Paradigm Extension of Faceted Search Techniques: A Review

Mohammed Najah Mahdi[1], Abdul Rahim Ahmad[2], Roslan Ismail[2]
[1]*College of Graduate Studies,*
[2]*College of Computer Science & Information Technology,*
University Tenaga National Malaysia
moh12004@gmail.com

***Abstract*—Modern populations rely heavily on the world-wide web in searching information because it is the largest human repository of knowledge. However, finding relevant information on the web is often challenging. In the current work, we review analyses and optimize the performance of exploratory and faceted search techniques. Search behavior that is characterized by a large amount of uncertainty about the goals of the search is common in exploratory search. On the other hand, faceted search technique refines search results by a faceted taxonomy in an iterative manner. In addition, facets provide an efficient way to analyze and navigate the search result space. However, we believe that facet selection has been guided by the properties of suboptimal facet and facet term. As a consequence, users may need technical support while searching information. Thus, this paper suggests a natural way of extending the current paradigm employed by traditional search systems, the exploratory search. Our main objective is to provide a framework or a platform which is extensible with plugins and able to provide instances tunable to a particular document collection of choice. In addition, this paper presents a research model based on the prototype that will be developed.**

***Index Terms*—Exploratory Search; Facet; Faceted Search; Search Engine.**

I. Introduction

The rapid growth of the internet has made it significantly more challenging for users to obtain accurate information, due to an insufficient amount of results relevant to the user, as well as the lack of result sorting to separate the relevant results from those that are irrelevant to the user's needs. Traditionally, the user interface of a search engine obtains input from a user and it would subsequently sort and display the outputs. Therefore, the user interface serves as the front end and any changes made would have a direct impact on the user [1]. The amount of information request is incredibly huge. As estimated by Baye, et al. [2] and Bughin [3], the internet today contains more than 3 billion documents, and search engines like Google can receive more than 3.5 billion queries per day. Unfortunately, only the initial listings are reviewed by users, leaving a bulk of collected information unchecked.

II. Related Work

In recent years, the field of faceted search has been gaining significant attention from researchers [4-6] It is widely applied in e-commerce such as eBay.com [7] and Amazon.com [8] as well as bibliographic databases including Faceted DBLP [9], the massive IEEE/IET Electronic Library [10] and ISI Web of Knowledge [11] databases. Adding to these are various multimedia libraries such as the Open Video Digital Library [12], Google Images [13] and a multitude of other databases. Currently, the effect of query ambiguity on the results has been studied and researchers have identified that the number of relevant results are simply insufficient [14]. Of course, queries that are less ambiguous would return more relevant results. Thus, the present work focuses on methods of result sorting.

The exploratory search involves information search in an unfamiliar domain. White and Roth, in 2009 [15], identified that exploratory searches encompass a much broader class of search activities than traditional IR, of course, while assuming that all the relevant information exists and it can be recovered by forming a proper query statement. Also, apart from obtaining the "correct" answer, White and Roth [15] pointed out that it is important to trigger the learning process when users are navigating to their destination during the exploratory search.

However, it is challenging for searchers to meet their search goals during exploratory searches, primarily as the significant levels of uncertainty in the search process and as well as the substantial user-specific information requirements. Proving features, including query suggestions, related searches and query auto-completions are typically not sufficient due to the highly exploratory nature of the task. As exploratory searches encompass numerous trials involving different queries, sources and information, the overall searcher behavior should be emphasized while providing a rigid analysis of the inherent search process in order to provide meaningful search paths [16].

Prior to the existence of web-based search engines that are commonly seen today, the effort to obtain information was usually relegated to librarians well-versed in operating those search engines. Nowadays, the web search engines are operated by common users, who may not have undergone proper training on the search engine. Therefore, these users tend to formulate inefficient queries as they do not understand the searching and sorting operations of a web search engine [17].

Therefore, this paper aims to investigate the current web-based information retrieval methods and highlight their weaknesses. Specifically, we intend to explore the available exploratory searches, the limitations of the current system, the need for developing the exploratory search system and the benefits of using an efficient exploratory search system. Finally, we wish to identify the elements or techniques that would improve the performance of the exploratory search system.

We adopt the current paradigm employed by the traditional search systems in the exploratory search by providing a framework or a platform extensible with plugins and it is able to provide instances tunable to a particular document collection of choice. It is important to note that the extension happens in a natural manner as the older settings are retained when migration occurs [18].

