A First Account on Stigmergic Information Systems and Their Impact on Platform Development

Juergen Musil[1], Angelika Musil[1], Dietmar Winkler[2], Stefan Biffl[2]
Christian Doppler Laboratory for Software Engineering Integration for Flexible Automation Systems
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11/188, 1040 Vienna, Austria
[1]{jmusil, angelika}@computer.org, [2]{dietmar.winkler, stefan.biffl}@tuwien.ac.at

ABSTRACT

Modern software platforms often depend on networks of user-generated content. Although such platforms are very popular, their architectural concepts and underlying system dynamics are not yet fully understood. This paper uses stigmergy, a form of indirect communication and self-organization, as basis for framing certain groups of software services in order to propose an architectural concept of stigmergic information systems and describe identified key fields (actor base, architecture, software ecosystem) influencing them. Thus enabling the creation of better solution stack configurations needed to support effective and efficient development of such services.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information networks*; H.3.5 [**Information Storage and Retrieval**]: On-line Information Services—*Web-based services*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Collaborative computing, Web-based interaction*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors, Human information processing*

General Terms

Design, Human Factors, Theory

Keywords

Collective intelligence, human computation, self-organization, software architecture, software ecosystems, stigmergy

1. INTRODUCTION

SaaS platforms involving contributions of connected users have been a hallmark technology of Web 2.0 proofing worthwhile and sustainable. Concepts like crowdsourcing, collective intelligence, network effects and user-generated content thus have subtly transformed our daily life. We follow the latest news on Twitter[1], get smartphone apps via an app store, stay connected with friends on Facebook[1], browse critiques of local businesses on Yelp[1] or look up facts in Wikipedia[1]. The inception of these platforms and their wide adoption among users' willingness to contribute within these services have brought an age of collective sharing. Although these services are very popular, their architectural concepts and underlying system dynamics are not yet fully understood. The authors observed a conceptual gap when building such services due to a lack of design principles and an architectural model that acts as basis for these services. In order to build more worthwhile and sustainable services and identify solution stack configurations needed to support effective and efficient development of such services, there is a need to better understand the internal processes and involved information structures of systems that shape these platforms. Aforementioned points provided the impetus for this paper which introduces stigmergic information systems (SIS).

The remainder of this paper will propose an architectural concept of SIS and describe identified key fields influencing them. Finally, it concludes and illustrates future research work.

This paper intends to provide first insights into stigmergic information systems and to show the potential of investigating the software architectural anatomy of these systems more precisely.

2. SWARM INTELLIGENCE

The term "architecture of participation" [18] coined a characteristic of software platforms which enabled their users to create and share content among each other (prosumer communities [20]) with the platform acting as the facilitator of this process (e.g. Wikipedia, Pinterest[1], YouTube[1]). An underlying key concept of these systems is self-organization which describes the phenomenon that "macroscopic patterns emerge out of processes and interactions defined at the microscopic level" [3, p.6]. Self-organization has been extensively investigated in computer science eventually leading to the research field of swarm intelligence. Swarm intelligence focuses on studying "large collections of relatively simple agents that can collectively solve problems that are too complex for a single agent or that can display the robustness and adaptability to environmental variation displayed by biological agents." [9]. There are two basic forms of swarm intel-

[1]www.{name}.com (all URLs last visited 2012/07/15).

WICSA/ECSA 2012. August 20-24, Helsinki, Finland.

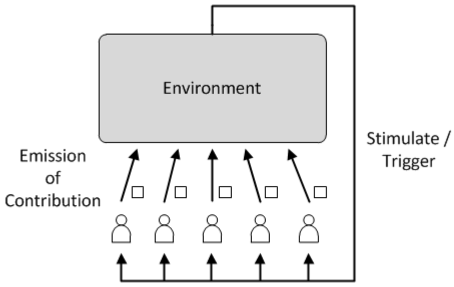

Figure 1: The stigmergy meta-process as reciprocation of environment configuration change and triggering of subsequent actor actions.

ligence which are swarm formation and stigmergy. Swarm formation focuses on the creation of groups and collective movement e.g. particle swarm optimization [12].

