Knowledge Graphs versus Hierarchies: An Analysis of User Behaviours and Perspectives in Information Seeking

Bahareh Sarrafzadeh
University of Waterloo
bsarrafz@uwaterloo.ca

Alexandra Vtyurina
University of Waterloo
avtyurin@uwaterloo.ca

Edward Lank
University of Waterloo
lank@uwaterloo.ca

Olga Vechtomova
University of Waterloo
ovechtomova@uwaterloo.ca

ABSTRACT

In exploratory search, how information is presented to the user and how the user interacts with the presented information heavily influence the user's success. In this paper, we examine two different spatial representations of search results: knowledge graphs and hierarchical trees. Through interaction logs we show that knowledge graphs can effectively reduce the need to read source content with no reduction in the quality of the information gathered by the user. Through qualitative interviews and thinkalounds we explore factors that influence user perception of different search results representations including biases, task, perceived structure of the data, and problem-solving approach. Overall, these results enhance our understanding of the role each of these representations can play in information seeking.

1. INTRODUCTION

In the domain of on-line search, the output of current search engines is normally sufficient for many well-defined online tasks, including navigational queries, transactional queries, and many types of informational queries. However, when information is sought to address broad curiosities, e.g. for learning and other complex mental activities, retrieval is necessary but may not be sufficient [35], [1].

One type of complex search that is of increasing interest to researchers is *exploratory search*, where the goal involves "learning" or "investigating", rather than simply "look-up" [20]. In characterizing searching that involves learning or investigating, Marchionini (referencing Bloom's taxonomy of educational objectives) notes that the goal of these searches involves "knowledge acquisition, comprehension of concepts or skills, interpretation of ideas, and comparisons or aggregation of data and concepts" [20].

There are many open research questions about how to design interfaces to support exploratory search using techniques that organize the retrieved information into meaningful structures. Search results presented by modern search engines are an example of an ordered list sorted by relevance, i.e. a *vectorial model* [19]. However, information seekers often express a desire for a user interface that organizes search results into meaningful groups to help make sense of the results, to infer relationships between concepts, and to help decide what to do next [13], [20]. As a result of this desire for organization, *spatial models* [23], i.e. hierarchies and networks, have also been used to organize information and support sensemaking [19]. While both hierarchies and networks have been shown to be useful in the structuring of content (e.g. [9], [22], [10]) little work has explored the similarities and differences between these two representations. The goal of this paper is to explore how two specific visualizations of information – Knowledge Graphs (or Knowledge Maps) and Hierarchical Trees – support exploratory search tasks.

We present the quantitative and qualitative results of a study contrasting participants' perspectives on the use of knowledge graphs versus hierarchical trees to support exploration of data for the purpose of developing an answer for informational queries. We describe the design of interfaces and our evaluation of the use of network and hierarchical data structures during exploratory search tasks. Log data indicates that knowledge graphs result in participants viewing source documents fewer times and spending less time reading those documents with no effect on overall quality of information gleaned to satisfy queries. Data from thinkalouds and a post-task interview are synthesized using a grounded theory approach, yielding observations on biasing factors, task effects, data relationships, and problem solving approaches which discriminate between use-cases for our hierarchical tree-based interface versus our knowledge graph interface. Overall, despite some statistical advantages of knowledge graphs in our study, our goal is not to argue that one presentation of information is better than another; information visualization research would argue that different visualizations serve different purposes [34] and our goal with this work is to better understand the purposes that each of these two interfaces serve with respect to exploratory search.

The remainder of this paper is organized as follows. First, we provide an overview of related work, focusing on different categorizations of search tasks and goals, organizing search results, and evaluation of different search results visualizations. Next, we briefly describe our interface design process, beginning with low-fidelity prototypes and culminating with the two tested interfaces. We then present our study design

CHIIR '16, March 13 - 17, 2016, Carrboro, NC, USA

ISBN 978-1-4503-3751-9/16/03…$15.00
DOI: http://dx.doi.org/10.1145/2854946.2854958

and our results. Finally, we conclude with a discussion on the implications of our work in the design of exploratory search interfaces.

2. RELATED WORK

2.1 Understanding Web Search Queries

Studies of user search behaviour have a long history in Information and Library Science. Specifically with respect to web search, Broder [5] proposed a taxonomy of Web Search in 2002. He was motivated by the idea that the traditional notion of an "information need" might not adequately describe web searching. Broder's taxonomy classifies web searches into navigational, informational and transactional. Similarly, Rose and Levinson [27] analyze user goals to classify web searches into Navigational, Informational and Resource. Drawing upon earlier work by Campbell [8] and Byström [7], web searches can broadly be classified into "Simple" and "Complex" searches. Simple search tasks are similar to "known-item" search tasks and usually involve looking up some discrete, well-structured information object: for example numbers, names and facts [20]. Complex search tasks, on the other hand, involve investigating, learning and synthesizing of information [36].

In contrast to Broder's and Rose and Levinson's taxonomies, Marchionini [20] focuses specifically on a process he terms *exploratory search*. Marchionini broadly separates web search into three categories: Look-up, which includes fact retrieval, navigation and transaction; Learn, which includes knowledge acquisition, comprehension, and comparison; and Investigate, which includes analysis, synthesis and evaluation. The latter two categories, Learn and Investigate, he groups under the umbrella of exploratory search. There are two activities which mediate the process of exploratory search: information foraging theory [25], which describes how searchers collect relevant pieces of information, and sensemaking [11], which describes the process through which people assimilate new knowledge into their existing understanding.

Marchionini notes that there are interactive aspects to exploratory search, rather than simply viewing the query satisfaction or information retrieval problem as optimally matching documents to a query. Characteristics of these interfaces, drawn from research in human-computer interaction, include the use of high-level overviews and rapid previews to facilitate sensemaking during the exploratory process. The incorporation of overviews argues for some organization of search results that both presents this overview and allows the user to explore the data and its interconnected relationships more fully through filtering and the examination of user-selected details [32].

