Proceedings Paper:
Yang-Turner, F, Lau, LMS, Dimitrova, V et al. (1 more author) (2013) Profiling exploratory browsing behaviour with a semantic data browser. In: AMCIS 2013 Proceedings. Americas Conference on Information Systems (AMCIS), 15-17 Aug 2013, Chicago, USA. AIS .

Profiling Exploratory Browsing Behaviour with a Semantic Data Browser

Fan Yang-Turner, Lydia Lau, Vania Dimitrova, Dhavalkumar Thakker

University of Leeds, Leeds, LS2 9JT, United Kingdom.
{f.yang-turner, l.m.s.lau, v.g.dimitrova, d.thakker}@leeds.ac.uk

ABSTRACT

Semantic Web technologies are increasingly being adopted for aggregating Web data. Tools such as Semantic Data Browsers have been proposed to assist users to access and make sense of the vast semantic space. However, further investigations are needed to understand how users make use of the additional semantic features provided by these new breed of browsers and their effectiveness in supporting exploration of a domain. Measurements of browsing behaviour in a semantic space are also needed. Using the log data from a semantic browser (MusicPinta) for the music domain, this paper takes the first step in profiling browsing behaviour of users in a semantic space and compares the outcome against their task performance. Two exploratory search tasks were designed for the experiment. Movements in terms of users traversing the provided semantic links in the browser were captured and the patterns of clicks between abstract and concrete concepts were analysed.

Keywords

Exploratory search, semantic data exploration, semantic data browser, exploratory browsing, browsing behaviour.

INTRODUCTION

Semantic Web technologies are increasingly being adopted for aggregating Web data. Tools such as Semantic Data Browsers (SDB) have been proposed to assist users to access and make sense of the vast semantic space, such as Tabulator (Berners-lee et al., 2006), Parallax (Huynh and Karger, 2009), VisiNav (Harth, 2009), FacetGraphs (Heim et al., 2010) and I-CAW (Thakker et al., 2012). Such browsers operate on semantically augmented data (e.g. tagged content) and lay out browsing trajectories using relationships from the underpinning ontologies. Understanding the value added by semantics in SDB remains challenging, as the interface of these browsers hides the complexity of the data infrastructure from their users. Capturing users browsing behaviour and comparing them with task performance may provide insight. This insight can assist us further exploit the semantics in the search space apart from the known benefit of being able to merge Web data from disparate sources. Hence, the first research question posed in this paper is

> "Can we model exploratory browsing behaviour in terms of semantic features of hyperlinks?"

This paper investigates the user behaviour of exploratory browsing. As classified by Bawden (1986), exploratory browsing is deliberately searching for inspiration. Since the Web is increasingly being used as a source for inspiration, investigation into possible browsing behaviour associated with effective outcome would be interesting. Hence, the follow-on research question posted in this paper is

> "Can we relate the possible browsing behaviour in a SDB with performance in exploratory tasks?"

A methodology is proposed for using log data to profile browsing behaviour. MusicPinta, a SDB which accesses Linked Open Data in the music domain, has been developed to capture user browsing behaviour by monitoring the clicks and their context.

In the following sections, we start with reviewing related work and introduce MusicPinta that is used in our study. We then propose the methodology for profiling browsing behaviour. Following the methodology, we present the user study and the analysis of results. Finally, we conclude with discussions and future work.

RELATED WORK

The interaction behaviours observed during an exploratory search are generally a combination of exploratory browsing and focused searching, with more emphasis on the former. People use browsing as a way to resolve the uncertainty and confusion that can occur as new information is encountered. As an important part of exploratory search, exploratory browsing enables users navigate through (and to) information that helps them develop powerful cognitive capabilities and leverage their newly acquired skills to address open-ended, persistent, and multifaceted problems. This behaviour of zooming in and out across spaces is attributed to the exploratory browsing behaviour in comparison to an iterative search strategy (White and Roth, 2009 in Figure 1). Within exploration, there may be some degree of progressive narrowing of search space as part of the exploration-enrichment-exploration trade-off (Patterson et al., 2001). Under this model, a search begins with the retrieval of a broad set of documents, such as one retrieved by a high-recall/low-precision query, then proceeds with narrowing that down into progressively smaller but higher-precision result sets, before reading the documents and extracting the information (for example, iteration 4, 5 and 6 in Figure 1).

