Link to publisher version: *http://dx.doi.org/10.4018/IJDST.2016010105*

Citation: Thakker D, Yang-Turner F and Despotakis D (2016) User Interaction with Linked Data: An Exploratory Search Approach. International Journal of Distributed Systems and Technologies. 7(1): 79-91.

Table of Contents

International Journal of Distributed Systems and Technologies

Volume 7 • Issue 1 • January-March-2016 • ISSN: 1947-3532 • eISSN: 1947-3540

An official publication of the Information Resources Management Association

Special Issue on Advances in Cloud for Smart Cities

The *International Journal of Distributed Systems and Technologies* is indexed or listed in the following: ACM Digital Library; Bacon's Media Directory; Cabell's Directories; Compendex (Elsevier Engineering Index); DBLP; Google Scholar; INSPEC; JournalTOCs; Library & Information Science Abstracts (LISA); Linguistics & Language Behavior Abstracts (LLBA); MediaFinder; SCOPUS; The Standard Periodical Directory; Ulrich's Periodicals Directory

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.

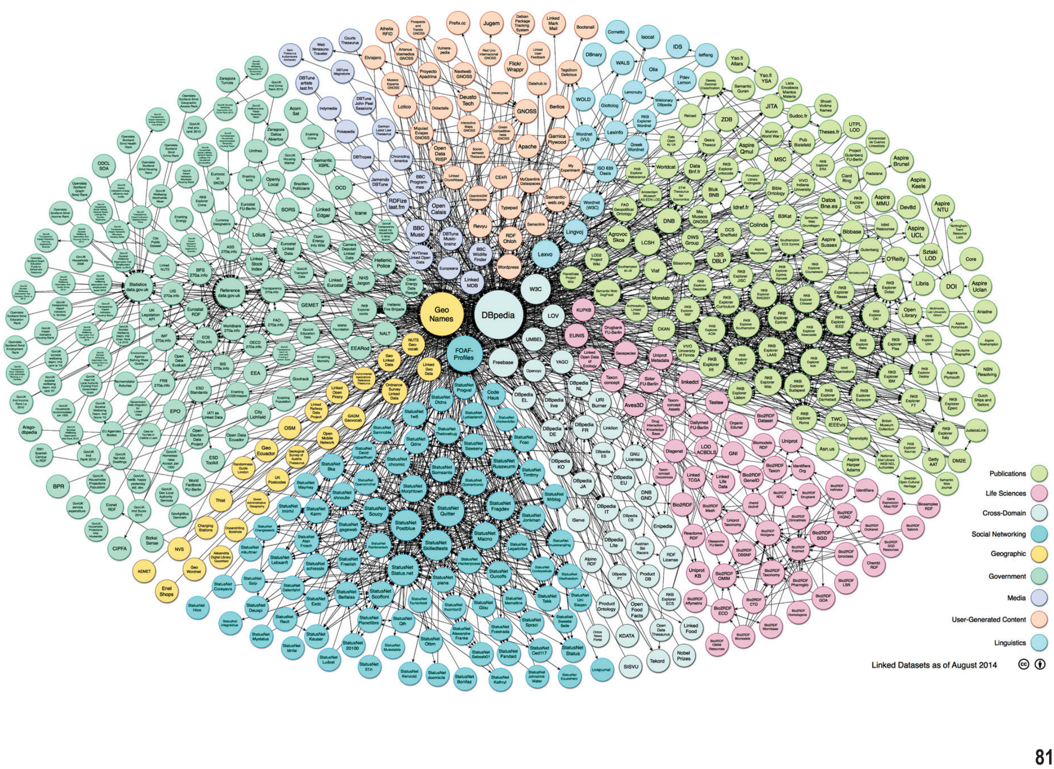
RDF (Resource Description Framework)(Klyne & Carroll, 2006) is a simple data model for semantically describing resources on the Web. SPARQL (SPARQL Protocol and RDF Query Language)(Pérez, Arenas, & Gutierrez, 2006) is a query language for RDF data, which supports querying diverse data sources, with the 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 consisting of approximately 3.9 billion facts. Figure 1 shows available Linked datasets. Each bubble in the figure indicates the name of an organisation or a dataset. The bi-directional arrows indicate the connection maintained by the datasets, where a dataset refers to data from another dataset. These datasets belong to various categories such as media, government, publications, life sciences, geographic, cross-domain, user-generated content, and social networking. A study conducted in 2014 found that the number of Linked datasets has approximately doubled between 2011 and 2014(Schmachtenberg, Bizer, & Paulheim, 2014).