Stigmergy (from Greek *stigma*: sting, and *ergon*: work) is a concept of indirect interaction [3] originally describing the dynamics of construction task coordination and regulation of termite nest architectures [10]. Bonabeau et al. note on stigmergy that "the coordination/regulation of building activities do not depend on the workers themselves but are mainly achieved by the nest structure: a stimulating configuration triggers the response of a termite worker, transforming the configuration into another configuration that may trigger in turn another (possibly different) action performed by the same termite or any other worker in the colony" [3, p.14]. Figure 1 shows the *stigmergy cycle* as indirect interaction and communication of individuals by modifying their environment which in turn stimulates responses by others [3].

Due to the circumstance that stigmergy does not provide "detailed mechanisms" by which the coordination has to be achieved [3, p.16], it can be used as a meta-process model for explaining situations where indirect interaction between actors over an environment takes place, and where this environment has a pro-active regulatory influence over subsequent actions of the involved actors, thus leading to emergent behavior of the overall system. Nowadays stigmergy in virtual space has gained particular importance, since it can be observed that major "prosumer"-centred platforms imbue stigmergic dynamics e.g. Wikipedia, eBay, Amazon [18, 21]. Architectures of participation find a strong ally in stigmergy, since stigmergy provides the meta-process model which is reproduced by these architectures. The authors suspect that such systems possess definite, characteristic architectural traits, which are inevitably necessary to maintain the stigmergic cycle's flow, thus proposing the term *stigmergic information system* (SIS) for software systems imbuing these traits.

Above paragraphs have introduced to the concepts of self-organization and stigmergy as a meta-process which can be found in physical and contemporary virtual environments as well as introduced the term of stigmergic information system which will be investigated in depth in the following section.

3. STIGMERGIC INFORMATION SYSTEM

Over the last decade there has been some amount of research in the field of virtual stigmergy, but not with regards to aspects of software architecture. This section describes the architectural concept and essential elements of stigmergic information systems which remodel the stigmergy cycle within digital information systems.

A stigmergic information system is a software platform, which facilitates the building of an information network by allowing actors to create/modify network elements and thereby share information among each other. Hence a SIS is a combined communication system [14] and information regulation system. The central principle behind SIS is harnessing collective intelligence by stimulating, aggregating, leveraging, and distributing user contributions.

A stigmergic information system consists of the three big key fields actor base, architecture and software ecosystem. Figure 2 illustrates the architectural concept: The actor base is responsible for contributing data into the system. The architecture instruments the actors' contributions in order to stimulate a subsequent action by other actors. This interplay leads to the flow of a typical stigmergy cycle. The architecture provides interfaces to allow external software services to work with the data in the information network in a limited way. SIS provide benefits for five major self-organization concepts as described by Floreano et al. [9]: (believe) aggregation [7], division of labor, clustering [1], collective construction and foraging.

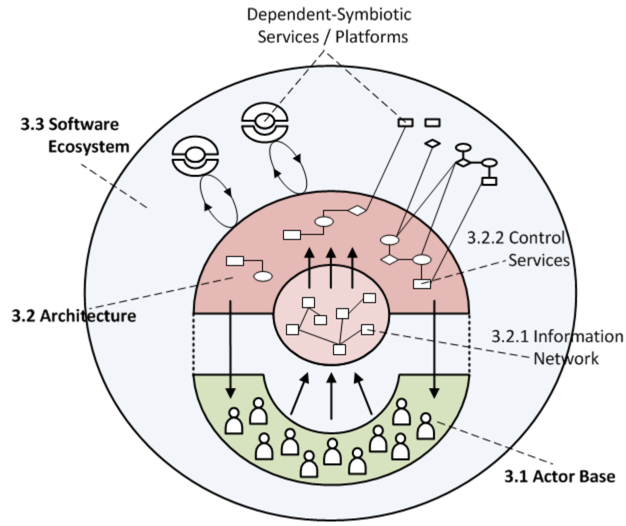

Figure 2: SIS architectural overview and stigmergy cycle: actor base (3.1) contributions are stored in the information network (3.2.1) and percolate through various control services (3.2.2) in order to be triggered back to the actor base. Software ecosystem (3.3) elements (dependent-symbiotic services / platforms) support components of the architecture.

