

VSDR: Visualization of Semantic Data Representation for Information Search over Semantic Web

N. Kanjanakuha, C. Techawut, R. Sukhahuta, and P. Janecek

Abstract—Many semantic searching tools for encyclopedia searching are presently available on the Internet. Most of them display results based on some textual representations. Some visual representations are rarely found by having visual interfaces mirroring link data structure from any sources. However, they hardly interact with a large number of searching results on a screen because of some complicated filtering tools and results unrelated to user requirements. This paper presents Visualization of Semantic Data Representation (VSDR) as a framework for displaying a 3-layer approach based on the hyperbolic tree model. It helps users to mainly focus on a linked data structure that hides some complex query scenarios from users occurring while searching through a semantic web. It can also leverage some difficulties of user's exploration, perception, and interpretation with direct manipulation. The prototype of VSDR is constructed and evaluated by means of user experiment in order to assess overall usability of the proposed 3-layer approach based on the hyperbolic tree model, and the relevance between results and user requirements.

Index Terms—Knowledge exploration, visual exploration, human-computer interaction, semantic search.

I. INTRODUCTION

Nowadays, most Encyclopedia information is accessed via semantic searching tools that have efficient algorithms to match the input keyword with conceptual structures from any sources. Users can see results in a textual representation with many hyperlinks to the target information [1], [2]. It causes some confusion, as it is easy to get lost, while users are navigating through hyperlinks in a semantic web. Semantic web searching tools are used to extend normal search to user exploration. Semantic data exploration can expand the user's view to see a set of related information. Users can see some visual representations as visual interfaces mirroring linked data structure from any sources. Users can explore, perceive, and interpret visual information. However, reaching the target information that users desire is still a concerning problem. The target information is not only a pool of keywords, texts, images, and audios, but it is also classified and linked in a structured form represented as concept, property, and relationship. Therefore, a simple semantic searching tool for user exploration and visual representation of a large set of related information with usability are required.

Manuscript received July 15, 2017; revised October 11, 2017. The Graduate School, Chiang Mai University, funds this research.

N. Kanjanakuha, C. Techawut, R. Sukhahuta are with the Department of Computer Science, Chiang Mai University, Chiang Mai, 50200 Thailand (e-mail: natanun.ka@cmu.ac.th, churee.t@cmu.ac.th, rattasit.s@cmu.ac.th).

P. Janecek is with the Think Blue Data Co., Ltd., Nonthaburi, 11120, Thailand (e-mail: paul.janecek@gmail.com).

Most researches have been conducted with the part of processing approach at the back-end to process the linked data structure, but fewer researches have been conducted with the part of representation approach at the front-end to visualize the linked data exploration. The survey of currently active semantic searching tools with visualizations (Adaptive VIBE or TaskSiev [3], Aemoo [4], Wivi [5], IntrospectiveViews [6], LinkedMDB [7], PrVIEW [8]) shows 2 basic types of interaction styles to support users in finding their target information on a visual representation as follows.

- Providing some filtering tools to narrow down some alternatives to target information.
- Providing a linked data structure as a visual structure based on SPARQL command and some alternative ranking.

Other difficulties are summarized as follows.

- Losing screen space with some filtering tools: Some filtering tools can obstruct user explorations, and do not help users generate incremental searching activity for reaching the target information.
- Disorganizing results: A large number of results displayed on a screen can make users confused. It is hard to know a relationship between alternative results and input keywords. Users may waste time judging the relevance of the data by looking for all keywords in a linked data structure.
- Continuing exploration without a solution: Navigating through a linked data structure is easy to get lost. Unfortunately, users may not find their target information.

Therefore, it is a challenge to improve a visual representation of a semantic searching tool. The contribution of this paper is the Visualization of Semantic Data Representation (VSDR), a framework for displaying a 3-layer approach based on the hyperbolic tree model, including:

- A visual representation for supporting user exploration, providing incremental searching activity and hiding a sophisticated procedure of a filtering tool.
- A visual representation with a decent organization of a linked data structure to provide cognition.
- A visual representation that displays a set of results related to user requirements.