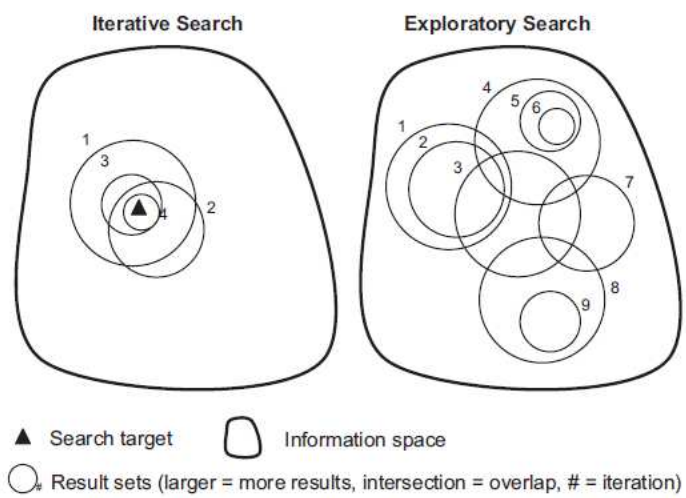

Figure 1: Iterative search versus exploratory search strategies (White and Roth, 2009, p. 21)

Motivated by this illustration, our work investigates how different users traverse the information space between abstract and concrete concepts. This is made possible by the nature of a SDB as the hyperlinks encapsulate the relationships between related concepts. Interaction logs are one of the resources for understanding and characterising user behaviour.

For web search, there has been a growing interest in using interaction logs for research (Facca and Lanzi, 2005; Murray and Teevan, 2007; Craswell et al., 2009). The study of users' interactions with the web browser enabled researchers to understand user behaviour according to the timing, query terms and URLs. A method was proposed to automatically discover user interests based on browsing behaviour (Velayathan and Yamada, 2007). This kind of research has the advantage of its large scale study but none has been conducted on SDBs.

Smaller scale studies on different types of SDB and associated evaluation have emerged (Şah et al., 2008; Wilson and schraefel, 2008). As highlighted by Uren et al. (2010), evaluation of semantic search systems had been sporadic and ad hoc. As these semantic tools were maturing, they called for a community effort in providing benchmarking mechanisms, in particular, the use of logging parameters for evaluating individual components of these search systems. Hoxha et al. (2012) recently proposed the addition of semantics into log data for analysing behaviour. Our work echoes this vision of using semantics to deepen our understanding of browsing behaviour from the log data.

MUSICPINTA: A SEMANTIC DATA BROWSER

Semantic data browser combines state of the art semantic web technologies for semantic augmentation, semantic query and representation. The semantic features not only allow users to easily tap into resources built from the Linked Open Data but also enable researchers to conduct semantic analysis on user interactions.

Interface and dataset

In our study, we used a SDB called MusicPinta. It provided an interface for browsing domain of music instruments through music ontology-driven datasets: DBpedia[1], MusicBrainz[2], Jamendo[3] and Megatunes[4]. In total, it had 2.4M entities and 19M triple statements. It used 2GB physical space, including 876 musical instruments entities, 71k entries for performances (albums, records and tracks) and 188k music artists. Amazon reviews were added as an example of social content in MusicPinta. Users of MusicPinta could explore the information space through moving from one entity to another entity. An entity could be a concrete concept (e.g. "bouzouki") or an abstract category concept (e.g. "string instrument") in the underlying ontology. As "bouzouki" is a kind of "string instrument", MusicPinta had the ability to link these two entities together and presented to the users.

