On the Collaborative Formalization of Agile Semantics Using Social Network Applications

Hans-Georg Fill and **Tania Tudorache**
Stanford Center for Biomedical Informatics Research
Stanford University
251 Campus Drive, Stanford 94305
{fill|tudorache}@stanford.edu

Abstract

In this position paper we investigate the opportunities of using functionalities provided by social network sites for the collaborative formalization of semantics in the domain of health. In particular we identified benefits in regard to communication support, economic benefits, and technical opportunities. The implementation of the functionalities are illustrated by describing a use case from an ongoing project with the World Health Organization.

Introduction

Over the past years, the health care sector has been characterized by increasing global collaboration and intensive use of information technology. Thereby, huge amounts of information are generated, which originate not only from care related activities that are part of the electronic health care records such as operation reports, discharge letters, pathology reports or prescriptions (Haux 2006). They also result from admininstrative tasks such as billing and the provision of health related statistics as well as clinical research and regulatory activities. This information is exchanged with third parties on a global scale, including collaborating medical personnel and public and international authorities that monitor the well functioning of the health care sector.

In order to process this health related information, internationally aligned standards and guidelines have been used for a long time. The International Classification of Diseases (ICD) for example is one of the central classifications in this area. It is used in over 100 countries around the world and its current revision comprises more than 20,000 terms. Its purpose is the classification of diseases and other health problems which allows to store and retrieve diagnostic information for clinical, epidemiological, and quality purposes[1]. Another example are the guidelines and standards issued by ICH[2]. Their objective is to increase international harmonization of technical requirements to ensure that safe, effective, and high quality medicines are developed and registered in the most efficient and cost-effective manner. They are based on a formal alignment process that involves regulators and research-based industry representatives of the European Union, Japan, and the USA.

To keep the content of such classifications and guidelines up to date with the constantly changing scientific information and enable new forms of analyses, continuous adaptations are necessary. Due to the size and complexity of these artifacts as well as their international impact, the adaptations are carried out by large international communities of experts. These experts have to agree both on the content as well as on the procedures for conducting the updates. Traditionally, these agreements have been achieved through series of meetings by particular committees that were based on internal and external consultations. The resulting documents were then made publicly available. Recent attempts have been made to formalize the semantics of the results in a machine readable form by using ontologies and thus enable the automatic verification of their consistency and the linkage to other biomedical terminologies and ontologies (Tudorache et al. 2010). With the maturing and general availability of social web approaches, new forms of interaction emerge which have the potential to improve the agility of handling updates and incorporating previously unused knowledge. In the following we will therefore explore the opportunities of using social network applications for the collaborative formalization of semantics in such settings and give an outlook on how this will be realized in a concrete use case.

The remainder of the paper is structured as follows: At first we will briefly outline the foundations of the collaborative formalization of semantics and social network applications. This will then allow us to explore the opportunities of using social network applications in this setting. Finally, we will illustrate a use case and give an outlook on the future research.

Foundations

As there are today several approaches available for the collaborative formalization of semantics and the realization of social network approaches, we will briefly outline some foundations to base our subsequent elaborations on.

Collaborative Formalization of Semantics

When formalizing semantic information it can be distinguished between *semi-formal* approaches where mainly the

[1] See http://www.who.int/classifications/icd/en/

[2] ICH: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, see: http://www.ich.org/

syntax of the information is formally described and *formal* approaches where both syntax and semantics are formally described. Both directions are characterized by the use of *schemas* and *instances* that populate these schemas. As an example for semi-formal approaches consider the various types of conceptual modelling languages that are today widely used to describe business processes, workflows or data models (Karagiannis et al. 2008). In most cases only the syntax of their schema is formally specified and the formal semantic definitions are added as needed. In this way they can be used for a number of applications including the representation and simulation of organizational processes or as a basis for the configuration of automated systems. As an example for formal approaches consider ontologies and rule based systems. The term "ontology" has been used to characterize a wide range of formal representations including taxonomies and hierarchical terminology vocabularies as well as detailed logical theories describing a domain (Noy and Klein 2004). Using formal semantics is an integral part of ontologies, which has allowed to successfully apply them to conduct automatic classifications, verify consistencies and act as a formal reference. Rule-based systems are characterized by a knowledge base that consists of rules and facts and an inference engine. The rules can be used to define actions or infer conclusions based on a number of antecedents that are described with the facts (Hayes-Roth 1985). The semi-formal and formal approaches may also be combined. For example by annotating elements of conceptual modelling languages using concepts from ontologies, the natural language descriptions in models may be made available to verify the consistency of the models (Lautenbacher, Bauer, and Seitz 2008). Similarly, ontologies and rule-based systems may be joined to use the concepts of the ontology as facts for applying rules and deriving additional additional information for the ontology (Horrocks et al. 2004).

