

Introducing Ontologies*

1 Internet and the World Wide Web

The Internet is a worldwide, decentralized network of interconnected computers: a network of networks. Via the Internet, connections are established amongst individual Internet users themselves and institutions such as universities, governmental and commercial organizations. The Internet offers a multitude of possibilities in the field of communication and information, such as:

- Electronic mail (e-mail): to send and receive mails in a digitalized format. Particular examples are discussion and mailing lists. There exist ten thousands of discussion lists on very diverse topics. By subscribing on a list, one is able to participate in the discussions.
- Short Message Service (SMS): to send text messages amongst cell phones or amongst a computer and a cell phone.
- Chat boxes: enabling instantaneous exchange of information through keyboard and screen.
- Telnet or remote login: to remotely log on to another computer.
- File Transfer Protocol (FTP): to exchange files between a remote and a local computer.
- Video-conferencing: to organize meeting with several persons connected through the Internet, supported by video and sound equipments.
- World Wide Web (WWW): to locate and consult documents or files online.

*Based on [SMJ02, VA04, Usc03]

The World Wide Web was invented in the early nineties of the twentieth century by Tim Berners-Lee and Robert Cailliau, working at the European Organization for Nuclear Research (CERN) in Geneva. At that time, many documents were placed on the Internet, merely by universities. However, the traceability of these documents was a major issue. Berners-Lee was convinced that the storage, consultation and in particular the retrieval of documents should happen more efficient, which resulted in the World Wide Web.

The popularity of the Internet is due the deployment of the World Wide Web, which is a multimedia information system whereby various types of documents (text, images, video, sound, software...) are presented via web pages.

The access to the Internet is orchestrated and partly controlled by Internet Service Providers (ISP's). The numerous marketing strategies (free subscriptions, free installation, introduction of broadband) deployed by those ISP's combined with the evolution of the Internet towards an interactive and social medium, lay the foundation of the success of the large-scale acceptance and adoption of the Internet. Everyone has the possibility of publishing information via the World Wide Web, which resulted in an information overflow.

Information contained within web pages is open to different interpretations. Whether or not one might correctly interpret this information depends largely on the experience, background, education... of the natural person reading this information, but also on the information's context. As an example, consider an online application form containing an "address" field on the web site of an international company with branches in Brussels, New York, Paris and Tokyo. This field can be interpreted as referring to the residence of the online applicant, or as the address of the branch that is most appealing to the applicant. Because of the ambiguous interpretation of information presented on web pages, the World Wide Web is sometimes referred to as the "eyeball web", since the the structuring and interpretation of the information is oriented towards humans only [VA04].

2 The Semantic Web

Considering the exponential increase of information on the World Wide Web, it is nearly impossible for a human to identify and analyze all relevant web pages given a topic of interest. Therefore, research efforts are oriented towards the evolution of the World Wide Web into the Semantic Web. Despite the widespread use of the term "Semantic Web", it does not yet exist except

in isolated environments such as research labs. According to World Wide Web Consortium (W3C) [Wor01]:

*the Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.*¹

Berners-Lee envisions that [BLHL01]:

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

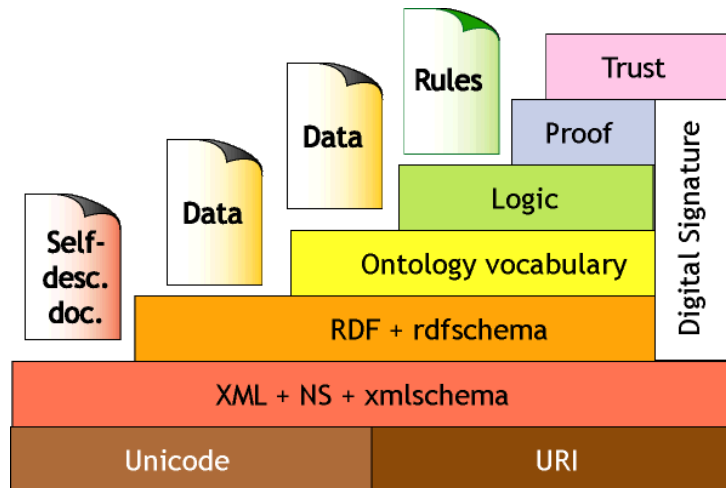


Figure 1: The Semantic Web layers (reproduced from [KM01]).

