BASIC ARCHITECTURE OF THE „SEMANTICALLY ENRICHED E-WORKING PLATFORM“

Thomas Kern[a], Gerald Petz[b]

[a] University of Applied Sciences Upper Austria, School of Informatics/Communications/Media, Softwarepark 11, A-4232 Hagenberg, AUSTRIA
[b] University of Applied Sciences Upper Austria, School of Management, Wehrgrabengasse 1-3, A-4400 Steyr, AUSTRIA

[a] thomas.kern@fh-hagenberg.at, [b] gerald.petz@fh-steyr.at

ABSTRACT
The term crowdsourcing denotes the outsourcing of tasks for division of labour to a large number of individuals who offer their skills, capacities and intelligence via the internet. Currently these are mostly intrinsically motivated amateurs, who often complete diverse jobs with no or low remuneration. The main intention of the project Semantically Enriched E-Working Platform (SEEWOP) is to make crowdsourcing to an attractive offer for enterprises as e-working method for special fields of activity in addition to conventional working methods. For this purpose Web 2.0 will be combined with Semantic Web technologies, enabling the annotation of tasks and their specifications, e-workers and their qualifications as well as results and their quality in machine processable form. Therefore, adequate vocabularies and ontologies have to be designed to enable automatic and intelligent division of tasks within the e-worker community. Additionally, a semantic description of processes and rules for routine jobs will be developed.

Keywords: e-work, crowdsourcing, semantic web

1. INTRODUCTION
In June 2006, Jeff Howe as one of the first authors used the term crowdsourcing in the Wired Magazine (Jeff Howe 2006) and argued that the concept of crowdsourcing mainly depends on the fact that it is an open call to an undefined group of persons, and those win, who run tasks best, solve complex problems and contribute the most relevant and freshest ideas. For example, the public will be invited to develop a new technology, to implement a new design (community-based design, design by democracy), to refine or perform an algorithm (human-based computation), or to help to collect large amounts of data, to systematize or analyse (citizen science). Therefore, crowdsourcing has become synonymous for the cooperation of the masses enabled by Web 2.0.

Viewed from the perspective of the research field of e-mobility, these platforms provide an interesting potential that can bring benefits to so-called e-workers as well as to businesses by transition to flexible employer-employee relationships. E-Mobility in this context is not understood in terms of electrical mobility using electric or hybrid vehicles, and even not primarily in terms of wireless communication with mobile devices such as smart phones – although these may of course play a role in the technical implementation –, but refers to the execution of tasks and business processes via the internet, both, to minimize the need for physical mobility to increase energy efficiency in times of rising energy costs, and on the other hand, to allow greater flexibility of individuals with regard to time and place of work performance and an improved accordance of work with leisure and family.

The European Commission already determined in 1998 that in future a hybrid form of work will establish itself, in which physical work and telecommuting – employees do at least some of the work outside the premises of the employer – are indistinguishable (Bundeskanzleramt Österreich 2010).

As part of our research, we deal with e-mobility in the context defined above. To answer the research question, *"What can be done to implement e-mobility for all involved parties (e-managers, e-workers and their environment) successfully?"*, an integrated software solution will be designed that allows both, companies (requesters) and individuals (employees, contractors), to cope with the new challenges encountered in planning, coordination and communication in business as well as in the private environment. The management of e-work shall be enabled by an intelligent crowdsourcing platform – the *Semantically Enriched E-Working Platform (SEEWOP)* – which supports the management of complex tasks by semantic technologies.

2. STATE OF THE ART
For over a decade, on the one hand crowdsourcing established itself as one of the key Web 2.0 applications in many different areas. On the other hand the World Wide Web Consortium (W3C) promotes the development of the Semantic Web, to improve interoperability between different Internet data sources by means of controlled vocabularies, and to allow automated pro-

cessing of tasks that were previously subjected to human users only. These two components form the basis for the SEEWOP system architecture and are therefore described in this section.

2.1. Crowdsourcing Platforms

The world's largest collection of knowledge on the Internet *Wikipedia (http://www.wikipedia.org/)* is one of the first crowdsourcing platforms emerged. Since 2001 thousands of people gather daily information and make their contributions freely available to the public (Miscellaneous Authors 2012).

Another typical example follows a completely different intention. The web-based job-marketplace *Amazon Mechanical Turk (http://www.mturk.com)* exists since November 2005 and allows requestors (companies) to find those people who are able to perform tasks that cannot or just poorly be solved by software (e.g.: transcription of audio or video files, data verification or cleanup, generation of large test data sets, sentiment analysis, translation, etc.). These tasks are called *Human Intelligent Tasks (HITs)* and are divided into very small sections. Their processing requires very little time, so usually compensation is only a few dollar cents per HIT (Amazon Mechanical Turk 2012).

