A Topic Modeling Approach for Web Service Annotation

Leandro Ordonez-Ante[*1], Ruben Verborgh[†2] and Juan Carlos Corrales[‡1]

[1]Grupo de Ingeniería Telemática (GIT), Universidad del Cauca , Calle 5 No 4-70, Popayán, Colombia
[2]Multimedia Lab, Ghent University , Gaston Crommenlaan 8/102, 9050 Ghent, Belgium

April 9, 2014

Abstract

This is a sample document which shows the most important features of the Standard LaTeX Journal Article class.

1 Introduction

During last two decades we have been witnessing how the Web has evolved from being a text and image repository on its early stages, to provide a huge offer of both information-providing and world-altering services. The current Web has posed a paradigm that revolutionized the generation and consumption dynamics of this kind of resources, encouraging its users, not merely to consume these services, but also to build and publish them [1].

This dominant paradigm of the current Web, has inspired the conception of initiatives into other communities such as the Telco providers, which include the GSMA's One API [?], and the ECMA-348 [?] and ECMA-323 [?] standards. Such initiatives promote for network operators to expose their capabilities and information via Web service interfaces, easing this way for users (service designers and developers) to create and deploy new telecom services, with a reduced time-to-market and tailored to their specific needs.

Thus, the service offering inside the Web is diversifying and steadily growing, so it is necessary to provide the users with increasingly intelligent mechanisms

---

[*]leandro@unicauca.edu.co
[†]ruben.verborgh@ugent.be
[‡]jcorral@unicauca.edu.co

for services search and retrieval, identifying in a truthful way the functionality provided by such resources while being able to deliver relevant services to the customer. The above has meant the transition from the traditional keyword-based or table-based search methods [?], to approaches supported on semantic Web technologies [?] which provide meaning for both the services specifications, and user queries, through a formal and machine-readable specification of knowledge (e.g. ontologies, taxonomies, lexical database and so on).

In practical terms, however, the actual implementation of semantic-based mechanisms for service retrieval has been restricted precisely due to the expensive procedure involved in the formal specification of services. Such a procedure comprises a regular, time-consuming task of semantic annotation, performed by hand by service developers, who additionally require specialized knowledge on models for semantic description of services (e.g. OWL, WSMO, RDF-S), as well as the aforementioned formal specifications of knowledge.

In order to overcome this limitation, currently some approaches are considered to tackle the problem of semantic service annotation, by applying knowledge discovery and emergent semantics techniques over a huge corpus of service descriptors, which in some cases already contains annotations made by consumers in a collaborative way. Those approaches however, has serious limitations in terms of the reliability of the users feedback they are built upon, which impacts the precision of search and selection tasks. Therefore it's considered necessary to develop mechanisms that enable the automation of semantic service annotation tasks. This paper introduces our proposal for service annotation, based on processing existing web service documentation resources for extracting information regarding its offered capabilities. By uncover the hidden semantic structure of such information through statistical analysis techniques, we are able to associate meaningful annotations to the services operations, and to group those operations into non-exclusive semantic related categories.

Based on this approach we have build Topicalizer, a tool that allows the user to process a bunch of SOAP API descriptors (WSDL documents), in order to group the technical information they contain into semantic categories, and specifying such categorization as RDF statements stored in a Sesame triple-store, to which users may access and issue SPARQL queries.

2   Motivation and Background

3   Related Work

4 Proposal

5 Experimentation

6 Discussion and Conclusions

References

[1] J. C. Yelmo, J. M. del Álamo, R. Trapero, and Y.-S. Martín, "A user-centric approach to service creation and delivery over next generation networks," *Comput. Commun.*, vol. 34, pp. 209–222, Feb. 2011.

[2] GSMA, "Gsma–oneapi for developers," 2013.

[3] ECMA-International, "Standard ecma-348 web services description language (wsdl) for csta phase iii," 2012.

[4] ECMA-International, "Standard ecma-323 xml protocol for computer supported telecommunications applications (csta) phase iii," 2011.

[5] A. Bernstein and M. Klein, "Towards high-precision service retrieval," in *Proceedings of the First International Semantic Web Conference on The Semantic Web*, ISWC '02, (London, UK, UK), pp. 84–101, Springer-Verlag, 2002.

[6] S. Chance, "Semantic service oriented architecture: An overview," 2009.