The Semantic Web principles are implemented in the layers of Web technologies and standards. The layers are presented in Figure 1 (reproduced from [KM01]):

- The *Unicode* and *Uniform Resource Identifier (URI)* layers make sure that we use international characters sets and provide means for identifying the objects in Semantic Web, respectively. The most popular URI's on the World Wide Web are Uniform Resource Locators (URL's).

¹Note that “machine” refers to computers or computer programs that perform tasks on the Web. These programs are commonly referred to as *software agents* or *softbots* and are found in Web applications.

- The *XML layer* with *namespace* (NS) and *schema* definitions make sure we can integrate the Semantic Web definitions with the other XML based standards.
- With the *Resource Description Framework (RDF)* (W3C standard) and *RDFSchema (RDFS)* it is possible to make statements about objects with URI's and define vocabularies that can be referred to by URI's. In this layer, types are attributed to resources and links.
- The *Ontology* layer supports the evolution of vocabularies as it can define relations between the different concepts (see also section 3).
- A digital signature is an electronic signature that can be used to authenticate the identity of the sender of a message or the signer of a document. The *Digital Signature* layer ensures that the original content of the message or document is unaltered.
- The top layers *Logic*, *Proof* and *Trust*, are currently being researched and simple application demonstrations are being constructed. The Logic layer enables the writing of rules while the Proof layer executes the rules and evaluates together with the Trust layer mechanism for applications whether to trust the given proof or not.

The main intent of the Semantic Web is to give machines much better (read: autonomous) access to information resources in order to become fully fledged information brokers in support of humans. In order to carry out their required tasks, intelligent software agents must communicate and understand meaning. They must advertise their capabilities, and recognize the capabilities of other agents. They must locate meaningful information resources on the Web and combine them in meaningful ways to perform tasks. They need to recognize, interpret, and respond to communication acts from other agents. In other words, when software agents communicate with each other, there needs to be some way to ensure that the meaning of what one agent “says” is accurately conveyed to and understood by the other agent.

The main hurdle to agent communications is that different agents adopt different repositories of terms, and even if they adopt the same terminology, it cannot be assumed that the terms will mean the same thing. Every agent will need to publicly declare exactly what terms it is using and what they mean. To achieve this goal, an appropriate machine-readable way to describe these repositories and their meaning is required. Many alternatives have been developed, however, there is not a consensus to use one or a small set

of them yet. Important approaches as data structures, relational data bases, mathematic logic, procedures, taxonomies, and ontologies amongst others, have been proposed by researchers in the domains of Artificial Intelligence and database semantics.

3 Ontologies

It is possible to find in the literature several definitions of ontologies. The most cited one is proposed by Gruber [Gru93]: an ontology is a formal, explicit specification of a shared conceptualization. This definition identifies four main concepts involved: an abstract model of a phenomenon termed “conceptualization”, a precise mathematical description hints the word “formal”, the precision of concepts and their relationships clearly defined are expressed by the term “explicit”, and the existence of an agreement between ontology users is hinted by the term “shared” [Fen01].

Some essential aspects of ontologies are:

- ontologies are used to describe a model of a specific domain,
- its *concepts* and *relations* are unambiguously and formally defined by axioms² and definitions stated in a formal language (e.g. OWL), such as logic or some computer-oriented notation that can be translated to logic,
- there is a mechanism to organize the concepts³ by means of relationships, which might be hierarchical or non-hierarchical⁴,
- there is an agreement between users of an ontology in such a way that the meaning of the concepts is used consistently by all of them (see figure 2).

²A well-formed formula which is taken to be true without proof in the construction of a theory.

³Concepts are regarded as abstract or general ideas inferred or derived from specific instances, and are, in contrast to instances, not tangible. E.g., “Robert Meersman” is an instance of the concept “full professor”.