2.2. Linked Data based Exploratory Search Systems

Exploratory search is different from regular search. 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). One class of systems developed to facilitate exploratory search focuses on faceted search using Linked data. Faceted search works by suggesting restrictions as facets, i.e. selectors for subsets of the current set of items (Ferré & Hermann, 2011). The work presented in this paper examines the role of semantic tags and their effect while browsing and learning in another class of applications called "semantic linked data browsers" or "Linked data browsers" for short. Linked data browsers have emerged from a collective effort in the semantic research community. Such browsers operate on semantically augmented data (e.g. tagged content) and layout browsing trajectories using relationships in the underpinning ontologies. Tabulator (Berners-Lee et al., 2006) can be considered the first Linked data browser that enables users to browse data by following semantic links to resources. Two types of linked data browsers have since emerged – (i) pivoting (or set-oriented browsing) and (ii) multi-pivoting. Ina pivoting browser, a many-to-many graph browsing technique is used to help a user navigate from a set of instances in the graph through common links (Popov, Schraefel, Hall, & Shadbolt, 2011).Exploration is often restricted to a single start point in the data

Figure 1. "Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/"



and uses ‘a resource at a time’ to navigate anywhere in a dataset(Araujo, Schwabe, & Barbosa, 2009). This form of browsing is also referred as uni-focal browsing. A second type of browsers supports multi-pivoting that allows a user to start from multiple points of interest. For example, PolyZoom enables multi-focus exploration of maps for a user to zoom various parts of the map at the same time (Javed, Ghani, & Elmqvist, 2012).Different interfaces have also been proposed in recent work. Notable attempts are Parallax (Huynh & Karger, 2009), VisiNav(Harth, 2009), Facet Graphs(Heim, Ertl, & Ziegler, 2010) and I-CAW(Thakker, Despotakis, Dimitrova, Lau, & Brna, 2012; Thakker, Osman, Gohil, & Lakin, 2010).This paper presents a traditional Linked data browser that provides a uni-focal interface for browsing through several linked datasets. Our contribution to Linked data browsers is considering the user behaviour with the browser and, notably, examining how Linked data browsers can support exploratory search. We also identify key challenges for exploratory search users face and derive requirements on how to extend a browser to address these challenges.

2.3. User Studies with Exploratory Search Systems

The exploratory search has been trialled in different domains. Nguyen et al. (Nguyen, Bodenreider, Minning, & Sheth, 2011) provided a web tool with a graphical interface for interactive knowledge exploration within the biomedical domain (which integrates the experimental data with the knowledge extracted from the PubMed articles). An empirical study (Heo & Yi, 2008) was conducted to examine the impact of browsing an ontology-driven information system on users’ ability to learn domain knowledge. While the expected learning effect did not occur among differently trained participant groups, previous online search experience showed a positive correlation with participant performance. A mixed approach supporting exploratory search through a simple keyword-based search interface over diverse sources such as Linked data, Wikipedia, Twitter, and Google News was provided in (Musetti et al., 2012). An application of Linked data browser in learning was developed in the recently completed mEducator project (Dietze, Yu, Pedrinaci, Liu, & Domingue, 2011).

Despite these initial prototypes, more case studies and experimentation are needed to identify which features in a Linked data browser may benefit or hinder the exploratory process. For this line of investigation, experimental and methodological tool set developed by information retrieval and human-computer interaction research on exploratory search (Parameswaran, Sarma, Garcia-Molina, Polyzotis, & Widom, 2011) can provide useful guidelines. Metrics that emerged from brainstorming and breakout sessions at a workshop on exploratory search (White & Roth, 2009) indicate how to assess the performance of exploratory search systems, considering: engagement and enjoyment, information novelty, task success, task time, learning and cognition. However, as raised by an evaluation of an exploratory search tool (CoSen) (Qu & Furnas, 2008), when people want to change or grow their current knowledge structure, they may not be able to specify precisely what is needed. Not only is this structural-information need hard to express, there is also a lack of effective search mechanisms for finding appropriate structures. This supports the assumption that semantics may be able to fill this gap and underpins the approach and the findings in this paper.

An evaluation study of a faceted search interface (Kules, Capra, Banta, & Sierra, 2009) showed that facets played a major role in the browsing process, accounting for about half of the time spent looking at actual results. This underscores the importance of facets, which is adopted in the design of the Linked data browser presented in this paper. We further examine the role of semantic facets in exploratory search tasks. Our study focuses specifically on the effect of the browser on users’ ability to complete exploratory search tasks in an unfamiliar domain and identifies aspects for further intelligent support to facilitate fruitful exploration. We present requirements elicited by involving users in interaction with the system, which complements the requirements in (Uren et al., 2010). Notably, the user study steps on conducting user studies for deriving requirements for exploratory search interfaces.

There have been workshops on the topic of challenges of user interactions, e.g. Semantic Web User Interactions (SWUI), Intelligent Exploration of Semantic Data (IESD) and Exploratory Search series. These events have produced useful guidelines and shared experiences about outstanding challenges, requirements and methodologies to follow, e.g. (White, Muresan, & Marchionini, 2006). Our work contributes to these ongoing efforts by focusing specifically on the effect of the browser on users' ability to complete exploratory search tasks and identifying requirements for further intelligent support to facilitate fruitful exploration.