With the VSDR framework, users can have some good sequences or dialogs through their exploratory search on the hyperbolic tree model. Users can explore more of the input keywords on a well organized link data structure, showing a graph of some focused or distorted nodes. In addition, hiding some complicated filtering tools with users' direct manipulation on a graph can automatically bring users to their target information.

This paper firstly presents the 3-layer approach based on the hyperbolic tree model, a concept for a visual representation of a semantic searching tool. Secondly, an overview of the VSDR framework is presented to demonstrate the implementation of the 3-layer approach based on the hyperbolic tree model. Thirdly, the evaluation of the prototype constructed by VSDR framework is shown, and the results from user experiment are discussed and summarized. Finally, conclusion and future work are presented.

II. 3-LAYER APPROACH BASED ON HYPERBOLIC TREE MODEL

Hyperbolic tree model [9] is a technique for visualizing all information in one page on a limited screen size with hierarchies and well-organized nodes. Major characteristics of the hyperbolic tree, according to the contributions of this paper, are 1) Layout of graph has a focused node at the center. Some distorted and related node is placed from near to far with no overlapping of nodes. 2) Transformation of graph can change the focus by clicking on a node. The visual presentation will rotate to focus on the clicked node and distort related nodes from near to far.

Hyperbolic tree provides a sense of exploratory search [10], [11]. Learning and investigating are the most common user activities. Users spend time analyzing, comparing, and making conclusions. The 3-layer approach additionally defines a number of layers of the hyperbolic tree and some specifications with the intention to reduce some ambiguity and confusion as shown in Fig. 1. The representation has three circular layers. Each layer contains related tree nodes with different meanings.

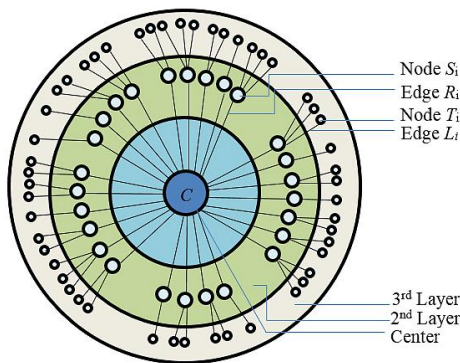


Fig. 1. The 3-layer approach based on hyperbolic tree model.

The input node C and the second layer nodes S_i are connected with the edges R_i between them, if and only if nodes S_i and node C are related. It can be described as the triple $\{C R_i S_i\}$. A node in the second layer can be linked to its full document or can be set to the center for a focus changing. The third layer displays a set of related nodes with the second layer. The third layer nodes T_i are connected to the second layer nodes S_i with the edges L_i . Each layer has a set of nodes with the broader semantic relationship from the center. Node S_i is more specific than node T_i , or in other words, node S_i is a subset of node T_i . It can be described as the triple $\{S_i L_i T_i\}$.

The center of the tree is a focused node, usually an input keyword from users. The second layer has a set of nodes semantically related to the input keyword. The third layer has a set of nodes related to the first and the second layers. From the third layer, we can see the relationship paths of nodes starting from the center to the second layer and third layer. Compared to other tree models, the three-layer approach has the advantage of displaying a large number of nodes on the same size of the display area by providing effective interaction and navigation. It can reduce users' exploration time by providing an overview of all related results in each layer and looking into details with the focus and context method. Node comparison with 3 layers is suitable. Users can see some related paths with their focused node from the center to the second and the third layers. They can easily maintain their focused nodes with alternative nodes one by one through each layer until they find the target node of information. Having more than 3 layers may cause users more confusion because users can easily get lost while they are traversing on different levels of a large tree. Whereas having only 2 layers is not ideal because the two layers require more searching interaction of the focus and context method to reach the target information compared to the 3 layers. In other words, with 2 layers, users have to conduct more time consuming tasks of node comparison compared to 3 layers.

III. VSDR FRAMEWORK

VSDR framework can be applied to the system architecture of a semantic searching tool connecting to DBpedia [12] as a web service to acquire the Encyclopedia information. Fig. 2 shows the VSDR framework consisting of 3 basic parts: input, processing, and output.