A common characteristic of all the abovely mentioned approaches is that their successful application is usually based on the involvement of several persons in order to integrate as much knowledge as possible. Furthermore, ontologies and rule-based systems often become so large and specialized that it is impossible for a single person or even a centralized group to develop them effectively (Noy and Tudorache 2008). By providing distributed, web-based editors, experts and users can be brought together to share their information on a formal basis. To support this collaboration several requirements have been derived in the area of collaborative knowledge construction (Noy and Tudorache 2008; Noy, Chugh, and Alani 2007): In order to facilitate the *discussion and reach consensus* about artifacts in the formal schemas it is essential to provide tools that are tightly integrated with the editors. Possible tools can be chats and discussion forums that are able to process linkages to the schemas and instances. To document the *provenance of information* in such large user groups, changes in the information have to be tracked together with their reasons and when and by whom they were conducted. Another important aspect for large groups of people who do not know each other is to provide mechanisms to establish *trust and credibility*. This can be accomplished for example by explicitly assigning authority to particular users. To address users with different levels of expertise, *view concepts* can be used that offer different presentations of the underlying information. Similarly, fine-grained *access controls and user roles* should be available to define which users have read and write permissions on which parts of the schemas and instances. And finally, the definition of *workflows* that define how new information is created and updates are conducted is necessary to make the underlying processes transparent and provide opportunities for the automated assignment of tasks.

Social Network Applications

Since the first services on the web came up that relied on social interactions to determine additions, annotations, or corrections from a multitude of potentially minor user contributions, the field has dramatically evolved (Auer and Ives 2008; Boyd and Ellison 2008). Especially, communication-centered approaches such as Facebook, Linkedin or myspace have recently gained much attention with Facebook currently taking the lead by hosting more than 500 million active users[3]. These social network sites typically allow individuals to create public or semi-public profiles, articulate a list of other users with whom they share a connection, and browse their list of connections and those made by others in the network (Boyd and Ellison 2008). Apart from creating this experience for users, social network sites also offer interfaces to access the contained information programmatically (Nazir, Raza, and Chuah 2008; Boyd and Ellison 2008). This information can then be used to create new applications based on the social network information and support the bi-directional exchange of information between the application and the network. Currently, the Facebook developer platform can not only be considered as the pioneering platform in this field but also as the most advanced one (Nazir, Raza, and Chuah 2008). However, similar functionality is for example also provided by the Linkedin developer network[4], the myspace developer platform[5] or the OpenSocial API by Google that can access several different social network sites[6].

Using Social Network Applications for the Collaborative Formalization of Semantics

When investigating the integration of a collaborative formalization of semantics with the functionalities of social network applications, three major areas could be identified that can offer benefits. They are based on the requirements of collaborative formalization as outlined above and a number of questions that have been put forward by Kapetanios to set up a methodology for using the potential of collective intelligence as a problem solver (Kapetanios 2008). Thereby it is assumed that a model, a rule editor or an ontology editor is made available as a social network application that

[3] See http://www.facebook.com/press/info.php?statistics

[4] See http://developer.linkedin.com/

[5] See http://developer.myspace.com/

[6] See http://www.opensocial.org/

can retrieve the social network data of a particular user who accesses the application. The areas are: *Communication support*, *economic benefits*, and *technical opportunities*.

