

MASTER THESIS PROPOSAL

An Architectural Design Approach for Software Platforms to Improve Knowledge Sharing and Self-Organization in Crowd Worker Communities

Julia Filler (1225408, E066 926)

supervised by
Ao.Univ.Prof. Dipl.-Ing Dr. Stefan Biffl

July 7, 2017

1 Problem Definition

For more than a decade, crowdsourcing has established itself as a popular form of digital work. Initially mentioned by Howe [1] in the term is a *portmanteau* of the terms crowd and outsourcing. In detail, the crowdsourcing mechanism works as follows [2]: A requester (a company) offers (many little) tasks (*microtasks*) via a digital platform (crowdsourcing platform) to anonymous workers ('the crowd'), who perform these tasks for a compensation. These tasks can be either location-independent (cloudwork) or location-dependent (gigwork) [3]. In recent years, the term crowdwork has been used more often to refer to crowdsourcing, as more work-related aspects have become particularly dominant in the societal discourse. On the positive side crowdwork is beneficial for workers and requesters: Workers can easily search for jobs online, and take on work when and how much they like. Requesters can offer jobs on the platform in an automated way, and can offload risks and costs to the workers. The crowd workers can decide which job they want to do and companies can choose who should do the task. Crowdwork is a socio-technical application domain that can be investigated from numerous scientific fields and perspectives, whereby two of those perspectives are particularly interesting.

Fristly, from a societal perspective: A recent study [4] on the crowdwork landscape in Austria funded by *Austrian Chamber of Labour* identified the following negative aspects: Workers only get paid for the output regardless of the actual time invested in finishing the task. It may happen that the worker does not get paid though he made the work, because the requester did not accept the work product. Also, crowd workers predominantly work as independent contractors, thus they need to pay social security themselves. In another recent work Risak and Lutz [5] investigated the legal challenges of crowdwork. Among many aspects, they identified a lack of transparency of internal platform mechanisms (like reputation systems) as particular issues.

Secondly, from a technical perspective: In the past, platforms have been proposed to support crowd workers to organize [6] and share knowledge [7]. However with respect to software architecture, these platforms typically are designed in an ad-hoc manner and are poorly documented. This leads to error-prone designs and implementations, but most importantly, it makes it difficult for other software architects and designers to systematically assess, replicate and adapt the designs of such platforms for their system-of-interest and their crowdwork community, its needs and concerns.

To address these technical limitations, this thesis presents design principles and criteria that help software architects and systems designers of collaborative platforms to support crowdwork communities on the societal level with transparency and knowledge sharing capabilities. The presented principles address conceptual and architectural aspects that are specific to the crowdwork domain. Furthermore they are technology and implementation-agnostic, thus are not prescriptive to a particular application framework or platform. We apply and evaluate the identified principles in the context of a prototype system, called CrowdSOS together with stakeholders from Austrian worker organizations and subject matter experts from the crowdwork industry.

The remainder of this work is structured as follows: Section 2 describes expected results and the research questions that should be answered in the context of this work. Section 3 details the methodological approach, and Section 4 discusses the state of the art. Finally, Section 5 discusses how the presented project is related to the curriculum

of business informatics.

2 Expected Results

The overall goal of this thesis is identifying a set of software architectural principles and mechanisms (architectural patterns, reference models, domain-specific viewpoints or description languages) that are appropriate and that support software architects in the design and technical realization of socio-technical software platforms that support crowd workers with lightweight knowledge sharing and self-organization capabilities. A major benefit of this approach is, that it supports software architects with design and decision support of assistive digital platforms that improve the transparency and collaboration among various crowd worker communities and domains.

This thesis contributes (1) a catalogue of software architectural principles and criteria for collaborative platforms in the domain of crowdwork and (2) a prototype called *CrowdSOS*. The prototype will be developed in a case study to evaluate the catalogue and its usefulness to the stakeholder groups of crowd workers and worker organizations. In this work we address the following research questions:

RQ 1: What are relevant requirements, capabilities and business processes of software systems that support crowd workers and worker organizations in knowledge sharing and self-organization?

In order to answer RQ 1, we will first conduct a literature review in the area. The findings of the review will inform the elicitation of the needs of the crowd workers, worker unions and overall system requirements, capabilities, and, in particular, a comparative analysis of existing platforms and alternatives.

RQ 2: What are major conceptual software architecture design principles of social-collaborative platforms for crowdworkers?