3.1 Actor Base

The actor base is the collectivity of all, predominantly human, actors using a SIS in a consuming or producing manner. The circumstance that human actors are the actual workforce of the system makes a SIS a textbook example of the human computation paradigm, which states a computational problem solving approach where the computer relies on humans to perform certain processing tasks [22]. However other than ordinary human computation services in the narrower sense (e.g. Amazon Mechanical Turk[2]), SIS use human actors to aggregate, select [13] and contextualize a structurally limited and quantitatively unlimited amount of information.

An actor uses a SIS primarily, because she seeks to rely on the contribution of others or share her own. Key motivators for contributing to collective systems according to Malone et al. [16] are recognition, enjoyment, socializing with others, contributing to a larger cause, improvement of professional reputation and promise of financial gain. Another important point is that users expect that they still own their contributed data, unless not explicitly stated otherwise, and are able to decide how it is processed (informational self-determination).

An important technical aspect with regards to the actor base are user profile models which are used to describe users, their relationships to objects and each other. Popular specifications include Friend-Of-A-Friend[3] (FOAF), XHTML Friends Network[4] or OpenSocial[5].

3.2 Architecture

The architecture is the operational locus of any SIS and consists of an information network and a set of control services operating on the network.

3.2.1 *Information Network*

The information network continuously stores all user contributions as well as information about the usage behavior of every single user, thus over time becoming to the most valuable asset in the system. The information network is organized as a self-organizational, scale-free network. Barabási et al. [2] identified two key characteristics of such networks: (1) the network continuously grows by the addition of new vertices and (2) a new vertex is more likely to be linked to highly connected existing vertices (preferential connectivity).

By these characteristics an information network can be regarded as a textbook example for big data which describes a collected mass of data sets that cannot be managed and processed in a usual way [8]. New generated data sets are also captured continuously so that the system can never complete the process which requires to focus on parts of the data set [8]. Often data sets contain heterogeneous data e.g. images, videos, geographical movements and communication. The generated amounts of structured and unstructured data provide a lot of insights when executing, usually time-consuming, analyses. In order to address the time-intensity of analysing there is an emerging trend of big data analytics in the cloud e.g. Google Big Query[6]. Besides the challenges of storing, managing and analysing, there exist ethical and privacy concerns such as analyses of anonymously collected data [5].

[2]https://www.mturk.com (last visited 2012/07/15).
[3]http://xmlns.com/foaf/spec (last visited 2012/07/15).
[4]http://gmpg.org/xfn (last visited 2012/07/15).
[5]http://docs.opensocial.org (last visited 2012/07/15).
[6]http://cloud.google.com/products/big-query.html (last visited 2012/07/15).

3.2.2 *Control Services*

The control services are responsible for keeping the stigmergy cycle flowing, that is the information exchange from the actors to the platform and vice versa, as well as the information exchange to external services and other SIS. Besides generic services which are required to provide common CRUD functionality, the following special service types can be identified:

1. Assimilation services: They assist in linking external content into the information network. They are normally connected with low friction plugins/tools which allow users to share easily content on-the-fly (e.g. sharing buttons, bookmarklets).

2. Access control (AC) services: Control services also grant restricted access to the information network to external dependent-symbiotic services by providing dedicated APIs. It can be assumed that a well-rounded set of AC service endpoints is crucial for establishing a solid ecosystem of dependent-symbiotic services in the medium term (e.g. Twitter API[7], Facebook Open Graph API[8]).

