

PROV-SAID: a W3C PROV Extension for Information Diffusion on Social Media

Tom De Nies¹, Io Taxidou², Peter M. Fischer²,
Erik Mannens¹, Rik Van de Walle¹

¹ Ghent University - iMinds - Data Science Lab, BE

² University of Freiburg, DE

tom.denies@ugent.be, taxidou@informatik.uni-freiburg.de, peter.fischer@informatik.uni-freiburg.de, erik.mannens@ugent.be, rik.vandewalle@ugent.be

Abstract. In this short position statement, we argue that there is a gap in the W3C PROV Data Model, concerning the capturing of influence and information diffusion in message-based social media. In a collaboration between Ghent University - iMinds - Data Science Lab and University of Freiburg, we have proposed an extension to PROV to fill in this gap. As we are continuously evaluating and updating the specification of this extension, we welcome all feedback from the provenance community.

Keywords: provenance, information diffusion, social media

1 Motivations and Prior Work

In 2015, we introduced PROV-SAID, an extension to PROV, allowing us to model the provenance of information diffusion on social media [3]. Whereas information diffusion typically models how information *travels forward* through a network, provenance looks *backwards*, and models where the information came from. Guille et al.[2] show that information diffusion is the subject of a significant research effort. Furthermore, researchers are often limited to the tools and APIs provided by the social media themselves, leaving little room for interoperability across use cases. Therefore, we argue that this topic deserves the attention of the provenance community, and a means to share interoperable provenance about interactions on social media is needed.

In essence, PROV-SAID introduces the following extensions to the PROV-DM:

1. four subtypes of Entity, modeling various types of Messages;
2. three subtypes of Influence, modeling various types of Influence on social media.

3. one subtype of Activity modeling Message emission, and three modeling InfluenceActivities;
4. six roles for usage by and association with InfluenceActivities;
5. one subtype of Derivation, modeling indirect derivation between Messages.

More recently, we have started capturing the Twitter stream during certain conferences we attended, and made a number of observations regarding *explicit* and *implicit* interactions. In [4], we discuss how next to explicit interactions such as re-emitting messages and replying, quoting, and mentioning other users, implicit interactions are also possible, which are not exposed by social media APIs. For example, users could propagate similar messages due to an *external influence*, such as an event. Users could also re-propagate their own messages, for example to modify earlier statements or for promotional purposes, and thus exert a certain *self-influence*. These additions have since been made to the model.

2 Next Steps

Since the social media landscape is very dynamic, it is likely that further additions to our extension will be necessary to match the new functionality and ways of influencing each other.

Furthermore, we are keen to investigate the interplay of this extension to other advances in the community around PROV. For example, as a next step, we will investigate in what ways Gamble & Goble's work on quantifying influence[1] can be applied to our model, and be used in the context of social media.

References

1. Gamble, M., & Goble, C. A. (2014). Influence Factor: Extending the PROV Model With a Quantitative Measure of Influence. In TAPP.
2. Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. ACM SIGMOD Record, 42(2), 17-28.
3. Taxidou, I., De Nies, T., Verborgh, R., Fischer, P. M., Mannens, E., & Van de Walle, R. (2015, May). Modeling information diffusion in social media as provenance with W3C PROV. In Proceedings of the 24th International Conference on World Wide Web Companion (pp. 819-824).
4. Taxidou, I., Fischer, P. M., De Nies, T., Mannens, E., & Van de Walle, R. (2016, April). Information Diffusion and Provenance of Interactions in Twitter: Is it only about Retweets?. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 113-114).