However, there are other platforms that use crowdsourcing for the provision of more complex, collaborative creative services, usually in the form of competitions, from concept development to the design of technical equipment. Thus since 2000, on the website *Threadless (http://www.threadless.com/)*, hundreds of users daily face up voluntarily and without charge to a weekly design competition to let their own T-shirt design ideas be assessed by a community. Winning designs that are actually produced may expect a cash prize by the company. However, the real winner is the company itself that only brings products to market, which enjoyed a great response in advance. Thus, according to the Forbes business magazine, in 2009 the company was able to generate a sales volume of 30 million $ by 50 employees. The about 1.8 million members of this internet community have two important functions: they act both as an outsourced design department, continuously developing new designs virtually free of charge and secondly, they are the customers who enthusiastically buy those T-shirts (Burkitt 2010).

Meanwhile, also crowdsourcing platforms for scientific issues and for the development of innovative solutions are emerging. This group, for example, includes the portal InnoCentive (http://www.innocentive.com/), a mediation platform for scientific services existing since 2001. Under the self-chosen slogan "Challenge-Driven Innovation" the platform primarily addresses scientists and technicians (so-called problem solvers) to apply with their innovative solutions to problems described by companies in the context of challenges. Selected solutions are rewarded with a previously announced amount. According to information of the operating company currently more than 250,000 solvers from 200 countries are available (InnoCentive 2012).

These four examples show the wide range of activities that are now covered by crowdsourcing. They illustrate but also accept that the great advantage of such platforms is currently closer to the requesters who, compared to conventional employment relationships, can generate a multiple of content, ideas and solutions at low costs, while e-workers usually can expect only a very small reward for their efforts.

2.2. Semantic Web

Sir Tim Berners-Lee applies with his definition "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." (Berners-Lee et al. 2001) as founder of the Semantic Web. His vision is that electronic content (the Web) can be read and processed not only by humans but by machines. The World Wide Web Consortium (W3C) defines the Semantic Web as follows: "The Semantic Web is a 'web of data' that enables machines to understand the semantics, or meaning, of information on the World Wide Web" (World Wide Web Consortium 2012) and defines with the so called Semantic Web Stack (see Figure 1) a collection of technologies to store data, to build vocabularies, and to describe rules for querying and automated reasoning.

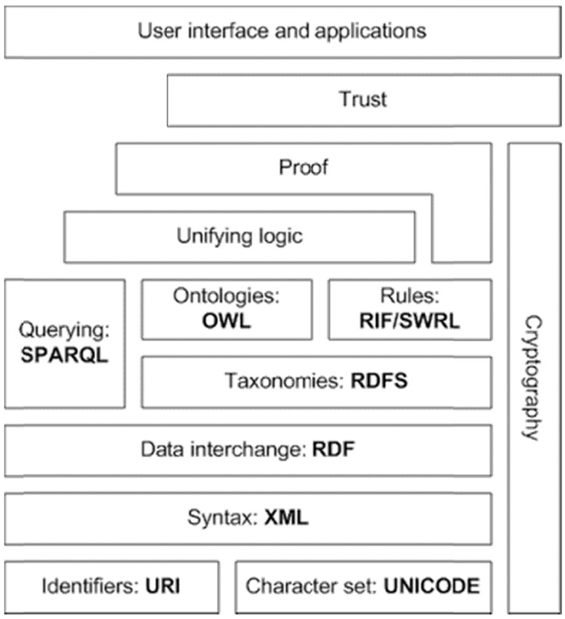

Figure 1: Semantic Web Stack (Obitko 2007).

The Semantic Web Stack basically consists of three layers which are still in development:

1. **Hypertext Web Technologies:** *Uniform Resource Identifiers (URI)* provide a clear identification of resources (e.g. HTTP, FTP or e-mail addresses) within the World Wide Web. *Unicode* is an international standard character set eliminating incompatibilities between different languages and countries. With the *Extensible Markup Language (XML)* text data can be represented in a hierarchically structured

form. The primary benefit of XML is to provide a platform, programming language and operating system-independent exchange of data between different applications.