2.2 Organizing Search Results

Any system that supports information seeking must structure information to make it accessible. The way information is organized and made available affects the strategies used to access this knowledge and thus information-seeking performance [21]. For example, Capra et al. [9], in their study on the relationships between search tasks, information architecture and interaction style, note, among other observations, that users gain benefits from support for facets and topic organization implemented in a flexible style.

Given the observation that organization of search results benefits users, one might then ask what organizations of search results exist. A taxonomy of techniques for organizing search results was proposed by Wilson et al. [37]. They identify two main classes of approaches: (1) Using annotations or classifications to organize results into groups (e.g. faceted search which uses a hierarchy structure to enable users to browse information by choosing from a pre-determined set of categories). (2) Directly organizing results which visualizes a result set to help users find the specific results they are looking for. As well, Ltifi et al [19] proposed a classification of visualization techniques for knowledge discovery including visualizations for linear data (e.g. timelines), multi-dimensional data (e.g. scatterplots), vectorial models (e.g. relevance-ordered results), hierarchies (trees, tables of contents) and networks (e.g. knowledge graphs). Given the non-numerical characteristic of web search results, the latter three types of visualizations (vectorial, hierarchical, and network) are particularly useful for displaying search results. While the vectorial representation presents results as a ranked list, hierarchical and network representations can be used to display grouping, similarity or relationships among search results.

Fully elaborating on all of the organization techniques or visualizations for search results is beyond the scope of this paper, and the interested reader is referred to the above taxonomies. However, some visualizations of search result data are specifically salient to our research, in particular, Ltifi et al.'s [19] vectorial model, hierarchies and networks. Trees are a common tool for representing hierarchies. A hierarchical structure is mainly made up of organizational links that organize the information into categories (topics) with no or few cross-links between categories. Google's "Knowledge Graph" enhances basic search results with structured data, essentially presenting a network organization of search results [33]. Google claims the knowledge graph enhances search in three main ways: query disambiguation, a summarization of related facts, and exploratory search suggestions (based on what other users explored next).

Most network visualizations tend to provide a global perspective on a graph by attempting to represent an overview of the information space so no information is missing and the data can speak for itself. Most of these techniques are based on Shneiderman's Visual Information Seeking Mantra [32]: "Overview first, zoom and filter, then details on demand". For example Sanchez and Llamas [29] followed this principle to visualize a large combination of concept maps to distinguish between an interface for the author and an interface for the end user that facilitates the exploration tasks.

2.3 Evaluating Search Results Visualizations

Novick and Hurley [23], working in the field of education, performed extensive research on the use of spatial models such as networks, hierarchies, and matrices. In particular, they were interested in the properties of these spatial models that were particularly suited to problem solving. Our work differs in its focus on information retrieval and the representation of search results. As well, our work differs in that Novick and Hurley do not develop interfaces that support problem solving; instead, they use questionnaire data to elicit from participants which representations participants think might best support information representation.

More recently, researchers in information retrieval have performed evaluations of techniques for representing search results, examining both hierarchical structures (e.g. [9], [22],

[12]) or networks ([31]). However, these results investigate how different properties of one structure may affect users' behaviour, whereas our work aims at understanding the type of support provided by two inter-related structures for different types of search tasks on the Web.

Other recent work in search results representation focuses on a single visualization (e.g. concept maps) that seeks to represent both hierarchies and networks ([3], [2], [10]) to support information seeking and finding. However, the focus of this research was on comprehension of the representations through a quantitative study. Our focus is on understanding how hierarchical versus network representations support different types of search tasks.

Most similar to our work, Sarrafzadeh et al. [31] investigated the effects of combining a knowledge graph with textual documents. Their goal, like ours, was to understand user behaviour with respect to different search tasks. They argue that a hybrid approach that combines the coherent content of text with the organized structure of graphs may better support information finding and sense making. They conclude that utilizing graphs of concepts and relationships which are derived from documents can be effective for finding relevant information when the information need is well defined. Their findings also demonstrate that providing meaningful relations that explain how different entities of a domain are connected are crucial for supporting more complex search task. Our work broadens this work by looking specifically at the contrast between hierarchical representations (e.g. trees) and network representations (e.g. knowledge graphs).

3. APPLICATION DESIGN

One challenge with any application that presents search results from an exploratory query to users is that the goal is rarely a static representation of the content returned by a user's query. Instead, the goal is to develop an interface that allows a user to interact with the content, to filter and select specific content, essentially to explore the information returned. As a result, the representation is linked to the interface that contains it and supports manipulation and exploration of it [13]. To develop an interface for exploratory search that would allow us to explore the characteristics of hierarchical and network visualizations, we engaged in an iterative process using a series of walkthoughs, thinkalouds, and pilot studies.

3.1 Prototype Development

To develop our representations and interface, we began with a low-fidelity design, where paper prototypes were used to explore user perception of representations of data and user interaction with those representations. We initially designed two low fidelity interfaces. The first interface employed a graph structure in which the entities from each article were the nodes and the sentences describing a semantic relation between them were the edges. In order to investigate how users navigate through large graphs to find information, our knowledge-graph-based prototype was designed such that the user would start from the overview page that contained all the nodes that had a high number of connections to other nodes in the graph. These nodes could be considered as representatives of different components of the graph and would help distribute different sub-graphs into different pages. The user was able to expand any of the nodes on the overview page and would proceed to a new page that contained the selected node and all the nodes that had a link to this node. The user could expand a new node on this page or collapse the expanded node and go back to the previous page.

The second interface utilized a hierarchy (or a tree) structure to organize headings and sub-headings of the articles, as observed in each page's table-of-contents. Each tree was in a collapsed format initially and the user would expand and collapse nodes to drill down into document content. We also created an overview node that linked all the trees in our collection. Interfaces were seeded with data gleaned from Wikipedia pages on Canadian capital cities.

