International Journal of Distributed Systems and Technologies
Volume 7 • Issue 1 • January-March 2016

User Interaction with Linked Data:
An Exploratory Search Approach

Dhavalkumar Thakker, School of Electrical Engineering and Computer Science, University of Bradford, Bradford, UK
Fan Yang-Turner, School of Computing, University of Leeds, Leeds, UK
Dimoklis Despotakis, School of Computing, University of Leeds, Leeds, UK

ABSTRACT

It is becoming increasingly popular to expose government and citywide sensor data as linked data. Linked data appears to offer a great potential for exploratory search in supporting smart city goals of helping users to learn and make sense of complex and heterogeneous data. However, there are no systematic user studies to provide an insight of how browsing through linked data can support exploratory search. This paper presents a user study that draws on methodological and empirical underpinning from relevant exploratory search studies. The authors have developed a linked data browser that provides an interface for user browsing through several datasets linked via domain ontologies. In a systematic study that is qualitative and exploratory in nature, they have been able to get an insight on central issues related to exploratory search and browsing through linked data. The study identifies obstacles and challenges related to exploratory search using linked data and draws heuristics for future improvements. The authors also report main problems experienced by users while conducting exploratory search tasks, based on which requirements for algorithmic support to address the observed issues are elicited. The approach and lessons learnt can facilitate future work in browsing of linked data, and points at further issues that have to be addressed.

KEYWORDS

Big Data, Exploration, Exploratory Search, Linked Data, Smart Cities

1. INTRODUCTION

The development of smart cities involves a multitude of technologies and processes (Deakin, 2013). Linked data (Shadbolt et al., 2012) is one such technology. Linked data contributes to the goal of Semantic Web to extend current human-centric Web with machine-interpretable data to process information automatically. One of the major factors for the success of the linked data technologies has been the availability of large amount of data in various formats and domains, often from the government departments or citywide sensors. The data generated from smart city censors is complex due to the heterogeneity of data, the rate of its generation and sheer volume. Linked data technologies allow processing data for increasing interoperability, easing data integration and providing support for information retrieval and knowledge discovery tasks (Roche, Lecue, Llaves, & Corcho, 2015). In parallel with engineering solutions for the seamless integration of linked data, efforts have been made to facilitate user interaction with such data to support smart city goals of helping users to learn and make sense of complex and heterogeneous data. There are arguments that linked data can be utilised to enable user-oriented exploratory search systems for the future systems including smart city applications (Thakker, Dimitrova, Cohn, & Valdes, 2015; Waitelonis, Knuth, Wolf, Hercher, & Sack,

DOI: 10.4018/IJDST.2016010105

2010). Stepping on such arguments, this paper aims to apply a systematic experimental methodology to examine user behaviour when interacting with a linked data browser to learn lessons to build better user-oriented exploratory search systems as smart city applications.

In contrast to regular search, exploratory search gives a more complete overview of a topic. Exploratory search is open-ended, multi-faceted, and iterative in nature and is commonly used in scientific discovery, learning, and sense making (Marchionini, 2006; White & Roth, 2009). Exploration demands more time, effort and creativity from the user but rewards the user with deeper knowledge (Marchionini, 2006). Exploratory search is particularly beneficial for ill-structured problems and more open-ended goals, with persistent, opportunistic, iterative query processes. Exploratory tasks inherently have uncertainty, ambiguity and discovery as common aspects. Linked data appears to offer a great potential for exploratory search. For example, earlier studies suggested that tags (Kammerer, Nairn, Pirolli, & Chi, 2009) or some form of presentation of the knowledge space structure (Qu & Furnas, 2008) could benefit browsing and learning. The work presented here starts from these claims (which did not exploit semantics and linked data), and examines the role of semantic tags and their effect on browsing, and sense-making in a class of applications called linked data browsers.

We argue that the time is ripe to start experimentation with exposing linked semantic data to end users, which requires carefully selected domains and systematically designed studies. The paper presents such a case study and points at the potential of Linked data browsers to facilitate exploratory search. We follow linked data tenets and exploit available linked datasets in the application domains we experiment with. A linked data browser shell, called Peruse, which provides auni-focal faceted exploration of linked semantic data, was developed (Section 4). The paper focuses on an instantiation of Peruse in a Music domain– which links music datasets from the Linked data and social content from Amazon reviews about musical instruments. The main contribution of this paper to linked data research is a systematic study (Section 3) to get an insight of main issues related to exploratory search in linked data, an analysis of the role of semantics, and an indication of heuristics to facilitate user exploration (Section 5). The paper concludes by outlining main contributions.

2. RELATED WORK

2.1. Linked Data

Semantic Web is as an evolution of the current Web that consists largely of documents for humans to read to one that includes data and information for computers to manipulate(Shadbolt, Hall, & Berners-Lee, 2006).Linked data is considered as a key enabler technology of this vision. However, linked data advocates the simplest form of semantics, and has thus far focused on promoting the publication, sharing and linking of data on the Web (Domingue, Pedrinaci, Maleshkova, Norton, & Krummenacher, 2011). There are four simple principles for publishing linked data (Berners-Lee, 2011):

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things.

RDF (Resource Description Framework)(Klyne & Carroll, 2006) is a simple data model for semantically describing resources on the Web. SPARQL (SPARQL Protocol and RDF Query Language)(Pérez, Arenas, & Gutierrez, 2006) is a query language for RDF data, which supports querying diverse data sources, with the results returned in the form of an RDFgraph(Domingue et al., 2011).

Linked data principles have been adopted by many organisations resulting in a large number of datasets available as Linked data. The most recent statistics[1] show that there are such 3308 datasets