1. INTRODUCTION

Deep Web is considered to be Web content, which is not indexed by search engines. Although the content itself is not directly indexed, this often does not apply to descriptions of the endpoints used for retrieving the content. Such endpoints like Web forms, database services, Web services and alike have been already exploited by major search engines for partially accessing the Deep Web content. However, not much is known about how this retrieved content can be combined to generate added value. Neither is much known how the deep Web is organized from the viewpoint of data usage. At the same time recent advances [3, 8, 5] in automated semantic annotation of Web services interface descriptions in WSDL and schema matching in general have made it feasible to match effectively large quantities of Web services interfaces, which describe endpoints to surface content from the deep Web. Based on these matchings, Web services networks can be constructed and used for analysing aspects of Web services from network theoretic perspective [2, 7, 4].

Information diffusion is defined as the communication of knowledge over time among members of a social system [9]. This phenomenon has been studied between and within biosphere, microblogs [11], social networks [1, 10] and other domains [9], where the network structure is present. These studies have turned to be useful for revealing intrinsic properties of particular real world phenomena. Similarly analysis of information diffusion patterns in Web services networks is expected to reveal certain phenomena in the deep Web as well.

One of the important concerns of services providers is the way to increase the visibility and usability and ultimately the service/content popularity within or cross commodity. For example, it is quite reasonable belief that services facilitating their composition with other services (through intermediaries, semantic annotation,...) gain quite high popularity. It is also quite known that services operating cross boundaries of two different business, organizations or sectors are particularly important [6]. Moreover, the issues such as impact of services in or cross commodity, flow of information across commodities, correlation between service functionality and its usability impact, to mention a few, could be explored in the Web service domain. The result of such analysis not only is beneficial for service providers, but also provides valuable feedback to Web services community. However, to the best of our knowledge, there has been no work addressing the analysis of web service domain exploring the aforementioned issue.

In this paper we present a model for analysing information diffusion in Web services networks. The presented work is the initial step to realize the aforementioned analysis objectives. More specifically, we investigate how to discover information diffusion among categories of Web services given the network of Web services. Our hypothesis in this context is that analysis of information diffusion in Web services networks is the key mechanism, which can unlock the so far hidden reality of how services in different service industries have been designed. And since each service, which we use in our case study, reflects a fraction of the deep Web, we reveal the first findings about the usability of the deep Web as well.

The proposed approach relies on a set of semantically annotated and categorized web services to first construct a Web services network, then transform it into a category network, and finally compute a diffusion matrix. The diffusion matrix captures the volume of potential information flow between Web services categories. The captured information diffusion volume reflects collaboration between different service industries. As a case study, we evaluate the proposed approach on a collection of approximately 27000 Web services’ WSDL descriptions collected from different repositories of the Web.

2. APPROACH AND PRELIMINARY RESULTS

In the interest of measuring information diffusion between different categories of Web services, we rely on network presentation of semantically annotated and categorized Web services where services are described by WSDL interfaces. We define information diffusion in terms of information flow from output parameter(s) of a Web service operation to input parameter(s) of another Web service operation, provided that interoperability between the correlated input and output element holds. A semi-automatic ontology learning and annotation scheme [5] is recruited to annotate the collected Web service corpus. Then we exploit a semantic matching scheme to create an intermediate network out of semantically annotated Web service corpus. In parallel to annotation and matching process, Web service categorization is
Figure 1: Visualization of matrices of category weights between different categories of Global web services (on the left (+Null) categories and on the right (-Null) categories). Each entry is shaded according to a normalized Z-score representing whether the density of information flow is higher or less than expected at random. Darker shading indicates higher Z-scores. The diagonal represents information flow within the same category.

3. REFERENCES


1http://www.soatrader.com/web-services/