WS 12/13

Seminar „Vertrauenswürdige Systeme“

Organisatorisches

Ansprech-partner	**Dr. Edzard Höfig** Raum 018, Tel. 838 75277 edzard.hoefig@fu-berlin.de
Termine	Voraussichtlich acht Termine im Zeitraum 19.10.12 bis 15.2.13. Termine am 28.12.12, 4.1.13 und 8.2.12 fallen aus, es gibt eine Pause (drei oder vier Termine) vor Beginn der Vorträge.

Geforderte Leistungen	• **Vortrag** über 30 Minuten + 15 Minuten Diskussionsleitung
	  • Einreichen der Folien eine Woche vor Vortrag
	• Schriftliche **Ausarbeitung** mit 10-15 Seiten in LaTeX
	  • LaTeX Vorlage wird gestellt
	  • Abgabe der Ausarbeitung spätestens am 5.4.13 per EMail an Edzard Höfig

Kursform	Seminar mit 2 SWS, bzw. 4 ECTS Punkten
Sprache	Vortragssprache und Sprache für die schriftliche Ausarbeitung wahlweise Englisch oder Deutsch.
Fehlzeiten	Maximal zwei Fehltermine sind möglich.

Zur Auswahl der Themen

Jeder Teilnehmer und jede Teilnehmerin stellt einen Artikel aus der folgenden Liste vor. Bitte sortieren Sie die 16 Artikel nach Ihren Präferenzen: Erstellen Sie zur nächsten Veranstaltung eine Liste mit 16 Artikelnummern. Die oberste Nummer auf der Liste gehört zu dem Artikel, den Sie am liebsten vorstellen möchten. Die unterste Nummer zu dem Artikel, den Sie überhaupt nicht vorstellen möchten. Wir werden in der nächsten Veranstaltung diese Vorauswahl verwenden, um die jeweiligen Vorstellenden zu bestimmen.
Die vollständigen Artikel können sie in einer passwortgeschützten Zip-Datei von der Veranstaltungs-Homepage runterladen. Bitte beachten Sie, dass die Artikel urheberrechtsgeschützt sind und nicht weitergegeben werden dürfen. Das Passwort wird während der Veranstaltungstermine bekannt gegeben.

1. Tracing the provenance of linked data using void

Author

Omitola, Tope and Zuo, Landong and Gutteridge, Christopher and Millard, Ian C and Glaser, Hugh and Gibbins, Nicholas and Shadbolt, Nigel

Booktitle

WIMS '11: Proceedings of the International Conference on Web Intelligence, Mining and Semantics

Year

2011

Pages

7

Abstract

In the open world of the (Semantic) web, a world where increasingly diverse materials from disparate sources of different qualities are being made available, an automatic mechanism for the provision of provenance information of these sources is needed. This paper describes voidp, a provenance extension for the void vocabulary, that allows data publishers to specify the provenance relationships of their data. We enumerate voidp's classes and properties, and describe a use case scenario. A wider uptake of voidp by dataset publishers will allow data consuming tools to take advantage of these metadata providing consumers with the origin, i.e., the provenance, of what is being consumed.

2. Untangling attribution

Author

Clark, D D and Landau, S

Booktitle

Proc. Workshop on Deterring Cyberattacks

Year

2011

Pages

25-40

Abstract

An overview of attribution from a cybersecurity perspective.

3. Linked provenance data: A semantic Web-based approach to interoperable workflow traces

Author

Ding, Li and Michaelis, James and McCusker, Jim and McGuinness, Deborah L

Journal

Future Generation Computer Systems

Year

2010

Pages

1-9

Abstract

The Third Provenance Challenge (PC3) offered an opportunity for provenance researchers to evaluate the interoperability of leading provenance models with special emphasis on importing and querying workflow traces generated by others. We investigated interoperability issues related to reusing Open Provenance Model (OPM)-based workflow traces. We compiled data about interoperability issues that were observed during PC3 and use that data to help describe and motivate solution paths for two outstanding interoperability issues in OPM-based provenance data reuse: (i) a provenance trace often requires both generic provenance data and domain-specific data to support future reuse (such as querying); (ii) diverse provenance traces (possibly from different sources) often require preservation and interconnection to support future aggregation and comparison. In order to address these issues and to facilitate interoperable reuse, integration, and alignment of provenance data, we propose a Semantic Web-based approach known as Linked Provenance Data, where: (i) the Web Ontology Language (OWL) can be used to support complex domain concept modeling, such as subtype taxonomy and concept alignment, and seamlessly connect domain extensions to OPM core concepts; (ii) Linked Data can enable open and transparent infrastructure for provenance data reuse.