RQ 2 contributes to improving the understanding of conceptual architecture principles for collaborative platforms in the domain of crowd work with a particular focus on transparency and information sharing among worker communities. The results elicited for RQ 1 will provide the foundation for the definition of a catalogue of design principles, success criteria and quality attributes.

RQ 3: Do the identified principles support the design of collaborative crowd-worker platforms?

We answer RQ 3 from a software architect's perspective and a user perspective in the context of a prototypical system that realizes the main principles, criteria and quality attributes identified in RQ 2. From the software architect's perspective we will evaluate how the identified principles support the architecture design process of a platform with regards to utility, applicability, and completeness. Secondly, from a user perspective we will evaluate how well the created system, and in particular the components that are related to the design principles address the quality attributes and key performance indicators associated with the needs and business processes of key stakeholder groups.

3 Methodological Approach

The methodological approach to fulfill the research questions will consist of the following steps:

1. **Literature Review**

The first step is to get an overview about the topic of social media architecture and on the other hand to get deeper into the world of crowd working. That implies a research about the use of different types of computer-supported cooperative work (*CSCW*)-like systems. The goal is to get an insight view of all relevant fields which is the foundation for the following steps.

2. **Elicitation of Crowd Worker’s Needs and System Requirements**

In this step, the needs of the crowd workers will be identified. In addition input will be collected from worker unions and subject matter experts of crowdwork. A core aspect of this step is to collect information about what system capabilities will be suitable for certain stakeholder groups regarding the identified needs. The focus in this part is on a comprehensive set of stakeholder, concerns and processes from studies and interviews. The consolidated results from step 1 and 2 are going to answer RQ 1. On the basis of these findings, system requirements will be determined.

3. **Creation of Design Principle Catalogue**

Based on step 1 and 2 the creation of design principles catalogue will be derived. Catalogue documentation is compliant with the well-established ISO/IEC/IEEE 42010:2011 standard on software architecture descriptions. Being standard compliant is an important criteria for the catalogue, as this enables it of being integrated along-side other more rigorous architecture approaches (like architecture frameworks, domain-specific languages or model-driven engineering methods) in the future.

4. **Architectural Design of CrowdSOS Prototype**

The applicability and usefulness of the proposed catalogue from a software architecture perspective will be investigated in a case study of the CrowdSOS platform. CrowdSOS is a social platform that enables crowd workers to share work related information (profitable tasks, routes), but also to report and collect experiences about misconduct, work quality and labour rights related issues. In this step the design principles catalogue from step 3 will be applied in conjunction with architecture approaches like domain-specific architecture frameworks and viewpoints. Elicitation of requirements that the system-to-be should satisfy. Comparative analysis of alternatives and existing systems.

5. **Technical Implementation of CrowdSOS Prototype**

Based on the architecture description from step 4, the prototype will be implemented using an agile software development process and state-of-the-art web development frameworks (e.g. Ruby on Rails) and cloud-based SaaS infrastructure (e.g. Heroku).

6. Evaluation

The CrowdSOS prototype will be empirically evaluated in a qualitative user study in order to identify completeness, utility and usefulness in addition to specific quality attributes in order to answer RQ 3. For the design of the user study we follow the seminal guidelines for conducting case study research in software engineering of Runeson *et al.* [8]. The planned target groups will consist of stakeholders from the groups of crowd workers and worker organizations, as well as subject matter experts from the field of crowdwork. Depending of the results of RQ 2, a set of quality attributes [9] (e.g. performance, interoperability) and measures associated with the respective attributes will be derived. The study procedure, quality attributes, measures and the procedures how trace evidence is analyzed will be reported in the study protocol, which follows the checklists for reading and reviewing case studies provided by Runeson *et al.* [8].

7. Refinement of principles catalogue

Finally, based on the findings of the evaluation will be integrated in the prototype. This prototype will be a platform which covers the needs of the crowd workers and fulfills required system specifications.

4 State of the Art

In the last few years, the topic of crowdworking gained in importance. Recent studies and papers **concentrate on** this new kind of work. Schmidt [3] **analyzes** the aspects of crowdwork and gigwork. The paper **concretizes** the field of cloud-, crowd- and gigwork and their relationship to employment protection and social insurance and resulting calls for action in politics.

Leimeister *et al.* [10] **deal with the** world of the crowd workers and their environment. The main focus lies on the analysis of the working environment and the crowdsourcing market. Additionally, several crowdsourcing platforms are emulated and questionnaires are created. Based on these questionnaires the crowd and findings will be evaluated.

