

A Modelling Based Notation for the Automated Extraction and Analysis of Social Networking Data

Samantha J. Dixon, Mark Dixon, Edward Halpin, and Colin Pattinson

Leeds Metropolitan University, Leeds, UK

{s.j.dixon,m.dixon,e.halpin,c.pattinson}@leedsmet.ac.uk

Abstract. There is a growing need for, often non-technical, organisations to analyse valuable information stored within often separate social networking systems (SNSs). Open architectures provide programmatic access to most SNSs permitting the creation of applications which may leverage information, for example statistics regarding the impact of marketing campaigns or new product or service announcements. This type of information is necessary for the development of sound evidence based social media strategies. Software products are available which provide this type of information, though for organisations to be able to tailor these to their specific needs, solutions are often very expensive. One solution would be for organisations to have the facility to build their own systems. This paper describes a research programme that will investigate developing, amending or extending a modelling notation, capable of being used by non-technical people for the development of systems to extract and analyse social networking data.

Keywords: Social network, domain specific notation, modelling notation, model driven development.

1 Background

The availability and usage levels of social networks have increased vastly over the last 10 years [1]. Research undertaken on behalf of Meeting Professionals International (MPI) [2], has shown that there is a growing need for non-technical organisations to gain access to and analyse the valuable information stored within the numerous SNSs. Many social networks provide programmatic access to their systems allowing the building of third party applications. Most major social networks such as Facebook, Twitter and LinkedIn provide specific application programming interfaces (APIs) [3] which allow automatic extraction and analysis of site usage data. These open architectures provide the opportunity for the development of systems which leverage available information, for example commercial organisations can be provided with useful statistics regarding the impact of marketing campaigns or new product and service announcements.

2 The Research Problem

Products are already appearing which provide exactly the type of information described above [4], though for organisations to have the flexibility with regard to the

type of information that is retrieved solutions are often extremely expensive. A major obstacle for many ‘would be’ developers is that the published APIs of the more popular sites are perceived to be not very well designed or user friendly [5]. This combined with the speed of technological development and the growing number of available social networking APIs has forced organisations down one of three routes:

1. Organisations get little or no social networking analytical information.
2. Organisations choose a low cost solution, though these are often limited and inflexible.
3. Organisations opt for a high cost social media solution.

Currently many organisations have either limited or no social media strategies in place [6]. If organisations had the facilities to build their own social networking analytical solutions then social media strategies could be devised based on their own sound historical data.

2.1 Aims and Objectives of the Proposed Research

The primary aim of the proposed research programme is to develop, extend or amend a modelling notation, specifically for use by non-technical people¹, for the development of custom applications to extract and analyse cross platform social network data. The key objectives are to:

1. Provide an abstract mechanism of accessing a diverse set of social network information, with limited regard for the technical requirements of the underlying APIs.
2. Identify and examine useful metrics which are commonly required for typical social network analysis scenarios.
3. Provide pre-defined services which would allow such common metrics to be applied to extracted data in real-time.
4. Develop a modelling notation, which adheres to current software standards, that combines both graphical and textual elements². Reasons for using a combination of the two include:
 - (a) When used at an abstract level graphical notations are easier to understand and faster to grasp [7], this is important as users may be non-technical and not using the notation on a regular basis. Graphical notations are typically less concerned with *how* a particular element will perform its operation and more with *what* it shall do [8].
 - (b) As the notation will be ultimately used to produce a working system the addition of textual annotations will add clarity where needed, especially to rule out ambiguities.

¹ Non-technical people are defined as those within an organisation who are not software developers but have the ability to understand modelling concepts.

² After the requirements analysis stage the notation may change from being graphical and textual.

3 Suggested Approach to the Problem

In defining a modelling language work will need to be undertaken to identify which of the currently available modelling techniques would provide a starting point for the development of a model based approach. Procedural and data oriented systems are traditionally modelled using data-flow diagrams (DFDs) and entity relation models (ERDs) whereas object oriented systems are predominantly modelled using the Unified Modelling Language (UML). It is likely however that the proposed research would be better supported by techniques designed specifically for Service Oriented Architectures (SOAs), such as the Business Process Modelling Notation (BPMN) since this provides an approach to modelling systems via orchestration of other available services. This would inevitably lead to an examination of other related technologies such as the Business Process Execution Language (BPEL) and the Web-Services Description Language (WSDL). When obtaining requirements suitable metrics to be supported by the modelling technique would need to be identified. Initial research has identified several appropriate measurements [9-11]. The existing modelling techniques would need to be analysed and extended in a manner which would allow automated extraction of social networking data. This would then need to be coupled with the application of appropriate metrics on that data. Previous academic studies that have addressed a similar problem include ones which explore the use of WebML [12] for model driven development of SNSs. These papers, though relevant to this research, are more concerned with the development of social network platforms such as Facebook or Twitter, rather than the development of systems to extract and analyse the information from within these platforms.