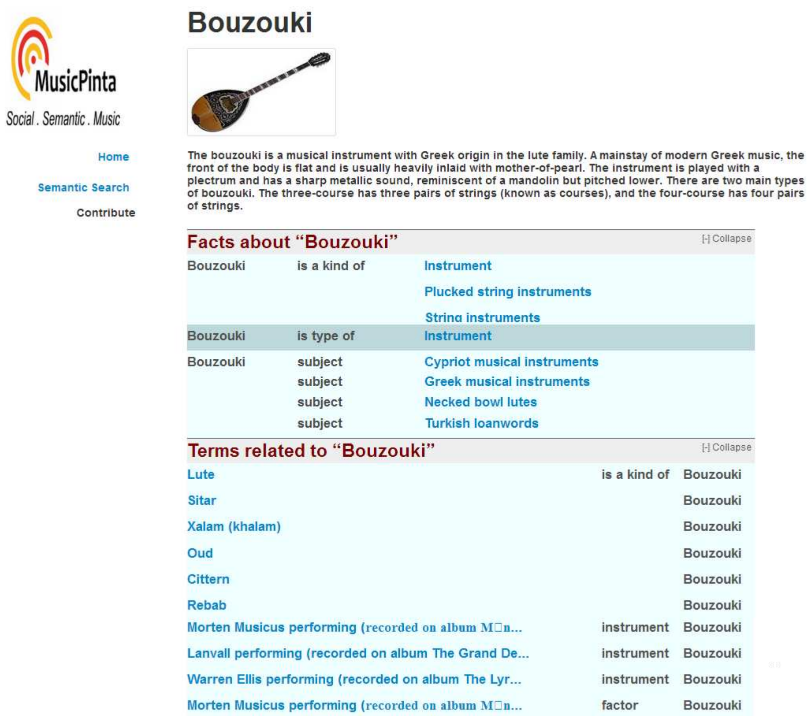

Figure 2: An example page of MusicPinta

A typical interface in MusicPinta is presented in Figure 2. Each 'entity page' started with the name of the entity and a description (including a picture if available). Three main facets were extracted from the aggregated datasets (semantic repositories): (i) facts about the focus entity; (ii) terms related to the focus entity; and (iii) content related to the focus entity (cropped in the figure). Facts and related terms for the focus entity consisted of triples from the semantic repositories, which included hierarchy links (denoted as *is a kind of*), membership (denoted as *is a type* of) and object properties (denoted as *other*). Hyperlinks pointed to further details for the retrieved objects.

User interaction with MusicPinta

It was expected that a user queried MusicPinta for a particular music topic, for example, an instrument. The user interaction with MusicPinta is illustrated in Figure 3. With a list of query result provided, user could select an entity page to browse. Sometimes, a page might include content, which were extracted from Amazon reviews. Solid lines in the figure show the basic interaction leading to a result. Dashed lines show the possible browsing path.

---

[1] http://dbpedia.org/About

[2] http://musicbrainz.org/

[3] http://www.jamendo.com/en/

[4] http://magnatune.com/

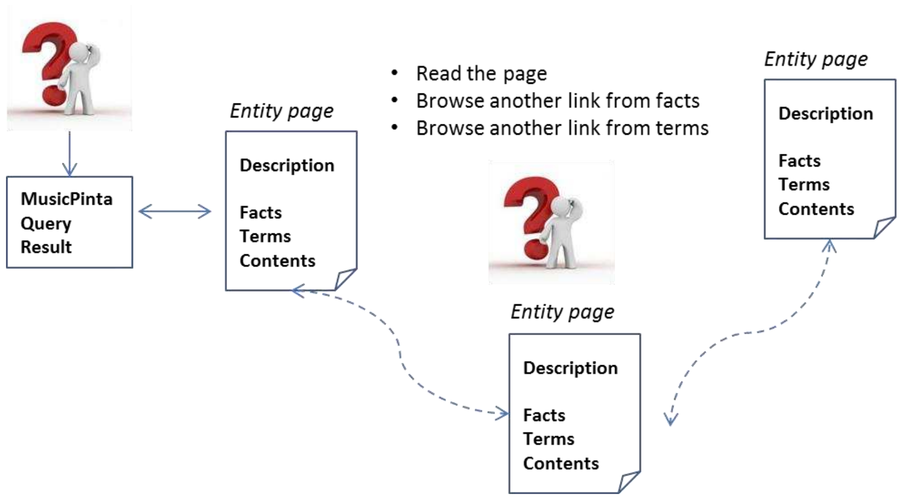

Figure 3: User Interaction with MusicPinta

METHODOLOGY FOR PROFILING BROWSING BEHAVIOUR

Our approach for profiling exploratory browsing behaviour with a SDB is shown in Table 1, with MusicPinta case study as an example. There are three stages: preparation, analysis and profiling.