⁴The generalization/specilization relationship between two concepts is regarded as the most widely accepted example of a hierarchical relationship between concepts, and typically referred to by the term “is-a”. E.g. *full professor* is-a *senior academic employee*. As an example of a non-hierarchical relationship, consider: *full professor* heads *research group*.

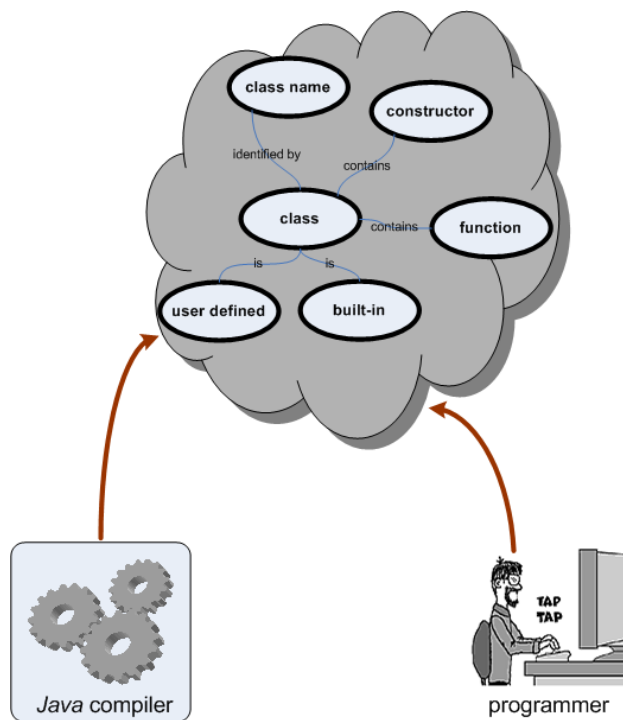


Figure 2: Representational model of a sub-domain of *computer sciences*. All ontology users agree on the meaning of involved concepts and their relationships represented in the model.

Semantical relationships in ontologies are machine understandable, in such a way that they enable making statements and asking queries about a subject domain due to the use of a conceptualization, which describes entities and their relationships. The usefulness of ontologies in agent based systems can be briefly summarized as they enable knowledge interoperation. In other research areas, ontologies support shared understanding, interoperability between tools, systems engineering, reusability and declarative specification [FFR97].

On the other hand, ontologies are used to build knowledge bases [NM01]. Knowledge bases are formed by an ontology and a set of individual instances of its classes, and can be queried by agents in order to enrich, reuse and maintain them.

Ontologies are able to operate as repositories to organize information for specific communities. They can be used as a tool for knowledge acquisition,

while teams can use ontologies as a common support to classify the knowledge of an organization. Ontologies allow users to reuse knowledge in new systems. They can form a base to construct knowledge representation languages [Nie03].

Within integration systems, ontologies are proposed as mediators in order to offer a unified access to such resources [Wie92]. A mediator is an information source that bridges the heterogeneities between two or more resources. Specifically, a mediator employs a terminology and maintains a unified knowledge scheme which suits the needs of its potential users, but does not maintain a database of objects. Instead, the mediator disposes of a set of articulations to the underlying resources. An articulation to a source is a set of relationships between the terms of a mediator and the terms of that source [TSC01] (see figure 3). Mediation recognizes the autonomy and diversity of data resources, information services and the user applications utilizing them. The autonomy of the participants enables the overall system to grow, since new sources and novel information processes can be inserted.

In information retrieval applications, ontologies serve to disambiguate user queries, to elaborate taxonomies of terms or thesaurus in order to enhance the quality of retrieved results [Fou01]. Machine-learning techniques are also used to extend ontologies based on users interactions.

4 Developing Ontology-Guided Mediation of Agents (DOGMA)

Unlike traditional Artificial Intelligence ontology engineering proposals, which consider an ontology as being a single unit, holding both conceptual related objects and rules together, the approach to ontology engineering adopted by STAR Lab's core research project termed DOGMA [Mee99b, Mee99a] is based on the principle of a double articulation [DBSM03], decomposing an ontology into (see Figure 4):

- an **ontology base** (see section 4.1), which holds intuitive conceptualizations (that is: concepts and relationships between concepts) of a domain,
- and a **commitment layer** (see section 4.2), where each commitment contains a set of domain rules applied to a specified subset of the ontology base [JDM03] and a set of mappings between ontological elements and application elements [DV03].