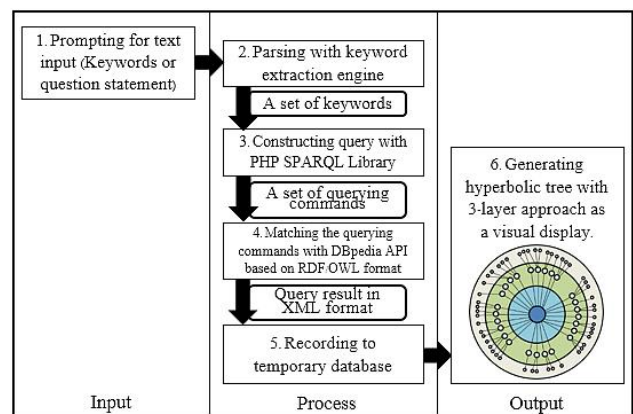


Fig. 2. VSDR framework for displaying a 3-layer approach based on hyperbolic tree model.

Input part provides a text box for input keywords (single or multiple keywords). Processing part starts with the engine making queries by using the SPARQL command to search through the RDF schema, a knowledge representation model with classes, properties and meta-description of ontology. The RDF resources are stored as a triple and can be accessed by using SPARQL command. There are many properties of RDF resources available for querying, but we select rdfs:label, which is the instance used to provide a

human-readable version of a description as a general resource's topic to define the relationship. The engine then checks the input with `rdfs:comment` or `owl:sameAs` and some other properties to confirm that the input matches with these resources, meaning that they are related. Hence, the engine collects a set of the triples $\{C R_i S_i\}$. Next, the engine selects `dct.subject` and collects a set of triples $\{S_i L_i T_i\}$. The output part displays all nodes in each layer. Every node is ordered alphabetically in a clockwise direction and has a number on them.

Our prototype of a semantic searching tool demonstrates the VSDR framework and implements JavaScript that can run on any web browser. Fig. 3 shows a screenshot of our prototype after entering keyword "History of England" in a textbox located at the bottom of the page. The 3-layer approach based on the hyperbolic tree model is displayed at the center area of the page. The center node C is "History of England". There are nodes S_i on the second layer, which relates to node C , for example, the node S_i is "A child's history of England". It can be written as (1).

$$\{C R_1 S_1\} = \{ \text{History of England, rdfs.label, A child's history of England} \} \quad (1)$$

Users can click on the link at the right side of the page if they want to read the detail information or they can click on any nodes S_i to change focus. The third layer nodes T_i contain

the result which are broader than S_i , for example, the node T_{163} is "Children's history books". It can be written as (2).

$$\{S_1 L_3 T_{163}\} = \{ \text{A child's history of England, dct.subject, "Children's historybooks"} \} \quad (2)$$

IV. EVALUATION

The goals of our evaluation are 1) to assess overall usability of the proposed 3-layer approach based on the hyperbolic tree model, and 2) to assess the relevance between results and user requirements. We choose 2 basic measurements for this evaluation that are consistent with our goals. System Usability Scale (SUS) [13] is used to assess the overall usability of the proposed 3-layer approach based on the hyperbolic tree model and a patterns-based technique [14] with a method-level pattern to capture the relevance between results and input keywords.

We constructed the user experiment with 30 subjects. Fifteen subjects were software developers and IT officers, and 15 subjects were persons using a computer everyday (such as internet searching). Subject's age ranged from 15 to 45 years old. We selected a group of subjects with computer skills and experiences with some software tools, so they could do trade-offs and give some useful comments. We were looking to gain some feedback for further development.