Communication Support

Through the provision of personal user information in the social network together with the user's social environment a *social clustering* of certain groups of users of the formalization tools can be conducted. Examples for the user information that can be used for this purpose are name, date of birth, education, spoken languages, employer, workplace, etc. The social environment can be characterized by the connected other people in the network, interests, attended events, group memberships, or identifications on photos. This exceeds the possibilities that have been in place for traditional collaboration approaches and can be used to offer more personalized configurations of the tools. Together with the possibility of tracking a user's actions in the tools, clusters of expert users and standard users for particular parts of models, rules, or ontologies can be derived. In addition, this information can be used to automatically assign users to groups with specific user roles and delegate tasks to certain expert user groups such as quality assurance or the testing of specific functionality. For example this would allow to automatically create a group of users who are interested only in a particular branch of an ontology, a specific set of rules or certain types of models. Thereby communication is supported on different levels of expertise and potentially even cultural and multi-lingual backgrounds. In this way also the notification mechanisms of social network sites can be accessed to inform users about other concepts that might be of interest for them and point them to existing tutorials or specialized information for particular tool functionalities. Through the integrated discussion forums for user groups that are available on most social network sites, users may also freely discuss general topics with their group. All this is however achieved at little cost for the user as most users already have made their personal data available in the social network and are therefore not required to re-enter it for these specific purposes.

Economic Benefits

The reduced effort of using the concept of social network applications also translates into economic benefits. Through the structured involvement of potentially giant user bases, tasks such as quality assurance and the testing of new functionalities can be easily distributed in the sense of crowdsourcing. This can either be handled manually or through the use of automated workflows. In contrast to the traditional feedback mechanisms in the context of open source software this however permits to track in detail which user conducted which task and also take into account for example the level of expertise of a particular user based on the abovely mentioned social clustering. Furthermore, as the largest part of the social network sites and their application interfaces can be accessed for free no licensing costs have to be incurred whereas at the same time a highly scalable infrastructure is provided. This infrastructure also enables the use of viral marketing approaches where the dissemination of concrete scientific results such as new model or ontology information can be easily realized by each individual user at no additional cost and targeted at a huge potential audience. Thereby it is also envisaged to inform large numbers of people about the opportunities of semantic formalization who have not been involved in these technologies before.

Technical Opportunities

Last but not least such an integration offers several technical opportunities. With the authentication protocols for the social network platforms separate authentication procedures and the management of user profile information do not have to be maintained anymore by the formalization tools. Instead the authentication is left to the social network sites and is only verified by accessing authentication tokens. Also the detailed permission handling for accessing user information can be left to the social network sites. Thereby, the privacy of the user data can be guaranteed based on the user settings and users may also easily revoke or change permissions for accessing their personal data. In addition, the social network application interfaces are supported for a number of programming languages and several open source APIs, which makes it easy to integrate them in any existing web based application. It has to be noted however that due to the currently high dynamics of the social network sites that brings about new features almost every day developers may have to adapt quickly in order to keep their applications working smoothly.

Outline of a Use Case

In 2007, WHO initiated the work on the 11th revision of the International Classification of Diseases (ICD-11) with the mission "to produce an international disease classification that is ready for electronic health records that will serve as a standard for scientific comparability and communication."[7]. To keep up to date with the scientific advances, WHO produces major ICD revisions every decade or so. Previous revision processes were conducted behind closed doors, with experts from different countries exchanging their list of disease by regular mail and then resolving the conflicting issues in face-to-face meetings. For the 11th revision of ICD, WHO envisions a more open process that involves the larger expert community in the creation of the ICD-11 content. Their vision is to use a social platform as the main communication means in which experts around the world enter change proposals. These proposals will be aggregated, reviewed and then submitted to the shared copy of ICD-11.

This kind of ambitious undertaking that involves the community at large in creating an agreed-upon content for ICD-11 is only possible by using a Web-based social platform. We consider the role of the social platform as crucial. Its main function is in forming and aggregating a community of experts, and keeping them involved throughout the entire revision process. One of the major challenges in creating a social process for the ICD-11 revision is building a critical mass of interested users that would contribute content. We are convinced that a social platform is the ideal environment for this purpose. We envision using a social platform, such

[7] http://sites.google.com/site/icd11revision/home

as Facebook, to make the ICD-11 revision process known to the community of experts. We plan to use the "friends" of domain experts from Facebook to identify other people possibly interested in the revision process, and sending them invitations to participate. We plan to use the social platform features to create networks of trust among the users and for identifying key experts in a domain as users who are central in the friends-of-a-friend network. These key users may have a have a more influential role when it comes to making decisions about the proposals.

One of the concerns with opening up the ICD-11 revision to a large community is how to handle a high number of incoming proposals. We envision using a "community" filtering, in which groups of users interested in a particular topic (e.g., mental health) may form specialized communities on the social platform. Any new proposal made in a particular area will have to pass first the "judgment" of the specialized community, and only after acquiring its approval, the proposal will pass to next phase in the process, in which a more formal review will be performed.