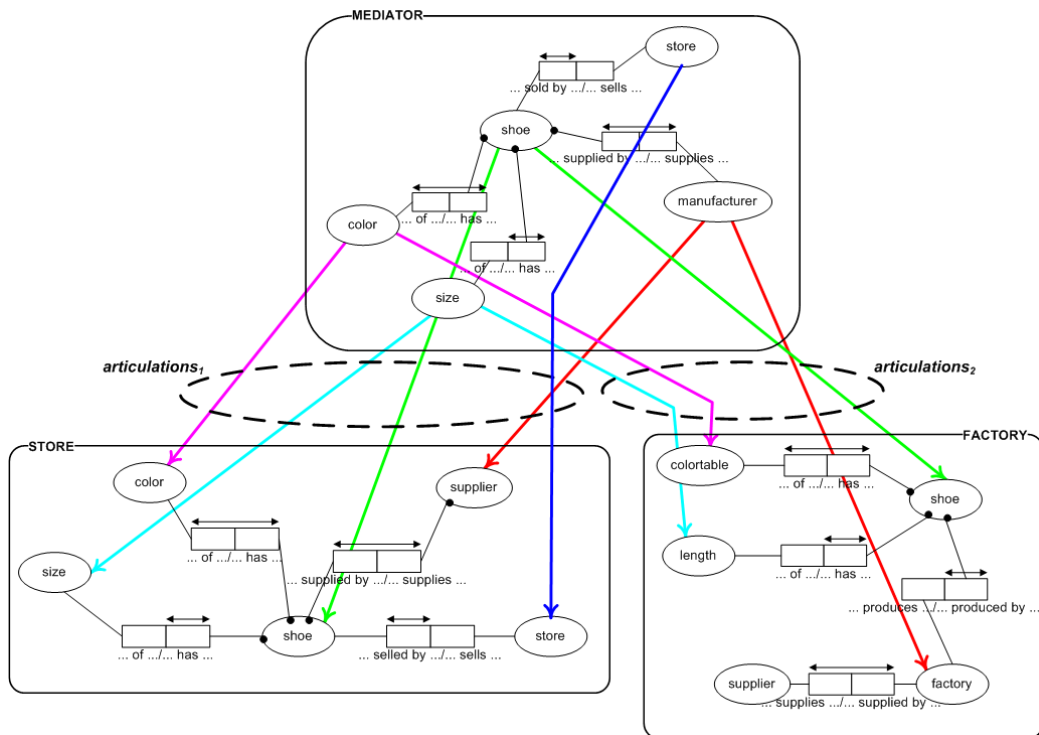


Figure 3: Representation of an integration system containing a mediator and sets of articulations towards participating sources. The mediator employs an agreed and shared terminology and knowledge scheme in order to offer a unified access to the heterogeneous sources. In order to maintain the autonomy of the underlying resources, the sets of articulations that define relationships between elements of the mediator and elements of the sources, are stored at the integration system.

This strategy is based on experiences that parties committing to an ontology reach consensus about the objects and relationships involved, but fail to impose a single set of agreed rules and constraints [SM02].

4.1 Ontology Base

Currently, the ontology base holds binary conceptual relations, called *lexons*, which are intuitive⁵ statements or facts within a specific real world domain [Mee99b, Mee99a]. A lexon is formally described as $\langle \gamma, term_1, role, co-role, term_2 \rangle$, where γ is an abstract context identifier, used (i) to refer to a

⁵By “intuitive”, we assume a shared agreement about the meaning of its terms.



Figure 4: The double articulation of a DOGMA ontology (reproduced from [JDM03]).