Table 1: Methodology of profiling browsing behaviour with semantic data browser

General Approach		MusicPinta Case Study
Preparation	Extend semantic data browser with logging features	MusicPinta was programmed with the ability to capture user interaction with user id, timestamp and the entity that a user clicked.
	Encode log data with semantic annotations	Hyperlinks with different levels of "abstract concept" or "concrete concept" were encoded for the study.
	Choose one or more human factors to profile	User performance for each task was chosen to be profiled.
Analysis	Count the total number of links user clicked	Total number of links of each user was counted.
	Examine the number of links based on semantic coding	The number of links encoded "abstract concept" and "concrete concept" of each user was examined.
Profiling	Profile browsing behaviour in terms of human factors and semantic features	The relationships amongst number of clicks, links of abstract/concrete concepts and task performance were examined.

This proposed approach provides a method for profiling browsing behaviour with interaction log data and externalising the behaviour in an explicit form. With the profile in an explicit form, we can conduct comparison of user behaviour across different versions of SDB which may provide useful insight while experimenting new ways of using semantics.

USER STUDY

Participants

Twelve participants (seven male and five female) were recruited on a voluntary basis (Table 2). Half of the participants were native speakers and the other half spoke and communicated in English fluently. They were from different age groups and educational background. A screening questionnaire was used to check the participants' knowledge of music and interests with musical instruments. Ten participants played one or more instruments but only one participant practised weekly. Nine participants listened to online music sites daily, weekly or monthly. Half of the participants visited sites with music information weekly or monthly. All of them claimed to be unfamiliar with the instruments used in the study, confirming that they were conducting an exploratory search task.

Table 2: Information of Participants

Participants ID	Sex	Native Speaker	Age Group	Education	Use of music web site	Play Instruments	Currently practicing	Listen/download online music
p01	Male	no	40-50	Graduate	occasionally	no	no	occasionally
p02	Male	no	20-30	Postgraduate	occasionally	yes	occasionally	weekly
p03	Female	yes	20-30	PhD	weekly	yes	no	monthly
p04	Male	no	20-30	Postgraduate	weekly	yes	weekly	daily
p05	Female	yes	>50	High School	occasionally	yes	occasionally	monthly
p06	Female	yes	>50	High School	monthly	yes	occasionally	occasionally
p07	Male	yes	20-30	PhD	occasionally	yes	no	daily
p08	Male	no	30-40	PhD	occasionally	yes	occasionally	monthly
p09	Male	yes	20-30	Postgraduate	never	yes	no	occasionally
p10	Male	no	20-30	Postgraduate	weekly	yes	no	daily
p11	Female	yes	20-30	Undergraduate	weekly	yes	no	daily
p12	Female	no	20-30	Graduate	monthly	no	no	monthly

Tasks

Two representative exploratory search tasks were designed for the study. Task1 was analytical which requires exploring, comparing, finding similarities and differences within a pool of knowledge items. Task2 was more creative with open-ended outcomes. An advertising scenario for a hypothetical music shop was used as the context. Participants were asked to explore the information on musical instruments within MusicPinta. In both tasks, the participants were given an entry point to the browser and a form to fill in their answers. The details of the tasks are shown in Table 3.

Table 3: Two tasks in MusicPinta user study

Task1	Task 2
The music shop is extending its collection of instruments with international musical instruments. You work in an advertising agency which has been asked to prepare an advertisement script for some of the new instruments that will appear in the shop. A key part of the preparation of the advertisement script is the research of the product. You have been asked to conduct a research on one of the new instruments, called bouzouki, using the information available in MusicPinta. You have to identify: • The main characteristics of bouzouki; • Up to five similar instruments to bouzouki; • Features that make bouzouki distinctive from the similar ones you have chosen. Go to 'Semantic Search' in MusicPinta and type bouzouki. Browse the content and follow links. Complete the provided form.	The music shop wants to increase the sales of its traditional musical instruments, such as electric guitars. It intends to do this by adding links to album recordings with electric guitars which are available in creative commons, together with some interesting information about these albums to inspire customers to play/buy electric guitars or any other musical instruments. Furthermore, when displaying its electric guitar items, the shop wants to highlight key features people look for when purchasing electric guitars. You are asked is to conduct a research to address the above requirements by using information provided in MusicPinta. You have to review the information about electric guitar and identify: • Three interesting album recordings that include electric guitars and specify what is interesting; • Key features that people look for when purchasing an electric guitar. Go to 'Semantic Search' in MusicPinta and type electric guitar. Browse the content and follow links. Complete the provided form.