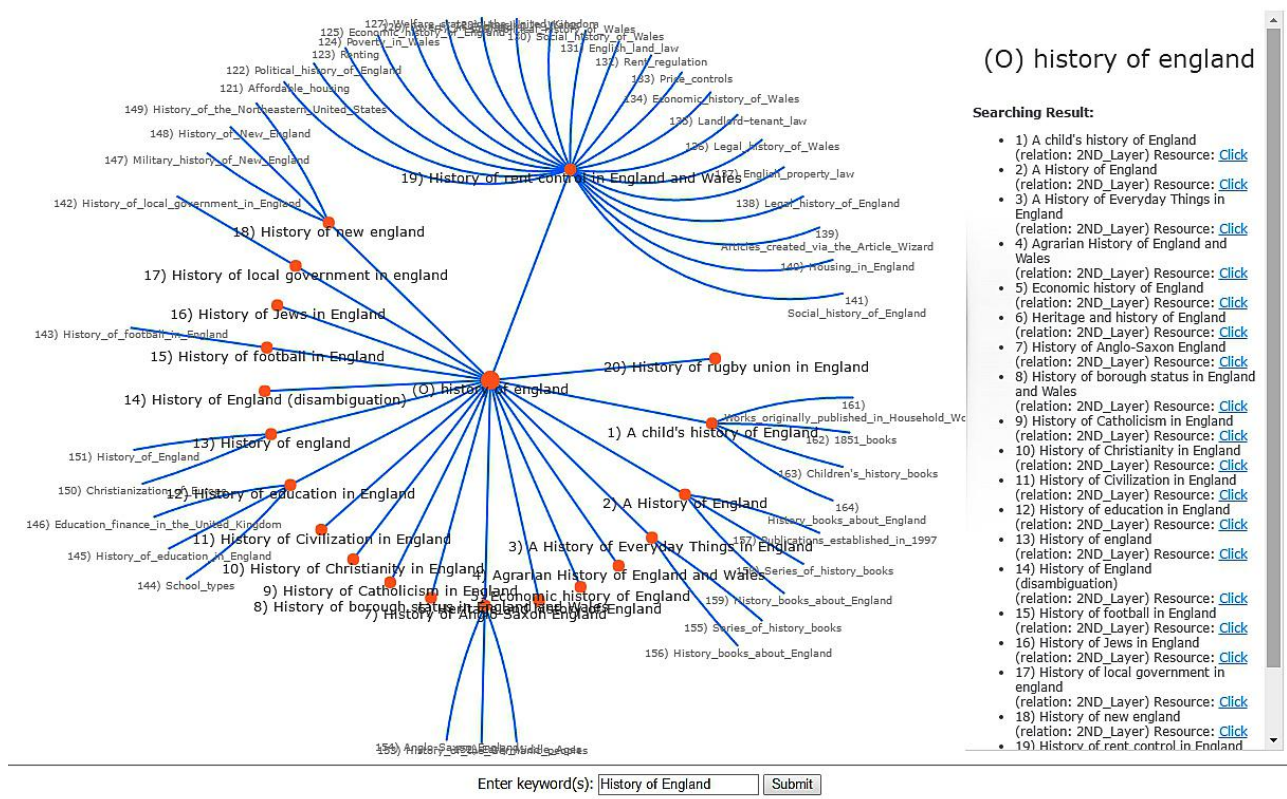


Fig. 3. VSDR prototype showing the searching result of the input keyword "History of England".

A. Evaluation with SUS

To assess overall usability of the proposed 3-layer approach based on the hyperbolic tree model, we measured user satisfaction via SUS questionnaire. It was commonly

used to measure the ease of use, learnability, and robustness by 10 questions of odd-numbered for positively-worded questions, and even-numbered for negatively-worded questions. The item scores were calculated to compute the overall SUS score.

We gave the general idea of the VSDR and the goal of the experiment to the users without any explanation of how to use it. Then, we gave them 20 minutes to use the searching tool and answered the questionnaires. Table I shows the summary result of the survey. Rating score of positively-worded and negatively-worded questions showed the validity of user satisfaction. Overall SUS score was 85.92%. It showed that

users were highly satisfied with the overall usability of the proposed 3-layer approach based on the hyperbolic tree model. In other words, the prototype was easy to use, understandable and provided robustness. Non-experience users familiar with IT could use the tool even when it was not explained to them.

