

Sentence Templating for Explaining Provenance

Heather S. Packer^(✉) and Luc Moreau

Web And Internet Science Group, University of Southampton, Southampton, UK
{hp3,L.Moreau}@ecs.soton.ac.uk

Abstract. Disseminating provenance data to users can be challenging because of its technical content, and its potential scale and complexity. Textual narrative and supporting images can be used to improve a user's understanding of provenance data. This early work aims to support the exploration of provenance data by allowing users to query provenance data with a *provenance subject* (either an entity, activity or agent) recorded in it.

1 Introduction

Provenance data can be hard for both expert and non-expert users to understand because of its technical content, and its potential scale and complexity. In order to address these obstacles, we propose allowing users to explore provenance data via an explanation service. This service requires provenance data and a single subject (either an entity, activity or agent) described in the data, as parameters. It returns a description of the subject, which includes text and a provenance graph.

Section 2 describes the sentence templates used to generate a textual narrative of a *provenance subject*. Following that, Sect. 3 describes the provenance graphs generated to support the textual narrative. We then provide conclusions and discuss future work in Sect. 4.

2 Sentence Templating

The explanation service uses sentence templates to explain the provenance types defined in the PROV [1] W3C's standard for provenance (see Table 1 for examples). The sentence templates are strung together to form a paragraph describing the entities, agents and activities which relate to the *subject*. For example, the following paragraph has been generated about the *provenance subject* **rs:/rideRequest/1** from the provenance in Fig. 1.

The [rs:/rideRequest/1](#) is a [UserInput](#) entity. It was generated by the activity [rs:post_ride_request_103](#). It was attributed to the agent [rs:/users/agent2](#). It was used by the activity [rs:store_ride_request81935](#).

```

entity(rs:/rideRequest/1, [prov:type='prsm:UserInput'])
activity(rs:post_ride_request_103, -, -, [prov:type='prsm:Provide_Info'])
activity(rs:store_ride_request81935, -, -, [prov:type='prsm:StoreData'])
agent(rs:ride_server, [prov:type='prsm:webservice'])
agent(rs:/users/agent2, [prov:type='prsm:loggedInUser'])
wasGeneratedBy(rs:/rideRequest/1, rs:post_ride_request_103, -)
used(rs:store_ride_request81935, rs:/rideRequest/1, -)
wasInformedBy(rs:store_ride_request81935, rs:post_ride_request_103)
wasAttributedTo(rs:/rideRequest/1, rs:/users/agent2)
wasAssociatedWith(rs:post_ride_request_103, rs:/users/agent2, -)
wasAssociatedWith(rs:store_ride_request81935, rs:ride_server, -)

```

Fig. 1. Provenance extract used for the ride share example.

3 Provenance Graphs

As well as providing sentences which describe a subject, we also generate provenance graphs highlighting the agents, activities, entities and relationships which are in the textual narrative. The graphs represent the provenance data described in the narrative using the standard colours used in PROV and greys out data that was not used. In our example from the previous section, Fig. 2 shows which items were used in the narrative.

Table 1. Template Sentence examples, where items in {} are variables which can be either a single item or a list.

Prov Types	Sentence template
Entity	The {subject} is a {type} entity
Agent	The {subject} is a {type} agent
Activity	The {subject} is a {type} activity
Alternate	It was an alternate of {alternate/s}
Association	It was associated with agent/s {agent/s}
Attribution	It was attributed to agent/s {agent/s}
Collections	It was a member of the {collection/s} collection/s
Communication	It was informed by {agent/s}
Delegation	It acted on behalf of {agent/s}
Derivation	The {subject} was derived from the entity/ies {entity/ies}
Specialization and Revision	The {subject} was a specialisation of the entity {entity}, and is a revision of {entity/ies}
Generation	It was generated by the activity {activity/ies}
Usage	It was used by the activity/ies {activity/ies}

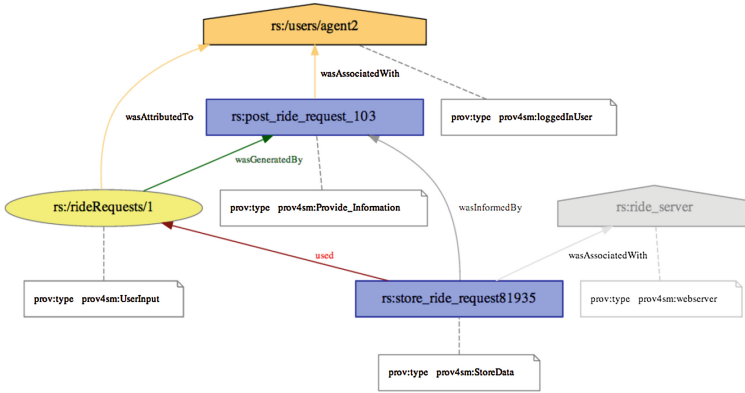


Fig. 2. A graph generated using the ride share example.

4 Conclusion

The explanation service provides users with a description of a *provenance subject* using text and images. For future work, we plan to expand and develop two categories of sentence templates: inspection, which are used to describe facts; and comparison, which are to rank a subject against others. We will also consider policies to support privacy and security in the templates. Finally, we will explore how to foster trust using templates by adopting different authorial tones in the narrative and granularities of explanations about policies, such as privacy and how provenance data is used.

Acknowledgments. The research leading to these results has received partially funding from the European Community’s Seventh Framework Program (FP7/2007-2013) under grant agreement n. 600854 Smart Society: hybrid and diversity-aware collective adaptive systems: where people meet machines to build smarter societies <http://www.smart-society-project.eu/>.

Reference

1. Moreau, L., Missier, P. (eds.) Belhajjame, K., B’Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., Miles, S., Myers, J., Sahoo, S., Tilmes, C.: PROV-DM: The PROV Data Model. W3C Recommendation REC-prov-dm-20130430, World Wide Web Consortium, Oct 2013. <http://www.w3.org/TR/2013/REC-prov-dm-20130430/>