3. METHODOLOGY TO STUDY USER BEHAVIOUR

Peruse is described in Section 4. To get an insight of how Peruse can support exploratory search through Linked data, we conducted an experimental study following methodological recommendations for evaluating exploratory search systems (White & Roth, 2009). Linked data browsers, including Peruse, are unique systems that use ontologies, often multiple, to support the exploration. Hence, a task-based study approach was preferred over comparing the performance of the Peruse system to a baseline, search tools, such as Google or Wikipedia. To design the study tasks, we have followed the main characteristics of exploratory search tasks summarised in (Wildemuth & Freund, 2012): the main goal is learning and/or investigation of a musical instrument; there is a low level of specificity about the information needed and how to find it; search is open-ended, requires finding several items and involves a degree of uncertainty; tasks are 'not too easy' and include multiple facets. The study required participants to complete two tasks related to researching musical instruments and positioned within an advertisement scenario of a hypothetical music shop. In both tasks, the participants were given an entry point to the browser and a template form to fill in their answers. The two tasks have different goals and require different strategies for browsing. For example, completion of Task 1 required mainly browsing through the musical instrument classification and reading descriptions provided from DBpedia. For this task, participants produced a form with three sections: (i) characteristics of bouzouki; (ii) instruments similar to bouzouki; and (iii) summary of distinctive features of Abouzouki. In contrast, Task 2 required browsing through content about music albums and artists and reading through Amazon reviews. For this task, the participants produced a form with two sections: (i) interesting albums, and (ii) key features of electric guitar people look. After each task, the users were asked to fill out a short questionnaire to rate their subjective level of cognitive load using a modified version of the NASA-TLX questionnaire (Hart & Staveland, 1988).

The study addressed the following research questions:

Q1: How well can users without domain knowledge perform exploratory search tasks using Peruse, and what is the benefit/drawback of its features?
Q2: What further improvements, in the form of requirements, have to be fulfilled to make Peruse (and Linked data browsers in general) suitable for exploratory search tasks.

The study involved 12 participants recruited on a voluntary basis. Half of the participants were native speakers and the other half spoke and communicated in English fluently. All participants had IT background, experience in web search, and most participants had good experience in data analysis. Half of the participants have visited sites with music information regularly while the others did this only occasionally; 4 participants listened to online music sites daily. Half of the participants indicated that they currently practiced musical instruments. Each participant attended an individual session of about an hour conducted and observed by an experimenter. The participants were asked to think aloud; the experimenter kept notes of the interaction noting any interesting comments made. All browser clicks were logged for further analysis.

4. PERUSE: A UNI-FOCAL LINKED DATA BROWSER

This section presents the implementation of Peruse[2], a Linked data browser shell, which combines state-of- the-art semantic web technologies for semantic augmentation, semantic query and representation. The main goal of Peruse is to enable users to tap easily into resources built from the Web and, in particular, exploring the use of the Linked data paradigm (Bizer, Heath, & Berners-Lee, 2009). For the study, Peruse was instantiated in the music domain.

4.1. Architecture

Figure 2 depicts the generic three-layer architecture for Peruse: (i) the data layer consisting knowledge sources and content, (ii) the processing layer consisting modules for semantic augmentation and query, and (ii) the presentation layer for content browsing.