We conducted a thinkaloud study to evaluate our paper prototypes with six participants (two female) to gather a set of features required for these interfaces. Data was presented in both graph and tree form to each participant. We asked participants to think aloud about what the data represented and how they would interact with the data. We also collected qualitative data on different use cases of these interfaces with respect to different search tasks. From this initial study, we redesigned our interfaces.

3.2 Final Design

The qualitative data and participants' feedback helped refine both the design of our search result presentation and our interface for manipulating the representation of the search results. We used force and pack layouts (as part of the D3 library[1]) to visualize the graphs and trees respectively.

When the user launches the graph-based application (Figure 1, top), they are presented with a knowledge graph containing labelled nodes and unlabelled links between nodes. Nodes that represent entities with low frequency are hidden in the initial view, and only appear once a higher-frequency, connected node is clicked. This ensures that the graph does not become too cluttered. Once the user clicks on a node, that node and all connected nodes are highlighted, while the remainder of the graph is alpha-blended into the background. By hovering over any connected node in highlighted portion of the graph, the user can see the relationship(s) between the two nodes in the snippet window located on the left side of the interface (Figure 1. top). For each relationship in the snippet region, participants have a link that allows them to view the corresponding Wikipedia article.

The tree interface is shown in Figure 1, bottom. When the user launches the application, the user is presented with a fully expanded tree. By clicking on any node within the tree, that portion of the wikipedia document corresponding to the node is presented in the preview area at the left of the interface. Under the snippet in question, there is a link to view article, allowing users to access the article in question.

4. EXPERIMENTAL DESIGN

To detail our experimental design, we first discuss the data extraction method that we used to populate our interface with data. Next, we present the tasks in our study and describe our participant population. Finally, we describe the data we capture from each participant.

4.1 Data Extraction

To populate our interactive applications, we created two distinct data sets: one focusing on history and the second on global politics. For the history data set, we used the previous

[1] http://d3js.org/

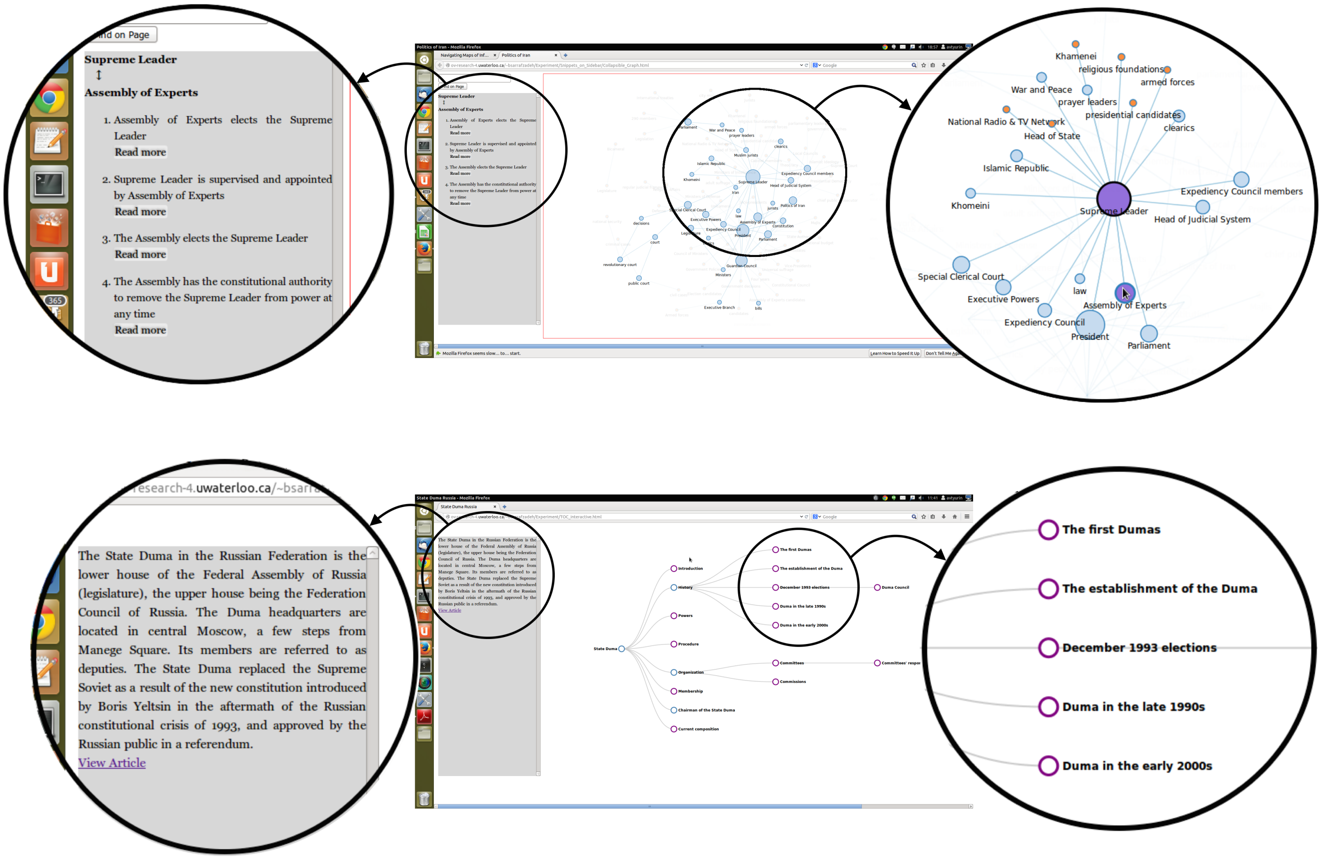

Figure 1: The graph and tree visualization interfaces. Note the callouts of nodes and document snippets.

search task exploring former capital cities of Canada. For the politics search task, we created a data set representing governmental structures in Iran and Russia.

To create this data set, we first collected a set of Wikipedia articles by querying the Web using a popular search engine. We retrieved the top 10 articles in Wikipedia based on their relevance to three queries corresponding to three topics: "Former Capital Cities of Canada", "Political System of Iran" and "Political System of Russia".