Procedure

Each participant attended an individual session for about an hour, conducted and observed by an experimenter. The structure of a session is shown below:

- Pre-study questionnaire [5 min] - collect information about the user profile and test his/her domain awareness;
- Introduction to MusicPinta [10 min] – introduce main features of MusicPinta using *tenor saxophone a*s an example;
- Task1 [20 min] - identify distinctive characteristics of the musical instrument *bouzouki*;
- Task2 [20 min] - identify usage and features of the musical instrument *electric guitar*;
- Post-study questionnaire [10 min] – test again the participant's domain awareness and gather usability feedback;
- Brief interview [5 min] – collect general feedback on using MusicPinta.

Data collection and performance assessment

The data collected in the study included: (i) the answers from the participants for Task1 and Task2; (ii) the pre- and post-experiment questionnaires and word association tests; (iii) system log data; and (iv) experimenter's notes. For this paper, we report on the findings from the answers for the tasks and system log data.

User performance

With the answers from the participants, the success rates of the two tasks were assessed by domain experts. A model answer based on the ontology and the content of MusicPinta was produced by the experts. The task performance of a user was rated (in percentage) by the experts according to how much overlap there was between the participant's answer and the model answer.

Log data

MusicPinta recorded every user click with following information: user id, timestamp and ontology entity of the link. The ontology entity of a link consisted of the source of dataset (e.g. *dbpedia* or *dbtune)* and the entity (e.g. *ukulele* or *plucked string instrument)* as shown in Table 4.

Table 4: Example of log entries of MusicPinta

User Id	Ontology entity of the link	Timestamp
p01	*dbpedia:ukulele*	*2012-12-08 22:03:24*
p01	*dbtune_instrument:plucked_string_instruments*	*2012-12-08 22:03:30*
p01	*dbtune_instrument:balalaika*	*2012-12-08 22:04:18*
p01	*dbtune_instrument:plucked_string_instruments*	*2012-12-08 22:05:04*

RESULTS

In this section, we report the outcome of the three stages of our approach in the study with MusicPinta.

Preparation

Encode log data with semantic annotations

Table 5: Semantic coding of entities of MusicPinta

Concepts	Abstraction level	Code	Examples
Abstract concepts	Classification – upper level	L5	Entity as a upper level concept e.g. instrument, performance, artists
	Classification – middle level	L4	Entity as a middle level concept e.g. string instrument, drums
	Classification – lower level	L3	Entity as a lower level concept e.g. plucked string instruments, hand drums
Concrete concepts	Concrete concepts	L2	Entity as a concrete item e.g. violin, electric guitar

To understand the semantic relationship among the links that users browsed, we coded ontology entity into four levels based on the abstraction levels in the ontology. This coding is defined in Table 5. With this coding, a snippet of log data on one participant (p01) after pre-processing is shown in Table 6.

Table 6: Example of log data with semantic coding for participant p01

User Id	*Ontology entity of the link*	*Abstraction Level*	*Timestamp*
p01	*dbpedia:Ukulele*	*L2*	*2012-12-08 22:03:24*
p01	*dbtune_instrument:plucked_string_instruments*	*L3*	*2012-12-08 22:03:30*
p01	*dbtune_instrument:balalaika*	*L2*	*2012-12-08 22:04:18*
p01	*dbtune_instrument:plucked_string_instruments*	*L3*	*2012-12-08 22:05:04*

Choose one or more human factors to profile

One of the important factors of evaluating an exploratory search system was task success (White and Roth, 2009). As the first step, we chose "user performance" as a human factor for behaviour profiling. We aimed to profile users' browsing behaviour and compare that against their task performance. The summary of user performance (Table 7) is ordered by average performance of task1 and task2 with top and bottom performers identified. As our study was to profile browsing behaviour against task performance, the extreme cases in terms of task success rate can allow us to separate clearly "successful behaviour" from clearly "unsuccessful behaviour".

Table 7: Top and bottom performers in the experiment study ordered by average performance of Task1 and Task2

Participant	*Task1*	*Task2*	*Average of Task 1& Task 2*	*Top and bottom Performers*
P01	***81%***	***66%***	*73.5%*	***3rd highest performer in Task1; 3rd highest performer in Task2.***
P05	***97%***	*50%*	*73.5%*	***Highest performer in Task1***
P04	*71%*	***75%***	*73.0%*	***Highest performer in Task2***
P07	*67%*	*67%*	*67.0%*	
p10	*82%*	*50%*	*66.0%*	
P06	*73%*	*58%*	*65.5%*	
P03	*74%*	*50%*	*62.0%*	
p12	*69%*	*50%*	*59.5%*	
p11	*71%*	*41%*	*56.0%*	
P02	***53%***	***42%***	*47.5%*	***2nd lowest performer in Task1; 4th lowest performer in Task2.***
P08	*64%*	***8%***	*36.0%*	***Lowest performer in Task2***
P09	***44%***	*17%*	*30.5%*	***Lowest performer in Task1***