“source” (in general a document) providing evidence for modelling this lexon, and (ii) to specify the intended concept denoted by $term_x$. This latter functionality is necessary, since a $term_x$ could refer to more than one concept, e.g. the term “*ladder*” can be associated to both of the following concepts⁶: “*steps consisting of two parallel members connected by rungs, for climbing up or down*” and “*a row of unravelled stitches*”. However, combining a γ and a $term_x$ will uniquely identify the intended concept.

Finally, the *role* (with *co – role* as its inverse) of a lexon defines the conceptual relation between the two intended concepts $(\gamma, term_1)$ and $(\gamma, term_2)$. As an example, Figure 5 defines a number of lexons valid within the context of computer sciences.

<computer sciences, class, inherits, inherited by, superclass>
 <computer sciences, class, contains, is part of, function>
 <computer sciences, language construct, contains, is part of, user-defined pointer>

Figure 5: Representation of lexons within the context of computer sciences.

In order to disambiguate the meaning of any term and role used in Figure 5, we provide following definitions⁷:

⁶According to the online lexical thesaurus WordNet [Fel98].

⁷Some of the definitions are aligned with definitions provided in WordNet [Fel98]. How-

- **language construct:** a syntactically allowable part of a program that may be formed from one or more lexical tokens in accordance with the rules of a programming language.
- **class:** a template for objects that defines the internal structure and the set of operations for instances of such objects.
- **superclass:** a class from which another class (called a subclass) inherits attributes and methods.
- **function:** a named section of a program that performs a specific task. In this sense, a function is a type of procedure or routine. Some programming languages make a distinction between a function, which returns a value, and a procedure, which performs some operation but does not return a value.
- **user-defined pointer:** a data object, defined by a user, whose data value is the address of another data object.
- **to inherit:** transmission of class characteristics from one or more of its superclasses.
- **to inherit by:** transmission of class characteristics to one or more of its subclasses.
- **to contain:** have as a component.
- **is part of:** something determined in relation to something that includes it.

All rules and constraints implied on the lexons in the ontology base are moved to the commitment layer. Therefore, all lexons in the ontology base are free of any specific interpretation.

4.2 Commitment Layer

The commitment layer, mediating between the ontology base and applications, contains a set of interpretations, also called commitments. Each commitment consists of:

ever, since WordNet merely provides general purpose terminologies, some definitions were extracted from technical and domain-specific thesauri such as American National Standard Dictionary of Information Technology (ANSDIT) [Ame00]

- rules that constrain and attribute specific interpretations to a selected subset of lexons contained within the lexon base [JDM03],
- and a set of mappings that link elements of this subset of lexons to elements of specific applications [DV03].

Each individual commitment within the commitment layer is a representation of the semantics of a specific application. The common ground for all these representatives is the ontology base containing shareable lexons. Shareable implies that a lexon being an element of a lexon subset of one commitment, might also be an element of another commitment's lexon subset.

Formal axioms are introduced to constrain the interpretation and well-formed use of commitments.⁸In this respect, an ontology is viewed as the explicit statements of a logical theory [Pre04].

4.2.1 Constraints

Each individual commitment rule represents the semantics of a specific application by applying constraints on a subset of lexons. Currently, most of the constraints occurring in the Object-Role Modelling (ORM) language, which is the conceptual modelling language of Halpin [Hal01], are supported. As an example, figure 6 contains some lexons regarding computer sciences and defines two commitments which accommodate for incomplete representations of the programming languages C++ and Java.

In Figure 6, although the commitment's constraints that accommodate for the C++ and Java programming languages are both individually self-consistent, they adopt conflicting opinions regarding the deployment of multiple inheritance[Che97] with respect to class types.

This strategy of separating the concepts and their relations (ontology base) from multiple, parallel interpretations (commitment layer) increases the overall reusability of the ontology.

4.2.2 Mappings

So far, a commitment is merely a secondary information source in the sense that it only applies semantic constraints to lexons contained in the Ontology

⁸For further details see [Rei].

Ontology Base

<computer sciences, class, inherits, inherited by, superclass>
<computer sciences, class, contains, is part of, function>
<computer sciences, language construct, contains, is part of, user-defined pointer>
...