To create our knowledge graph, we designed an Open Information Extraction system that processes a text collection and generates (entity-relation-entity) triples [30]. This module is implemented in four phases. During the first phase we create the input corpus by collecting retrieved documents based on a given query. Next, we extract entities from text using state-of-the-art entity taggers [2]. We then select the sentences that contain at least two entities in them and parse them using Stanford Dependency Parser. For each sentence, we extract meaningful relations between the entities by finding the shortest path in the corresponding parse tree. For example we extract *The Constitutional Act divided the Province of Quebec into Upper and Lower Canada* as a relationship between the entities *Constitutional Act* and *Upper Canada*. We constructed a set of patterns based on dependency triples that lead to semantically meaningful relations. In the final phase we generate labels for the extracted relations and rank them based on relevance to the query and the informativeness of the extraction. Once the knowledge graph is generated, we hand-tune some aspects of the graph by correcting minor errors caused by the extraction of entities and relations.

[2]https://cogcomp.cs.illinois.edu/page/software_view/NETagger

For the tree based interface, we extracted the Tables of Content (TOCs) embedded in each Wikipedia article. We then manually extended the table-of-contents by adding subheadings to each section in order to provide a richer structure for the trees. Overall, our goal was to create visualizations that could realistically be created by computer algorithms while ensuring equivalent, high-quality for each of the generated visualizations.

4.2 Search Tasks

We noted earlier that researchers have defined search queries as simple or complex. With respect to the complexity level, each participant performed one Simple (i.e. question answering) and one Complex (i.e. essay writing) task. We also used two different topics (i.e. History and Politics) to investigate the relation between the topic and content knowledge with the structure used to organize the retrieved information. The queries we asked people to find information to satisfy in our study were the following:

Simple Politics: What governmental body or bodies are involved in the impeachment of the President of Iran and of Russia?

Complex Politics: Imagine you are a high school student who is going to write an essay on the Political Systems of Iran and Russia. Knowing little about the presidents of these two countries, you wish to determine which president has more power. Find at least 3 arguments to justify your answer."

Simple History: As a result of which act were Upper and Lower Canada formed?

Complex History: Imagine you are a high school student who is going to write an essay on the History of Canada. Knowing little about Canadian History, you wish to know which cities have served as a capital for Canada. You would also like to understand the reasons behind moving the capital from one city to another.

To assess the study design, we piloted with four participants. The pilot ensured that the usability of the system was sufficient to support interaction and provided guidance on the semi-structured interview to collect qualitative data on distinctions and use cases of the designed interfaces.

To limit study length and ensure coverage of simple and complex queries within subjects, our final study design was a $2 \times 2 \times 2$ [interface, interface-topic, topic-complexity] mixed design with interface as a within subjects factor, topic to interface assignment and complexity to topic assignment as between subject factors. This resulted in eight different groupings of participants. Each participant saw both interfaces. In the first interface, they had either politics or history, with the other topic in the second interface. For the two topics, each participant saw a complex query on one topic and a simple query on the other. In order to control for order effects, we rotated the order in which the tasks and the interfaces were assigned to the participants. That is, participants were randomly divided into 8 groups.

4.3 Participants

Once the study design was final, we recruited 26 (13 female) participants from different areas of Science, Math and Engineering for this study, all of whom use the Internet on a regular basis to search for information.

4.4 Procedure and Data Collection

The study proceeded as follows. After a brief introduction to the study, participants were given an initial questionnaire that evaluated their knowledge of the first query's topic. Participants were then presented with their first interface, were given an introduction to the features of the interface demonstrating how each feature of the interface worked, and were then given some time to manipulate the interface.

Once participants had developed some comfort with the features of the interface (approximately three minutes), participants were given the query and told to manipulate the interface as if they did not know the answer to the query and wished to locate it. To capture data on participants' actions, participants were asked to "think aloud" during each task and share their thoughts and strategies with the researcher. For both tasks, the participants were given 15 minutes and were required to find relevant information by providing a reference sentence or sentences from the interface or document collection to justify their arguments or answers. The need to find specific information ensured that each participant manipulated the interface to find relevant information.

After providing an answer to the query, participants completed a post-task questionnaire that evaluated the experience they just had. We used questionnaires provided by TREC-9 Interactive Searching track [3] and modified them to fit into our experiment design. At the end of each task, via a semi-structured interview participants were asked to reflect on their experience with using the assigned interface for performing the assigned task. They were encouraged to think about the conceptual usability of the type of structure utilized for information organization as well as the technical usability of the application. At the end of the second task a semi-structured interview format was used to elicit comparison between the two interfaces with respect to the different types of search tasks and to reflect on the design of an "ideal" interface that could support them more efficiently. Participants received a \$10 incentive for their participation.

Data was captured in a variety of ways. Each interface was instrumented with a logger which monitored participants during the search sessions. Both movement on the computer screen and participants' interactions with the system were captured. Interactions we collected included clicking on nodes or edges, reading snippets, viewing articles, and the time they spent reading the articles. The activity logs for two of the participants were corrupted so we excluded their data from our activity log analysis. Experimental blocks and a post-task semi-structured interview were audio recorded. Finally, two assessors evaluated the quality of answers provided by the participants for each of the search tasks independently. Simple queries were rated as either correct or incorrect. To receive a correct rating, both answer and referenced document section were required to be correct. Complex questions were rated on a scale. Scores for all queries were normalized to reflect a value in the range [0, 1]. Inter-assessor reliability was evaluated using Pearson coefficient and an overall value of 0.8 for simple queries and 0.9 for complex queries was found.

4.5 Analysis

Log and Questionnaire Data were extracted and analyzed for statistical patterns. Interviews were all transcribed and analyzed using affinity diagramming. Data were clustered collaboratively by two researchers using open coding, producing 14 clusters. Clusters were then analyzed using axial coding to identify overall themes in the data.