One of principled goals of the revision process is to ensure a high quality of the ICD content. This goal will be achieved by using a peer-review mechanism similar to the one used in the scientific communities for reviewing journal articles. We may use the social platform for identifying recognized experts in a particular domain by automatically parsing the profile of users in the social platform, and cross-checking their interest and expertise with the specific reviewing needs in ICD-11. We may also use the friend-of-a-friend networks to asses the centrality of a user in the network of a given domain, and for identifying other possible reviewers for a particular topic.

The ICD-11 revision process already uses a collaborative ontology development tool, iCAT (Tudorache et al. 2010), for acquiring the initial content of ICD-11. We have done some preliminary work in which we have integrated the iCAT platform as an application in Facebook[8]. Users of Facebook can access the application directly from the social platform. In the near future, we will support a seamless integration of the user accounts of Facebook and iCAT: Users authenticated in one of the systems will be automatically authenticated in the other one. This will make the transitioning from the social platform to the authoring environment much easier. We also plan to support in the short term the publishing of comments made in iCAT also in Facebook to keep the users engaged in the process and encourage them to participate.

Conclusion and Outlook

Although several benefits of integrating formalization tools with social networks could be shown, the entire social process, the workflow and the interoperation between the authoring tool and the social platform are still under development and require further research. We expect that the workflow defining the interactions has to be defined in an agile way to be adaptable to the continuously changing domain and technical requirements. Furthermore, it will have to be investigated how social network applications may be combined with other formalization approaches such as models and rule-based systems.

[8]See http://apps.facebook.com/icatapp/

Acknowledgements

Parts of the work on this paper have been funded by the Austrian Science Fund grant number J3028-N23 and by the NIGMS Grant 1R01GM086587.

References

Auer, S., and Ives, Z. 2008. Enabling social semantic collaboration: Bridging the gap between web 2.0 and the semantic web. In Rech, J.; Decker, B.; and Ras, E., eds., *Emerging Technologies for Semantic Work Environments: Techniques, Methods, and Applications*. IGI Press.

Boyd, D., and Ellison, N. 2008. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13:210–230.

Haux, R. 2006. Health information systems - past, present, future. *International Journal of Medical Informatics* 75:268–281.

Hayes-Roth, F. 1985. Rule-based systems. *Communications of the ACM* 28(9):921–932.

Horrocks, I.; Patel-Schneider, P.; Boley, H.; Tabet, S.; Grosof, B.; and Dean, M. 2004. Swrl: A semantic web rule language combining owl and ruleml. http://www.w3.org/Submission/SWRL/ accessed 21-03-2007.

Kapetanios, E. 2008. Quo vadis computer science: From turing to personal computer, personal content and collective intelligence. *Data and Knowledge Engineering* 67:286–292.

Karagiannis, D.; Fill, H.-G.; Hoefferer, P.; and Nemetz, M. 2008. Metamodeling: Some application areas in information systems. In Kaschek, R., and et al., eds., *UNISCON*. Springer. 175–188.

Lautenbacher, F.; Bauer, B.; and Seitz, C. 2008. Semantic business process modeling - benefits and capability. In *AAAI Spring Symposium*. Stanford University, California: AAAI.

Nazir, A.; Raza, S.; and Chuah, C.-N. 2008. Unveiling facebook: A measurement study of social network based applications. In *IMC'08*. Vouliagmeni, Greece: ACM.

Noy, F. N., and Klein, M. 2004. Ontology evolution: Not the same as schema evolution. *Knowledge and Information Systems* 6(4):428–440.

Noy, N., and Tudorache, T. 2008. Collaborative ontology development on the (semantic)web. In *AAAI Spring Symposium*. Stanford University: AAAI.

Noy, N.; Chugh, A.; and Alani, H. 2007. The CKC Challenge: Exploring Tools for Collaborative Knowledge Construction. *Intelligent Systems* 23(1):64–68.

Tudorache, T.; Falconer, S.; Nyulas, C.; Noy, N. F.; and Musen, M. A. 2010. Will Semantic Web Technologies Work for the Development of ICD-11? In *accepted for: International Semantic Web Conference*.