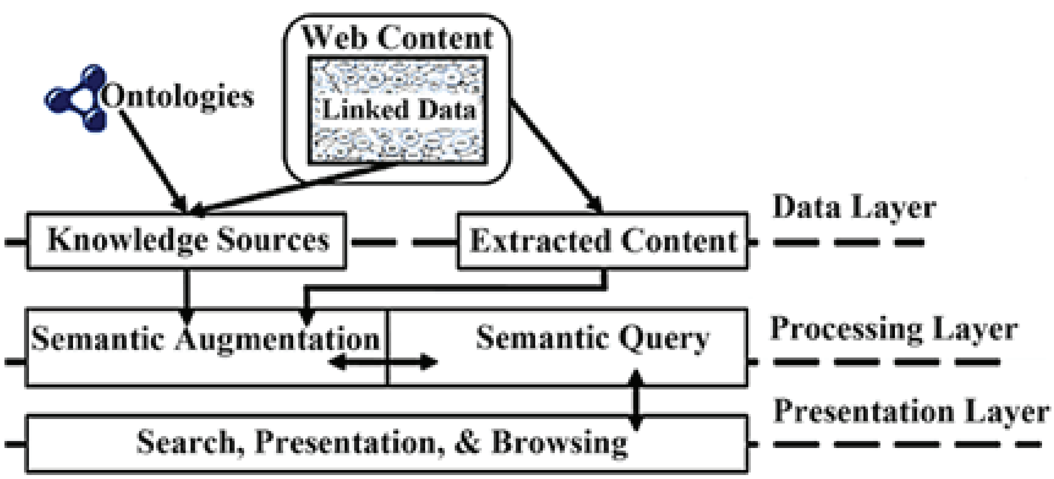
4.2. Data Layer

This layer contains domain-specific *ontological knowledge sources* and *content* assembled from the Web (Linked Data and other sources, as appropriate). The knowledge sources consist of graphs of ontological concepts relevant to the domain of interest. They provide the foundation for semantic augmentation of the content and the structure for semantic trajectories for browsing in the presentation layer. The content in Peruse is in atextual format and can be assembled from more than one online platform, e.g. blogs, reviews, comments (e.g.(Thakker et al., 2012)). For the study, Peruse utilised publicly available music related datasets and RDF version of Amazon Reviews (Thakker, Dimitrova, Lau, Yang-Turner, & Despotakis, 2013).

4.3. Processing Layer

This layer has two main services: (i) semantic augmentation of the assembled content and (ii) semantic queries to retrieve content for the presentation layer. Semantic augmentation(also known as semantic tagging) is a process of attaching semantics (in the form of “concepts” in an ontology) to a selected part of a piece of text. The augmented text can be automatically interpreted by another software component. The semantic augmentation module in Peruse includes semantic repository, information extractor and semantic indexer. *The Semantic Repository* combines the functionality of an RDF-based DBMS and an inference engine. *The Information Extractor* (using GATE- General Architecture for Text Engineering (Cunningham, 2002)) produces annotated sets of extracted entities with offset,

Figure 2. Architecture of the Linked data browser “Peruse”



ontology URI and type information. *The Semantic Indexer* (using Sesame SPARQL (Prud'Hommeaux & Seaborne, 2008) API) converts these annotated sets to RDF triples. Semantic Queries take a term(s) or concept(s) as keywords and output information relating to the matching concept(s) and content(s). Together with the semantic repository, semantic queries implement various concept/content lookup functionalities to find related and relevant concept(s) or content(s) from are pository.

4.4. Presentation Layer

This layer provides a front-end for the output of semantic queries from the processing layer. As a uni-focal browser, Peruse allows browsing one entity at a time. The entity that is the current focus is called a "focus entity". The layout template for a focus entity includes three main facets and a description extracted from the knowledge data sets for the *focus entity*: (i) Facet 1, called "features", includes facts about the focus entity; (ii) Facet 2 called "relevant information", includes terms related to the focus entity; and (iii) Facet 3 shows content related to the focus entity. Hyperlinks are provided to further details for the retrieved objects.

5. RESULTS AND DISCUSSION

Two musical instrument experts (one for Bouzouki, one for Electric guitar) have marked the outcome of participants for the two tasks. The marking is to measure how successful the participants have been in completing the tasks using Peruse.

For task 1, participants produced a form with three sections: (i) characteristics of bouzouki; (ii) instruments similar to bouzouki; and (iii) summary of distinctive features of Abouzouki. The expert examined the information about bouzouki (using both the description and the semantics presented in Peruse) and identified main characteristics. The participants' answers were compared with the experts' answer and scored according to the overlap. The learning outcome of all participants on the different components in task 1 is given in Table 1. All together, the participants identified 44 characteristics (70%, individual score median 4) from the description section of bouzouki (including the picture), and 19 descriptions (30%, individual score median 1.5) from the semantic tags. The difference between the two sets is significant (Wilcoxon test, $W=-60$, $p<0.01$), i.e. participants tended to use the instrument links in answering the task. Also, three participants pointed out characteristics from the bouzouki picture (e.g. elegant looking or tear-drop shape).

For task 2, the participants produced a form with two sections: (i) interesting albums, and (ii) key features of electric guitar people look. The electric guitar expert examined the electric guitar content presented by Peruse - which presents a total of 31 unique albums – with information of artist, biographical detail, photo, other albums by the same artist, country, and range of instruments. Against this knowledge, each album described by a user was marked considering: no attempt, superficial fact (i.e. can be obtained by one look at the page), interesting insight on one aspect, and the maximum score was for indication of deeper thinking on more than one aspects. For Amazon reviews, the expert conducted a content analysis on the Amazon reviews related to electronic guitar available in Peruse to extract features that people referred to. The extracted list was used to mark the list produced by each user. To compare task success, the scores for both tasks were computed as percentages.

The average performance on task 2 was 48%, which was significantly lower than task 1 (Wilcoxon test, $W=74$, $p<0.005$), yet there was no significant difference in the participants' confidence scores.

Table 1. Average percentage scores for task 1

Characteristics	Similar instruments	Similarity-difference	Distinctive features	Overall score
65%	80%	70%	68%	**70%**

Task 2 was more frustrating with border line significance (Wilcoxon test, W=-34, $p<0.05$). There was no correlation between scores and confidence – participants seemed confident that they did as much as they could in the given time.