4. Janus: from workflows to semantic provenance and linked open data

Author

Missier, P. and Sahoo, S. and Zhao, J. and Goble, C. and Sheth, A.

Booktitle

Provenance and Annotation of Data and Processes

Year

2010

Pages

129-141

Abstract

Data provenance graphs are form of metadata that can be used to establish a variety of properties of data products that undergo sequences of transformations, typically specified as workflows. Their use- fulness for answering user provenance queries is limited, however, unless the graphs are enhanced with domain-specific annotations. In this paper we propose a model and architecture for semantic, domain-aware prove- nance, and demonstrate its usefulness in answering typical user queries. Furthermore, we discuss the additional benefits and the technical impli- cations of publishing provenance graphs as a form of Linked Data. A prototype implementation of the model is available for data produced by the Taverna workflow system.

5. Provenance-based strategies to develop trust in semantic web applications

Author

Li, X. and Lebo, T. and McGuinness, D.

Booktitle

Provenance and Annotation of Data and Processes

Year

2010

Pages

182-197

Abstract

Linked data and Semantic Web technologies enable people to navigate across heterogeneous sources of data thus making it easier for them to explore and develop multiple perspectives for use in making decisions and solving problems. While the Semantic Web offers benefits for developers and users, several new challenges are emerging that may negatively impact users' trust in Web-based collaborative systems. This paper describes several use cases to illustrate potential trust issues faced by Semantic Web applications, and provides a concrete example for each using a specific system we built to investigate United States Supreme Court decision making. Provenance-based solutions are pro- posed to develop trust and/or minimize the distrust that is provoked by the situation. While these use cases address distinct situations, they are all described in terms of how a contradiction can arise between the user's mental model and the statements presented in the display. This commonality may be used to develop additional classes of trust-threatening use cases, and the proposed provenance-based solutions can be applied to many other Semantic Web Applications.

6. Publishing and consuming provenance metadata on the web of linked data

Author

Hartig, O. and Zhao, J.

Booktitle

Provenance and Annotation of Data and Processes

Year

2010

Pages

78-90

Abstract

The World Wide Web evolves into a Web of data, a huge, globally distributed dataspace that contains a rich body of machine- processable information from a virtually unbound set of providers cov- ering a wide range of topics. However, due to the openness of the Web little is known about who created the data and how. The fact that a large amount of the data on the Web is derived by replication, query processing, modification, or merging raises concerns of information qual- ity. Poor quality data may propagate quickly and contaminate the Web of data. Provenance information about who created and published the data and how, provides the means for quality assessment. This paper takes a first step towards creating a quality-aware Web of data: we present ap- proaches to integrate provenance information into the Web of data and we illustrate how this information can be consumed. In particular, we introduce a vocabulary to describe provenance of Web data as metadata and we discuss possibilities to make such provenance metadata accessible as part of the Web of data. Furthermore, we describe how this metadata can be queried and consumed to identify outdated information.

7. Reflections on Provenance Ontology Encodings

Author

Ding, L. and Bao, J. and Michaelis, J. and Zhao, J. and McGuinness, D.