3. Meta-heuristic services: They are responsible for controlling and adjusting the stigmergy flow, especially stimulating directly the actor base to contribute content (e.g. reviews or rates). Northrop et al. [17] define meta-heuristics as "a class of (often biologically inspired) search techniques that iteratively seeks an optimum solution within a landscape of possibilities that may be extremely complicated and even discontinuous". Common examples of meta-heuristics are genetic programming, simulated annealing, greedy algorithms and ant-colony optimization. A typical example of a meta-heuristic service is the friend recommendation functionality in social networks [6, 19].

4. Mining and analyzing services: Their task is to analyze data stored in the information network in order to provide knowledge about actor contributions. Solutions often consist of multiple components forming analysis stacks with Hadoop[9]-based solutions as a popular choice [15].

3.3 Software Ecosystem

The third key field is the software ecosystem (SECO) populated with services and cooperation-willing SIS which surrounds the architecture and the actor base. A SECO is defined as "a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artifacts." [11]. Stengths of SECOs lie in increased platform "stickiness", ability for open innovation thus shared cost of innovation and reduced TCO [4].

[7]https://dev.twitter.com (last visited 2012/07/15).
[8]http://ogp.me (last visited 2012/07/15).
[9]http://hadoop.apache.org (last visited 2012/07/15).

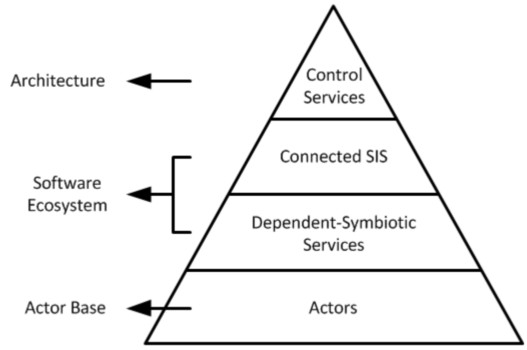

Figure 3: Pyramide of participation in a SIS.

Dependent-Symbiotic Services / Platforms

Dependent-symbiotic services describe third-party software services which are extensively dependent on accessing the information network of a host platform in order to function properly. By providing functionality which is not available at the host platform, dependent-symbiotic services are key for enriching the utility or reducing drawbacks of the host platform for the actor base. An example of such a service ecosystem is the Twitter ecosystem which has services for trend analyses, stream management, image hosting and relationship management. Some of these services eventually got reimplemented and so integral part of the host platform.

Platforms represents other connected SIS that have a benefit when accessing the information network of a host platform in order to utilize the actor base contributions for themselves. An example would be to allow posting of Twitter tweets also in Facebook, thus cascading actor contributions in Twitter also to Facebook.
Figure 3 depicts a theorized four step participation model of a SIS. The largest contribution amount comes from the actor base, followed by content from services around close vicinity of the platform. Moderate inbound traffic could be expected from collaborating SIS and few new content is expected from control services (mostly digests, rankings and trend analyses).

4. FURTHER RESEARCH AND CONCLUSIONS

This paper provided first insights into stigmergic information systems as well as an architectural concept of these systems and described its three key fields: actor base, architecture and software ecosystem. Discussion of these areas highlighted the importance of better understanding the interdependencies between those platforms, users and services. Hence, next research steps include conducting (1) a survey of existing SIS and their characteristics as well as (2) an expert interview study to better understand the anatomy and behavior of them. Based on the results the architectural concept will be constantly refined to address the identified gaps. The long term perspective should be to establish a software architecture perspective used to describe platforms with stigmergy-like systems dynamics in order to:

- investigate the software architectural anatomy of these systems and the related software ecosystems thus providing insights into their strengths and areas for potential improvements,
- better understand the information models which constitute the created networks, and
- support a classification and comparison of different SIS in a generic manner to have the perspective to draw conclusions that go beyond of each ecosystem.

