Knowledge Engineering in the Context of Human-Computer Interaction

Author: Ștefan Negru

Supervisor: Professor Dr. Henri Luchian
Knowledge Engineering in the Context of Human-Computer Interaction

Abstract

This thesis contributes to both Knowledge Engineering and Human-Computer Interaction research with important results in regards to three main directions: (1) Knowledge Based Systems, (2) Knowledge Visual Representation and (3) Knowledge Quality Evaluation.

Therefore a new approach which seeks to adapt Semantic Web technologies to the Human-Computer Interaction field is proposed. With the purpose of providing a knowledge-based approach on modeling the User-Centered Design process, this thesis focuses on two of its main components: the first one utilized for communicating user requirements (the Persona method) and the second one for evaluating a product’s ease of use (the Usability Testing method). It discusses the strengths and weaknesses of providing a conceptual model as a basis for structuring, extracting and linking data collected via these two methods. The approach consists of a set of HTML5 Microdata schemas and an OWL specification, both of them including concepts and properties used to model persona and usability testing domain.

In addition we proposed a novel visual notation for ontologies, which accommodates graphical representation for elements of the Web Ontology Language. Such a visual notation has an appealing potential when it comes to exploring, verifying and ultimately making sense of semantic data and the ontology that describes it. We investigated the potential issues and challenges of such a notation in two user studies. While the first study was focused on analyzing the graphical notation the second one sought to provide insight on how both knowledge engineer experts and domain experts perceive the structure of an ontology and the domain it describes, in the process of selecting representative concepts.

The last part of this thesis approaches particular Knowledge-Based Systems deployment scenarios, while practical applications
of Semantic Web data visualization are explored.

As a result the solutions we presented in this thesis should be of value to both Knowledge Engineering and Human-Computer Interaction research communities. It is also hoped that our work can provide the necessary foundation for further investigation and development of Human-Computer Interaction Knowledge Based Systems and Semantic Web visualization-centered applications.

**Keywords:** Human-Computer Interaction, Semantic Web, Knowledge Engineering, Knowledge Visualization, User-Centered Design, Description Logic, Microdata, OWL, Ontologies, Personas, Usability Testing, Evaluation, Knowledge Discovery
List of Published Articles

Original contributions included in our thesis [16, 17, 18, 19, 20, 25, 24, 23, 22, 21], in chronological order, are:


Other published articles, that are not part of this thesis:


Our experience as a Teaching Assistant for courses such as *Human-Computer Interaction* (2011,2012,2013), *Client-Side Web Application Development* (2011), *Network Technologies* (2011), as a reviewer for conferences such as International Conference on Web Engineering 2013, International Conference on Semantic Systems 2013, as well as a visiting researcher within the *Institute for Visualization and Interactive Systems (VIS)*, Stuttgart Germany contributed to reaching our research goals and ultimately finalizing our doctoral thesis.
Knowledge Engineering (KE) [4, 14] is a Computer Science field, focusing on the development of either software or hardware systems dedicated to solving tasks with an increased level of difficulty, which otherwise would have required human proficiency. Albeit this aspect does not imply that this specific type of systems are designed to replace humans, but rather to aid them in completing their tasks and increasing their overall performance.

Knowledge Engineering is commonly regarded as a multidisciplinary field, bridging concepts and methods from several domains such as Artificial Intelligence, Databases, Decision Support Systems, Information Retrieval and many more. Therefore researchers have adopted KE “best practices” in many domains (hardware manufacturing, decision support system, collaborative system, power grid solutions etc.), with the purpose of designing and developing systems able to facilitate communication and interaction between humans and machines, but also among humans. The purpose of this thesis is to bring KE to the Human-Computer Interaction field with the objective of augmenting various methods and processes, while in return utilizing them in the KE field.

Human-Computer Interaction [2, 10] is viewed as an interdisciplinary field, utilizing concepts and methodologies form Computer Science, Cognitive Science, Visualization, Design, Sociology and others. In addition to that, because it studies the “partnership” between humans and machines, it must embrace and support both perspectives, providing a broader vision in terms of
practical applications, but also in terms of research directions – main reasons why we decided to apply our research (in Knowledge Engineering) to this field.