Gray *et al.* [11] **define** the crowd in their papers a collaborative network. This is because the crowd workers are no independent, autonomous workers by themselves. They need other workers to fulfill technical and social needs. Another paper from Gray and colleges [12] **builds up** on the paper mentioned before and goes deeper into a specific crowdsourcing platform. They are mapping the worker's entire communication network on Amazon Mechanical Turk¹. Whiting *et al.* [13] **focus on crowd** guilds and how they help crowd workers. They found out that it is very important for crowd workers to interconnect because then they can learn from each other and have an impact to the platform they are using. **And there is also a recent study from** the Austrian Chamber of Labour that helps to get an understanding of the crowd working scene in Austria and the crowd worker's satisfaction [4]. Alkhatib *et al.* [14] **examine** parallels between crowdwork and gigwork and piecwork in the 20th century. They take a closer look on scholarship on piecework because nowadays in context of on-demand work there is a similar trend of work decomposition, distribution and payment. **This paper presents** a theoretical foundation of the most persistent questions in crowdwork by doing a research

¹<https://www.mturk.com/mturk/welcome>. Accessed 26.5.2017

of the mechanism that enables and limits piecework historically. Also they identify pitfalls and differences on on-demand work.

Irani and Silberman [7] analyze the technical relationship between the worker and the requester on Amazon Mechanical Turk. This is a popular human computation system and it relies on worker invisibility. In course of this assumption they introduce Turkopticon² as a response to the hazards of crowdwork. Turkopticon helps the worker to exchange information about the requester like the willingness to pay. They identified the challenges of the developing of Turkopticon and detect that these challenges are the same on developing real-world technologies that intervene in existing, large-scale sociotechnical systems.

Salehi *et al.* [6] engage with Amazon Mechanical Turk, too. They studied challenges to collective action efforts in relation to online labor. From an ethnographic fieldwork, they understand the workers barriers to collective action. In the study they talked to workers about the relationship to other workers and the collective action. The result is a framework called Dynamo³. This platform should help workers who have the aim for action and change. To support such collectives better, they also identified stalling and friction as twin failures and introduced a set of behaviours to keep this collective together.

A particular topic of interest in socio-technical systems is the concept of collective intelligence, whereby Bonabeau [15] discusses context and processes. Bonabeau's work analyzes whether collective intelligence leads to better decisions and discusses key issues which are followed by implementing such a system. Another focus lies on the mechanism design which decides for example if every person has an equal voice etc.

Musil *et al.* [16] deal with the architecture of collective intelligence systems. This article describes how effectively the framework supports stakeholder in the context of CIS concepts, applying the *stigmergy* coordination model, which deal with the indirect communication between individuals and getting started with a CIS in their organization. All these current publications analyze the hazards and problems of crowd working. There are also studies where on the workers and their working condition. The presented papers makes clear that there is a problem and gain more attention to societal stakeholders (politicians, unions) and do more work to support the workers to self-support them. This is the foundation of this thesis. The prototype which will be developed, should support the workers in knowledge sharing and self-organization. In addition researcher gain insights form the prototpye case about potentials and challenges, when addressing a societal problem form a software engineering/architectural perspective..

5 Relation to the Curriculum of Business Informatics

This thesis is highly relevant in the context of the study of Business Informatics. The aim of the module "Advanced Software Engineering" is to gain in-depth knowledge of designing and building an advanced software system [17]. This is relevant to use a CIS because these lectures provide a summary about different architecture styles and an overview on selected research and industrial topics [17]. The goal in the this project is to build a usable prototype [17]. Crowd working is a new form of organization. The module "Organization" deals with question on how organization can be designed and how do

²<https://turkopticon.ucsd.edu/>. Accessed 26.5.2017

³<http://www.wearedynamo.org/>. Accessed 26.5.2017

they interact in their environment [17]. It helps to understand complex organizational behavior and imparts knowledge in organizational theory and in the management of people and organizations [17]. “Service and Communities” is also a module which is important for developing a prototype in the context of CIS and social media. The lectures in this module focuses on the history and success criteria of social media and on products by applying social media [17]. A skill acquired in this module is the analysis of complex service systems which is relevant to this thesis for developing the system requirements of the prototype [17]. The basis of the prototype is to find out what the crowd worker needs on the one hand and on the other hand to define the optimal system requirements. Therefore, it is relevant to analyze existing crowdsourcing platforms. In the module “Web: Analysis and Search”, social network analysis is conducted and current trends in personalization and recommendation are illustrated [17].

