Towards dynamic provenance tracking in web feed Mash-Ups


DOI: 10.1109/CIT.2014.72

Abstract

A new breed of applications and services that combines data and content from diverse sources which we refer to as 'mash-ups' has grown significantly all across the Internet. One of the most popular mash-ups comes in the form of web feed mash-ups relying on syndication technologies such as RSS and Atom. This kind of mash-ups aggregates web feeds derived from multiple news websites or blogs and then timely presents them in a single web-based interface. In such systems, it is difficult to know exactly how feed results in data mash-ups are generated. In particular, it is difficult for users to make determinations about whether information is trusted, considering that source feeds have to pass through several processing operations before the feed results are presented. Therefore, it is crucial that web feed mash-ups have to support a mechanism for recording and querying provenance information - the information about the process that led to result data. In this paper, we propose a provenance tracking solution that enables provenance functionality to be facilitated in web feed mash-ups. We demonstrate how the provenance of feed mash-up results to be determined by means of a provenance query algorithm. Finally, we evaluate our provenance tracking solution in terms of storage consumption for provenance collection, demonstrating predictable storage consumption and reasonable storage overheads. © 2014 IEEE.

Author Keywords

Data provenance; Mash-up; Web feed

References

Buneman, P., Khanna, S., Tan, W.C.
Why and where: A characterization of data provenance

Freire, J., Silva, C.T., Callahan, S.P., Santos, E., Scheidegger, C.E., Vo, H.T.
Managing rapidly-evolving scientific workflows

Freitas, A., Knap, T., Oriain, S., Curry, E.
W3P: Building an OPM based provenance model for the web

Freitas, A., Legendre, A., O’riain, S., Curry, E.
Prov4j: A semantic web framework for generic provenance management
(2010) The 2nd International Workshop on Role of Semantic Web in Provenance Management (SWPM), 2010,

Gil, Y., Miles, S.
PROV model primer

Groth, P., Gil, Y., Cheney, J., Miles, S.
Requirements for provenance on the web

Groth, P., Jiang, S., Miles, S., Munroe, S., Tan, V., Tsakasou, S., Moreau, L.
(2006) An Architecture for Provenance Systems,
University of Southampton, Technical Report

- Groth, P., Miles, S., Moreau, L.
  **A model of process documentation to determine provenance in mash-ups**  

- Groth, P., Moreau, L.
  **Recording process documentation for provenance**  

- Hartig, O.
  **Provenance information in the web of data**  

  **The open provenance model core specification (v1.1)**  
  June

- Sansrimahachai, W., Weal, M., Moreau, L.
  **Stream Ancestor function: A mechanism for fine-grained provenance in stream processing systems**  
  Valencia, Spain

  **Vis-Mashup: Streamlining the creation of custom visualization applications**  

- Tan, W.-C.
  **Provenance in databases: Past, current, and future**  

**Document Type:** Conference Paper

**Source:** Scopus