

COEUR-SW: Concepts On Enriching, Understanding, and Retrieving the Semantics on the Web



Sari Hakkarainen, Arne Sølvberg,
Terje Brasethvik, Yun Lin, Jennifer Sampson, Guttorm Sindre,
Darius Strasunskas, Xiaomeng Su, and Csaba Veres

Norwegian University of Science and Technology
Sem Sælands vei 7-9, NO-7491 Trondheim, Norway
Name.Surname@idi.ntnu.no

Abstract. The Information Systems on the Web comply with the semiotic triangle. In the **COEUR-SW** triangle, a concept in a Universe of Thought is related to an uttered symbol in a Universe of Language, the symbol is related to a referent in a Universe of Structure, and the referent is related back to the concept. The concept, the symbol and the referents are related to a context in the Universe of Discourse (UoD), as embedded in the Web. The **COEUR-SW** program at IDI, NTNU in Norway, approaches UoD from three interrelated angles and thus seeks to capture the heart of the Semantic Web.

1 Introduction

The Web has three major roles in Information Systems; one role as the dominant medium for information dissemination, one as the tool for information compilation, and one as the evolving information repository. The roles comprise enterprise and user interface issues (UoT), information categorisation and interpretation issues (UoS), and information storage and information access issues (UoL).

In order to support both the dissemination and the compilation of selected information, the repositories may be organised according to the principles of integration or interoperability. In the former, the relevant information repositories are viewed as one large distributed database, organised according to a common schema. In the latter, every information system is considered to consist of autonomous subsystems.

The integration principle is dominant in the classical approaches to information systems design. The alternative principle of semantic interoperability is more compliant to the Semantic Web. For interoperable systems to communicate there must be some agreement on coding as well as meaning of data, but only for the data that is interchanged. The mutual understanding must be as wide as it is necessary for a particular data interchange, but need not to be wider.

2 Semantic interoperability

Data semantics is the relationship between the data and what the data stand for. In order to obtain mutual understanding of interchanged data, the actors have to share a

model of what the data stand for. Semantic interoperability is about how to achieve such mutual understanding.

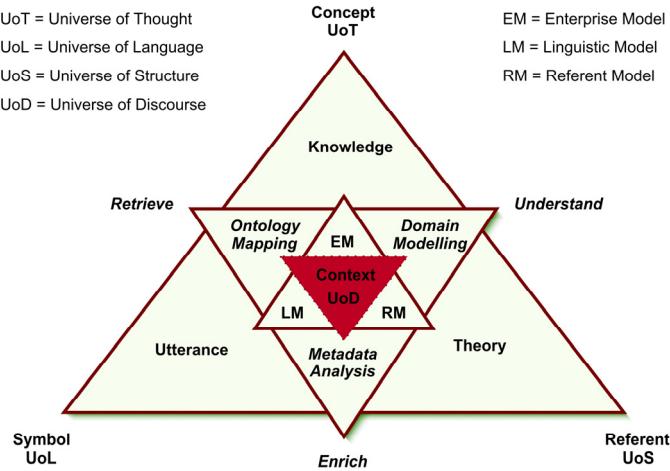


Fig. 1. The **COEUR-SW** program at IDI, NTNU in Norway, approaches UoD from three inter-related angles. Domain modelling is used in order to capture the knowledge within a UoT into a man/machine understandable theory in a UoS. Ontology Mapping is used in order to capture the utterances in a UoL into man/machine retrievable knowledge. Metadata analysis is used in order to capture the theories in a UoS into enriched man/machine generative utterances in a UoL. In all the approaches inter-related conceptualisation of UoD is used in order to capture the heart of the Semantic Web.

In order to achieve this we use concepts and symbols. The concept is the unit of thought. The symbol is the unit of language. Theories, such as Newton's theory of motion, are structures of thoughts, and referents like force, mass and acceleration are the units of these structures. Conceptual knowledge comes wrapped in symbols, e.g. mathematical notation, words or diagrams, which are the linguistic expressions of knowledge. In order to access the ideas of other people, i.e. the concepts in their UoT, we must understand the conceptual structures that are employed in the common UoS. Furthermore, we need to understand the relationship between the symbols and the ideas that they stand for.

3. Agent systems

Agents in a multi-agent system are characterised by abstraction, interoperability, modularity and dynamism. These qualities are particularly useful in that they can help to promote open systems, which are typically dynamic, unpredictable, and highly heterogeneous as is the Internet. Within such a multi-agent system, agents represent their 'view of the world' by explicitly defined semantics over the information and the services.

The interoperability of such a multi-agent system is achieved through the reconciliation of these views by a commitment to common ontologies that permit agents to interoperate and cooperate. However, such ontologies are developed and maintained independent of each other. Thus, two agent systems may use different ontologies to represent their views of the domain. This is often referred to as ontology mismatch. In such a situation, interoperability between agents is based on the reconciliation of their heterogeneous views by techniques that provide mappings for use in transformation or meaning negotiation between services.

