Chapter 4.12

Mailing Lists and Social Semantic Web

Sergio Fernández
Fundación CTIC, Spain

Diego Berrueta
Fundación CTIC, Spain

Lian Shi
Fundación CTIC, Spain

Jose E. Labra
University of Oviedo, Spain

Patricia Ordóñez de Pablos
University of Oviedo, Spain

ABSTRACT

Electronic Mailing lists are a key part of the Internet. They have enabled the development of social communities who share and exchange knowledge in specialized and general domains. In this chapter the auhtors describe methods to capture some of that knowledge which will enable the development of new datasets using Semantic Web technologies. In particular, the authors present the SWAML project, which collects data from mailing lists. They also describe smushing techniques that normalize RDF datasets capturing different resources that identify the same one. They have applied those techniques to identify persons through the mailing lists of open source communities. These techniques have been tested using a dataset automatically extracted from several online open source communities.

INTRODUCTION

Early forms of electronic mailing lists were invented almost as soon as electronic Mail (e-Mail) and are a cornerstone of Internet, allowing a lot of people to keep up to date on news related with their interests. Besides direct messaging between individuals, mailing lists exist as private or public forums for information exchange in communities with shared interests. Mailing list archives are compilations

of the previously posted messages that are often converted into static HTML pages for their publication on the web. They represent a noteworthy portion of the contents that are indexed by web search engines, and they capture an impressive body of knowledge that, however, is difficult to locate and browse.

The reason for this difficulty can be traced back to the translation procedure that run to transform the e-mail messages into static HTML pages. This task is fulfilled by scripts that create static HTML pages for each message in the archive. In addition, some indexes (by date, by author, by thread) are generated and usually split by date ranges to avoid excessive growth.

On the one hand, this fixed structure reduces the flexibility when users explore the mailing list archives using their web browsers. On the other hand, most of the meta-data that were associated to each e-mail message are lost when the message is rendered as HTML for presentational purposes.

We propose to use an ontology and RDF (Resource Description Framework, Klyne 2004) to publish the mailing list archives into the (Semantic) Web, retaining the meta-data that were present in the messages. Additionally, by doing so, the information can be merged and linked to other vocabularies, such as FOAF (Brickley and Miller, 2005).

The rest of the chapter is organized as follows: in section 2 we describe the main developments of Social Semantic Web related with mailing lists. In section 3, we explain several techniques to collect RDF datasets from mailing lists and other social sources. Section 4 contains a description of the SWAML project that collects those RDF datasets from mailing lists. In section 5, we describe several applications that consume that data. In section 6, we discuss some experiments that we have done over those datasets. Finally, in section 7 we present some conclusions and future work.

SOCIAL SEMANTIC WEB

The Semantic Web vision tries to develop new ways to integrate and reuse the information published on the web. To that end, the W3C has developed several technologies, like RDF, which enable to add metadata descriptions that contain meaningful values and global properties to resources. The resulting metadata forms a graph model which can be easily linked with other graphs (Berners-Lee, 2006) incrementing the knowledge represented by the original graph. Those values and properties formalize the knowledge of a particular. In 2004, the W3C consortium developed OWL (Patel-Schneider et al, 2004), a web ontology language which facilitates the definition of those formalizations, called ontologies. Based on description logics, OWL has been adopted as the standard ontology language with several available editors, reasoners and tools. There have been also a number of ontologies developed in OWL for different purposes and with different level of detail, from generic to domain-specific ones.

On the other hand, in the last years, the concept of Web 2.0 has attracted a lot of interest. One of the key aspects of Web 2.0 applications is the social part of the web. Users are not considered as mere consumers of information, but also as producers. People want to share knowledge, establish relationships, and even work together using web environments. It is necessary to develop people-oriented web technologies which can represent people interests and that enable the integration and reuse of people related information in the same way that the semantic web vision advocates. These technologies can be seen as social semantic web and we expect that there will be more and more applications making use of them.

One of the first developments is the FOAF vocabulary, which represents basic properties of people, like their name, homepage, etc. as well as the people they know. FOAF descriptions are very

Related Content

Analyzing the Traffic Characteristics for Evaluating the Performance of Web Caching

G.P. Sajeevand M.P. Sebastian (2011). *Web Engineered Applications for Evolving Organizations: Emerging Knowledge (pp. 196-210).*

www.irma-international.org/chapter/analyzing-traffic-characteristics-evaluating-performance/53061

Context-Aware Service Provisioning in Next-Generation Networks: An Agent Approach

Vedran Podobnik, Krunoslav Trzecand Gordan Jezic (2007). *International Journal of Information Technology and Web Engineering (pp. 41-62).*

www.irma-international.org/article/context-aware-service-provisioning-next/2636

Performance Testing: Reference Technology and Languages

B. M. Subraya (2006). *Integrated Approach to Web Performance Testing: A Practitioner's Guide (pp. 52-76).*

www.irma-international.org/chapter/performance-testing-reference-technology-languages/23974

Approaches and Principles for UX Web Experiences: A Case Study Approach

Fernando Almeidaand José Augusto Monteiro (2017). *International Journal of Information Technology and Web Engineering (pp. 49-65).*

www.irma-international.org/article/approaches-and-principles-for-ux-web-experiences/176908

On Cost-Aware Heterogeneous Cloudlet Deployment for Mobile Edge Computing

Hengzhou Ye, Fengyi Huangand Wei Hao (2022). *International Journal of Information Technology and Web Engineering (pp. 1-23).*

www.irma-international.org/article/on-cost-aware-heterogeneous-cloudlet-deployment-for-mobile-edge-computing/297968