Booktitle

Provenance and Annotation of Data and Processes

Year

2010

Pages

198-205

Abstract

As more data (especially scientific data) is digitized and put on the Web, the importance of tracking and sharing its provenance metadata grows. Besides capturing the annotation properties of data, provenance research also emphasizes interlinking relevant data. Therefore, it is desirable to make provenance metadata easy to access, share, reuse, integrate and reason with. To address these requirements, ontologies can be of use to encode expectations and agreements concerning provenance metadata reuse and integration. The Web is of use to support access and sharing. The Semantic Web, with its languages for representing terms and their descriptions,, such as RDFS and OWL, is of use for capturing expectations, agreements, and meaning. We are investigating best practices for providing Semantic Web encodings for provenance ontologies by analyzing a selection of popular Semantic Web provenance ontologies such as Open Provenance Model (OPM), Dublin Core (DC) Terms, and the Proof Markup Language (PML). In this paper, we will highlight a few findings which include: (i) similarities and

differences among existing provenance ontologies; (ii) popular approaches used to model provenance concepts and lessons learned from the usage of Semantic Web language features in representing provenance concepts; (iii) expressivity and tractability of representative provenance ontologies. The outcome of our study provides not only guidance to provenance ontology users but also insights to promote better collaborative provenance ontology development and scalable processing of provenance ontologies.

8. Securing Provenance-based Audits

Author

Aldeco-Pérez, R and Moreau, L.

Journal

Provenance and Annotation of Data and Processes

Year

2010

Pages

148-164

Abstract

Given the significant increase of on-line services that require personal information from users, the risk that such information is mis- used has become an important concern. In such a context, information accountability is desirable since it allows users (and society in general) to decide, by means of audits, whether information is used appropriately. To ensure information accountability, information flow should be made transparent. It has been argued that data provenance can be used as the mechanism to underpin such a transparency. Under these conditions, an audit's quality depends on the quality of the captured provenance in- formation. Thereby, the integrity of provenance information emerges as a decisive issue in the quality of a provenance-based audit. The aim of this paper is to secure provenance-based audits by the inclusion of cryp- tographic elements in the communication between the involved entities as well as in the provenance representation. This paper also presents a formalisation and an automatic verification of a set of security properties that increase the level of trust in provenance-based audit results.

9. SPARQL query rewriting for implementing data integration over linked data

Author

Correndo, G and Salvadores, M and Millard, I and Glaser, H and Shadbolt, N

Booktitle

Proc. 1st International Workshop on Data Semantics

Year

2010

Pages

11

Abstract

There has been lately an increased activity of publishing structured data in RDF due to the activity of the Linked Data community. The presence on the Web of such a huge information cloud, ranging from academic to geographic to gene related information, poses a great challenge when it comes to reconcile heterogeneous schemas adopted by data publishers. For several years, the Semantic Web community has been developing algorithms for aligning data models (ontologies). Nevertheless, exploiting such ontology

alignments for achieving data integration is still an under supported research topic. The semantics of ontology alignments, often defined over a logical frameworks, implies a reasoning step over huge amounts of data, that is often hard to implement and rarely scales on Web dimensions. This paper presents an algorithm for achieving RDF data mediation based on SPARQL query rewriting. The approach is based on the encoding of rewriting rules for RDF patterns that constitute part of the structure of a SPARQL query.

10. A model for sharing of confidential provenance information in a query based system

Author

Nagappan, M. and Vouk, M.

Booktitle

Provenance and Annotation of Data and Processes

Year

2008

Pages

62-69

Abstract

Workflow management systems are increasingly being used to automate scientific discovery. Provenance meta-data is collected about scientific workflows, processes, simulations and data to add value. There is a variety of workflow management tools that cater to this. The provenance information may have as much value as the raw data. Typically, sensitive information produced by a computational processes or experiments is well guarded. However, this may not necessarily be true when it comes to provenance information. The issue is how to share confidential provenance information. We present a model for sharing provenance information when the confidentiality level is decided by the user dynamically. The key feature of this model is the Query Sharing concept. We illustrate the model for workflows implemented using provenance enabled Kepler system.

11. Provenance and the Price of Identity

Author

Chapman, A. and Jagadish, H

Journal

Provenance and Annotation of Data and Processes

Year

2008

Pages

106-119

Abstract

As developers acknowledge that provenance is essential, more and more datasets are attempting to keep provenance records describing how they were created. Some of these datasets are constructed using workflows, others cobble together processes and applications to manipulate the data. While the provenance needs are the same, the inputs and set of processes used must be kept, the identity needs are very different. We outline several identification strategies that can be used for data manipulation outside of workflows. We evaluate these strate- gies in terms of time to create and store identity, and the space needed to keep this information. Additionally, we discuss the strengths and weaknesses of each strategy.

