Provenance-based Analysis of Crowd-sourcing Activities
Trung Dong Huynh, Amir Sezavar Keshavarz, and Luc Moreau
Electronics and Computer Science, University of Southampton

Overview

- Provenance of crowd-generated data captured in CollabMap.
- Analytical study on various network measures of the over 5,000 provenance graphs captured in the application.
- Learning from those network measures allows online classification of provenance graphs (and crowdsourcing tasks) into different groups requiring different levels of attention.
- Developed a generic framework for propagating annotations in provenance graphs, such as trust values, certainty values, and data privacy designations.

CollabMap Provenance Graphs

Crowdsourcing the identification of buildings and evacuation routes:

- City-wide mapping of buildings and evacuation routes for disaster-recovery simulations.
- Crowd-drawn buildings and routes cross-checked and voted by fellow contributors.
- Provenance recorded for auditing data quality.

Assessing Quality of Crowd-Generated Data

- Trust values for entities are calculated from votes and propagated in the provenance graph to entities without votes.
- An original provenance graph from a CollabMap task.

Provenance Standardisation

Southampton is leading the W3C Provenance Working Group in standardisation efforts defining a language for exchanging provenance information on the Web, including:

- The PROV Data Model (PROV-DM)
- The PROV Ontology (PROV-O)
- The PROV Notation (PROV-N)
- Constraints of the Provenance Data Model
- Provenance Access and Query

The family of specifications are expecting to become W3C’s Candidate Recommendations in December 2012 and has seen growing adoption from industry and academia.

References