The natural elements of exploratory search include facets, information visualization, and query-by-example and social search. For example, exploratory search is achieved by extending faceted search, supplemented by added functionalities of visualization, query-by-example and social search. In general, the faceted search can be performed by combining faceted navigation with full-text search.

The main contribution of the current work is to offer suggestions or recommendations in extending the current paradigm of faceted search into exploratory search. The search results and facets should be extended with different views. Query-by-example should be integrated with Bayesian Sets as it replaces the handling of complex content-based searches to choosing the right plugin. Finally, the system should serve as a framework to instantiate Exploratory Search Systems, whereby every component in the system is a community-driven plugin.

III. EXPLORATORY SEARCH AND SEARCH ENGINES

Exploratory searches encompass a combination of queries and collection browsing while collating information [15]. It adopts a sequence of multiple query and search sessions to allow the user searching for information to expand the knowledge of the task currently undertaken while in the process of identifying the useful information.

During the searching process, a series of complex cognitive tasks is taking place which would encourage the information seeker to learn, explore and acquire intellectual skills [19]. Searchers who are conducting exploratory searches are generally not familiar with the domain in which their goal lies, and as such in most cases would need to learn about a topic to have an understanding of how their goals can be achieved. This can, in turn, result in them being either unsure about the goals they seek, unsure about the method in which to achieve those goals, either through the technology or process available, or even a combination of both. Figure 1 presents the exploratory research process.

A. *Exploratory Search Systems*

Exploratory Search Systems (ESSs) gain advantages from new capabilities induced by the latest developments in technology as well as updated and more natural interface paradigms to promote interaction with search systems. Examples of ESSs include information visualization systems, document clustering and browsing systems, as well as intelligent content summarization systems.

According to White and Roth [15], the main goals of exploratory search systems (ESSs) are to facilitate learning and investigation during the searching process and to guide the users in exploring uncharted territories. In exploratory search, users are exposed to various collections information; therefore, it is important for an exploratory search system to summarize the information in appropriate categories in order to facilitate user exploration.

Via ESS, users are able to enhance the rate at which they gain information, determine appropriate navigational paths, and understand the encountered information. In addition, through interface features such as dynamic queries in ESS [20], users are able to identify the immediate impacts of their decisions made.

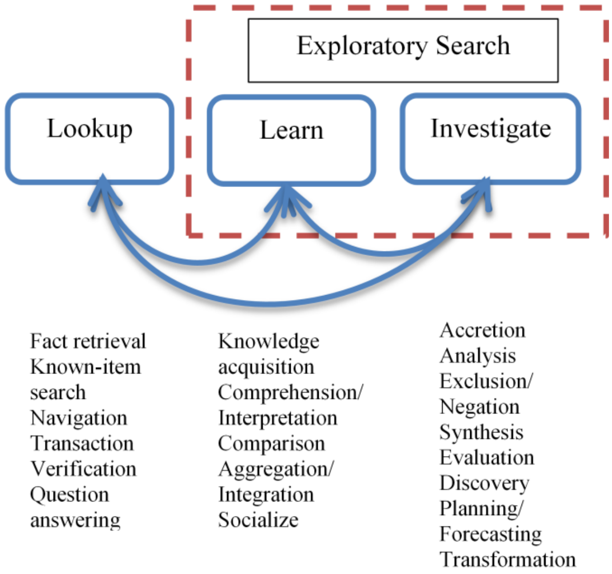

Figure 1: Exploratory research [21]

B. *ESS Features*

White and Roth [15] discussed eight key features of a powerful and successful ESS. Only four features, however, are examined in the current work. The first feature enables users to refine their queries. The users are able to design interactions in order to narrow down the results in order to seek for more relevant information. White and Roth promoted the idea of a dynamic query and highlighted that it enables users to generate and view results rapidly and continuously by using visual tools such as sliders and filter widgets. According to White and Roth, dynamic queries explore the data simultaneously on “multiple dimensions”; hence, it is efficient for hypothesis generation.

Another useful feature is faceted navigation, where facets are used to represent metadata related to the searched objects. Besides that, faceted navigation is able to filter desired results and explore the collection. White and Roth (2009) stressed the importance of visual representation of the available data because it can stimulate searchers' insight and guide them in the subsequent exploration. Therefore, the visual representation is considered in the current work as well by representing the amount of resources available per facet (resource count). In addition, the impact of these features on the primary goals of exploratory searches i.e. learning and understanding [15], is examined.