We found that saturation for our qualitative data occurred approximately mid-way through our participant sample – after 14 participants, no new clusters of information were identified. However, to allow statistical analysis of log data, we continued to cluster interview data for the remaining participants, particularly attuned to data that might expand our clusters or add nuance to our analysis.

5. MAIN OBSERVATIONS

In this section, we present an analysis of data collected during the study. We first present some numerical data collected from search logs which provides a broad overview of participants' contrasting behaviours given different interfaces and given queries of differing complexity. Next, we present the results of our qualitative analysis, clustered into four broad themes: Biasing Factors, Task Effects, Data Relations, and Problem Solving Approach.

[3] www-nlpir.nist.gov/projects/t9i/qforms.html

5.1 Validating Search Tasks

In any study where the goal is to explore search result representations for exploratory search, one concern is whether or not the search tasks are representative of exploratory search tasks. In our task design, we were guided by Marchionini's work on exploratory search [20]. Leveraging two task domains, politics and history, we created one look-up task and one exploratory search task within each domain using Marchionini's definitions, yielding four tasks overall. The politics tasks asked participants to compare two different governmental structures, Iran and Russia, rationalizing and providing citations for answers they provide. Similarly, the history tasks asked participants to discover something about the history of Canada and, again, rationalize and provide citations for their answers. For our complex tasks, in particular, we argue that the tasks combine aspects of knowledge acquisition or comparison (the *learn* subcategory of exploratory search) with analysis, synthesis, and evaluation (the *investigate* subcategory).

Another concern is whether the actual topics are of sufficiently similar complexity that topic effects do not overwhelm other factors in our results. To address this, beyond ensuring counterbalancing of topics, we analyzed topic effects vis a vis our dependent variables to determine whether either the history or politics task resulted in statistically significantly varying behaviours. Interestingly, our *look-up tasks* in both history and politics, where participants returned a factoid, differed in quality of answers, time reading, and document views ($F_{1,24} = 6.02, p < 0.05$ for quality; $F_{1,24} = 6.00, p < 0.05$ for reading; and $F_{1,24} = 21.22, p < 0.001$ for document views). However, for our *exploratory tasks*, i.e. our complex tasks where participants were asked to learn or investigate, scores did not differ significantly between the two topics areas of history and politics ($p > 0.05$ in all cases). Because our primary interest is supporting exploratory search, we argue that our complex tasks are of sufficiently similar complexity as to limit topic effects.

Finally, alongside care designing our search tasks and a analysis of topic effects on dependent numerical measures, we also examined our qualitative data to determine whether participants found the tasks to be aligned with their conceptualization of exploratory search. The comments made by our participants when they were presented by the tasks descriptions indicated that these tasks were indeed complex, i.e. that they were ambiguous and open ended in nature. As well, different participants interpreted the task descriptions differently and came up with different strategies based upon their interpretation, further validating the open-ended, exploratory nature of the search tasks.

5.2 Log Data Analysis

As noted earlier, our data logged all user action with the system. Of particular interest to us was information on the scoring of participant responses, the number of nodes clicked in each interface, the number of documents read, and the amount of time spent reading documents. Table 1 summarizes this data. The Mark column contains scoring of participant responses. Clicks is a count of the number of nodes clicked on. Views is the number of instances when a participants used the interface to view the actual Wikipedia document (as opposed to relying on the information contained in the interface). Finally, ViewTime is the amount of time in seconds spent reading documents (as opposed to manipulating the interface).

We performed a repeated measures ANOVA with interface (tree versus graph) as a within subject effect and query complexity as a between subjects effect. Dependent variables were scoring of query results, clicks with the interface, number of document views, and time spent viewing documents. Overall, RM-ANOVA indicated that interface had a statistically significant effect on dependent variables ($F_{4,20} = 5.83, p < 0.01, \eta^2 = 0.54$). Query complexity was not significant, nor was there any interaction between complexity and interface. Univariate tests of dependent variables with respect to interface (tree versus graph) show statistically significant effect on number of document views ($F_{1,23} = 26.29, p < 0.001, \eta^2 = 0.53$) and on time spent viewing documents ($F_{1,23} = 6.01, p < 0.05, \eta^2 = 0.21$). Marks and Clicks were not significant.

	Marks	Clicks	Views	ViewTime
Graph	0.74 (0.27)	18.7 (3.2)	1.6 (0.43)	131 (37)
Tree	0.43 (0.04)	17.9 (2.2)	4.9 (0.49)	1228 (444)

Table 1: Mean (Standard Deviation) values for marks (average independent evaluator scores), clicks on nodes, document views, and document view time.

Overall, our data indicate that the knowledge-graph visualization allows participants to glean more information from the data structure (67% fewer document views, on average) in less time (almost 90% less time reading documents). The knowledge graph is designed to represent the information in the document in a way that obviates the need to read extensively, and it was very successful at accomplishing this. Over half of all participants examined either 0 or 1 documents while using the knowledge graph (mean of 1.6 documents), whereas all except one participant examined at least three documents with the tree structure (mean of 4.9 documents). Qualitatively, we note that the knowledge graph also fared better in score, though not statistically significantly better. As well, the workload in both documents (as measured by node clicks) was very similar (18.7 versus 17.9 clicks per query on average).

5.3 Qualitative Data

Given the statistical advantage enjoyed by the graph representation, the next question we wished to explore involved participant perspectives on each of these representations of search results. How did they differ? What were the advantages and disadvantages of each from a user perspective? We present four themes arising from our qualitative data analysis in this section.

5.3.1 Biasing Factor: A Willingness to Explore

Exploratory behavior, defined by the National Library of Medicine as "the tendency to explore or investigate a novel environment", is driven by curiosity and is evident in most exploratory searches. Both lookup and exploratory searches use curiosity in their search models, though the actual curiosity which drives each type of search is slightly different [4]. *Specific curiosity* is the desire for a particular piece of information, as typified by an attempt to solve a problem or puzzle. *Diversive curiosity* is a more general seeking of stimulation or novelty, for example a television viewer flipping between channels. In information seeking, specific curiosity corresponds to well-defined goals and directed searching, while diversive curiosity corresponds to ill-defined goals and exploratory browsing [24].