2. **Standardized Semantic Web technologies:** the *Resource Description Framework (RDF)* enables the description of (true) statements about possible resources in the form of a directed graph, which is formed by subject-predicate-object triples. Through this formal representation, the information described and evaluated are machine read- and understandable. *RDF Schema (RDFS)* provides a basic vocabulary for RDF. Thus, for example hierarchies (taxonomies) are defined by classes and properties. The *Web Ontology Language (OWL)* extends RDFS to describe the semantics of RDF statements. It is based on description logic and enables the use of automated reasoning mechanisms. The *RDF Simple Protocol and Query Language (SPARQL)* is used to retrieve information from RDF-based data (including RDFS and OWL).
3. **Not standardized Semantic Web technologies:** By the aid of the *Semantic Web Rule Language (SWRL)* and the *Rule Interchange Format (RIF)* rules can be defined and exchanged with other rule-based systems. These technologies are especially relevant to relations that cannot be defined by the OWL description logic but only be evaluated by inference at run time. The *Trust* block on top of the Semantic Web Stack attempts to verify statements on their trustworthiness. This on the one hand can happen within the block *Cryptography* e.g. by digital signatures ensuring that information comes from trusted sources and, on the other hand, using formal logic during the investigation of new information within the block *Proof*. For the development of applications within the block *User Interface and Applications* different libraries in diverse programming languages are available by now.

3. BASIC ARCHITECTURE

To describe how a typical crowdsourcing platform architecture will be enriched with semantic technologies, a classical crowdsourcing workflow is shown in Figure 2. The workflow is started by a requester (a company) splitting its scope of work into small Human Intelligence Tasks (HITs) and applying them to the crowdsourcing system. In addition, qualification tests for more complex tasks can be created, which first have to be passed by the interested e-workers to be allowed to perform the corresponding HITs. The HITs are then offered to the e-worker community. It is also possible that the requester assigns HITs to experienced e-workers directly. The solutions produced by the e-workers are processed back to the system and assessed by the requester. If the requester accepts the result, a corresponding compensation is performed in form of money or other output-related incentives.

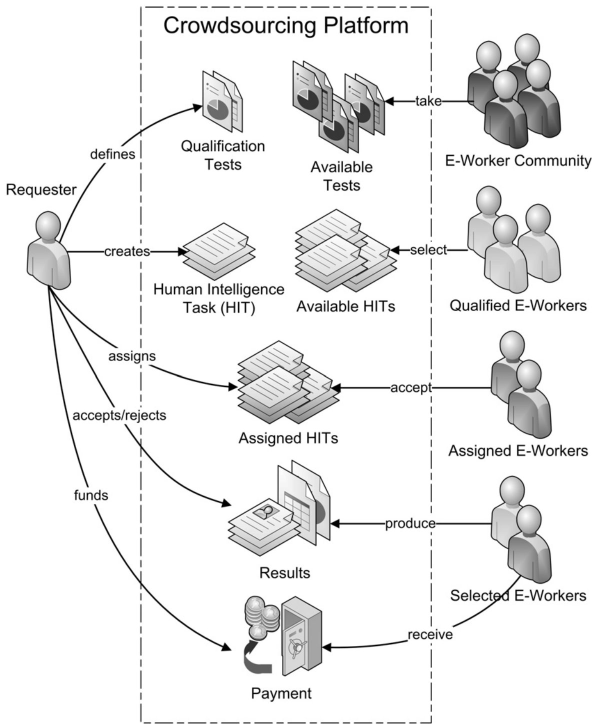

Figure 2: Crowdsourcing Workflow.

Especially for the requester a significant administrative burden arises from managing HITs which can be improved by a pervasive enhancement of crowdsourcing platforms through semantic technologies.

3.1. Semantic Enhancements

The biggest challenge for the requester is to divide complex tasks into separate HITs so that they can be distributed and processed through the crowdsourcing system. Especially workflows, which often have to be performed in a similar, but infrequently in the same way within a company (e.g. the preparation of offers for customers, the procurement of equipment or specific customization of an ERP system) cannot yet be assigned via crowdsourcing, because of the fact that the effort to define, manage and control the HITs and the merging of the individual results would be too big.

In order to make this execution possible and especially to support the requester systematically, an automated creation of HITs based on an adaptable process definition or a set of rules shall be supported. A modelling language called Event-driven Process Chains (EPCs) allows a graphical representation of business processes in principle, but for a further automatic execution it lacks of a semantic specification of the EPC-model elements themselves. The potential to solve this problem arises from ontologies.

According to Tom Grubers definition an ontology is an explicit specification of a conceptualization (Gruber 1993), i.e. an abstract representation of the real world. Ontologies also contain inference and integrity rules. The definition of an EPC-ontology using OWL offers many further advantages such as the possibility to query process models at the semantic level (using SPARQL), the reasoning of new facts by inference mechanisms (reasoners) at query time, or the possibility for user-defined enhancements of existing ontologies to include additional statements and rules (using SWRL) (i.e. to comply with law regulations or to describe technical details of the implementation of a process) (Thomas and Fellmann 2006).