Kules and Shneiderman [22] have described 8 design guidelines for future exploratory search systems. Again, they have emphasized on the benefits of information visualization by adopting the information hierarchy (e.g. parent-child relationship). This would enable a better understanding of the information space.

C. *Key Concepts in Faceted Search*

Figure 2 shows the user interface of faceted navigation at Soap [23]. This interface provides an intuitive impression of faceted search and serves as a good example that illustrates the concepts of facet, faceted taxonomy, and faceted search.

Facet: Ranganathan [24] first introduced this concept to describe the multidimensional properties of documents. He described five fundamental facets, these being Personality, Matter, Energy, Space and Time (PMEST). Based on these facets, the first faceted classification system, namely the Colon Classification, was developed.

Faceted Taxonomy: A faceted taxonomy encompasses a large group of taxonomies [25]. This is given as an example in Figure 2, which includes the facets of "BRAND", "TYPE" and others. A faceted taxonomy can be used to identify user intents [26] and instant overviews [27]. In a faceted taxonomy, the terms of different facets are indeed orthogonal, i.e. multiple appearances of one term in different facets is not allowed [28]. This would ensure the separability of a faceted taxonomy; therefore, other facets would not be affected if the term/structure of one facet is changed.

Faceted Search: In certain cases, a faceted search is also referred to as a faceted navigation or faceted browsing. This is an interactive, heuristic and progressive refinement of the search paradigm, allowing searchers to iteratively select various facets and terms to refine the search results [5, 29, 30].

In actual scenarios, only a small subset of a full faceted taxonomy is given in the interface. These taxonomies are dynamically generated (i.e. dynamic taxonomy) from the search results obtained after each iteration [31]. It is interesting to note that these dynamic taxonomies are able to summarize the current search results and provide a set of links leading to new search results [32].

IV. Literature Survey On Faceted Search

Numerous techniques for faceted search have been proposed in open literature. In the current work, all relevant methods or techniques are reviewed. The strengths and weaknesses of each method are highlighted in Table 1.

V. Observation And Remarks

Although faceted search has been intensively studied, we are aware that the implementation of faceted search is associated with a few challenges: Firstly, most existing faceted search systems have faceted taxonomies that are created and maintained manually.

This makes these approaches highly resource, time and labor intensive. Although initial studies have been done by Dakka, et al. [41] and Stoica, et al. [42] on automatic faceted taxonomy construction, these methods are still impractical. Thus, automatic extraction of the facet terms and hierarchical relations are preferable. Substantial research on extracting domain terms as well as their relations have been carried out within the ontology learning research community. However, these approaches cannot be applied towards efforts to extract facet terms and their relations due to two main reasons: (a) facet terms and domain terms deviate considerably from each other, as the domain terms are always domain-specific, as compared to the facet terms can be domain-independent; as well as (b) domain terms are typically spread-out in data objects (for example, domain text). However, the appearances of the top-level facet terms in data objects are rare [41].

The second challenge is that faceted search focuses mainly on closed datasets which are usually semi- or fully-structured. With the rapid development of web-based applications, a substantial amount of data produced is either unstructured or have complex structures. Thus, novel faceted classification and faceted navigation methods for these types of data and should also be studied. Additionally, visual clutter and change blindness are two unsolved problems in visualization. Usually, visual clutter occurs due to the substantially large number of facet terms and the limited size of the user interface. Thus, the resolving efforts such as adaptive displays, hiding, expanding and folding of facet and facet terms is worth for future investigation. Change blindness happens whenever an observer fails to notice visual changes [41]. In faceted search, change blindness refers to a scenario where the data objects disappear suddenly during a navigation process [43]. In order to alleviate this problem, the use of animated transitions of faceted search would be able to assist users in perceiving any alterations in the interface.

In addition to this, it is interesting to represent the search results in a form of a multi-dimensional cube because it improves the search efficiency and facilitates the integration of data mining, OLAP Cube, etc. as compared to one-dimensional cube [29, 44]

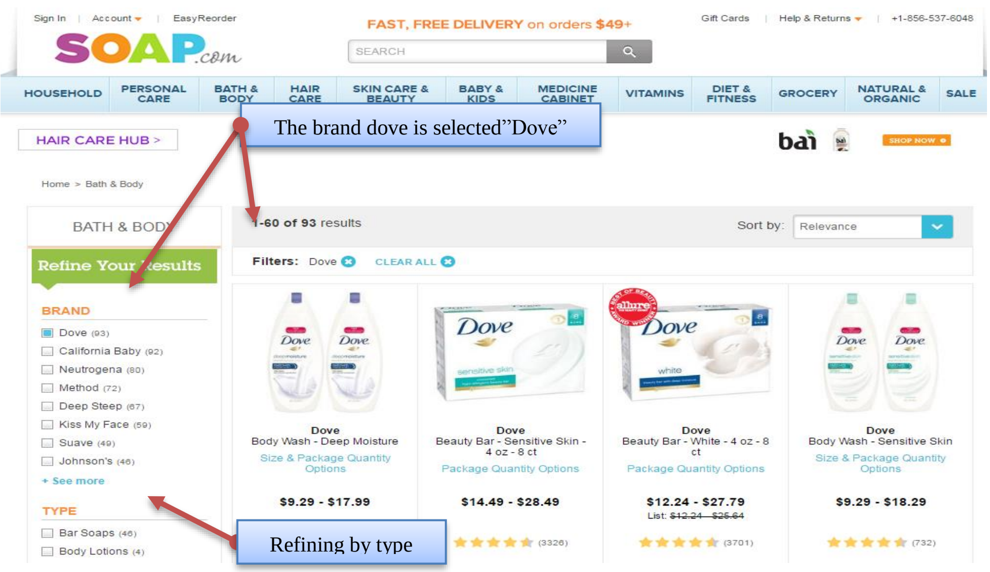

Figure 2: Faceted Navigation at Soap.com

Table 1
Critical Review Faceted Search

Author(s)/Year	Methods/Techniques	Brief Description	Strengths and Weaknesses
Stuart-Moore, et al. [33]	Developed an interface that was coded in PHP, MySQL, and Javascript. PHP whose function was to perform queries to a target database in a way that limits the resources needed by setting metadata facets.	The search approach uses a combination of combination of browsing and searching to obtain the desired results. The target was to create a browsing experience that would leverage on faceted metadata in the database catalog, allowing users to refine the catalog to a list of relevant items by placing various limits on the search, including requiring specific values as well as on a number of different metadata facets such as orthogonal fields like "Topic" and "Resource Type".	Authors suggest further improvement of the interface design and database query efficiency.
Simonini and Zhu [34]	This approach visualizes large amounts of data using a Bayesian suggestion algorithm. This approach was widely used in the enterprise search platform Solr.	This article presents a novel approach towards the visualization of large amounts of data based on the aforementioned Bayesian technique.	The suggested approach provides a dynamic and visual tool for big data activities. This approach is essentially a derivation of the popular open source enterprise search engine Apache Solr.
Seifert, et al. [35]	The proposed method here involves the combination of Voronoi subdivisions and tag clouds.	The approach here is a novel visualization technique that allows for the refinement and easier navigation of users through search results. From this, users can then construct complex Boolean search queries to narrow down the search space.	The authors put a spotlight on the fact that the ability to move facets out of focus would be of greater importance when the number of facets rises.
Ullah and Iftikhar [36]	Uses a customized prototype interface based on a specialized facet for children. The search engine developed by the authors using this technique was named Kidzz search.	The authors propose an alternate search mechanism that can be used by children that can overcome the natural limitations and complications expected from that approach.	The obtained results indicate that the proposed faceted searching mechanism is inherently better than the traditional query approach typically sued for children. However, a shortcoming of this approach is that the authors can only use the headings and subheadings in web pages as metadata. Although it is possible to obtain more useful and relevant metadata through more sophisticated techniques.
Whaling, et al. [37]	The study employed a *LENS Faceted Browser*	This paper presents the Lens system, a client-side interface for the faceted navigation of scholarly RDF data.	The authors provide an open source faceted browser that can be used by researcher profiling systems to enable deep exploration and the classification of the relationships of researchers with their research products.
Al-Aqrabi, et al. [38]	A faceted search framework based using cloud computing with OLAP querying on XML based data warehouses	The work here proposes a method develop a technology positioning map for implementing faceted searches in a business intelligence framework using cloud computing.	The authors highlight that through the use of facets, the user can define direct queries and rapidly identify the most relevant XML views required to generate the dashboards.
Bonino, et al. [39]	The model is developed in accordance with the Set Theory, and is able to express multiple of facet types, depending on the nature of the metadata presented to it.	This paper proposes the FaSet system, which is a representation model and search algorithm that supports the implementation of faceted search engines. The authors suggest that the FaSet system will rely on set theory, and therefore presents a good balance between expressive power and ease of implementation on web architectures.	A Drupal content management system module and can be used to drive a website aggregating expert reviews for accessible software. Extensions of this work will see performance evaluation of the FaSet implementation on realistic datasets, and the evaluation of user interface design for improving usability.
Wang, et al. [40]	The proposed system is able to extract and mine, from multiple syntactic and semantic facets, information for a candidate program with implemented elements in the form of a web-based tool MFIE.	This paper recommends a feature location approach that is able to support a multi-faceted interactive program approach towards exploration.	The authors highlight that the feature location practice of their approach can be enhanced through the use of a multi-faceted interactive exploration process, and thus recommend the investigation of more sophisticated techniques for generating more relevant facets.

VI. CONCLUSION

This paper reviews and summarizes a number of selected literature on faceted search. A comparison study on existing faceted search systems is undertaken, and while there is a multitude of faceted search systems applied in many domains, there are several unresolved problems that are worth for future investigation. Hence, four future research directions have been proposed, i.e. automatic faceted taxonomy construction, faceted interface and visualization, relevance evaluation of faceted search results, and faceted classification and navigation of complex/open data.

REFERENCES

[1] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "SAFQuery: a simple and flexible advanced Web search interface," *The Electronic Library,* vol. 34, pp. 155-168, 2016.

[2] M. R. Baye, B. De los Santos, and M. R. Wildenbeest, "Search Engine Optimization: What Drives Organic Traffic to Retail Sites?," Journal of Economics & Management Strategy, vol. 25, pp. 6-31, 2016.

[3] J. Bughin, "Big data, Big bang?," Journal of Big Data, vol. 3, p. 1, 2016.

[4] M. Hearst, "User interfaces for search," Modern Information Retrieval, pp. 21-55, 2011.

[5] D. Tunkelang, "Faceted search," Synthesis lectures on information concepts, retrieval, and services, vol. 1, pp. 1-80, 2009.

[6] L. Zhang and Y. Zhang, "Interactive retrieval based on faceted feedback," in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010, pp. 363-370.

[7] eBay, http://www.ebay.com/.

[8] Amazon, https://www.amazon.com/.

[9] Homoceanu, Silviu, et al. "Time-Based Exploratory Search in Scientific Literature." International Conference on Theory and Practice of Digital Libraries. Springer Berlin Heidelberg, 2013.

[10] B. Zhao, X. Lin, B. Ding, and J. Han, "TEXplorer: keyword-based object search and exploration in multidimensional text databases," in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 1709-1718.

[11] I. W. o. Knowledge, "ISI Web of Knowledge," 2016.

[12] G. Marchionini and G. Geisler, "The open video digital library," D-Lib Magazine, vol. 8, pp. 1082-9873, 2002.

[13] G. Images, "Google Images," 2016.

[14] I. Ounis, C. Macdonald, and R. L. T. Santos, "Search result diversification," Foundations and Trends in Information Retrieval, vol. 9, pp. 1-90, 2015.

[15] R. W. White and R. A. Roth, "Exploratory search: beyond the query-response paradigm (Synthesis lectures on information concepts, retrieval & services)," Morgan and Claypool Publishers, vol. 3, 2009.

[16] K. Athukorala, D. GÅ‚owacka, G. Jacucci, A. Oulasvirta, and J. Vreeken, "Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks," Journal of the Association for Information Science and Technology, 2015.

[17] K. Collins-Thompson, S. Y. Rieh, C. C. Haynes, and R. Syed, "Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies," in Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval, 2016, pp. 163-172.

[18] J. Teevan, K. Collins-Thompson, R. W. White, S. T. Dumais, and Y. Kim, "Slow search: Information retrieval without time constraints," in Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval, 2013, p. 1.

[19] M. Dörk, "Visualization for Search: Exploring Complex and Dynamic Information Spaces," Doctoral dissertation, University of Calgary, 2012.

[20] B. Shneiderman, Designing the user interface: strategies for effective human-computer interaction: Pearson Education India, 2010.