Analysis

Count the total number of links user clicked

Figure 4 shows the number of links browsed by each user in Task1 and Task2, sorted from the top clickers (who browsed the most number of links) to the bottom clickers (who browsed the least number of links). The top and bottom performers are marked in the figure.

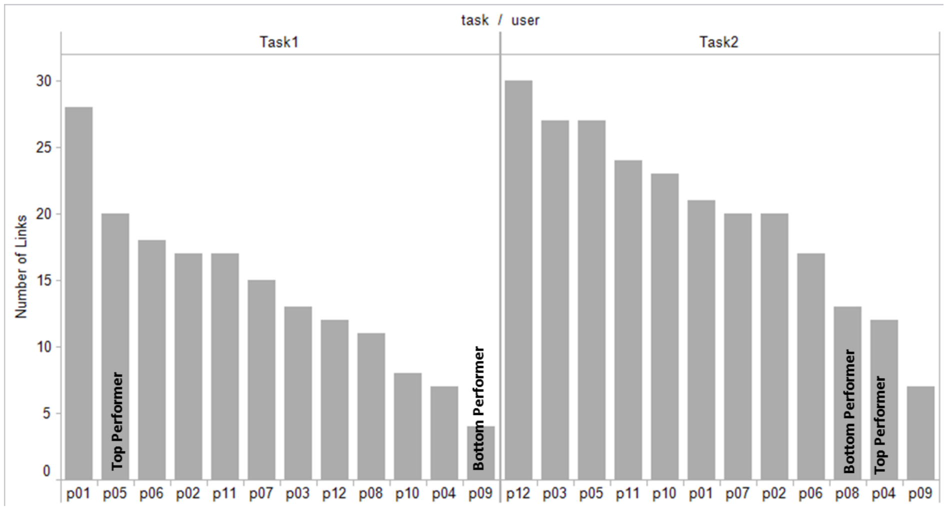

Figure 4: Number of links browsed by users in two tasks

Examine the total number of links based on semantic coding

Based on the semantic coding, two types of links are classified - abstract (L3, L4 and L5) and concrete (L2). To compare the browsing behaviour against users' overall performance (which is the average of Task1 and Task2), numbers of abstract and concrete concept links browsed by users in the two tasks are visualized in Figure 5. The user list is ordered by the total number of links browsed by users. On the left and middle part of Figure 5, the abstract concept links and concrete concept links are presented separately; while on the right the abstract concept links and concrete concept links are presented together to show the proportion of the two.