6 References

- [1] J. Howe, “The Rise of Crowdsourcing,” *Wired Magazine*, vol. 14, no. 06, pp. 1–5, 2006. [Online]. Available: <http://www.clickadvisor.com/downloads/Howe{-}The{-}Rise{-}of{-}Crowdsourcing.pdf>
- [2] E. Simperl, “How to Use Crowdsourcing Effectively: Guidelines and Examples,” *LIBER Quarterly*, vol. 25, no. 1, pp. 18–39, 2015.
- [3] F. A. Schmidt, “Arbeitsmärkte in der Plattform-ökonomie–Zur Funktionsweise und den Herausforderungen von Crowdwork und Gigwork,” Tech. Rep., 2016. [Online]. Available: <http://library.fes.de/pdf-files/wiso/12826.pdf>
- [4] U. Huws, S. Joyce, and University of Hertfordshire, “Österreichs crowdworkszene,” Tech. Rep. September, 2016.
- [5] D. L. Martin Risak, *Arbeit in der Gig-Economy*. ÖGB Verlag, 2017.
- [6] N. Salehi, L. C. Irani, M. S. Bernstein, A. Alkhatib, E. Ogbe, K. Milland, and Clickhappier, “We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers,” in *Proceedings of the ACM CHI’15 Conference on Human Factors in Computing Systems*, vol. 1, 2015, pp. 1621–1630. [Online]. Available: <http://dx.doi.org/10.1145/2702123.2702508>
- [7] L. C. Irani and M. S. Silberman, “Turkopticon: Interrupting worker invisibility in amazon mechanical turk,” in *Proceedings of the SIGCHI Conference on ...*, 2013, pp. 611–620. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2470742>
- [8] P. Runeson, M. Host, A. Rainer, and B. Regnell, *Case Study Research in Software Engineering: Guidelines and Examples*, 1st ed. Wiley Publishing, 2012.
- [9] L. Bass, P. Clements, and R. Kazman, *Software Architecture in Practice*, 3rd ed. Addison-Wesley Professional, 2012. [Online]. Available: <http://www.amazon.com/Software-Architecture-Practice-Edition-Engineering/dp/0321815734>
- [10] J. M. Leimeister, D. Durward, and S. Zogaj, “Crowd Worker in Deutschland,” Tech. Rep. 323, 2016. [Online]. Available: <http://www.boeckler.de/pdf/p{-}study{-}hbs{-}323.pdf>

- [11] M. L. Gray, S. Suri, S. S. Ali, and D. Kulkarni, “The Crowd is a Collaborative Network,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 2016, pp. 134–147. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2818048.2819942>
- [12] M. Yin, M. L. Gray, S. Suri, and J. W. Vaughan, “The Communication Network Within the Crowd,” in *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, 2016, pp. 1293–1303. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2872427.2883036>
- [13] M. E. Whiting, D. Gamage, S. S. Gaikwad, A. Gilbee, S. Goyal, A. Ballav, D. Majeti, N. Chhibber, A. Richmond-Fuller, F. Vargus, T. S. Sarma, V. Chandrakanthan, T. Moura, M. H. Salih, G. B. T. Kalejaiye, A. Ginzberg, C. A. Mullings, Y. Dayan, K. Milland, H. Orefice, J. Regino, S. Parsi, K. Mainali, V. Sehgal, S. Matin, A. Sinha, R. Vaish, and M. S. Bernstein, “Crowd Guilds: Worker-led Reputation and Feedback on Crowdsourcing Platforms,” 2016. [Online]. Available: <http://arxiv.org/abs/1611.01572> <http://dx.doi.org/10.1145/2998181.2998234>
- [14] A. Alkhatib, M. S. Bernstein, and M. Levi, “Examining Crowd Work and Gig Work Through The Historical Lens of Piecework,” in *Chi 2017*, 2017.
- [15] E. Bonabeau, “Decisions 2.0 : The Power of Collective Intelligence,” *MIT Sloan Management Review*, vol. 50, no. 2, pp. 45–52, 2009.
- [16] J. Musil, A. Musil, D. Weyns, and S. Biffl, “An Architecture Framework for Collective Intelligence Systems,” in *Proc. of the 12th Working IEEE/IFIP Conference on Software Architecture (WICSA'15)*. IEEE, 2015, pp. 21–30. [Online]. Available: <http://dx.doi.org/10.1109/WICSA.2015.30>
- [17] “Studienplan (curriculum) für das masterstudium business informatics an der technischen universität wien,” <http://www.informatik.tuwien.ac.at/studium/angebot/studienplaene/wirtschaftsinformatik/MasterstudienplanWinfOktober2013.pdf/>, 2013, [Online; accessed 23 April 2017].