The findings from the study provide insight into exploratory search behaviour when browsing through Linked data. We will summarise by revisiting the main questions of the study.

5.1. Use of Linked Data Browsers for Exploratory Search [Q1]

The study provides evidence that *Linked data browsers are suitable for exploratory search*. Users without domain knowledge could complete investigation tasks with Peruse fairly quickly and create meaningful answers (provided that there was suitable content). None of the users indicated that they were lost in the information space, and they used the browser facets to *navigate easily through Linked semantic data*. Users performed better and felt less frustrated, on the *analytical task1*that required browsing at the classification level (which was more about exploration on the structure of concepts/knowledge). Peruse *did not appear very beneficial for tasks requiring creativity and exploration at the content level, such as task 2* – see improvement discussed below.

5.2. Future Improvement to Realise the Potential of Linked Data Browsers for Exploratory Search [Q2]

The study indicated the need for further algorithmic support to achieve the exploratory search potential of Linked data browsers. The following observations are based on the study speculating this algorithmic support. Each observation was assessed to elicit requirements for supporting exploratory environments.

- **Observation 1. Abstraction Conundrum:** While browsing specific instruments, performances and performers, two participants clicked on abstract concepts, such as instrument, performance and performer, from the music ontology. In both cases, the participants were looking for concrete information (e.g. participant-12 clicked on instrument in task 1 when seeking for more detail about a musical instrument, while participant-05 clicked on performer and performance in task 2 when seeking more detail about an album). The aggregated datasets in Peruse have large number of instances for the abstract concepts (which is typical of Linked datasets), which led to confusion as the result was a long list of performers, performance and instruments, and the participants quickly pressed the back button on their browsers.
    - **Requirement 1. Offering Semantic Links at an Appropriate Level of Abstraction:** The above observation motivates consideration on identifying what can be algorithmically offered as the right level of abstraction on various browsing junctures. This is especially important when the abstract concepts have large amount of concrete instantiations. The main challenge here is what to suppress and what to display to the user; e.g. how to decide which performances out of 71k to display when a user is on the entity page of the abstract concept performance.
- **Observation 2. Exploring Entities/Content with Insufficient Information:** Another interesting case is the high number of 'empty clicks' - the user clicks on a link and is taken to a page with no information, sees that this link is not helpful and quickly returns to the previous page. In task 1, such clicks concerned similar instruments, e.g. there was no information about bajitar, xalam, rebab. In task 2 such clicks concerned performances (music albums) and happened quite often. 'Empty clicks' leading to pages with no information was seen as one of the main reasons for user's frustration. At the same time, may be due to their experience of links that lead to dead ends, some links were perceived as empty without exploring them further and the users missed to click on important for the tasks information (e.g. pages about musical instruments were abandoned, although there was useful information about relevant instruments; or interesting facts about an album artist were overlooked as the users did not click on the corresponding link). With Linked data, it is typical to find entities that do not have much explanation or links to other entities.

Similar issues were observed with content (Amazon Reviews in our user study). The textual content in Linked data browsers are semantically tagged and made available via one of the facets. Users clicked to view some of the Amazon reviews to find out more information about an instrument and its review. However, some of the reviews were deemed to have insufficient information to be useful. This observation is in line with relevant research conducted which concludes that not all reviews are equally helpful (for example, (Mudambi & Schuff, 2010) identifies useful reviews have considerable review depth compared to non-useful reviews).One can extrapolate such observations to be generic enough to be applied to social content and conclude that social content has a variety of usefulness levels, while being possible to find content that has insufficient information to be of help in browsing.

   - **Requirement 2. Reduce Entity Link Options:** Avoid showing entity links that do not lead to any new information. Reduce the number of entity links shown to the user based on their browsing value; allowing reduction of clutter and confusion. The challenge here is to define what 'browsing value' is and how to calculate it for an entity with respect to other entities from the same entity page.
   - **Requirement 3. Reduce Content Link Options:** Avoid showing content links that do not lead to any new information. Reduce the number of content links shown to the user based on their helpfulness/usefulness. The challenge here is to decide the parameters of helpfulness/usefulness of content.

- **Observation 3. Varied selection strategies while facing too many choices.** Both tasks (deliberately) put the users in situations where they had too many choices. This means that the users had a large number of links to review while in a focus entity page. For example, the bouzouki page included 12 different links in the facts facet (11 links to concepts in the middle classification level and 1 link to the abstract concept instrument) and 51 links in the terms facet (43 links to musical instruments and 8 links to performances).The entry point in task 2, the electric guitar page, included 18 links in the facts facet (to concepts in the middle and upper classification levels), 78 links to albums in the terms facet, and 8 links to Amazon reviews in the content facet. This is a typical situation with datasets from Linked data. For example, for the DBpedia dataset, which has 3.5M entities and 627M triples, on average, a user might have to review 192 links while exploring a focus entity.