[21] G. Marchionini, "Exploratory search: from finding to understanding," Communications of the ACM, vol. 49, pp. 41-46, 2006.

[22] B. Kules and B. Shneiderman, "Users can change their web search tactics: Design guidelines for categorized overviews," Information Processing & Management, vol. 44, pp. 463-484, 2008.

[23] Soap, " https://www.soap.com/".

[24] S. R. Ranganathan, Classification, coding and machinery for search: Unesco, 1950.

[25] E. Kopanaki, V. Karkaletsis, C. D. Spyropoulos, N. Avradinis, N. Fakotakis, T. Kalamboukis, et al., "MITOS: An integrated web-based system for information management," in 8th Pan-Hellenic Informatics Conference. Greek Computer Society, November, 2008.

[26] V. Jethava, L. Calderón-Benavides, R. Baeza-Yates, C. Bhattacharyya, and D. Dubhashi, "Scalable multi-dimensional user intent identification using tree structured distributions," in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, 2011, pp. 395-404.

[27] P. Fafalios, I. Kitsos, and Y. Tzitzikas, "Scalable, flexible and generic instant overview search," in Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 333-336.

[28] D. N. Joudrey, A. G. Taylor, and D. P. Miller, Introduction to Cataloging and Classification: ABC-CLIO, 2015.

[29] Zheng, Bweijunl, Wei Zhang, and Xiaoyu Fu Boqin Feng. "A survey of faceted search." Journal of Web engineering 12.1&2 (2013): 041-064.

[30] Voigt, Martin, et al. "Weighted faceted browsing for characteristics-based visualization selection through end users." Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems. ACM, 2012.

[31] Sacco, Giovanni Maria, and Yannis Tzitzikas, eds. Dynamic taxonomies and faceted search: theory, practice, and experience. Vol. 25. Springer Science & Business Media, 2009.

[32] Abello, Alberto, et al. "Using semantic web technologies for exploratory OLAP: a survey." IEEE Transactions on Knowledge and Data Engineering 27.2 (2015): 571-588.

[33] Papadakos, Panagiotis, and Yannis Tzitzikas. "Hippalus: Preference-enriched Faceted Exploration." EDBT/ICDT Workshops. Vol. 172. 2014.

[34] G. Simonini and S. Zhu, "Big data exploration with faceted browsing," in High Performance Computing & Simulation (HPCS), 2015 International Conference on, 2015, pp. 541-544.

[35] C. Seifert, J. Jurgovsky, and M. Granitzer, "FacetScape: A visualization for exploring the search space," in 2014 18th International Conference on Information Visualisation, 2014, pp. 94-101.

[36] K. A. Ullah and M. A. Iftikhar, "Information search for children using Faceted navigation," in Open Source Systems and Technologies (ICOSST), 2013 International Conference on, 2013, pp. 28-33.

[37] R. Whaling, T. Malik, and I. Foster, "Lens: A Faceted Browser for Research Networking Platforms," in eScience (eScience), 2013 IEEE 9th International Conference on, 2013, pp. 196-203.

[38] H. Al-Aqrabi, L. Liu, R. Hill, and L. a. L. Cui, Jianxin, "Faceted Search in Business Intelligence on the Cloud," in International Conference on and IEEE Cyber, Physical and Social Green Computing and Communications (GreenCom), 2013, pp. 842-849.

[39] Arenas, Marcelo, et al. "Towards semantic faceted search." Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014.

[40] J. Wang, X. Peng, Z. Xing, and W. Zhao, "Improving feature location practice with multi-faceted interactive exploration," in Proceedings of the 2013 International Conference on Software Engineering, 2013, pp. 762-771.

[41] Pai, Deepak, et al. "Automated faceted reporting for web analytics." Proceedings of the 4th international workshop on Web-scale knowledge representation retrieval and reasoning. ACM, 2013.

[42] Kong, Weize, and James Allan. "Extending faceted search to the general web." Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, 2014.

[43] Walton, Rachel. "Searching High and Low: Faceted Navigation as a Model for Online Archival Finding Aids (A Literature Review)." Journal for the Society of North Carolina Archivists 12 (2015).

[44] Pawliczek, Piotr, Witold Dzwinel, and David A. Yuen. "Visual exploration of data by using multidimensional scaling on multicore CPU, GPU, and MPI cluster." Concurrency and Computation: Practice and Experience 26.3 (2014): 662-682.