C++ commitment

constraints

each class inherits zero-or-more superclass
each class contains zero-or-more function
each function is part of exactly-one class
(each language construct contains zero-or-more user-defined pointer) AND
(each user-defined pointer is part of zero-or-more language construct)
each user-defined pointer is part of at-least-one language construct

Java commitment

constraints

each class inherits at-most-one superclass
each class contains zero-or-more function
each function is part of exactly-one class

Figure 6: Representation of commitment’s constraints that accommodate for incomplete interpretations of the programming languages C++ and Java, described by DOGMA’s commitment definition language Ω -RIDL [DV03].

Base. Since a concept is an abstract or general idea inferred or derived from specific instances, it is not tangible in contrast to its instances. The aim of introducing a DOGMA ontology is that software agents will autonomously use the conceptualizations as a “mindmap” for the exchange of real data (instances).

Instead of storing instances of involved concepts, a specific commitment relates to its instances by defining mappings between ontological concepts or relations and elements of data resources containing these instances.

Actually, these mappings link the agreed terminology employed within the Ontology Base and the different terminologies proper to the various applications (software agents, programs, data sources. . .). E.g., in figure 7, the element denoted by the term “member function” and “method” within the programming language C++ and Java, respectively, correspond to the agreed

Ontology Base

<computer sciences, class, inherits, inherited, superclass>

<computer sciences, class, contains, is part of, function>

<computer sciences, language construct, contains, is part of, user-defined pointer>

...

C++ commitment

mappings

map member function on function (is part of class)

Java commitment

mappings

map method on function (is part of class)

Figure 7: Commitment' mappings from ontological concepts towards the C++ and Java programming language specific elements, described by DOGMA's commitment definition language Ω -RIDL [DV03]

term "function" defined within the Ontology Base.

5 Summary

Through the widespread acceptance of the Internet a wealth of information became available to everyone connected to this powerful network. In correspondence with the decentralized nature of the Internet, little or no control was exerted on information content published on the World Wide Web, hereby interfering with its initial aim: providing a more efficient means of storing, consulting and in particular retrieving documents. Since humans are unable to process the massive amount of available information, software agents are assumed to support this process partially. However, the current World Wide Web does not offer the explicit and disambiguated features required by these agents. It is said that the current World Wide Web is oriented towards humans only.

In order to overcome these aforementioned inconveniences, research effort focus on the realization of the Semantic Web, which will be the successor of

the World Wide Web. There are many different views of what the Semantic Web is and how it may or should evolve. The most frequently quoted feature is machine usable web content. Fundamentally, this requires that machines should be aware how to recognize the content they are looking for, and what to do when they encounter it. This process of awareness requires access to the meaning or semantics of the content. Ontologies are one of the promising technologies aiming at fulfilling this challenge.

6 Conclusion

Gruber [Gru93] defined an ontology as a formal, explicit specification of a shared conceptualization. Ontologies provide a set of well defined, structured and agreed terms in order to disambiguate communication exchange between applications (software agents, programs. . .) We presented some important aspects and applications of ontologies, before highlighting STAR Lab's approach towards ontologies, within the context of its core research effort termed DOGMA.

DOGMA represents an ontology in a layered architecture: (i) the base layer, called the Ontology Base, contains concepts and relations organized into lexons, characterized by an agreement with respect to their meaning; (ii) and the Commitment layer containing individual commitments serving as representatives of different applications that adopt the ontology. This strategy of separating the concepts and their relations (ontology base) from specific (application-representative) interpretations (commitment layer) increases the overall reusability of a ontology.

The terminology used by any particular application has been unambiguously defined thanks to its commitment's mappings to the Ontology Base (see figure 8).

Although research efforts regarding ontologies and Semantic Web technologies are advancing, many hurdles still need to be surpassed in order to realize its effective deployment.

Acknowledgements: This introduction to ontologies is the result of fruitful remarks and scientific contributions on behalf of all STAR Lab members and its Director, Professor Robert Meersman.