We observed users following different strategies when presented with too many choices in the browsing interface: (i) clicking on the nearest classification link from the 'features facet' (e.g. plucked string instruments or string instruments) to see general characteristics in the case of bouzouki as part of task 1. However, users rarely clicked on links from the 'features' facet as part of task 2, as the task did not require this; (ii) clicking on instruments mentioned in the 'relevant information facet'– (e.g. lute and mandolin mainly in task2; (iii) clicking on something (e.g. 'an instrument') that 'sounds familiar' (e.g. sitar, banjo, pipa in task 1); (iv) click on something (e.g. 'an instrument or an album') that sounds interesting or unusual (e.g. oud, xalam in task 1 and noticing a women artist or something interesting in the album name in task 2); (v) clicking on something that looks important (e.g. an artist has several albums in task 2); and, (vi) clicking randomly (after exhausting other strategies).

This observation is in line with the latest research in search engines and HCI; increasing numbers of options can make designers and users feel less confident when deciding and less happy with the results (Oulasvirta, Hukkinen, & Schwartz, 2009; Schwartz & Kliban, 2004).To support the varied level of selection strategies, two requirements are derived.

   - **Requirement 4. Make Facet Selection Process Dynamic and Intuitive:** The use of different types of facets is useful but dependent on tasks. The challenge here is to give control to users and help them to decide or make it easier for them to choose the facet and when they would like to use it. For example, providing guideline on the utility of different types of facets in the system and allowing facets to be added or removed as required by the user.

- **Observation 4. Text and Media Information Influence User Experience and Performance:** For task 1, a great deal of performance owed to the use of textual description and images while identifying characteristics of an instrument. Hence, there is value in offering unstructured (textual and multimedia information) in conjunction with structured data (semantics) for exploration.
    - **Requirement 5. Offer Relevant Multimedia or Textual Information:** The exploration tools developers shall carefully select multimedia or textual information for the domain and make them available as part of the focus entity pages. For example, in Peruse instruments and performances related pages can contain YouTube videos of instruments or performances involving instruments.

6. CONCLUSION

We have presented a study with a traditional uni-focal linked data browser to observe the browsing behaviour of users while interacting with several linked datasets aiming at deriving requirements to inject intelligent features. We have found several intricate challenges that are applicable to typical interaction over linked semantic datasets. For example, the disparity of the options available while browsing from an entity. In some cases a large number of links available from an entity, hence posing too many options for the user to choose from and in other cases no links or information available making users frustrated. We have also observed and reported varied levels of selection strategies when a user is faced with too many options.

As a continuation of our work, we are implementing requirements identified in this study in Peruse. With this extension, we will conduct a comparative user study with the current system presented in this paper as a baseline. We also intend to involve a large number of participants to exploit appropriate quantitative analysis techniques.

REFERENCES

Araujo, S., Schwabe, D., & Barbosa, S. (2009). Experimenting with explorator: a direct manipulation generic rdf browser and querying tool. Visual Interfaces to the Social and the Semantic Web (VISSW 2009), Sanibel Island, Florida.

Berners-Lee, T. (2011). Linked data-design issues (2006). *URL* http://www. *w3. org/DesignIssues/LinkedData. html*.

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., . . . Sheets, D. (2006). *Tabulator: Exploring and analyzing linked data on the semantic web.* Paper presented at the Proceedings of the 3rd International Semantic Web User Interaction Workshop.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332. doi:10.1207/s1532690xci0804_2

Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities*, *36*(2), 223–254. doi:10.1023/A:1014348124664

Deakin, M. (2013). *Smart cities: governing, modelling and analysing the transition*. Routledge.

Dietze, S., Yu, H. Q., Pedrinaci, C., Liu, D., & Domingue, J. (2011). SmartLink: a Web-based editor and search environment for Linked Services The Semanic Web: Research and Applications (pp. 436-440): Springer.

Domingue, J., Pedrinaci, C., Maleshkova, M., Norton, B., & Krummenacher, R. (2011). *Fostering a relationship between linked data and the internet of services*. Springer. doi:10.1007/978-3-642-20898-0_25

Ferré, S., & Hermann, A. (2011). *Semantic search: Reconciling expressive querying and exploratory search The Semantic Web–ISWC 2011* (pp. 177–192). Springer.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. doi:10.1016/S0166-4115(08)62386-9

Harth, A. (2009). *Visinav: Visual web data search and navigation.* Paper presented at the Database and Expert Systems Applications. doi:10.1007/978-3-642-03573-9_17

Heim, P., Ertl, T., & Ziegler, J. (2010). *Facet graphs: Complex semantic querying made easy The Semantic Web: Research and Applications* (pp. 288–302). Springer.

Heo, M., & Yi, M. (2008). *An empirical study of the learning effect of an ontology-driven information system.* Paper presented at the Proceedings of the nineteenth ACM conference on Hypertext and hypermedia. doi:10.1145/1379092.1379137

Huynh, D. F., & Karger, D. (2009). *Parallax and companion: Set-based browsing for the data web.* Paper presented at the WWW Conference. ACM.

Javed, W., Ghani, S., & Elmqvist, N. (2012). *Polyzoom: multiscale and multifocus exploration in 2d visual spaces.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. doi:10.1145/2207676.2207716

Kammerer, Y., Nairn, R., Pirolli, P., & Chi, E. H. (2009). *Signpost from the masses: learning effects in an exploratory social tag search browser.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems. doi:10.1145/1518701.1518797

Klyne, G., & Carroll, J. J. (2006). Resource description framework (RDF): Concepts and abstract syntax.

Kules, B., Capra, R., Banta, M., & Sierra, T. (2009). *What do exploratory searchers look at in a faceted search interface?* Paper presented at the Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. doi:10.1145/1555400.1555452

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, *49*(4), 41–46. doi:10.1145/1121949.1121979

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *Management Information Systems Quarterly*, *34*(1), 185–200.

Musetti, A., Nuzzolese, A. G., Draicchio, F., Presutti, V., Blomqvist, E., Gangemi, A., & Ciancarini, P. (2012). Aemoo: Exploratory search based on knowledge patterns over the semantic web. *Semantic Web Challenge, 136*.

Nguyen, V., Bodenreider, O., Minning, T., & Sheth, A. P. (2011). The knowledge-driven exploration of integrated biomedical knowledge sources facilitates the generation of new hypotheses.

Oulasvirta, A., Hukkinen, J. P., & Schwartz, B. (2009). *When more is less: the paradox of choice in search engine use.* Paper presented at the Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. doi:10.1145/1571941.1572030

Parameswaran, A., Sarma, A. D., Garcia-Molina, H., Polyzotis, N., & Widom, J. (2011). Human-assisted graph search: It's okay to ask questions. *Proceedings of the VLDB Endowment*, *4*(5), 267–278. doi:10.14778/1952376.1952377

Pérez, J., Arenas, M., & Gutierrez, C. (2006). *Semantics and Complexity of SPARQL.* Paper presented at the International semantic web conference.

Popov, I. O., Schraefel, M., Hall, W., & Shadbolt, N. (2011). *Connecting the dots: a multi-pivot approach to data exploration The Semantic Web–ISWC 2011* (pp. 553–568). Springer.

Prud'Hommeaux, E., & Seaborne, A. (2008). SPARQL query language for RDF. *W3C recommendation, 15*.

Qu, Y., & Furnas, G. W. (2008). Model-driven formative evaluation of exploratory search: A study under a sensemaking framework. *Information Processing & Management*, *44*(2), 534–555. doi:10.1016/j.ipm.2007.09.006

Roche, S., Lecue, F., Llaves, A., & Corcho, O. (2015). *Special Issue on "The Role of Semantics in Smart Cities". Semantic Web - Interoperability, Usability*. Applicability.

Schmachtenberg, M., Bizer, C., & Paulheim, H. (2014). *Adoption of the linked data best practices in different topical domains The Semantic Web–ISWC 2014* (pp. 245–260). Springer.

Schwartz, B., & Kliban, K. (2004). *The paradox of choice: Why more is less*.

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic web revisited. *IEEE Intelligent Systems*, *21*(3), 96–101. doi:10.1109/MIS.2006.62

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., & schraefel, . (2012). Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, *27*(3), 16–24. doi:10.1109/MIS.2012.23

Thakker, D., Despotakis, D., Dimitrova, V., Lau, L., & Brna, P. (2012). *Taming digital traces for informal learning: a semantic-driven approach 21st Century Learning for 21st Century Skills* (pp. 348–362). Springer. doi:10.1007/978-3-642-33263-0_27

Thakker, D., Dimitrova, V., Cohn, A. G., & Valdes, J. (2015). *PADTUN-Using Semantic Technologies in Tunnel Diagnosis and Maintenance Domain The Semantic Web. Latest Advances and New Domains* (pp. 683–698). Springer. doi:10.1007/978-3-319-18818-8_42

Thakker, D., Dimitrova, V., Lau, L., Yang-Turner, F., & Despotakis, D. (2013). *Assisting user browsing over linked data: requirements elicitation with a user study Web Engineering* (pp. 376–383). Springer.

Thakker, D., Osman, T., Gohil, S., & Lakin, P. (2010). *A pragmatic approach to semantic repositories benchmarking The Semantic Web: Research and Applications* (pp. 379–393). Springer. doi:10.1007/978-3-642-13486-9_26

Uren, V., Sabou, M., Motta, E., Fernandez, M., Lopez, V., & Lei, Y. (2010). Reflections on five years of evaluating semantic search systems. *International Journal of Metadata*. *Semantics and Ontologies*, *5*(2), 87–98. doi:10.1504/IJMSO.2010.033280

Waitelonis, J., Knuth, M., Wolf, L., Hercher, J., & Sack, H. (2010). *The path is the destination–enabling a new search paradigm with linked data.* Paper presented at the Proc. of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly.