TABLE I: EVALUATION RESULT OF VSDR PROTOTYPE BY SUS

Question statements (Question number)	Rating Score					Average	St. dev.	SUS item score
	1	2	3	4	5			
I think that I would like to use this searching tool frequently. (1)	0	0	8	10	12	4.13	0.82	94
I thought the searching tool was easy to use. (3)	0	0	0	26	4	4.13	0.34	94
I found the various functions in this searching tool were well integrated. (5)	0	0	5	17	8	4.1	0.66	93
I would imagine that most people would learn to use this searching tool very quickly. (7)	0	0	1	9	20	4.63	0.55	109
I felt very confident using the searching tool. (9)	0	0	0	2	28	4.93	0.25	118
I found the searching tool unnecessarily complex. (2)	10	13	7	0	0	1.9	0.75	93
I think that I would need the support of a technical person to be able to use this searching tool. (4)	20	10	0	0	0	1.33	0.47	110
I thought there was too much inconsistency in this searching tool. (6)	13	12	5	0	0	1.73	0.73	98
I found the searching tool very cumbersome to use. (8)	17	12	1	0	0	1.46	0.49	106
I needed to learn a lot of things before I could get going with this searching tool. (10)	27	2	1	0	0	1.13	0.43	116
Grand total								1031
Overall SUS score								85.92

B. Evaluation with Pattern-Based Technique

Patterns-based technique is a set of categories used to capture evaluation practices with the idea of tried-and-test. We applied the patterns-based technique to mainly assess the relevance of the input and the result.

TABLE II: RATING SCORE OF THE RELEVANCE TESTING BETWEEN THE OUTPUT RESULTS AND INPUT KEYWORDS

Topic for input	Relevance score*	Not relevance (No. of times)	No. of satisfied result (Out of 150)
Philosopher	3.30	8	125
Writer	3.20	22	100
Movie	3.50	3	132
Music	3.50	7	126
Cartoon	3.30	18	115
Place	2.80	58	56
History	2.70	51	55
Animal	3.80	2	141

*Notes. Meaning of relevance score ranges from 0 to 4.

0 means the output result does not display on the visualization.

1 means the output result displays on the visualization but there is no relevance to the input.

2 means the output result displays on the visualization and input keyword shows in the result detail, but it is not satisfied.

3 means the output result displays on the visualization and input keyword relates to the output result, but some of them are moderately satisfied.

4 means the output result displays on the visualization, input keyword relates to the output result, and it is very satisfied.

We gave users some tasks to search on the VSDR prototype with 8 specific topics (such as philosopher, writer, movie, music, cartoon, place, history, and animal). We asked them to specify any 5 input keywords for each topic without any explanation of how to use the VSDR prototype. The total number of input keywords for each topic was 150 keywords. While they were using VSDR prototype, we asked them to rate the relevance between the output result and their input keywords. Table II shows the relevance score result of this survey ranging from 2.7 to 3.8 depending on each topic. It

showed that the VSDR prototype gave moderately satisfied output results. It meant that only a few output results did not relate to the input keywords.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed the VSDR as a framework for displaying a 3-layer approach based on the hyperbolic tree model. VSDR framework can be implemented on any semantic searching tools to bring up some prominent issues and solve some limitation issues from related works with overall usability.

For our future work, we will further develop and implement our VSDR framework to better provide the following:

- A proper mechanism of the 3-layer approach based on the hyperbolic tree model: It will generate potentially related output results on a screen based on a specific algorithm to rank output semantically and simplify user cognition.
- The meaningful representation of the 3-layer approach based on the hyperbolic tree model: It should increase users' memorability while users keep searching along the hyperbolic tree. Distortions in hyperbolic space may cause some disappearance of previous focused nodes. Users should obviously see their traversed paths and easily find the way back to the path which they have passed.
- A generation of high responsiveness and task conformance in VSDR framework: The engine will map input keywords with querying commands by the additional configuration of advanced SPARQL commands.

ACKNOWLEDGMENT

The authors would like to acknowledge the staff from

Information Technology Service Center, Chiang Mai University for participating in the experiment, Mr. Nicolas Garcia Belmonte for the suggestion of visualization techniques, and thank you Dr. Jakramate Bootkrajang and Dr. Prakarn Unachak for some good advice on the manuscript.