12. Recording the Context of Action for Process Documentation

Author

Wootten, I. and Rana, O.

Booktitle

Provenance and Annotation of Data and Processes

Year

2008

Pages

45-53

Abstract

In reviewing evidence about real world processes, being aware of the context in which activities within such processes are performed enables us to make more informed judgements. It is necessary to distinguish between the environment in which a process occurs, and the sequence of activities which form part of the description of that process. Each of these types of information is complementary to understanding the other and therefore making associations between them is also important. Our work has been exploring the use of context whilst documenting a process and working toward a solution which incorporates the two. We present an approach to automatically relating properties of workflow actors to the documentation of the process within which these actors are involved.

13. Data provenance: A categorization of existing approaches

Author

Glavic, B and Dittrich, K R

Journal

Proc. Datenbanksysteme in Business, Technologie und Web

Year

2007

Volume

7

Number

12

Pages

227-241

Abstract

In many application areas like e-science and data-warehousing detailed information about the origin of data is required. This kind of information is often referred to as data provenance or data lineage. The provenance of a data item includes information about the processes and source data items that lead to its creation and current representation. The diversity of data representation models and application domains has lead to a number of more or less formal definitions of provenance. Most of them are limited to a special application domain, data representation model or data processing facility. Not surprisingly, the associated implementations are also restricted to some application domain and depend on a special data model. In this paper we give a survey of data provenance models and prototypes, present a general categorization scheme for provenance models and use this categorization scheme to study the proper- ties of the existing approaches. This categorization enables us to distinguish between different kinds of provenance information and could lead to a better understanding of provenance in general. Besides the categorization of provenance types, it is important to include the storage, transformation and query requirements for

the different kinds of provenance information and application domains in our considerations. The analysis of existing approaches will assist us in revealing open research problems in the area of data provenance.

14. Implementing a secure annotation service

Author

Khan, I. and Schroeter, R. and Hunter, J.

Booktitle

Provenance and Annotation of Data and Processes

Year

2006

Pages

212-221

Abstract

Annotation systems enable ``value-adding'' to digital resources by the attachment of additional data in the form of comments, explanations, references, reviews and other types of external, subjective remarks. They facilitate group discourse and capture collective intelligence by enabling communities to attach and share their views on particular data and documents accessible over the Web. Annotation systems vary greatly with regard to the types of content they can annotate, the extent of collaboration and sharing they allow and the communities which they serve. However many applications share the need to authenticate the source of annotations and restrict access to them - in order to protect intellectual property rights or personal privacy. This paper describes a secure, open source annotation system that we have developed that uses Shibboleth and XACML to identify and authenticate users and restrict access to annotations stored on an Annotea server.

15. Issues in automatic provenance collection

Author

Braun, U. and Garfinkel, S. and Holland, D. and Muniswamy-Reddy, K.K. and Seltzer, M.

Booktitle

Provenance and Annotation of Data and Processes

Year

2006

Pages

171-183

Abstract

Automatic provenance collection describes systems that observe processes and data transformations inferring, collecting, and maintaining provenance about them. Automatic collection is a powerful tool for analysis of objects and processes, providing a level of transparency and pervasiveness not found in more conventional provenance systems. Unfortunately, automatic collection is also difficult. We discuss the challenges we encountered and the issues we exposed as we developed an automatic provenance collector that runs at the operating system level.

16. Why and Where: A Characterization of Data Provenance

Author

Buneman, Peter and Khanna, Sanjeev and Tan, Wang-Chiew

Booktitle

Proc. International Conference on Database Theory

Year

2001

Pages

316-330

Abstract

With the proliferation of database views and curated data- bases, the issue of data provenance - where a piece of data came from and the process by which it arrived in the database - is becoming increasingly important, especially in scientific databases where understanding provenance is crucial to the accuracy and currency of data. In this paper we describe an approach to computing provenance when the data of interest has been created by a database query. We adopt a syntactic approach and present results for a general data model that applies to relational databases as well as to hierarchical data such as XML. A novel aspect of our work is a distinction between ``why'' provenance (refers to the source data that had some influence on the existence of the data) and ``where'' provenance (refers to the location(s) in the source databases from which the data was extracted).