OPEN MODELING FOR USER-GENERATED CONTENT

Extended Abstract

Parsons, Jeffrey, Memorial University of Newfoundland, St. John's, CAN, jeffreyp@mun.ca

Lukyanenko, Roman, Florida International University, Miami, USA, rlukyane@fiu.edu

Abstract

User-generated content (UGC) has attracted considerable interest in many contexts, including such diverse examples as the use of crowdsourcing to generate product ideas for businesses, the use of citizen science to collect or analyse data, and the use of technologies by governments to engage the public in public policy discussions and decision making (Brabham, 2013; Doan et al., 2011; Lukyanenko et al., 2014; Susarla et al., 2012). Notwithstanding the power of crowds in contributing to predetermined and well-defined uses, we believe the full potential of UGC can only be realized by tapping into the ability of contributors to provide unanticipated content that cannot be shoehorned into a predetermined conceptual view of important domain concepts and relationships among them.

Traditional approaches to modelling the concepts in a domain are ill-suited for exploiting the potential of UGC to support innovation and discovery. First, traditional modelling approaches assume that the intended uses of data are known and stable (Chen, 2006; Olivé, 2007). Second, domain models are expressed as abstractions of the similarities among known phenomena of interest and the attributes they have in common (Mylopoulos, 1998; Smith and Smith, 1977). Third, domain models are assumed to be accurate and complete representations of the relevant phenomena in the domain (Olivé, 2007). Finally, traditional modelling is based on communicating with known, representative users in determining the relevant abstractions (Land and Hirschheim, 1983; Mumford and Henshall, 1979).

The desired characteristics of UGC violate all these assumptions. First, innovation and discovery require using information in ways not anticipated when it was collected. Second, insights from users often come in the form of the differences among phenomena of interest, rather than their similarities. Third, different users may view a domain in terms of different concepts; there is no 'gold standard' or correct way of classifying phenomena of interest. Finally, in most UGC settings, it is impossible to communicate with all (or even representative) contributors to determine what perspectives or abstractions might be relevant in a particular setting.

To address these tensions and provide a foundation for open modelling to support UGC, we propose several principles. First, models of UGC should contain minimal, low level, and stable abstractions. In particular, basic level categories are generally viewed as widely shared abstractions with well-understood meaning (Lassaline et al., 1992; Mervis and Rosch, 1981; Rosch, 1974). Second, the fundamental construct for UGC is the instance (Parsons and Wand, 2000). While classes emphasize the commonality of instances, we argue that innovation and discovery in UGC are best served by modelling the unique properties of instances, rather than their shared properties. Third, classes should be modelled as contextual assignment of meaning to instances, rather than fundamental abstractions. This means that different users can attach (or detach) classes to the same instances (and these might change over time) without affecting the underlying instance-based representation of the instances themselves. Fourth, instance models should support redundancy in attributes and classes attached to instances. This allows a particular instance to be viewed from very different perspectives by different contributors. Finally, contributors should be encouraged to define or specify the semantics of any class(es) they attach to instances. This allows data to be understood explicitly from the perspective of the contributor.

2015 Developmental Workshop on Open Research and Practice in Information Systems,
May 26, 2015, Münster, Germany.

We have developed these principles in the context of a citizen science project to collect UGC based on sightings of flora and fauna. We are currently working on incorporating the principles in a modelling grammar that will be designed to capture UGC.

References

Brabham, D.C., 2013. Crowdsourcing. MIT Press, Cambridge, MA.

Chen, P., 2006. Suggested Research Directions for a New Frontier – Active Conceptual Modeling, in: ER 2006, Lecture Notes in Computer Science. pp. 1–4.

Doan, A., Ramakrishnan, R., Halevy, A.Y., 2011. Crowdsourcing systems on the World-Wide Web. Communications of the ACM 54, 86–96.

Land, F., Hirschheim, R., 1983. Participative systems design: rationale, tools and techniques. Journal of Applied Systems Analysis 10, 15–18.

Lassaline, M.E., Wisniewski, E.J., Medin, D.L., 1992. Basic Levels in Artificial and Natural Categories: Are All Basic Levels Created Equal?, in: Barbara, B. (Ed.), Percepts, Concepts and Categories: The Representation and Processing of Information, Advances in Psychology. North-Holland, pp. 328–378.

Lukyanenko, R., Parsons, J., Wiersma, Y., 2014. The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-generated Content. Information Systems Research 25, 669 – 689.

Mervis, C.B., Rosch, E., 1981. Categorization of Natural Objects. Annual Review of Psychology 32, 89–115.

Mumford, E., Henshall, D., 1979. Designing participatively: A participative approach to computer systems design: A case study of the introduction of a new computer system. Manchester Business School, Manchester, UK.

Mylopoulos, J., 1998. Information modeling in the time of the revolution. Information Systems 23, 127–155.

Olivé, A., 2007. Conceptual modeling of information systems. Springer, Berlin Heildelberg New York.

Parsons, J., Wand, Y., 2000. Emancipating Instances from the Tyranny of Classes in Information Modeling. ACM Transactions on Database Systems 25, 228–268.

Rosch, E., 1974. Basic-Level Objects in Natural Categories. Bulletin of the Psychonomic Society 4, 246–246.

Smith, J.M., Smith, D.C.P., 1977. Database abstractions: aggregation and generalization. ACM Transactions on Database Systems 2, 105–133.

Susarla, A., Oh, J.-H., Tan, Y., 2012. Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube. Information Systems Research 23, 23–41.