In our thinkaloud data and in our follow-up interview data, we identified specific versus diversive curiosity as a factor that influenced participants' perceptions of each web interface. Essentially, some participants preferred an interface over the other based on the amount of time they were willing to spend in exploratory browsing. Linking to specific curiosity, if an interface is effective in accomplishing a search task but required extensive time browsing, the participant would rather use a different interface. Participants patience with the search task was influenced by the tension between the drive to solve the problem (specific curiosity) versus the tolerance for browsing (diversive curiosity):

"For specific questions, it depends on how much time I'm willing to spend. If I have more time I'd like the tree, because it's more scattered and I can learn more objectively." [P4]

"So I feel like the Tree would be good if I wanted to sit down and spend time reading about a topic and I wasn't looking for something specific. Whereas if I was looking for something very specific, for that, I think I would like the other one [graph] better. Cause it was already doing the keyword search and it was easier to pick out things." [P8]

"If I need a fast way, I go to the graph. I use the tree only when I'm learning deep about a new domain." [P4]

This observation is in line with the initial work on information foraging; Pirolli and Card [25] defined the profitability of an information source "as the value of information gained per unit cost of processing the source." Cost is defined in terms of time spent, resources utilized and opportunities lost when pursuing a search strategy instead of others [28]. We find that diversive curiosity biases toward the tree structure, whereas specific curiosity biases toward the graph.

5.3.2 Task Effects: Finding Versus Learning

The Web has provided the opportunity to browse and navigate through an extensive information space. However, beyond simply finding basic answers, web searchers also engage in learning and discovery [20].

As noted in our study design, we incorporate two tasks with different levels of complexity. Given these two levels of complexity, in post-experiment interviews the participants were able to compare the two interfaces based on the specificity of the information they were looking for. Interestingly, however, participants were divided on which interface was better for simple versus complex search tasks.

Overall, most participants found the graph interface more practical for finding specific information and simple question answering tasks. Both question-answering and keyword-based tasks were typically perceived of as advantaging the graph structure:

For the question-answering task I'd rather use the graph. Because I want to know exactly if this word is linked to that word. If there are two words appear in the same sentence you can quickly find an answer and I don't have to read the whole article. [P9]

When I was searching for specific keywords, with the tree interface I actually had to go to the article itself to search. so it wasn't useful. Whereas the other one [graph] actually gives me access to the keywords. [P14]

To learn I think the hierarchy interface is good if I want to learn say about history of Canada, because then you start from step 1 and the you go to the next level. [P10]

This is not to say that our participants were universal in their beliefs about data visualizations. Some participants found that the tree was significantly better for finding a specific piece of information. P2, P3, and P6 all articulated variants of this belief:

But when you are trying to find a specific answer to a question, then the tree structure is good, because it helps you traverse from the root to a leaf node. [P2]

I like the tree better for specific questions. It categorizes things better.[P3]

Tree is pretty good for finding exact information. [P6]

To try to understand this phenomenon better, we looked at other demographic data collected from our participants, and found an intersection between the belief that a tree was better suited to search tasks and our participants self-rated prior knowledge of the topic being examined. Participants were biased toward a tree structure for broad learning of the task domain particularly when they had low prior knowledge. This result seems to replicate findings by Amadieu et al. [3] on the use of network structures versus hierarchical structures in the education domain, i.e., that low knowledge learners benefited from hierarchical structures in free recall performance and exhibited reduced disorientation, whereas high knowledge learners performed better and followed a more coherent reading sequence given a network structure. Participants, too, noted this phenomenon:

So if you are an expert in a domain, you want the view very focused [knowledge graph]. But if you don't know much about a domain, you want to see an overview first [tree]. [P6]

5.3.3 Data Relations: Derivative Versus Multifaceted, Local Versus Global

Visualizations of structures, i.e. of entities and relationships, inherent in large data sets can help users understand the structure of the data and make information more accessible. However, participants may perceive a domain to have a derivative/hierarchical structure or a multi-faceted structure. If the representation of search results mimics that perceived structure, participants prefer that structure:

If you are searching for something that is already structured and we already know the names of these categories, then the tree is helpful. [P2]

For the [tree] interface, if I was using it for a topic like Geography, then I'm looking for continents, countries, cities, states, capitals, ... Then I know the headings and then I know which path to take.[P3]

I think [tree] would also be useful if you had some sort of notion of how things are laid out, like if there was a chronological order. Yeah, if that was chronological that would be nice cause you could gauge where you needed to click. Like you saw something that was really-really previous and two nodes down to find something more recent, even if you don't know exactly which one.[P8]

To clarify, the data relations in question are those perceived to be salient *by the participant*. If salient relationships are viewed as derivative or hierarchical (e.g. 'is-a' relationships), then a tree can best capture this view of data, whereas if salient relationships are more heterogeneous and resist structure as a hierarchy, that advantages the graph-based representation.

Beyond the specific relationships between entities, another theme that appeared in our data involved the scope of information required to satisfy a query. The tree structure

seems to provide a comprehensive overview for the information space. Even if we provide groupings and overviews for our graphs, the graph interface best serves exploration at the entity-relationship level. As a result, several participants liked the tree structure for cases where they needed a comprehensive overview of the domain:
If I'm learning about a new domain, in the case that I want to cover the entire domain and get a general understanding of everything but at the surface, I'd like the tree. [P7]

5.3.4 Problem Solving Approach: Depth-First Versus Breadth-First

According to Brown [6] information seeking is a goal-driven activity in which needs are satisfied through problem solving. This view is comparable to Wilson's model of information seeking [38], which considers information seeking as a problem-solving process with the goal of reducing uncertainty about the information being searched.
... if you have a large amount of data then you're kind of confused, you don't know which part to look at, which connection to look at. ... I would use hierarchy to get an idea of how everything is organized and then maybe I go and try to dig in more, find out the relations between terms.
- For digging more would you use the graph?
- Yes. But again, even in the graph, it should be the specific, the focused one. Not the whole thing. [P21]