White, R. W., Muresan, G., & Marchionini, G. (2006). *Report on ACM SIGIR 2006 workshop on evaluating exploratory search systems.* Paper presented at the ACM SIGIR Forum.

White, R. W., & Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, *1*(1), 1–98. doi:10.2200/S00174ED1V01Y200901ICR003

Wildemuth, B. M., & Freund, L. (2012). *Assigning search tasks designed to elicit exploratory search behaviors.* Paper presented at the Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval. doi:10.1145/2391224.2391228

ENDNOTES

[1] http://stats.lod2.eu/
[2] http://scim.brad.ac.uk/~dthakker/peruse/

Call for Articles

International Journal of Distributed Systems and Technologies

Volume 7 • Issue 1 • January-March 2016 • ISSN: 1947-3532 • eISSN: 1947-3540
An official publication of the Information Resources Management Association

MISSION

The mission of the **International Journal of Distributed Systems and Technologies (IJDST)** is to be a timely publication of original and scholarly research contributions, publishing papers in all aspects of the traditional and emerging areas of applied distributed systems and integration research (including data, agent, and mining technologies) such as next generation technologies, next generation grid, and distributed and throughput computing concepts, as well as related theories, applications, and integrations at various levels. Targeting researchers, practitioners, students, academicians, and industry professionals, IJDST provides an international forum for the dissemination of state-of-the-art theories, practices, and empirical research in distributed systems and prompts future community development as a means of promoting and sustaining a network of excellence.

COVERAGE AND MAJOR TOPICS

The topics of interest in this journal include, but are not limited to:

Distributed concepts, systems, applications, technologies, and advanced paradigms for early warning systems, evacuation processes, climate changes, disaster management, and threat detection including geographical information systems and health informatics • Distributed strategies and practices for e-society, global economies, developing countries, and education • Distributed systems, system architecture and design, technologies and applications, and their integration methods and tools for performance analysis, verification, testing, and benchmarking and their empirical results • Distributed technologies for science or business process, models, and application integration and management such as in operational research, economic models, and supply chain • Distributed technology applications and system integration solutions in diverse collaborative environments including but not limited to science, engineering, management, and business • Forensics, security, cryptography, threat detection, business continuity, trust, identity management, policies, and quality of service management • Innovative developments in distributed computer, operating systems, databases, middleware, networking, and application architecture including design and analysis of distributed algorithms, programming languages, compilers, software tools and middleware environments, autonomy services, IoT, smart environments, cloud computing, service oriented, grid, peer-to-peer, throughput and cluster programming interfaces, scheduling and fault-tolerance service and algorithms, and synchronization and concurrency theory and programming • Matching, mapping, and other novel techniques for the integration, management, and interoperability of distributed data, systems, technologies, services, architectures, applications, and legacy systems • Novel algorithms, uses, and implications of distributed concepts, models, architectures, technologies, and practices for example in Internet, wireless communications, mobile, ad-hoc networks, and sensors • Representation techniques like semantics, meta-data, tagging, clouds, ontologies, and knowledge bases in distributed environments • Theories and applications of distributed computing such as algorithms and services, and their implementation in Internet, pervasive and utility computing, IoT, smart environments, cloud computing, peer-to-peer, grid, next generation grid, and next generation technologies • Theories, applications, and technologies such as click stream, data mining, databases, data warehouses, Web houses, Web data centers and sites, mashups, intra-/inter-tagging, IoT, smart environments, cloud computing for managing and integrating linked, streaming, media rich, multimedia, spatial, temporal data, services, and resources • Theory and use of distributed technologies for collaborative work, resource sharing, and problem solving • Theory of distributed systems and technologies such as parallel or integration algorithms • User and community led systems and application development, software engineering modeling, architecture, description, deployment, packaging, interfaces, and distribution • Other implications, including scientific or business models, and human and/or market dynamics influencing decisions and transitions within small and large virtual communities or organizations

ALL INQUIRIES REGARDING IJDST SHOULD BE DIRECTED TO THE ATTENTION OF:

Nik Bessis, Editor-in-Chief • IJDST@igi-global.com

ALL MANUSCRIPT SUBMISSIONS TO IJDST SHOULD BE SENT THROUGH THE ONLINE SUBMISSION SYSTEM:

http://www.igi-global.com/authorseditors/titlesubmission/newproject.aspx

IDEAS FOR SPECIAL THEME ISSUES MAY BE SUBMITTED TO THE EDITOR(S)-IN-CHIEF

PLEASE RECOMMEND THIS PUBLICATION TO YOUR LIBRARIAN

For a convenient easy-to-use library recommendation form, please visit:
http://www.igi-global.com/IJDST