REFERENCES

- [1] A. J. Zitzelberger, D. W. Embley, S. W. Liddle, and D. T. Scott, "HyKSS: Hybrid keyword and semantic search," *Journal on Data Semantics*, vol. 4, issue 4, pp. 219-229, Dec. 2015.
- [2] J. A. Mac ís, "Enhancing interaction design on the semantic web: A case study," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 42, issue 6, pp. 1365-1373, Nov. 2012.
- [3] J. Ahn and P. Brusilovsky, "Adaptive visualization for exploratory information retrieval," *Information Processing & Management*, vol. 49, issue 5, pp.1139-1164, Sep. 2013.
- [4] A. Nuzzolese, V. Presutti, A. Gangemi, A. Musetti, and P. Ciancarini, "Aemoo: Exploring knowledge on the web," *Proceedings of the 5th Annual ACM Web Science Conference*, pp. 272-275, 2013.
- [5] S. Lehmann, U. Schwanecke, and R. Dörner, "Interactive visualization for opportunistic exploration of large document collections," *Information Systems*, vol. 35, issue 2, pp. 260-269, April 2010.
- [6] F. Balakov, B. König-Ries, A. Nauerz, and M. Welsch, "Introspective views: An interface for scrutinizing semantic user models," in *Proc. International Conference on User Modeling, Adaptation, and Personalization*, 2010, pp. 219-230.
- [7] F. Florenzano, D. Parra, J. Reutter, and F. Venegas, "A visual aide for understanding endpoint data," in *Proc. the Second International Workshop on Visualization and Interaction for Ontologies and Linked Data*, 2016, pp. 102-113.
- [8] S. Zainab, Q. Mehmood, D. Zehra, D. Rebholz-Schuhmann, and A. Hasnain, "PrEVIEW: Clustering and visualising PubMed using visual interface," in *Proc. the Second International Workshop on Visualization and Interaction for Ontologies and Linked Data*, 2016, pp. 17-27.
- [9] J. Lamping, R. Rao, and P. Pirolli, "A focus+context technique based on hyperbolic geometry for visualizing large hierarchies," in *Proc. the Second International Workshop on Visualization and Interaction for Ontologies and Linked Data*, 1995, pp. 401-408.
- [10] G. Marchionini, "Exploratory search: from finding to understanding," *Communications of the ACM*, vol. 49, issue 4, pp. 41-46, April 2006.
- [11] V. Dimitrova, L. Lau, D. Thakker, F. Yang-Turner, and D. Despotakis, "Exploring exploratory search: A user study with linked semantic data," in *Proc. the 2nd International Workshop on Intelligent Exploration of Semantic Data*, 2013, no. 2.
- [12] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," in *Proc. 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference*, 2007, pp. 722-735.

[13] J. Lewis and J. Sauro, "The factor structure of the system usability scale," in *Proc. International Conference on Human Centered Design*, 2009, pp. 94-103.

[14] N. Elmqvist and J. S. Yi, "Patterns for visualization evaluation," in *Proc. the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization*, 2012, no. 12.



interests is visualization techniques.

Natanun Kanjanakuha is the manager of Learning Innovation Center, Information Technology Service Center, Chiang Mai University, Thailand, and is currently studying a Ph.D. degree at the Department of Computer Science, Faculty of Science, Chiang Mai University. He graduated with the major of information technology and management from the Graduate School, Chiang Mai University, Thailand. His research



Churee Techawut is an assistant professor at the Department of Computer Science, Faculty of Science, Chiang Mai University. She completed her Ph.D. in computer science from the Asian Institute of Technology (AIT), Thailand. Her research interests are human computer interaction (HCI) and ontology engineering.



Rattasit Sukhahuta is the director of Information Technology Service Center, Chiang Mai University, Thailand, and an associate professor at the Department of Computer Science, Faculty of Science, Chiang Mai University. He completed his Ph.D. in computer science from the University of East Anglia, England. His research interest is natural language processing (NLP).



(HCI) and information visualization.

Paul Janecek is the founder and CEO at Think Blue Data Co., Ltd. He was an assistant professor at the Asian Institute of Technology (AIT), Thailand, a researcher of university labs in Europe and an engineer at Intel. He completed his Ph.D. in Computer Science from the École Polytechnique Fédérale de Lausanne (EPFL), a research institute in Lausanne, Switzerland. His research interests are human computer interaction