In unpacking this quote, we note that the process of directly addressing a problem is essentially a depth-first process where the knowledge graph allows a focused exploration of a region. On the other hand, with confusion the breadth-first or tree-based exploration is beneficial as it allows the user to iteratively reduce confusion, obtain an overview, and only slowly exploit detail. Many participants indicated similar concepts of confusion, nervousness, or inadequacy as a rationale for their preference for the tree structure:
[The graph interface] is not friendly. Too many things! [P2]
Because I get frustrated jumping from one node to the other for a while and don't get any information I want. ... If the graph is too big, I get scared of it! ... Too many things, so I don't know where to look [P26]
when the graph is too big, I don't know where to look ... and I don't also know where to start. Because I'm not familiar with most of the information .. the Councils, the positions, the names, ... So I don't know where to start. [P26]

More generally, many research domains argue for overview and structuring of content to permit sense-making and reduce confusion. Information visualization is founded on the techniques people use to structure and cluster visual stimuli (see, for example, [34], Chapter 1). Problem solving research in psychology connects aspects of visual perception and structuring of content to comprehension [26]. Designing for visually impaired readers argues for well-structured hierarchical content to allow more rapid sense-making [14], even in the absence of vision. Essentially, overviews are invaluable when people feel a need to orient themselves within data.

Alongside confusion and the need to orient ones-self within a domain, one's problem-solving strategies may bias behaviour. Research into problem solving strategies has a long history in the psychology domain. One well-known characterization of *coping* or problem-solving strategies identifies two groups of individuals: Problem-focused and Emotion-focused [18]. Problem-focused individuals tend to directly address a problem, whereas those with an emotion-focused strategy seek to reduce the effects of the problem. In web search, Kim [17] found that problem-solving style had some impact on navigational patterns. Emotion-focused subjects traversed several layers of nodes before returning to the starting page (i.e. a depth-first navigation), whereas problem-focused subjects spent more time checking nodes available in the same level (i.e. a breadth-first navigation). Acknowledging the lack of personality-testing in our questionnaires [17], the link between confusion, nervousness, or fear and a desire for a hierarchical structure that allows depth-first exploration may merit further inquiry.

6. DISCUSSION

6.1 Understanding Tree Versus Graph Visualizations

As we note in the introduction, our goal with this research was to explore the differences between graph and tree visualization, specifically to understand their similarities and differences with respect to the search process. Our results explore these differences, triangulating both quantitative data from log files and qualitative data from participant interviews to understand how search result representations influence search behaviour.

From our log data, we note that the hierarchical structures in our study serve as pointers to passages in a document due to their similarity to tables-of-contents. Essentially, they simplify the process of locating topics, but the monotonic relationship that they represent – for example an is-a relationship – limits the information they can represent. The end result is that hierarchies result in a greater need to read the document rather than find the information contained within the visualization, shown, in our log data, by more instances of reading documents, and a longer period of time reading documents. Specifically, participants read documents three times more frequently and spent almost ten times more time reading. Our data also highlights the advantages and disadvantages of gleaning information from a knowledge-graph versus finding the relationship within source material and generating an abstract version of the knowledge graph for oneself. However, it is also clear that one representation is not better than the other in any subjective sense. Many of our participants expressed a need for combining both interfaces into one interface which enables switching between a global and a local view of the information space.

6.2 Design Implications

When designing search interfaces, the process of creating a view of search results remains challenging. Information visualization tools such as the InfoVis Toolkit[4], SpotFire[5] and InfoZoom[6] typically support multiple representations of search results. Our work does not dispute the accepted practice of recognizing that heterogeneous, interactive visualizations are the best way to allow exploration of a data set generated by search queries.

Our study highlights the complementary nature of hierarchical structures and knowledge graphs as representations of data. Our data indicates that hierarchies allow a more grad-

[4]http://sourceforge.net/projects/ivtk/
[5]http://spotfire.tibco.com/
[6]http://www.softlakesolutions.com/

ual depiction of and immersion into the domain, essentially fostering sense-making of the overall content (see Section 5.3.3). On the other hand, participants note that graphs are “more engaging” (P4), yield “more control over exploration” (P8), or are “similar to my mind” (P16). This, then begs the question of whether hierarchies and knowledge-graphs could be combined, but the challenge with combining hierarchical structures with knowledge-graphs is that hierarchies represent topics within a corpus, whereas nodes in a knowledge graph represent entities and their relations. Any one entity in a knowledge graph can map onto several topics in a hierarchy: For example, a political figure or governmental structure (e.g. the Guardian Council) or a historical event (e.g. the War of 1812) may be mentioned in all topics in a hierarchy, depending on how pervasive that entity is to the overall corpus.

On the other hand, within our knowledge graphs, nodes have different prominence based upon the number of edges they connect to. Entities that are more pervasive in the document have more connections and, hence, can be assumed to be more central to the topic. One alternative to hierarchical structures is to consider central entities within a knowledge graph, those entities that have a higher connectivity to the graph. By setting thresholds, one might be able to structure a multi-level view of a knowledge graph around central entities. We are exploring this option as one way to effectively combine the advantages of knowledge graphs and hierarchies into a single view. Rather than breaking down topics or concepts as in our tree view or in concept maps, the multi-level view of knowledge graphs focusing on central entities simply introduces information seekers to those entities or objects most central to a retrieved corpus.

6.3 Limitations

In designing any study, compromises must be made. In this section, we discuss three potential confounds: interface effects versus information representations; the effect that hand-tuning may have had on results; and the generalizability of our results given corpus size and topic/task selection. In this section, we address each of these concerns.