2.1 Structure

The thesis is organized around seven individual chapters.

The first chapter presents a brief overview of the research domain along with some of the most important principles. At the same time we point out the novelty of the thesis by highlighting its objectives, its structure and our contributions.

Chapter 2 establishes the general framework for our thesis by presenting the fundamental formal/theoretical and practical notions of both Knowledge Engineering and Human-Computer Interaction. Emphasis is placed on Semantic Web technologies, Description Logic and establishing the background for optimizing knowledge discovery and management in a HCI context.

The Semantic Web (SW) [1, 11, 27] and Semantic Web technologies can be regarded as an extension of the Web, offering a new approach to managing information and processes, based on creating, categorizing and utilizing machine-readable semantic metadata [6].

In regards to our objectives, it provides the technical framework for bridging machine knowledge with human knowledge, by enabling machines to understand complex human requests and take into account both their background knowledge and also the environment in which they make the request. By making use of
such SW technologies we expect to improve not only how information is presented but also how information is integrated from heterogeneous sources.

An example which reflects the way information is processed and presented using the formal notions from SW is presented below.

**Example 2.1.1.** A PhD. student which is advised by a professor, has published conference papers and journal articles.

Such knowledge can be described using Description Logic, as follows:

\[ \text{Student} \sqcap \neg \text{Undergraduate}(\exists \text{hasAdvisor}.\text{Professor}) \sqcap (\forall \text{hasPublications}.(\text{ConferencePaper} \sqcup \text{JournalArticles})) \]

Chapter 3 proposes a series of methods and models for managing information and discovering knowledge in the User-Centered Design process.

User-Centered Design (UCD) can be considered the most frequently used process from HCI, and since its introduction in 1985 by Gould et al. [9] it is regarded as a fundamental approach on designing usable products and systems, by placing the user at the center of the techniques, procedures and methods utilized in the system design process – see Figure 3-1. Still, incorporating the user preferences proves to be more challenging than it seems. One of the key aspects is achieving a shared understanding between a user and a computer on how their communication should take place – and ultimately how the interface should look and feel.

Our aim in this third chapter is to explore new directions by focusing upfront on modeling two widely used HCI methods: Personas and Usability Testing. Both of them are part of a User-Centered Design (UCD) process: one is used for communicating user requirements, whereas the other one is used to evaluate a
product’s ease of use by observing the users behavior. Although both methods are used to collect data regarding user needs, preferences and behavior, little research has been conducted linking the collected data with other pieces of data concerning the product development and design process.

Figure 2-1: User-Centered Design process overview according to ISO 13407 (ISO/IEC 1999), image adapted from [13]

Chapter 4 presents a visual notation for the Web Ontology Language (OWL), that provides an integrated view on the classes and individuals, which are part of that ontology. The notation makes use of a small number of graphical elements and it is aimed at visualizing the property relations of either the classes or some selected individuals from an ontology. A comparative test has shown the notation to be suitable for an in depth understanding of additional characteristics of property relations specific to a
certain ontology.

Visualizations have an appealing potential when it comes to exploring, verifying and ultimately making sense of semantic data and the domain that encompasses it. Although visualizing Semantic Web data and ontologies have been a frequently discussed topic in literature, existing solutions are limited by their context of use or objective.

Moreover, the majority of existing approaches represent concepts and instances in a mutually exclusive manner, often lacking an integrated representation of the ontology. Separating between concepts and instances raises the level of abstraction and requires the users to be familiar with ontology-related formalisms. Furthermore even the underlying DL framework and OWL specification point out that both of them need to be considered when in any knowledge representation.

Less experienced users who would greatly benefit from an effective and uniform visual representation for ontologies [7].

Example 2.1.2. Following the context of Example 3.1.1, we devised an example represented below and graphically in Figure 3-2, on how such a notation can be utilized.

A *ConferencePaper* is a subclass of *Paper*, with the restriction that is presented at exactly 1 *Conference* (which coincidently is a subclass of *Event* – although not reflected in this example). The *presentedAt* property: domain *ConferencePaper* or *WorkshopPaper* and the range *Conference*. This is reflected in listing below.

```plaintext
Class: ConferencePaper
    SubClassOf: Paper
    presentedAt
        exactly 1 Conference
ObjectProperty: presentedAt
```
Ontology based systems have attracted considerable attention in the recent years, as ontologies became a fundamental model for representing knowledge within a domain. Still one of the challenges of designing such ontology-based systems is choosing the right ontology from a high number of ontologies that describe the required domain or a similar one. When faced with such a difficult decision users (both expert and non-expert) seek to understand and compare ontologies – a complicated task considering that some of them contain hundreds to tens of thousands (or even higher) of concepts and relations. Thus, the need for methods and techniques that provide users with a summarized version of the ontological model and to aid them in making sense of the domain, becomes an important aspect. Furthermore, being able to grasp the inner structure represents a step towards assessing an ontology’s quality and it raises its chances of reuse and further development.
Chapter 5 addresses the challenge of identifying key concepts in ontologies. Determining these concepts which offer an accurate summarization of both the modeled domain and the overall structure of the ontology is essential for evaluating the quality of an ontology. State-of-the-art solutions are reviewed in order to identify underlying characteristics of key concepts. At the same time we report on the results of our own empirical user study on how both knowledge engineer experts and domain experts select representative/important concepts.

A main advantage of our approach takes is that it takes into consideration the user’s domain and knowledge expertise and in order to identify and rank key concepts according to his understanding of a certain ontology.

In Chapter 6 we study some of the scenarios in which the techniques and approaches we presented in this thesis can be applied. At the same time we present some of the applications which contributed to the development and optimization of the solutions proposed in our thesis.

In the last chapter we expose a set of conclusions and future directions of research based on the ideas presented in our thesis.

2.2 Contributions

The primary objective of our thesis is to enhance various processes and methodologies from the Human-Computer Interaction (HCI) field (such as Persona, Usability Testing, generation of adaptive User Interfaces etc.), by applying knowledge engineering methods and technologies. We then proceed to further observe and analyze the impact of applying these engineering methods and technologies, in order to gain insight on some of the issues and challenges and propose feasible solutions for them.
As a secondary objective we focus on utilizing these enhanced processes, methods and methodologies in order to optimize and augment knowledge extraction and visualization, as well as knowledge quality evaluation.

Following our objectives, we identified and addressed three main research areas: Knowledge Engineering, Knowledge Visual Representation and Knowledge Quality Evaluation, all of them particularized to the Human-Computer Interaction field.

Overall our contributions can be summarized as follows:

• the PersonasOnto ontology along with the Personas and Usability Testing extensions to schema.org, utilized for extracting structured data from the User-Centered Design process;
• semantic annotated templates for the Persona and Usability Testing methodologies;
• a visual notation for complete and comprehensive graphical representing of OWL ontologies – which follows the Descriptions Logics separation of concepts, properties and relationships;
• insights into designing and developing practical applications for visualizing Semantic Web data;
• an investigation into identifying key concepts from ontologies, supported by an empirical user study.

The present thesis is partly developed and structured on the articles highlighted prior to this chapter. On top of these publications, a series of resources, documents and applications are available on the Web for public (re)use:

• Personas Template –
  http://blankdots.com/open/personas/;
2.3 Knowledge Engineering Perspective

As previously stated, our main objective is focused on optimizing and augmenting HCI processes by taking advantage of Knowledge Engineering techniques and methodologies.

Hence a first step would be to have a clear understanding of how knowledge is defined (in regards to the Computer Science field), what it consists of and how can we extract it from any given data source. While this first step is centered around the understanding phase, the second step would be to determine the practical (and useful) implications of using knowledge engineering techniques as means to enhance and refine existing processes and methodologies (while these techniques can be generalized, in our thesis we particularly focus our research towards the Human-Computer Interaction field).