4 Justification of Research

This work will contribute to existing knowledge by providing a validated modelling technique that will allow non-technical staff to develop custom applications to analyse cross platform social network data. This will not only provide organisations with the ability to measure their usage of social networking sites, e.g. event organisers and advertisers could view the impact of their campaigns, but it will also provide a framework in which new metrics could be validated, e.g. does Klout [13] tell us anything useful? These models will provide the abstract layer in a multi-layered architecture and will sit above a physical mapping layer where the elements of the notation are mapped to the various underlying APIs.

5 Research Methodology

The research programme will consist of three main steps. The first will be to determine user requirements using interviews and surveys. The second will be to define a modelling language and the third will be to evaluate that language. Given the technical nature of the suggested research a scientific method is most appropriate in which a

mixture of qualitative and quantitative primary data collection will be used. Potential users of the proposed notation from a variety of industries will be selected for in-depth, semi-structured and structured interviews, questionnaires and surveys. Targeted organisations will either be actively using or considering using social networks. Qualitative in-depth and semi-structured interviews shall be used together; with the in-depth interviews helping to form questions for the structured interviews, questionnaires and surveys as well as adding to the overall pool of data to be analysed. The main purpose of data collection will be to:

- Find out why social networks are used for business.
- Investigate how end users normally work with social networks.
- Identify a detailed list of requirements for a social network/media data extraction and analysis system.

An interview schedule will be produced which will be updated as new requirements are identified.

Literature reviews will be carried out to determine a suitable modelling technique to be either extended or amended in order to allow support for the extraction and analysis of social network information, including the application of metrics. Secondary research will also be performed to understand how to formally define models; this will include an examination of the Meta-Object Facility (MOF) [14].

For validation purposes an organisation shall be identified and a case study carried out to apply the proposed modelling notation, this will provide a basis for some repeatable experiments. Staff will be selected to use the notation for modelling specific scenarios. The resulting models will be analysed against the original requirements of the organisation, this will include assessing the ease of modelling for the end users.

6 Future Work

Following the successful completion of the research programme future extensions may include:

- Use of the notation to devise a metrics validation framework.
- The development of a complete CASE tool to support the notation directly.

There will also be a number of smaller improvements that could be made in the short-term, for example, a richer set of pre-defined functions available for use within the notation.

References

1. Boyd, D.M., Ellison, N.B.: Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13(1), 210–230 (2007), <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full>

2. MPI, MPI Home (2013), <http://www.mpiweb.org/Home> (accessed March 11, 2013)
3. Hawker, M.D.: Developer's Guide to Social Programming: Building Social Context Using Facebook, Google Friend Connect, and the Twitter API. Addison-Wesley (2010)
4. The Brick Factory, ImpactWatch, Impact Watch: monitoring made simple, The Brick Factory (2012), <http://www.impactwatch.com/> (accessed June 25, 2012)
5. Perez, S.: Facebook Wins "Worst API", Techcrunch (2011), <http://techcrunch.com/2011/08/11/facebook-wins-worst-api-in-developer-survey/> (accessed June 25, 2012)
6. Grant, M.: 76% of Companies Do Not Have a Social Media Policy. Social Business News (2012), <http://www.socialbusinessnews.com/76-of-companies-do-not-have-a-social-media-policy/> (accessed January 25, 2013)
7. Petre, M.: Why looking isn't always seeing. Communications of the ACM 38(6), 33–44 (1995), <http://rtsys.informatik.uni-kiel.de/teaching/ws03-04/s-synth/papers/p33-petre.pdf> (accessed January 28, 2013)
8. Groenniger, H., et al.: Text-based Modeling. In: Proceedings of the 4th International Workshop on Software Language Engineering, Nashville, TN, USA, Johannes-Gutenberg-Universitat Mainz, p. 2 (October 2007), http://www.se-rwth.de/~rumpe/publications20042008/Groenniger_et_al_ATEM_07.pdf (accessed January 28, 2013)
9. Chapron, P., Sibertin-Blanc, C., Adreit, F.: Analysis of Power Networks among the Actors of a Social Organisation. In: Kazakov, D., Tsoulas, G. (eds.) AISB 2011 Social Networks and Multiagent Systems, April 4–7, pp. 2–7. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, York (2011)
10. Chun, H., Kwak, H., Eom, Y., Moon, S., Jeong, H.: Comparison of online social relations in volume vs interaction: a case study of cyworld. In: Papagiannaki, K., Zhang, Z. (eds.) Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, October 20–22, pp. 57–70. ACM New York (2008)
11. Golbeck, J., Hendler, J.: Inferring binary trust relationships in Web-based social networks. ACM Transactions on Internet Technology (TOIT) 6(4), 497–529 (2006)
12. Brambilla, M., Mauri, A.: Model-Driven Development of Social Network Enabled Applications with WebML and Social Primitives. In: Grossniklaus, M., Wimmer, M. (eds.) ICWE 2012 Workshops. LNCS, vol. 7703, pp. 41–55. Springer, Heidelberg (2012)
13. Klout, Inc. The KLOUT Score (2008), <http://klout.com/corp/kscore> (accessed 26, 2012)
14. Object Management Group, OMG's MetaObject Facility (MOF) Home Page (2012), <http://www.omg.org/mof/> (accessed February 12, 2013)