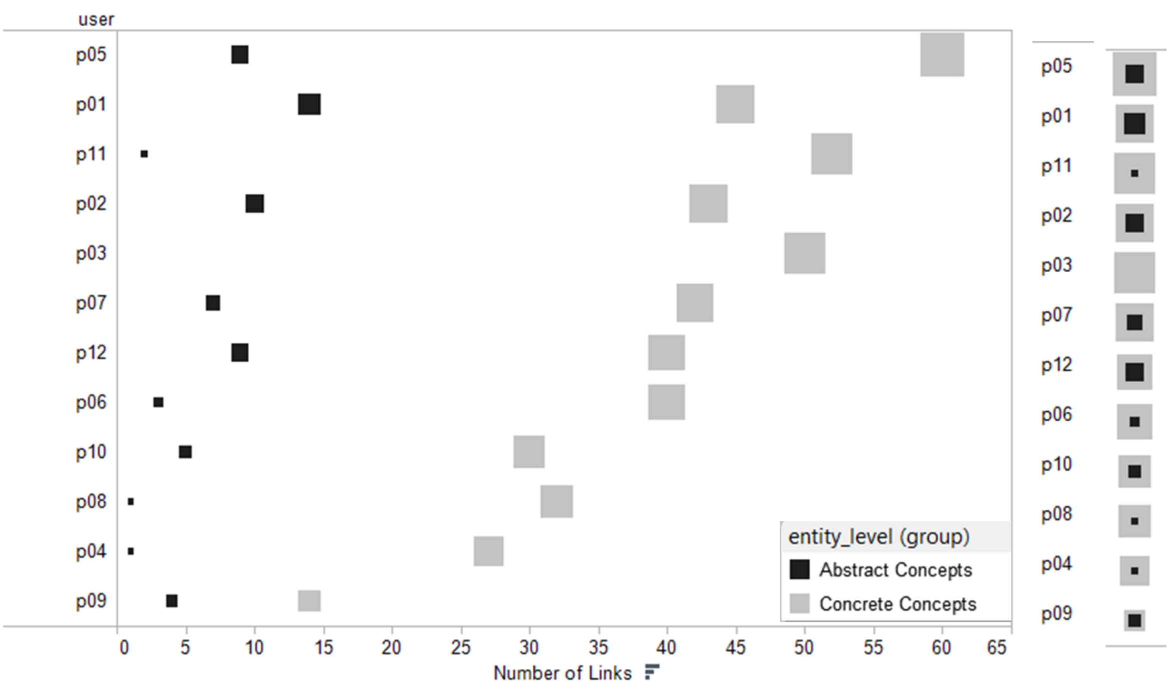

Figure 5: Number of abstract and concrete concept links browsed by users in the two tasks

Profiling

Browsing behaviour (number of links and the proportion of the concrete and abstract links) with overall task performance is visualized in Figure 6. With this figure, we are able to examine if there is any correlation between performance, the number of clicks and the level of abstract/concrete browsing. The participants are presented with different sizes of square, showing the number of clicked links. The squares are filled with two colours: black for the number of clicks on abstract concepts and grey for the number of clicks on concrete concepts. As seen in the figure, there is no obvious pattern detected amongst the top performers (p01, p04, p05) and amongst the bottom performers (p02, p08, p09). However, we can profile the users with an explicit form shown in Table 8. Although there is no strong correlation between the profile and the performance, this explicit form has established a way to externalise the user behaviour according to the semantics.

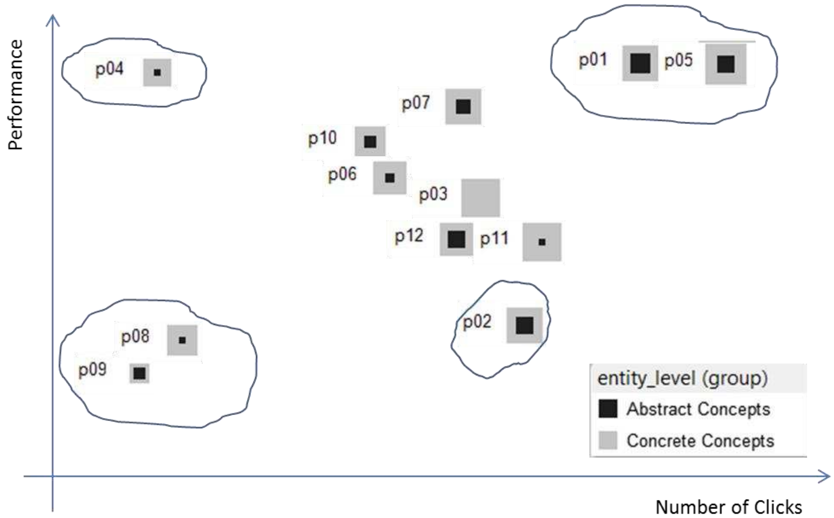

Figure 6: Browsing behaviour with overall task performance

Table 8: Profile of browsing behaviour of MusicPinta

Type of clickers	Description	Examples of top performers	Examples of bottom performers
Top clickers	Ranking top 3 in 2 tasks based on the numbers of clicks	p01 and p05	
Bottom clickers	Ranking bottom 3 in 2 tasks based on the numbers of clicks	p04	p09
Concrete clickers	About 90% clicks were concrete concepts	p04 and p05	p08
Mix clickers	About 25% clicks were abstract concepts and 75% clicks were concrete concepts.	p01	p02

DISCUSSION AND FUTURE WORK

In this research-in-progress paper, we examined users' interaction with semantic links and their task performance in order to understand users' behaviour with semantic data browsers. We reached an answer to our first research question: "can we model exploratory browsing behaviour in terms of semantic features of hyperlinks?" in two steps. Firstly, we proposed a methodology to profile browsing behaviour with log data and user performance; and secondly, conducted a study with a Semantic Data Browser (SDB), MusicPinta. In addition to the number of clicks as a measure of browsing behaviour, the semantic links in the log were annotated to gauge the level of zooming in and out of abstraction – a feature of ontological structure. The analysis was conducted by visualising the proportion of abstract and concrete level browsing. There is certainly potential to exploit further the semantic relationships of the hyperlinks for deeper insight. In future, more detailed semantic coding can be added, for example the type of content such as "descriptions" or "images" at the concrete level browsing.