In the same manner HITs resulting from the process can be represented. Their representation should be understood clearly and easily not only by humans (e-workers), but also be offered as automated and targeted as possible to the "appropriate" e-workers (assigned e-workers) to achieve a high quality solution. As a result of these requirements it is necessary to describe both the requirements of HITs and the qualifications of an e-worker, along with other information such as her/his availability, scheduling, etc. in the form of so-called semantic e-portfolios (using RDF/S) . Therefore, in particular the requirements for flexible working hours are taken into account. Even the quality of the results assessed by the requester should be part of e-portfolios, so they can on the one hand be considered in the selection of e-workers and, on the other hand, can be used to perform focused skill trainings for the e-workers, and thus to continuously improve the work environment.

To realize the SEEWOP software as a first step existing crowdsourcing platforms are now being investigated for their suitability and extendibility. Then, the required ontologies to describe the individual knowledge components will be implemented in an iterative development process by the aid of Computer Aided Knowledge Engineering (CAKE) tools and gradually integrated into the selected platform. This knowledge models and the detailed system design will be described in detail in a dissertation in progress and several accompanying publications.

4. SUMMARY

The paper describes the proposed SEEWOP system architecture as a combination of classic Web 2.0 with Semantic Web technologies. The enrichment using annotations, defined by vocabularies and ontologies in RDF/S and OWL, provides a machine-processable description of tasks, job profiles and e- worker portfolios. Processes and rules are defined in OWL and SWRL and can be evaluated at run time by inference mechanisms and SPARQL queries. Together with optional assessments for the results an intelligent management of e-work via a crowdsourcing platform is provided.

ACKNOWLEDGEMENTS

The described research activities are part of the research project 4EMOBILITY, carried out at the University of Applied Sciences Upper Austria (UAS), School of Management in Steyr together with the School of Informatics, Communications and Media in Hagenberg. The project is financed by the European Regional Development fund (EFRE/Regio 13) as well as by the Federal State of Upper Austria.

REFERENCES

Amazon Mechanical Turk (2012) Amazon Mechanical Turk. https://requester.mturk.com/. Accessed 17 Feb 2012

Berners-Lee T, Hendler J, Lassila O (2001) THE SEMANTIC WEB. Scientific American May 2001 (Vol. 284 Issue 5): 34–44

Bundeskanzleramt Österreich (2010) Wirtschaft und Arbeit (e-Business/e-Work). http://www.austria.gv.at/site/cob__9064/4544/default.aspx. Accessed 09 Jun 2010

Burkitt L (2010) Need To Build A Community? Learn From Threadless - Forbes.com. http://www.forbes.com/2010/01/06/threadless-t-shirt-community-crowdsourcing-cmo-network-threadless.html. Accessed 13 Feb 2012

Gruber TR (1993) A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2): 199–220

InnoCentive (2012) Facts & Stats | InnoCentive. http://www.innocentive.com/about-innocentive/facts-stats. Accessed 13 Feb 2012

Jeff Howe (2006) The Rise of Crowdsourcing. http://www.wired.com/wired/archive/14.06/crowds.html. Accessed 13 Feb 2012

Miscellaneous Authors (2012) Wikipedia. http://en.wikipedia.org/wiki/Wikipedia. Accessed 15 May 2012

Obitko M (2007) Semantic Web Architecture - Introduction to ontologies and semantic web - tutorial. http://obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html%20Semantic%20Web%20Architecture. Accessed 13 Feb 2012

Thomas O, Fellmann M (2006) Semantische Integration von Ontologien und Ereignisgesteuerten Prozessketten. In: Nüttgens M, Rump Frank J., Mendling Jan (eds) EPK 2006 Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten. Proceedings, pp 7–23

World Wide Web Consortium (2012) Semantic Web - W3C. http://www.w3.org/standards/semanticweb/. Accessed 13 Feb 2012

AUTHORS BIOGRAPHY

Thomas Kern is head of the Research Center Hagenberg, School of Informatics, Communications and Media, University of Applied Sciences Upper Austria. He finished his studies in Software Engineering in 1998. After industrial work experience he started as research associate and lecturer for University of Applied Sciences Upper Austria in autumn 2000. Since 2003 he conducted several application oriented R&D projects in the field of Computer Science focusing on design and implementation of knowledge based systems. Since 2010 he is member of the 4EMOBILITY research project.