This publication is based on ideas and structures from [SMJ02, VA04,

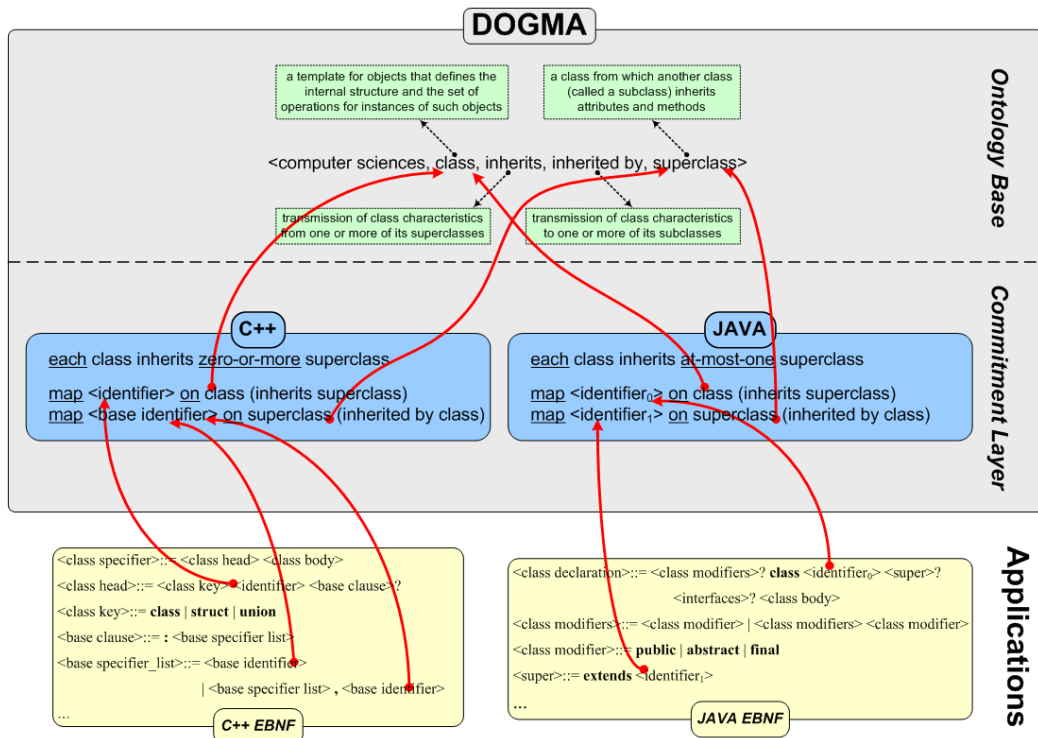


Figure 8: Graphical overview of the DOGMA framework. The C++ and Java commitments within the commitment layer link the agreed terminology within the Ontology Base to the different terminologies and interpretations adopted by the C++ and Java programming language grammars, respectively. An EBNF (extended Backus-Naur Form) is a meta-syntactic notation used to specify the syntax of programming languages, command sets, and the like... [ASU86] The EBNF terminals are shown in bold.

Usc03, Nie03].

References

- [Ame00] American National Standards Institute, and National Committee for Information Technology Standards, and Technical Committee on Vocabulary K5. American National Standard Dictionary of Information Technology (ANSDIT). http://www.ncits.org/tc_home/k5htm/Ansdit.htm, 2000.