The answer to our second research question "Can we relate the possible browsing behaviour in a SDB with performance in exploratory tasks?" was mixed. On one hand, the results clearly showed the power of visualising browsing behaviour against task performance. On the other hand, no conclusive patterns were found from this study which indicated that either a larger sample of users would be needed or the parameters chosen for analysis might have no impact on task performance. For the latter, new tasks with new parameters could be designed for a repeat study, for example, more human factors (such as user learning style, knowledge and skills, curiosity index etc.) can be selected for the profiling.

REFERENCES

1. Bawden, D. (1986) Information Systems and the Stimulation of Creativity. J. of Information Science, vol. 12, pp. 203-216.
2. Berners-lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D. (2006) Tabulator: Exploring and analyzing linked data on the semantic web. In Procedings of Semantic Web User Interaction Workshop SWUI06.
3. Craswell, N., Jones, R., Dupret, G. and Viegas, E. (eds). (2009) Proceedings of the 2009 Workshop on Web Search Click Data. ACM, Barcelona, Spain, p. 95.
4. Facca, F. M. and Lanzi, P. L. (2005) Mining interesting knowledge from weblogs: a survey. Data Knowledge Eng., 53, pp. 225–241.
5. Harth, A. (2009) VisiNav: Visual Web Data Search and Navigation. In Database and Expert Systems Applications, pp. 214-228. Springer.
6. Heim, P., Ertl, T., Ziegler, J. (2010) Facet graphs: complex semantic querying made easy. In Proceedings of the 7th international conference on The Semantic Web: research and Applications – ESWC2010, Berlin, Heidelberg, pp. 288-302.
7. Hoxha, J., Junghans, M. and Agarwal, S. (2012) "Enabling Semantic Analysis of User Browsing Patterns in the Web of Data." In 2nd International Workshop on Usage Analysis and the Web of Data (USEWOD2012) in the 21st International World Wide Web Conference (WWW2012), Lyon, France, April 17th, 2012.
8. Huynh, D. F., Karger, D. R. (2009) Parallax and companion: Set-based browsing for the data web. In Proceedings of WWW 2009 Conference. ACM.
9. Murray, G. C. and Teevan, J. (2007) Query log analysis: social and technological challenges (WWW 2007 Workshop Report). ACM SIGIR Forum, 41, pp. 112–120.
10. Patterson, E. S., Roth, E. M., and Woods, D. D. (2001) Predicting vulnerabilities in computer supported inferential analysis under data overload. Cognition Technology and Work, 3(4), pp. 224–237.
11. Şah, M., Hall, W. and de Roure, D. C. (2008) SemWeB: A Semantic Web Browser for Supporting the Browsing of Users using Semantic and Adaptive Links. Proceedings of the 5th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer-Verlag, pp. 431 – 436.
12. Thakker, D., Despotakis, D., Dimitrova, V., Lau, L., Brna, P. (2012) Taming digital traces for informal learning: A semantic-driven approach. In Proceedings of EC-TEL 2012.
13. Uren, V., Sabou, M., Motta, E.,Fernandez, M., Lopez, V. and Lei, Y. (2010) 'Reflections on five years of evaluating semantic search systems', Int. J. Metadata, Semantics and Ontologies, Vol. 5, No. 2, pp. 87–98.
14. Velayathan, G. and Yamada, S. (2007) Investigating User Browsing Behavior. Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 195-198.
15. White, R.W. and Roth, R.A. (2009) Exploratory search: beyond the query response paradigm, Synthesis Lectures on Information Concepts, Retrieval, and Services, Morgan&Claypool Publishers.
16. Wilson, Max L. and schraefel, m.c. (2008) Improving Exploratory Search Interfaces: Adding Value or Information Overload? In, Second Workshop on Human-Computer Interaction and Information Retrieval, Redmond, WA, USA, pp. 81-84.