The proposed architecture, like any self-organization-based model, assumes that complex behavior of user communities, as it is facilitated by some software architectures, may be explained "in terms of simple interaction processes" [3, p.6]. The stated assumptions will then be gradually refined, so if "an explanation fails that more complex assumptions will be put into the model" [3, p.6]. Thus this work provides the seed for an architectural agenda which itself grows continuously in depth and width: a control single service becomes a SOA compound, a platform matures to a web within the web. We think that the incomplete understanding about involved models of information structures as well as internal system dynamics combined with the mix of current research areas, is making the scientific investigation of SIS a challenging and yet promising undertaking.

5. ACKNOWLEDGMENTS

This work has been supported by the Christian Doppler Forschungsgesellschaft and the BMWFJ, Austria.

6. REFERENCES

[1] C. C. Aggarwal. *Social Network Data Analytics.* Springer New York, 2011.

[2] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, Oct. 1999.

[3] E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press, 1999.

[4] J. Bosch. From Software Product Lines to Software Ecosystems. In *Proceedings of the 13th International Software Product Line Conference (SPLC '09)*, pages 111–119. Carnegie Mellon University, 2009.

[5] D. Boyd and K. Crawford. Six Provocations for Big Data. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, Sept. 2011. Available: http://ssrn.com/paper=1926431 (last visited 2012/07/15).

[6] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, pages 201–210. ACM, 2009.

[7] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press, 2010.

[8] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker. Interactions with big data analytics. *interactions*, 19(3):50–59, May 2012.

[9] D. Floreano and C. Mattiussi. *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies.* MIT Press, 2008.

[10] P.-P. Grassé. La reconstruction du nid et les coordinations inter-individuelles chez Bellicositermes natalensis et Cubitermes sp. La théorie de la stigmergie: Essai d'interprétation du comportement des Termites constructeurs. *Insectes Sociaux*, 6(1):41–80, 1959.

[11] S. Jansen, S. Brinkkemper, and A. Finkelstein. Business Network Management as a Survival Strategy: A Tale of Two Software Ecosystems. In *Proceedings of the 1st International Workshop on Software Ecosystems (IWSECO '09)*, pages 34–48. CEUR-WS, 2009.

[12] J. Kennedy, R. C. Eberhart, and Y. Shi. *Swarm Intelligence*. Morgan Kaufmann, 2001.

[13] A. Kosorukoff. Human Based Genetic Algorithm. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 3464–3469. IEEE, 2001.

[14] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*, pages 591–600. ACM, 2010.

[15] J. Lin and A. Kolcz. Large-Scale Machine Learning at Twitter. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (SIGMOD '12)*, pages 793–804. ACM, 2012.

[16] T. W. Malone, R. Laubacher, and C. Dellarocas. Harnessing Crowds : Mapping the Genome of Collective Intelligence. Working Paper No. 2009-001, Feb. 2009.

[17] L. Northrop, P. Feiler, R. P. Gabriel, J. Goodenough, R. Linger, T. Longstaff, R. Kazman, M. Klein, D. Schmidt, K. Sullivan, and K. Wallnau. *Ultra-Large-Scale Systems: The Software Challenge of the Future*. Software Engineering Institute, Carnegie Mellon University, 2006.

[18] T. O'Reilly. What is Web 2.0, 2005. Available: http://oreilly.com/pub/a/web2/archive/what-is-web-20.html (last visited 2012/07/15).

[19] N. B. Silva, I.-R. Tsang, G. D. C. Cavalcanti, and I.-J. Tsang. A graph-based friend recommendation system using Genetic Algorithm. In *IEEE Congress on Evolutionary Computation (CEC '10)*, pages 1–7. IEEE, 2010.

[20] D. Tapscott and A. D. Williams. *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio, 2006.

[21] H. Van Dyke Parunak. A Survey of Environments and Mechanisms for Human-Human Stigmergy. In *Proceedings of the 2nd International Conference on Environments for Multi-Agent Systems (E4MAS '05)*, pages 163–186. Springer Berlin Heidelberg, 2006.

[22] L. von Ahn. *Human Computation*. Phd thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, Dec. 2005.