4 Web Resources

Recently, several general-purpose models for description Web resources have emerged. The intention of the industry-driven initiatives is to provide a metadata description framework for interconnected resources. A resource can be viewed as bibliographic document (DC, MARC), electronic Web resource (RDF, HTML) multimedia object (SMIL), ontology (OIL, DAML), database concept (MDIS, OIM), or case tool structure (XMI). Many of the models result from standardisation work as carried out in industry-driven coalitions like W3C, MDC, and OMG.

Academic research work has adapted and applied some of the models provided by the industry coalitions. Metadata is recognised as equally important for describing the components from which we build services and systems, regardless of whether they are on the Internet or not. The research so far has focused on technical aspects of the frameworks, semantic coverage of the models, and reasoning and querying mechanisms. The area is becoming mature enough for comparative studies and prescriptive theories on the proposed modelling frameworks and environments.

5 Ontology standards

During the last decade, a number of ontology representation languages have been proposed. The so-called traditional ontology specification languages include: CycL, Ontolingua, F-logic, CML, OCML, Telos, and LOOM. There are Web standards that are relevant for ontology descriptions for semantic Web applications, such as HTML, XML and RDF. Finally, there are the XML/HTML-based Web ontology specification languages such as XOL and SHOE, and those that form basis for the layered architecture for the semantic Web, such as OIL, DAML+OIL, and OWL. The latter group are the so-called semantic Web enabling languages (SWEL) for ontology building which are in the focus of the COEUR-SW program at IDI.

There already exist several methodologies to guide the process of Web ontology building that vary both in level of generality and granularity. The methodologies describe an overall ontology development process yet not the ontology creation itself. Primarily, they are intended to support the knowledge elicitation and management of the ontologies in a basically centralised environment. The methodologies provide a life cycle in an overall ontology development process but only a few user guidelines for carrying out the steps and for actually creating the ontology. In order to increase the number and the scale of practical applications of the semantic Web technologies,

the developers need to be provided detailed instructions and general guidelines for the actual ontology creation.

6 Activities

The overall intention of the COEUR Semantic Web research program¹ at IDI, NTNU is to explore and exploit existing research results in the areas of conceptual modeling, information systems analysis and design, agent technology, and information systems architectures. Our attacking point is semantic interoperability. Our main armoury consists of enriched descriptions and representations for each of the Universes of the Semantic Web triangle. Our ultimate target is to enable information systems development for Semantic Web applications in the areas such as information services, e-commerce, knowledge management, and cooperative systems.

Currently, several activities are underway within this program:

- the Web Information Service Modeling project, WISEMOD
- the Referent Modeling Language project, RML
- the Semantic Enrichment for Ontology Mapping project, iMAPPER
- the Semantic Modeling of Documents project, SeMD_{oc}
- the Conceptual Metadata Analysis and Design in Information Service Development project, COMAD
- the Adaptive Distributed information Service project, ADIS
- the Product Fragment Management in Collaborative Model Driven Development project, CoSy

Key References

- Terje Brasethvik and Jon Atle Gulla. Natural language analysis for semantic document modeling. *Data & knowledge Engineering*, 38(1):45–62, 2001.
- Sari Hakkarainen, Lillian Hella, Stine Tuxen, Guttorm Sindre. Evaluating the quality of web-based ontology building methods: a framework and a case study. In Proceedings of 6th International Baltic Conference on Databases and Information Systems (Baltic DBIS'04), Riga, Latvia, 2004.
- Yun Lin, Sari Hakkarainen, Dariusz Strasunskas. Towards the Two-World Modeling. In Proceedings of IASTED International Conference on Software Engineering and Applications (SEA'03). Marina Del Ray, CA, USA, 2003.
- Jennifer Sampson. An Agent Based Infrastructure for facilitating decision making in healthcare. CAiSE'04 Forum. Riga, Latvia, 2004.
- Arne Solvberg. Data and what they refer to. In P. P. Chen, editor, *Concept Modeling: Historical Perspectives and Future Trends*. Springer Verlag, 1998.
- Xiaomeng Su and Jon Atle Gulla. Semantic enrichment for ontolog mapping. In Proceeding of the 9th International Conference on Applications of Natural Language to Information System (NLDB'04), Springer Verlag, Manchester, UK, 2004.
- Xiaomeng Su, Sari Hakkarainen, and Terje Brasethvik. Semantic enrichment for improving systems interoperability. In Proceeding of the 2004 ACM Symposium on Applied Computing, pages 1634–1641, ACM Press, Nicosia, Cyprus, 2004.
- Xiaomeng Su, Mihail Matskin, and Jinghai Rao. Implementing explanation ontology for agent system. In Proceeding of the 2003 IEEE/WIC International Conference on Web Intelligence (WI 2003). IEEE Computer Society Press, 2003.

¹ <http://coeur-sw.idi.ntnu.no/>