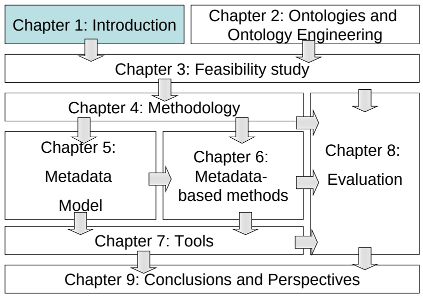
1 Introduction



This chapter gives an overview of our work. We start with an extended description of the area of research to which the solutions elaborated in this thesis are applicable (Section 1.1). The current development trends on the Semantic Web and the experiences in adjacent engineering disciplines motivate the need for the core contributions of this work, which are elaborated in Section 1.2. Our goal is to provide methodological and technological means to enhance knowledge reusability on the Semantic Web. In this context we outline the research questions, describe the research narratives and point out the main contributions of the thesis. The subsequent sections contribute to the localization of the proposed solution in a broader research context; Section 1.3 aims at a clarification of the relevant terminology, while Section 1.4 surveys related research areas. The organization of the remainder of the thesis is elaborated in Section 1.5

1.1 Problem Description

The Semantic Web initiative is developing tools and technologies for a new generation of the World Wide Web [12]. It envisions a machine-processable network of semantically annotated documents and semantics-aware Web Services. For this purpose the Semantic Web community strives the standardization of Web-compatible knowledge representation languages and promotes the deployment and the dissemination of *ontologies*, regarded as a means for a shared understanding of knowledge and a way to (formally) represent real world domains.

By *using* ontologies, services across the Web—be that search engines, shopping bots or travel agencies—can better adapt to the individual information needs of their users, as they are provided with background knowledge in the respective application domain. By *sharing* ontologies, these services are able to communicate with each other. Finally, just as in software or knowledge engineering, *reusing* existing ontologies diminishes implementation costs while improving the quality of the corresponding ontological content. The latter is justified by the fact that reused models are likely to contain less errors, as multiple usage contexts offer more opportunities to detect them.

Ontologies are expected to play a significant role in various application domains on the emerging Semantic Web [60, 205]. Confirming these expectations, to date we record an increasing number of industrial project initiatives choosing to formalize application knowledge using ontologies and Semantic Web representation languages. The emerging ontologies, however, seldom reflect a consensual or at least application-independent view of the modelled domain. Even when using Semantic Web representation languages, these ontologies are, in the sense of knowledge management, simple means to represent knowledge [45], without any claim of being *formal*, *application-independent* or the result of a *common agreement* within a community of practice. Paraphrasing the well-known definition of Gruber [84] they might be indeed *specifications of conceptualizations*, but they are rarely built to be *shared and reused*. Consequently, the relatively large number of qualitatively heterogeneous ontologies modelling the same or related domains of interests, though ubiquitously accessible, are not being used beyond the boundaries of their originating context. Their limited impact seriously impedes the realization of a real-world Semantic Web.

This situation is the result of a series of factors, some of which specific to the Semantic Web context, others related to more general knowledge representation and management problems.

A common pre-condition for the reusability of an ontology is that it should be conceived and developed *independently from its usage context* [86]. Consequently, reusable ontologies tend to be over-generalized and to omit relevant domain knowledge, thus needing considerable modifications before being re-utilized. On the other hand, the more committed an ontology is to a specific domain and task, the less its terminological elements can be generalized and reused beyond this scope [2, 35, 106]. A reusable ontology ideally achieves a balance between the specification and over-generalization without compromising its cross-application usability. Furthermore, the difficulties associated with the realization of a *shared* ontological commitment are inherently related to the fact that the domain experts involved in the conceptualization process posses, even when belonging to the same community of practice, different, personal or organization-specific views upon the domain to be modelled, or are not necessarily willing to exchange and communicate the domain knowledge across the engineering team [14].

A third factor contributing to the described state of the art is related to the still young nature of the Semantic Web field. Even if knowledge representation structures have been around in computer science for many decades, ontologies, and in particular *Semantic Web ontologies*, have emerged so far in stand-alone systems—targeted at closed user communities—or in proof-of-concept applications. In addition to this, adequate methodological and technological support for ontology reuse processes, as for ontology engineering in general, is still under development.

Discovering existing ontologies has been considerably improved in the context of the Semantic Web, which pre-defines the usage of URIs and Web-accessible ontological sources, and with the emergence of ontology-specific search engines such as Swoogle.[1] However, discovering these resources is only the first step towards their real, operationalized reuse: both ontology designers and ontology users need means to *understand and evaluate existing ontologies*, which, as aforementioned, are based on neither application-independent, nor community-agreed conceptualizations. These issues are vital especially for large-sized ontologies, describing knowledge-intensive domains (e.g., in medicine, biology, legislation). In this case the evaluation can not be performed completely manually due to the complexity and the dimensions of the modelled field. In the same time the usage of existing sources in these domains is inevitable because of the high costs associated with a new implementation. A third requirement in this respect is related to reliable methods and tools that are able to *manipulate*

[1] `http://swoogle.umbc.edu/` last visited in May, 2006

and customize existing ontologies in new application settings. Once ontology designers have assigned the relevance of the candidate source ontologies with respect to a set of application requirements, the ones selected for reuse might require some form of customization in order to be integrated into the target ontology. This necessitates not only methods to match and merge overlapping ontologies, but also tools to extract semantically-consistent and complete fragments from ontologies which are only partially relevant to the application setting.

The dimensions of the mentioned problems can be illustrated with the help of a simple application scenario from the medical domain, in which a considerable number of ontologies have already been built and have demonstrated their utility. Representative for this category is UMLS[2], which partially integrates over 100 independent medical libraries and contains over 1,500,000 terms representing approximately 300,000 medical concepts. The size of UMLS is not an isolated case for the medical domain: ontologies like SNOMED,[3] DigitalAnatomist[4] or NCI [78] incorporate tens of thousands of concepts as well. Though these ontologies cover important parts of the medical domain, they can not be directly integrated into a medical information system as stated for instance in [33, 73, 174, 175, 182, 192, 194]. The main reason for this situation is the lack of information and tools that would allow ontology designers to evaluate and adapt the ontologies to particular application needs (such as a restricted application domain e.g., an ontology describing lung diseases, cf. [175]). In order to select the appropriate sources to be reused the engineering team currently has to "read through" over 100 extremely comprehensive ontologies and to gather additional information about them in an uncontrolled, ad-hoc manner. While information about the content and the usage of these ontologies might be available in text form on the Web, the search process is tedious in the absence of systematic presentation means for the potentially large amount of informative material. Provided a list of useful ontologies, the reuse processes goes on with the identification of relevant sub-ontologies and their extraction. Due to the huge size of common medical ontologies, a specific medical application would probably require certain parts of the source ontologies, which are to be integrated into the final one. The identification of the fragments to be extracted is scarcely possible, since there is no available information about the structure and the main parts or about the domain covered by the studied ontologies. This leads to the adoption of domain- and ontology-specific heuristics, whose results can not be easily monitored or evaluated by the domain experts and hence have a low acceptance rate in the application setting [169, 174]. Furthermore, matching, merging and integrating ontologies—tasks which are envisioned to be performed automatically—still necessitate considerable manual pre- and post-processing because of the lack of generality of the existing approaches in this area [173]. They are also confronted with serious scalability and performance problems. We will illustrate these aspects in more detail by means of two case studies in building ontology-based Semantic Web applications in the domains of medicine and human resources in Chapter 3.

[2] `http://www.nlm.nih.gov/research/umls` last visited in May, 2006
[3] `http://www.snomed.org` last visited in April, 2004
[4] `http://sig.biostr.washington.edu/projects/da/` last visited in December, 2005

1.2 Our Approach

1.2.1 Objectives

Committing to the general notion of research as *"activity that contributes to the understanding of a phenomenon"* [119, 124], the first aim of this thesis is to perform a deep analysis of ontology reuse and to identify associated costs and benefits, as well as critical factors and methods that affect its quality or maybe guarantee its success. The results of our case study research complemented by experience reports in adjacent engineering areas indicate that sharing and reusing ontologies can be enhanced by the realization of support tools and methods which enable Semantic Web ontologies to be discovered, evaluated and managed more efficiently. A second aim of this thesis is therefore to conceive an ontology reuse framework which provides

- the methodological support required to systematize and monitor reuse processes, and
- an inventory of methods, techniques and tools contributing to their efficient operationalization.

The results proposed in this thesis are relevant for scenarios in which *available ontologies are reused in the context of a particular application or scope*. Therefore, we aim at methodologies, methods and tools which guide ontology engineers and users under these circumstances by *improving the poor reusability of existing resources*—as opposed to methods addressing the more general (and non-trivial) issue of building new, more reusable knowledge components.

1.2.2 Research Methodology

For the achievement of these goals we organized our research narratives in terms of the general methodology for design science research as illustrated in Figure 1.1 (refer to [99, 217] for a discussion on this topic):

1. **Awareness of problem**: we conducted systematic case study and literature research in order to analyze the requirements of current ontology reuse processes.
2. **Suggestion**: these investigations clearly evidenced that the heterogeneity of ontological sources (with respect to the content, implementation and application dimensions) influences the quality of the outcomes of specific reuse stages and implicitly of the overall reuse process.
3. **Development**: we created a context-oriented ontology reuse methodology and methods and tools to amend the quality of several core reuse activities, both with respect to outcomes and costs. The technological framework *contextually* uses *additional information about* the ontologies involved in the reuse process in order to simplify the access of humans and machines to these resources. We developed a model to represent ontology metadata within information systems and demonstrated its utility within activities like ontology evaluation and matching. Orthogonal to the information model

at the core of the approach, we elaborated methods to create, manage and exchange ontology metadata. The approach is supported by a prototypical implementation.

4. **Evaluation**: We evaluated the various implications of our work according to several evaluation methodologies. Primarily we applied the case study validation approach [231] to compare the operation of reuse processes with and without the proposed methodology in real-world case studies in ontology engineering. We demonstrated that for a class of applications which is considered to be representative for the success of Semantic Web technologies, i.e. information retrieval and semantic annotation [113, 115, 161, 163, 215], our approach improves ontology reuse with respect to the invested efforts, the user-perceived process operation efficiency and the fitness of use of the reuse outcomes in the target application setting. Further on, we situated our research in the broader context of ontology reuse (goal-free evaluation [101]). Different stages of the work were subject to professional reviews [101], which estimated its technical quality and originality. The prototypical implementation was tested on on pre-defined test sets against method-specific quality criteria.

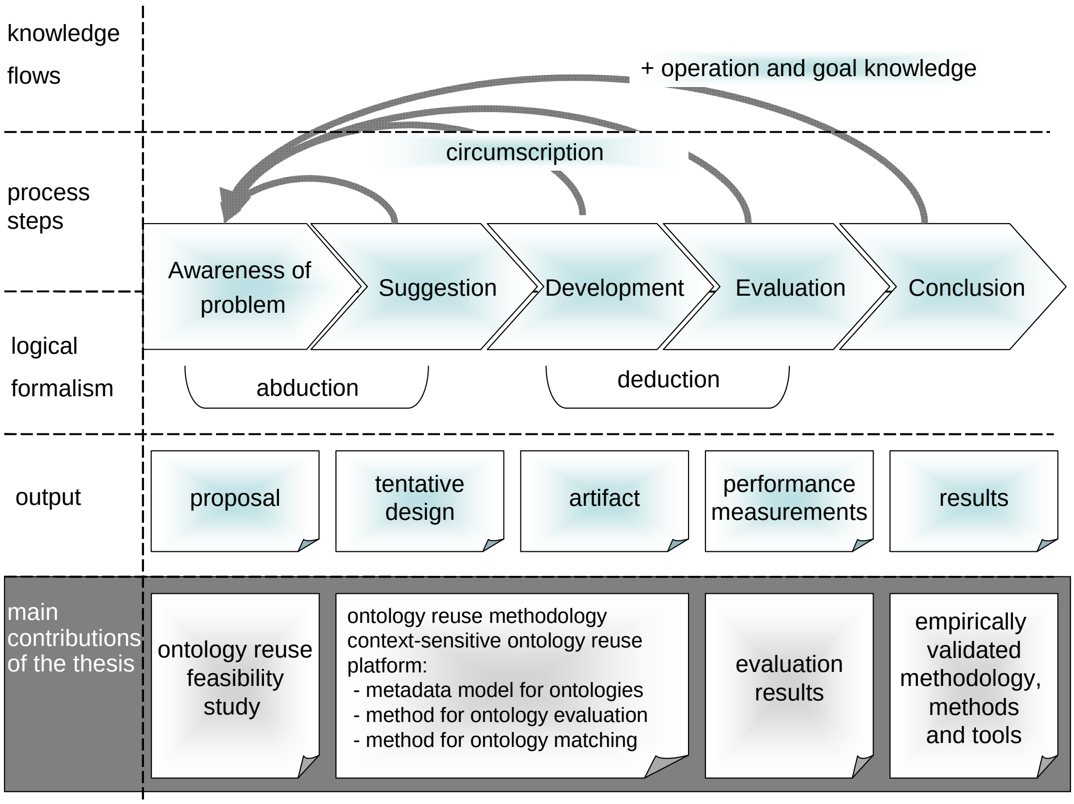


Figure 1.1: Research Narratives of Our Work and Its Results (cf. [217])

1.2.3 Main Contributions

The contributions of this work can be divided into four components:

1. **State of the art**: an in-depth analysis of the feasibility of existing methodologies, methods and tools targeted at the completion of specific reuse activities, concluded by a comprehensive requirements specification.

2. **Methodology**: a detailed account of the characteristics that are important for successfully reusing Web ontologies and a fine-grained description of the process model with an explicit focus on the reuse context and on its tool-supported usability in different user communities. The methodology resulted in an inventory of requirements for methods and tools which might be used to ease its instantiation.

3. **Methods**: a metadata model for the description of ontologies, which has been agreed upon within a representative European Network of Excellence in the Semantic Web area and methods using this model for evaluating and comparing ontologies for reuse purposes.

4. **Tools**: a context-sensitive, extensible reuse platform architecture and prototypical implementations of the proposed methods.

In the remaining of this introductory chapter we give an overview of the thesis-related research areas. After introducing the relevant terminology, we briefly describe the problem field of the thesis, which is reuse and reusability in the particular context of Web ontologies, and continue by summarizing some of the most important initiatives in the solution space of our approach, loosely localized at an interdisciplinary research area studying ways to represent and use context and metadata.

1.3 Terminology

This section explains the way we define and understand the central notions of our work.[5] The relevant terminology can be divided into five categories. As the results of this thesis are applicable to ontology engineering, the first three sub-sections build up the required terms starting from a general engineering perspective and continuing with their meaning in the context of ontologies. Further on, we restrict the scope of the terminology to ontology reuse and introduce our understanding of these notions. The last category is justified by the solution space of our context-oriented approach. We specify how the terms *"metadata"* and *"context"* are used in this thesis, as they have been, also because of their interdisciplinary nature, defined in a plethora of ways in the literature.

[5]With respect to the articulation of these explanations we sometimes resort to existing definitions in the field, which are referenced accordingly. The remaining ones are explained through self-created definitions, without any claim for completeness.

1.3.1 General Engineering Process

Methodology: *"a comprehensive, integrated series of techniques or methods creating a general systems theory of how a class of thought-intensive work ought be performed"* [103]. In particular a methodology includes a description of the process to be performed and of the participants/roles involved in the process, assigns responsibilities to activities and people and gives recommendations in form of best practices and guidelines. Most sciences have their own specific methodology. The term is sometimes used synonymously with *"method"*, particularly a complex method or body of methods, rules, and postulates employed in a discipline.

Method: a codified series of steps taken to develop a product or to perform a service [103]. In computer science its meaning is similar to that of the terms *"technique"*, *"algorithm"*, *"function"* or *"procedure"*, which will be used alternatively in this thesis.

Tool: (software) implementation of a method. From a general point of view the term *"tool"* can be employed synonymously to *"means"* or *"method"*.

1.3.2 Ontologies

Ontology: defines a vocabulary of domain terms and constrains their meaning by indicating how ontology concepts denominated by these terms are defined and are inter-related within a specific domain structure [225].

Ontological primitive: a concept, property, axiom, rule or instance in a knowledge level ontology.

Conceptualization/Conceptual model: an ontological model describing a domain of interest independently of any particular implementation language. When referring to a general *"model"* the meaning of the term is similar to that of *"conceptualization"* or *"conceptual model"*. These will be used interchangeably within this thesis.

Implementation: a conceptualization implemented in a particular knowledge representation language. The representation language does not necessarily posses a machine-understandable semantics, as in the context of the Semantic Web. The implementation can also resort to languages whose semantics is implicit i.e. human-understandable.

1.3.3 Ontology Engineering

Ontology/Ontological engineering: *"the set of activities that concern the ontology development process, the ontology life cycle, and the methodologies, tools and languages for building ontologies"* [80].

Ontology building: the set of activities required to develop an ontology for a particular scope and purpose. This includes manual building, as well as knowledge acquisition and ontology reuse activities.

Ontology reuse: the set of activities required to find and evaluate the application usability of existing ontologies and to integrate the results to the ontology building process.

Ontology evaluation: the process of assessing the general-purpose or application-oriented quality of an ontology.

Ontology merging/integration: given a set of ontologies, the activities required to unify their contents to a single resource. The distinction between merging and integration is defined in [180] with respect to the domains of the ontologies involved.

Ontology matching: the process of finding similarities between (groups of) ontological primitives in an ontology or a group of ontologies. By contrast, *ontology-based matching* can be defined as the process of computing similarities between arbitrary information items with the help of ontologies, which are used to describe these resources.

Ontology alignment/mapping: the process of specifying relationships between matching elements in ontologies. The two terms are used interchangeably in the literature.

Ontology customization: the set of actions required to adapt an ontology to a particular application context. This includes operations at implementation, conceptual and linguistic level: translations to other representation languages, re-labeling of ontological primitives, deletion or revision of concepts, properties, axioms or instances.

1.3.4 Ontology Reuse

Reusability: *"the degree to which a software module or other work product can be used in more than one computing program or software system"*[103].

In a specific context the term is similar to *"usability"* or *"utility"*.

Ontology reusability: the adaptation capability of an ontology to *arbitrary* application contexts.

Reuse: the process of finding, selecting, adapting and integrating existing components to a new usage context.

Ontology reuse: the process in which existing ontological knowledge is used as input to generate new ontologies.

Usability/Utility: the likelihood a component can be re-used in a *specific* application context. This concept can be also termed as *"utility"*. Both terms will be used alternatively in this thesis.

A second sense of the term is related to the ease of use of a system or component. As defined by the IEEE in [103] usability is *"the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component"*. This meaning is, however, only partially relevant for the scope of this thesis.

Ontology usability/utility: the effort required to customize the ontology so that it can be used by humans or machines in a given application context.

Use/Usage: the process of using a component in a new application context (possibly after some customization).

Ontology use/usage: the process in which an ontology is applied to complete particular tasks in an application. Sometimes the process is referred to as *"ontology application"*. In this thesis however, we understand the term *"ontology application"* as the system using the ontology for a particular purpose.

1.3.5 Context and Metadata

Context: the Free Online Dictionary of Computing holds a very general view on the meaning of context. This is considered to be everything *"that which surrounds and gives meaning to something else"*.[6] The term is often used synonymously to *"situation"* or *"setting"*. In this respect the Merriam Webster Online Dictionary explains *''context"* as *"the interrelated conditions in which something exists or occurs"*.[7].

Ontology reuse context: the information space regarding the reuse process (e.g., its stages and participants), the ontologies to be reused and the application setting embedding the final ontology. Which information and how it is being used depends on the information user and the purpose of using it.

Context-sensitive system: a system whose behavior is influenced by contextual information.

Metadata: data about data. In this thesis we differentiate between context and metadata. While metadata objects are primarily static, the decision upon which information can be considered contextual depends on the situation in which the information is being utilized.

Metadata model/ontology: a declarative representation of the metadata information.

Ontology metadata: data about an ontology.

1.4 Review of the Related Research Areas

1.4.1 Knowledge Reuse

Reusability issues have been intensively analyzed in areas like software engineering and knowledge engineering [35, 48, 65, 71, 118, 145, 207]. [207] advances the idea of decomposing knowledge levels into reusable components by means of a componential framework. Building different levels of abstraction and modularization are often recommended across disciplines as means to aid reusability [35, 118]. Increased reusability can also be achieved by isolating thematic clusters or sub-ontologies, or by conceptualizing application-dependent and pure domain knowledge in separate modules. Due to the difficulties related to achieving this goal an alternative to improve the understandability of a specific component to be reused is often considered in the form of various description models, be that contexts, metadata or documentation, whose main role is to offer a compiled representation of the most significant

[6] `http://foldoc.doc.ic.ac.uk/foldoc/index.html` last visited in January, 2005
[7] `http://www.m-w.com/` last visited in May, 2006

properties of the component [29, 77, 117, 118, 145]. Orthogonal to the content issue, the reusability of a knowledge component depends also on technical aspects, such as the representation language used, the readability with respect to the human users and the availability of tools to process and extend it.

This topic has been marginally explored in the ontology engineering community so far. A high-level analysis of the reuse process and the definition of its main stages are provided in [180, 224]. Case studies on this topic are described for example in [5, 81, 187, 222, 224]. Though most of the available ontology engineering methodologies mention the possibility of *reusing* existing knowledge sources in building new ontologies, the reuse and re-engineering of ontologies has been rather poorly explored in this field. For example, Uschold and King [226] describe in detail how to build an ontology from scratch, but on the matter of reusing existing ontologies they only give a general explanation of the approach. Several methodologies address this issue only in the context of ontology customization: extracting relevant fragments of very comprehensive ontologies [179, 216].
Several research initiatives address subjects related to the reuse of ontologies such as *ontology evaluation* and *ontology comparison*. Ontology evaluation is usually considered from an ontological perspective or in conjunction with a specific application system. [114] gives an up-to-date overview of the methods. Ontology comparison approaches rely mainly on basic syntactic features of ontologies e.g., number of classes or on matching algorithms taking into account the vocabulary and/or the taxonomical structure of an ontology [132].

1.4.2 Metadata

The term *"metadata"* has gained increasing popularity with the rise of the World Wide Web. However, the underlying concepts have been in use for as long as collections of information have been organized: a wide range of metadata standards have already emerged in a variety of computer sciences areas, such as digital libraries, image retrieval or data management and maintenance systems [112].

Among the various ways to understand this concept, we find a popular dictionary-oriented definition focusing on the original meaning of the Greek prefix *"meta"* translated as *"among"*, *"together with"* or *"after"*. This emphasizes the relative nature of metadata, which is always used to describe some thing in the world one can talk about. Other definitions are more application-oriented: literally defined as *"data about data"*, metadata answers who, what, when, where, why, and how questions about every facet of the data that are being documented in a consistent and precise format, with the aim of simplifying data users the access to and the interaction with the data [153]. In conjunction with the emergence of the Semantic Web we find a definition according to which metadata consists of a set of independently authored assertions about URI-identifiable resources [122]. In this thesis we rely on the application-oriented view upon this concept and introduce a metadata model for ontologies *for reuse purposes*.

A recent overview of metadata standards for describing electronic information items is provided in [153]. From the impressive number of existing metadata formats we mention some of most important standards capturing information about documents, including those ones relevant for the Semantic Web. The **Dublin Core** metadata standard is a simple, yet effective element set for describing a wide range of networked resources.[8] It differentiates among two levels of detail: simple (with fifteen elements) and qualified, including an additional element, as well as a group of element refinements (or qualifiers) that adapt the semantics of the elements for resource discovery purposes. The international Text Encoding Initiative **TEI** developed guidelines to mark up textual documents.[9] The schema contains basic bibliographic information, as well as details about the way the text was transcribed, its revisions etc. The Metadata Encoding and Transmission Standard **METS** extends the scope of the described objects and captures information about complex digital library contents.[10] It includes structural and technical metadata.

The **Reference Ontology** [5] is a domain ontology that gathers, describes and has links to existing ontologies. Its focus is to characterize ontologies from the user point of view in terms of property-value pairs. The Semantic Web search engine **SWOOGLE** [49] makes use of an implicitly defined metadata schema, which covers information that can be extracted automatically from ontology implementations. Further on, the issue of creating metadata for ontologies is addressed by various ontology repositories initiatives. However, the majority of these repositories rely on a restricted, implicitly declared vocabulary, whose meaning is not machine-understandable. The **DAML ontology library** is a catalog of DAML ontologies that can be browsed by different properties.[11] The **FIPA ontology service**[212] defines an agent wrapper of open knowledge base connectivity. The **SchemaWeb Directory** is a repository for RDF schemas expressed in RDFS, OWL and DAML+OIL.[12] Metadata for ontologies has been superficially considered at language level for OWL and RDFS [22, 176]. Accounting for the need for a dedicated model [112] analyzes several general-purpose metadata models in order to derive an inventory of content and technical requirements which should be fulfilled by an ontology metadata schema.

1.4.3 Context-sensitive Computing

Contextual aspects play a decisive role in a variety of disciplines, including the topic of knowledge reusability in computer science and beyond. In this regard the function of context is a descriptive one, since this type of information, just as metadata, is recognized to be useful in simplifying the user interface to the components it is associated to. The primary distinction to metadata information is localized in the dynamics and the structure of the contextual information. While metadata are static information structures used to describe a specific item, context can be defined only in relation to a situation [21, 213]. Consequently, given an item and the associated information about this item i.e. the metadata, the question of which parts of the metadata belong to a context can not be answered without additional knowledge about tasks and users.

Although the importance of context and contextual reasoning has been often emphasized in different research communities, there is no common definition for this concept. Previous definitions of *"context"* have emerged in cognitive psychology, philosophy and also in spe-

[8] `http://dublincore.org/` last visited in December, 2005
[9] `http://www.tei-c.org` last visited in April, 2006
[10] `http://www.loc.gov/standards/mets` last visited in April, 2006
[11] cf. `http://www.daml.org/ontologies/` last visited in May, 2006
[12] `http://www.schemaweb.info` last visited in May, 2006

cific areas of computer science, especially when dealing with natural language processing or automatic reasoning. Artificial intelligence research adopted three main directions when dealing with context: it can be either a situation in the general sense of the term, a part of a knowledge base or both of them [7, 90, 137]. However, since these definitions have proved to be hardly applicable in other computer science domains, numerous context-sensitive applications define context by enumerating potentially application-relevant information sources or by choosing synonyms for context (as in the area of mobile computing [190]). A primary concern in mobile computing is *context-awareness*, i.e. the awareness of the physical environment surrounding users and their mobile devices. Both in real time systems and mobile applications the focus is placed on the acquisition and processing of contextual information, while a clear definition of context is of secondary importance [36]. In the Semantic Web community the most frequent definitions regard this concept as a specific ontology standing for a domain of discourse, as a view upon a domain or as a mediation layer between ontologies—positions compliant to previous approaches of contextual/local reasoning. We conclude this section with an (incomplete) definition of context provided by the Free On-line Dictionary of Computing[13]: context is *"that which surrounds and gives meaning to something else"*. This definition, without claiming to tap the full potential of the underlying concept, represents quite well the way humans think about context apart from any scientific considerations. Besides, it outlines two aspects that are significant to classify information as *"contextual"*.[14] The term *surrounds* reflects the relative nature of context. Context is information/knowledge which can be used *additionally* to perform some task. The second part of the definition mentions the term *meaning*, as a synonym for *relevance*. Since every available piece of information could be considered contextual [20], additional parameters are required to specify the *relevant* fragments. The current task and the relevance criteria imposed by the user of the context operate as filters to distinguish between *related* information and *relevant related* information about the context target. Some of the mentioned context definitions are used in our work in order to formally represent contextual information about ontology reuse processes, and theoretically ground the underlying conceptual model.

1.5 Organization of the Thesis

The remainder of this thesis is organized as follows:

Chapter 2 Ontologies and Ontology Engineering: in this chapter we elaborate on the notion of *"ontology"*. We pinpoint the major research results achieved in the last decades in this community, while emphasizing the areas of *"ontology engineering"* and *"ontology-based applications"*. We give an overview on the most important methodologies and methods targeted at building and managing ontologies and on the mainstream applications of Web ontologies, including the Semantic Web, since knowledge on these two dimensions is fundamental in order to detect current pitfalls in *"ontology reuse"* and to understand how this situation could be improved through our approach.

[13] `http://foldoc.doc.ic.ac.uk/foldoc/index.html` last visited in May, 2006

[14] These two parameters are not the only choice in the attempt of defining context, but they are significant in order to identify contextual information and to differentiate among types of contexts.

Chapter 3 Ontology Reuse Feasibility Study: following the considerations from the previous chapter, we describe the current state of the art in ontology reuse and highlight the most important reuse practices having emerged in the last decades. We further exemplify this state of the art with two self-conducted extensive case studies in the areas of eHealth and eRecruitment. Within these application scenarios we aimed at developing Semantic Web ontologies for different scopes and purposes by resorting to existing ontological knowledge on the Web. We scrutinize the requirements identified within the case study examinations to cope with the limitations of existing technologies and techniques. In particular, we emphasize the need for a context-sensitive treatment of ontologies, both from an engineering and a usage perspective, and typical phases of reuse processes which could considerably profit from such an approach.

Chapter 4 Ontology Reuse Methodology: in this chapter we introduce the methodological backgrounds for our approach to ontology reuse. Building upon the findings of the feasibility study we identify a set of application scenarios which are expected to take benefit from using ontology-driven technologies and describe the way reuse-oriented ontology engineering should be operated under these circumstances. For each process stage we introduce the participants and their activities, as well as the expected results and required technological means. The chapter concludes with an inventory of requirements for methods and tools which are likely to ease the methodology application. Besides conventional ontology management tools like ontology repositories, ontology search engines, ontology editors or matching and merging services, we claim the need for a systematic representation of ontology metadata. The availability of a uniform structure to summarize the most important features of Web ontologies has a direct impact on the human-perceived understandability of the ontological content, facilitates its evaluation and forms the basis for novel support methods and tools.

Chapter 5 Ontology Reuse Metadata Model: accounting for the need for a systematic and in the same time user-friendly access to and interaction with existing ontological resources on the Web, we elaborate a formal metadata model for capturing reuse-relevant information about ontologies. We describe the process through which the metadata model was developed, and its future directions of evolution as part of the ontology metadata model **OMV**.[15]

Chapter 6 Metadata-based Methods Supporting Ontology Reuse: in this chapter we demonstrate the usage of the proposed metadata model as an enabling technology for particular reuse activities. The first section of the chapter is dedicated to the topic of ontology evaluation. Though this task is not intended to be performed automatically, it can take advantage of operationalized decision-support heuristics. The method we propose is semantics-driven; evaluation criteria referencing concepts in the metadata ontology are aggregated to a weighted semantic query which is processed with the help of reasoning services and ontology-based similarity measures on the metadata repository in order to detect potentially adequate reuse candidates. The second section sketches how ontology metadata can be used to overcome some of the limitations of current ontology matching approaches.

[15] `http://omv.ontoware.org` last visited in May, 2006

Chapter 7 Tools Supporting Ontology Reuse this chapter describes the implementation of the proposed methods. Firstly we introduce the tool PROMI (**P**latform for the **R**euse of **O**ntologies through **M**erging and **I**ntegration), which prototypically realizes the metadata-based methods for ontology evaluation and merging/integration, respectively. The platform provides a context-aware behavior to ontology reuse in that it adapts the presentation and content of particular reuse-relevant methods depending on pre-defined user groups, application scenarios and ontology characteristics. Using a declarative model to represent contextual information it offers a simple revision and extension mechanism by which new methods and metadata models can be included to the platform. Secondly, we introduce OntoMeta, a service for the automatic acquisition of ontology metadata. The tools are evaluated on pre-defined test cases by aligning the results of the methods to human judgement.

Chapter 8 Evaluation: this chapter describes the results of the validation of our approach to ontology reuse. The first section introduces the evaluation methodology, which includes three evaluation strategies. The remaining sections describe the evaluation process and its results, summarize the lessons learned and scrutinize possible directions of research.

Chapter 9 Conclusions and Perspectives: we close this thesis by revisiting the contributions made. We analyze their impact in the Semantic Web field, present several extensions of our work, which are currently being developed, and speculate on further research directions.

The content of the thesis and the way it is structured are depicted in Figure 1.2. The figure contains the 9 chapters of the thesis and their core parts (represented by rectangles). Dependencies between chapters or chapter parts within or across chapters are represented by block arrows. Within the same chapter arrows symbolize a logical dependency between chapter parts.

Within the thesis, every chapter is preceded by an introductory overview briefly explaining the goal of the chapter and the contents of each section. This overview is accompanied by a simplified version of the picture above highlighting the current chapter and how it fits in the overall thesis.

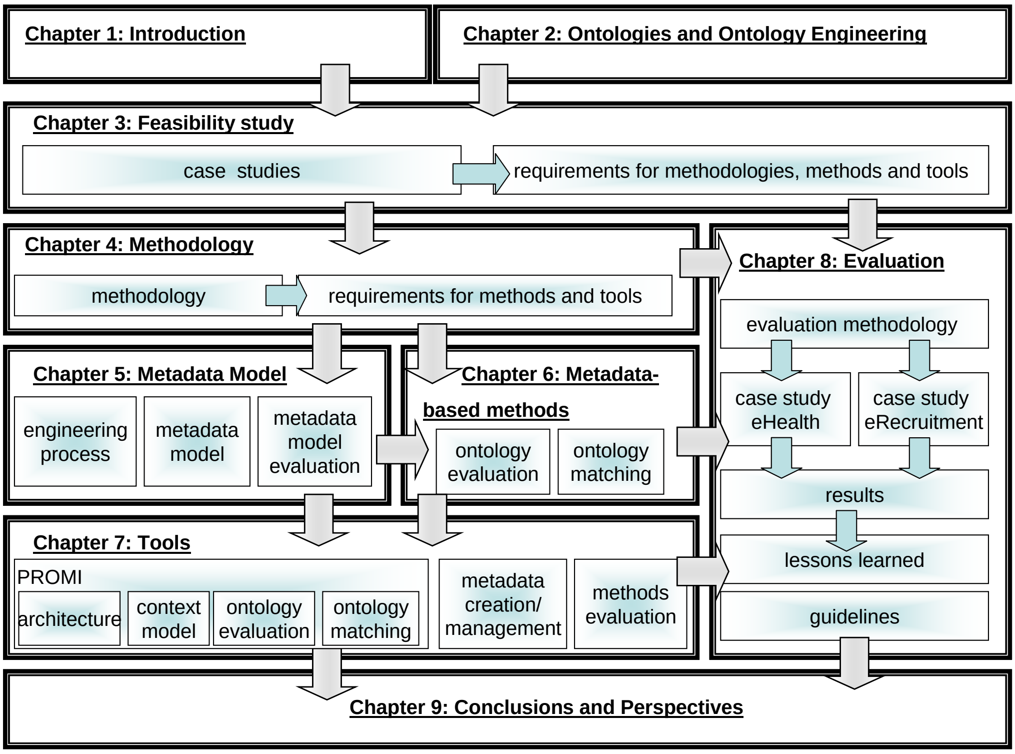


Figure 1.2: Organization of the Thesis