Consequently, these two steps draw our attention on the first three levels of the DIKW (Data, Information, Knowledge and Wisdom) model proposed Liebowitz et al. [15] – see Figure 3-3 (the figure reflects our own perspective/adaptation of the DIKW
model) – more precisely on the Data, Information and Knowledge layers. By directing our (knowledge) engineering process onto these layers, we aim to propose models of knowledge discovery and visualization, but also models for knowledge quality evaluation.

Figure 2-3: The DIKW model adapted from [3, 15].

One hypothesis of our DIKW model adaptation is the existence of multiple data sources (much like humans store and organize data/information), and although most of these data sources are diverse, some of them share certain features (e.g. numerical data, textual data etc.), therefore the Data layer could be partitioned into several data stores. Overall the main source of
data could be regarded as an Environment – digital (e.g. Web of Data) or physical (real world resources generating data – ‘Internet of Things’ [12]).

“Knowledge is a fluid mix of framed experience, values, contextual information, expert insight and grounded intuition that provides an environment and framework for evaluating and incorporating new experiences and information” [5]. Hence knowledge is defined by the complexity of experiences, but also by the fact that is can be expressed as a pattern, which ultimately will become embedded in certain processes, practices and routines performed by a certain (specialized) category of users.

Although the relationship between data, information, knowledge and wisdom and the conversion process appears to be under a bottom up approach (the pyramid model in Figure 3-3), it does not imply that the transition from data to information, or from data to knowledge and so on is unidirectional. Instead this transition is cyclic in its nature as old data, information, knowledge and even wisdom is constantly (re)used.

2.4 Human-Computer Interaction Principles

Whereas Knowledge Engineering deals with computational intelligence, Human-Computer Interaction deals with human intelligences. Our goal is to combine the best of these two fields by analyzing the impact of this interdisciplinary research and at the same time finding methodologies for integrating them into practice.

In broader terms, HCI can be viewed as the field centered around the study of the interactions and relationships between humans and computers [10]. Another characteristic of HCI is that it is a multidisciplinary field, combining the expertise and
research from many fields, as Helander et al. [10] point out.

In its earlier history, HCI was focused on designing interfaces (graphical user interfaces – GUIs – using windows, icons, menus, and pointing devices – WIMPs), and as Fischer [8] argues next step was to “improve the way people use computers to work, think, communicate, learn, critique, explain, argue, debate, observe, decide, calculate, simulate, and design”. Thus a focus on the understanding and justifications of the actions/tasks (overall the interactions), not just the interface – this translates to building systems that are tailored in accordance to the user’s needs.

Figure 2-4: The knowledge-based HCI model, image adapted from [8].

For this reason Fischer [8] presents a knowledge-based (see Figure 3-4 diagram) HCI model, which has the goal to facilitate the use of sophisticated interaction techniques and processes (e.g. Ubiquitous Computing interfaces [28], touch devices interaction,
sound and haptic interfaces etc.), by focusing on the machine-user (knowledge-based) communication.

At its core the model identifies two types of communication channels: a basic *explicit communication channel* that characterizes the relationship between humans and machines – it features the interaction techniques with a certain type of user interface (graphical, natural, haptic etc.), and the *implicit communication channel* – between a human area of expertise and the knowledge base contained within a machine [8].

Concurrently users also want to ‘make sense of the data’ processed in such a knowledge-based system, with the purpose of having a clear overview over future improvements. This need can be managed by an immediate connection with both the (digital) data and the (physical) environment data they generate – with regards to their privacy. Chapter 4 seeks solutions to this issue by exploring and analyzing an approach to knowledge visual representation.
Bibliography


65
[23] Stefan Negru and Sabin Buraga. A knowledge-based approach to the user-centered design process. *Communications in Computer and Information Science*, Accepted paper:TBD, 2013. (Cited on pages 1, 23, and 45.)