- [ASU86] A.V. Aho, R. Sethi, and J.D. Ullman. *Compilers - Principles, techniques and tools*. Addison-Wesley, 1986.
- [BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, May 2001.
- [Che97] F.F. Chew. *The Java/C++ Cross Reference Handbook*. Prentice Hall PTR, 1997.
- [DBSM03] J. De Bo, P. Spyns, and R. Meersman. Creating a "DOGMAtic" multilingual ontology infrastructure to support a semantic portal. In R. Meersman and R. Tari, editors, *On the Move to Meaningful Internet Systems 2003: OTM 2003 Workshops*, volume 2889 of *LNCS*, pages 253–266. Springer Verlag, 2003.
- [DV03] T. Deray and P. Verheyden. Towards a semantic integration of medical relational databases by using ontologies: a case study. In R. Meersman and et al. Tari, Z., editors, *On the Move to Meaningful Internet Systems 2003: OTM 2003 Workshops*, volume 2889 of *LNCS*, pages 137–150. Springer Verlag, 2003.
- [Fel98] C. Fellbaum. *WordNet: An Electronic Lexical Database* (<http://www.cogsci.princeton.edu/~wn>). Cambridge, US: The MIT Press, 1998.
- [Fen01] D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer, 2001.
- [FFR97] A. Farquhar, R. Fikes, and J. Rice. The Ontolingua Server: A tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, 46(6):707–727, 1997.
- [Fou01] Foundation for Intelligent Physical Agents. FIPA Ontology Service Specification. <http://www.fipa.org/specs/fipa00086/XC00086D.html>, 2001.
- [Gru93] T.R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5:199–220, 1993.
- [Hal01] T. Halpin. *Information Modelling and Relational Databases*. Morgan-Kaufmann, 2001.
- [JDM03] M. Jarrar, J. Demey, and R. Meersman. On Reusing Conceptual Data Modeling for Ontology Engineering. In K. Aberer, S. March,

- and S. Spaccapietra, editors, *Journal on Data Semantics*, volume 2800 of *LNCS*, pages 185–207. 2003.
- [KM01] M.R. Koivunen and E. Miller. W3C Semantic Web Activity. In E. Hyvonen, editor, *Proceedings of the Semantic Web Kick-off Seminar*, pages 27–44, Finland, 2001.
- [Mee99a] R. Meersman. Semantic Ontology Tools in Information System Design. In Z. Ras and M. Zemankova, editors, *Proceedings of the ISMIS 99 Conference, LNCS 1609*, pages 30–45. Springer Verlag, 1999.
- [Mee99b] R. Meersman. The use of lexicons and other computer-linguistic tools in semantics, design and cooperation of database systems. In Y. Zhang, M. Rusinkiewicz, and Y. Kambayashi, editors, *Proceedings of the Conference on Cooperative Database Systems (CODAS 99)*, pages 1–14. Springer Verlag, 1999.
- [Nie03] M.A.M. Nieto. An Overview of Ontologies. Mxico: Center for Research in Information and Automation Technologies, Interactive and Cooperative Technologies Lab - Universidad De Las Amricas Puebla, 2003.
- [NM01] N.F. Noy and D.L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology. Technical report, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics, 2001.
- [Pre04] A.J. Pretorius. Visualising Binary Fact Types in Ontology Bases. Master’s thesis, Free University of Brussels (VUB), 2004.
- [Rei] R Reiter. *On Conceptual Modelling*, chapter Towards a Logical Reconstruction of Relational Database Theory, YEAR = "1982", pages = "191-233", .
- [SM02] A. Sheth and R. Meersman. Amicalola Report: Database and Information Systems Research Challenges and Opportunities in Semantic Web and Enterprises. *SIGMOD Record: Special Issue on Semantic Web and Data Management*, 31(4):98–106, December 2002.
- [SMJ02] P. Spyns, R. Meersman, and M. Jarrar. Data modelling versus Ontology engineering. *SIGMOD Record: Special Issue on Semantic Web and Data Management*, 31(4):12–17, 2002.

- [TSC01] Y. Tzitzikas, N. Spyrtatos, and P. Constantopoulos. Mediators over ontology-based information sources. In *Second International Conference on Web Information Systems Engineering, WISE 2001*, volume 1, 2001.
- [Usc03] M. Uschold. Where are the Semantics in the Semantic Web? *AI Magazine*, 24(3):25–36, 2003.
- [VA04] S. Van Acker. Information management door middel van ontologies. Nieuwsbrief Faculteit Wetenschappen 20, 7-11, 2004.
- [Wie92] G. Wiederhold. Mediators in the architecture of future information systems. *IEEE Computer*, 25:38–49, 1992.
- [Wor01] World Wide Web Consortium. Semantic Web Activity Statement. <http://www.w3.org/2001/sw/Activity>, 2001.