Any time one conducts a user study comparing two artifacts, it is always possible to bias the study through selective design. A poor user interface or poor interaction design can disadvantage one experimental option, leading to biased results. To limit this confound, we conducted multiple rounds of pilot studies and made modification to ensure that each representation was sufficiently rich that participants could perform a significant portion of the information seeking task within the visualization. In analyzing our data, we found that participants in our study data indicated no dissatisfaction with the interaction within the visualizations, and, instead, focused on the visualizations themselves. Even on probing during de-briefing interviews, participants would frequently discuss the advantages and disadvantages of knowledge representations (hierarchies versus graphs) when asked to comment on each interface.

A second concern revolves around the ecological validity of our results, particularly in light of hand tuning. As we noted in our experimental design section, we used automated algorithms to generate knowledge graphs [30] and extracted hierarchies from tables-of-contents or headings within documents. However, we then performed some refinement of the hierarchies (adding low-level sectioning to documents) and knowledge graphs (mainly refinement of coreferencing). We address this point in two ways. First, arguably, to ensure that confounds are *not* present in our results, hand-tuning (or at least manual verification) is essential. Otherwise, error-prone algorithms and poorly structured data could influence the effectiveness of any individual representation of search results, focusing the data around the algorithmic failures as opposed to the nature of hierarchies versus graphs. Second, it is important to note that the manual refinement we performed was very limited. In hierarchies, we created a richer set of leaf nodes, but did not modify the overall structured content of the document; in knowledge-graphs, a small set of entities (less than 10%) needed to be combined when coreference resolution failed. As research in automatic summarization and coreference resolution continues, these problems will hopefully be addressed by researchers working in natural language processing.

Finally, task and corpus has been a concern in past iterations of this paper. Our tasks and topic effects are discussed in Section 5.1. We argue that the 10 most relevant documents from Wikipedia represents a set of documents similar to the number explored in real-world web searching tasks. First, while web searches return more results, work on information seeking argues that the effective size of a *relevant* document set for web search results is significantly smaller that all documents returned – on the order of six documents – hence the importance of ranking algorithms in information retrieval [16] [15]. Second, not every retrieved document is directly relevant to any specific information seeking task. A user may look within any individual document within a set of top ranked documents, and he or she may also combine information from multiple sources to satisfy his or her information needs.

7. CONCLUSION

In this paper, we present the results of a study evaluating knowledge graphs and trees as spatial representations of Web search results. Our analysis includes both log data gleaned from participant interactions with data representations and qualitative interview data gleaned from thinkalouds and semi-structured interviews. Overall, we find that knowledge graphs are effective in capturing the entities and relationships in a corpus in a way that reduces participant reliance on actual retrieved documents, i.e. participants viewed significantly fewer documents for significantly less time. As well, the quality of participant responses to pre-specified queries (a measure of how effective visualizations are at representing data) was statistically unaffected by representation. Finally, from the perspective of our participants, we find that tree-based representations are better suited to learning, provide better overviews of a domain, and are more approachable for participants who are confused. Graphs, in contrast, work best for directly seeking answers, and appear to be a more playful mechanism for exploring the details of individual entities and their relationships.

8. REFERENCES

[1] J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012. In *ACM SIGIR Forum*, volume 46, pages 2–32. ACM, 2012.

[2] F. Amadieu and L. Salmerón. Concept maps for comprehension and navigation of hypertexts. In

Digital Knowledge Maps in Education, pages 41–59. Springer, 2014.

[3] F. Amadieu, A. Tricot, and C. Mariné. Interaction between prior knowledge and concept-map structure on hypertext comprehension, coherence of reading orders and disorientation. *Interacting with computers*, 22(2):88–97, 2010.

[4] D. E. Berlyne. Conflict, arousal, and curiosity. 1960.

[5] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.

[6] M. E. Brown. A general model of information-seeking behavior. In *Proceedings of the ASIS Annual Meeting*, volume 28, pages 9–14. ERIC, 1991.

[7] K. BystrÖm. Information and information sources in tasks of varying complexity. *Journal of the American Society for information Science and Technology*, 53(7):581–591, 2002.

[8] D. J. Campbell. Task complexity: A review and analysis. *Academy of management review*, 13(1):40–52, 1988.

[9] R. Capra, G. Marchionini, J. S. Oh, F. Stutzman, and Y. Zhang. Effects of structure and interaction style on distinct search tasks. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 442–451. ACM, 2007.

[10] M. Carnot, P. Feltovich, R. Hoffman, J. Feltovich, and J. Novak. A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. 2003.

[11] B. Dervin. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management*, 2(2):36–46, 1998.

[12] N. Ducheneaut and V. Bellotti. E-mail as habitat: an exploration of embedded personal information management. *interactions*, 8(5):30–38, 2001.

[13] M. A. Hearst. Clustering versus faceted categories for information exploration. *CACM*, 49(4):59–61, 2006.

[14] M. Hersh and M. A. Johnson. *Assistive technology for visually impaired and blind people*. Springer Science & Business Media, 2010.

[15] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proc. of SIGIR*, pages 154–161. ACM, 2005.

[16] D. Kelly and L. Azzopardi. How many results per page? a study of serp size, search behavior and user experience. 2015.

[17] K.-S. Kim. Searching the web: Effects of problem solving style on information-seeking behavior. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, volume 1999, pages 1541–1542, 1999.

[18] R. S. Lazarus and S. Folkman. Stress. *Appraisal, and coping*, 725, 1984.

[19] H. Ltifi, M. Ben Ayed, A. M. Alimi, and S. Lepreux. Survey of information visualization techniques for exploitation in kdd. pages 218–225. IEEE, 2009.

[20] G. Marchionini. Exploratory search: from finding to understanding. *CACM*, 49(4):41–46, 2006.

[21] G. Marchionini and B. Shneiderman. 3.1 finding facts vs. browsing knowledge in hypertext systems. *Sparks of innovation in human-computer interaction*, page 103, 1993.

[22] K. L. Norman and J. P. Chin. The effect of tree structure on search in a hierarchical menu selection system. *Behaviour & Information Technology*, 7(1):51–65, 1988.

[23] L. R. Novick and S. M. