CC2001. The following list of foundational topics, reflecting previous discussions in REWERSE and in Knowledge Web, includes foundational topics from both categories mentioned above:

- Knowledge Engineering and Ontology Engineering
  - Methodologies,
  - Ontology population/generation,
  - Maintenance and versioning (dynamics),
  - Mapping/translation/matching/aligning (heterogeneity),
  - Validation,
  - Interoperability/Integration,
  - Modularization and Composition,
  - Tools;

– Web information technologies
  - XML (including Namespaces, Schema Languages, XML query and transformation languages, XML programming techniques),
  - Web data integration,
  - Security,
  - Web services,
  - Personalization techniques,
  - Web data extraction/information extraction,
  - Architecture of Web Information Systems.

Notice that the above topics do not address explicitly the Semantic Web. However, development of the Semantic Web relies to large extent on the use of ontologies and on the use of the above listed web technologies.

3 Structuring of the Semantic Web body of knowledge

This section presents the proposed structure of the Semantic Web body of knowledge which should be used as a basis for development of a recommendation for graduate Semantic Web curriculum and options. The structure is already used in REASE for classification of learning units. The REWERSE deliverable E-D7 shows how the topics in this structure are covered by the learning units in REASE and by the Semantic Web courses offered in the ERASMUS MUNDUS supported European Master Program in Computational Logic.

i. Knowledge Engineering / Ontology Engineering
   1. Methodologies
   2. Ontology population / generation
   3. Maintenance and versioning (dynamics)
   4. Mapping / translation / matching / aligning
   5. Validation
   6. Interoperability / Integration
   7. Modularization and Composition
   8. Tools

ii. Knowledge Representation and Reasoning
   1. Logic
   2. Logic Programming
   3. Reasoning

iii. Basic Web information technologies
   1. XML
   2. Web data integration
   3. Security
   4. Web services RM35
   5. Personalization techniques
   6. Web data extraction
   7. Architecture of Web Information Systems

iv. Resource Description Framework / RDFSchema

v. Semantic Web Query and Update Languages
    1. Query Languages
    2. Update Languages
vi. Ontologies for the Semantic Web
    1. Ontology representation / Ontology languages / OWL
    2. Ontology Engineering
    3. Ontology reasoners
vii. Semantic Web Rules + Logic
    1. Rule languages
    2. Rule Markup
    3. Reasoning languages
    4. Rule reasoners
viii. Proof in the Semantic Web
ix. Security / trust / privacy in the Semantic Web
x. Semantic Web Applications
    1. Knowledge Management
    2. e-learning
    3. Bioinformatics
    4. Multimedia
    5. e-health
    6. e-business
    7. Law
    8. Engineering
    9. e-government
xi. Semantic Web Special Topics
    1. Natural language processing / human language technologies
    2. Social impact of the Semantic Web
    3. Social networks and Semantic Web
    4. Peer-to-peer and Semantic Web
    5. Agents and Semantic Web
    6. Semantic Grid
    7. Semantic Web Services
    8. Outreach to industry
    9. Benchmarking and scalability
    10. Design and testbed case studies

4 Conclusions and Future Work

We attempted to identify a body of knowledge in the field of Semantic Web, to structure it and to identify its links to Computer Science. We referred to IEEE/ACM CC2001 document for identifying prerequisites for graduate Semantic Web education.

The proposed structure is preliminary and will be subject of further discussion in REWERSE and in Knowledge Web. Also, revisions reflecting development of the Semantic Web will certainly be needed. We encourage all REWERSE members to participate in the discussion.

We notice that according to the information at hand some of the topics in the structure are not yet supported by existing learning units/learning material in REASE. In some cases this may be caused by insufficient development of the field. For example this seems to apply to some of the listed topics in the area "x. Semantic Web Applications", or to the area "vii. Semantic Web Rules" which only recently became a subject of W3C activities. Also the importance of some topics in the foundational areas "i. Knowledge Engineering/Ontology Engineering" and "iii. Basic Web information technologies" might not have been yet sufficiently explored in the context of the Semantic Web. The "proof level" ("viii. Proof") postulated in the original Semantic Web vision of Tim Berners-Lee was not yet sufficiently explained. Proofs are inherent both on ontology level and rule level and there may be no need for special proof level.

The future work includes:

- **Refinement of the proposed structure.** The discussion preceding preparation of this deliverable will be continued, taking into account recent developments in the field, especially W3C activities, and new contributions to REASE. This may lead to revision of the proposed structure. The structure will be refined by suggesting the recommended content for the foundational topics not covered by CC2001 and for the core topics of the Semantic Web. In particular the recommendation should consider different options in graduate Semantic Web education, identifying the prerequisites and the elements of the structure to be covered by a defined option.
- **Supporting the structure by new learning units in REASE.** This applies in the first hand to the topics in our structure which are not covered by the existing units, and/or to the topics which are closely related to REWERSE research with particular focus on recent developments. These criteria have been taken into account while preparing the programme of the REWERSE Summer School 2006 (see REWERSE deliverable E-D8-1). The materials of the school will be uploaded in REASE. Among others they will address the following topics in the structure:
  - Bioinformatics and the Semantic Web,
  - Semantic Web query and update languages with particular focus on recent developments,
  - Rule languages for the Semantic Web with particular focus on W3C Rule Interchange Format and on integration of rules and ontologies postulated in the original vision of the Semantic Web architecture.
  - Outreach to industry, not yet addressed by REWERSE will be a subject of two contributions.

  REWERSE will also encourage uploading to REASE learning materials of Erasmus Mundus supported European Master Program in Computational Logic.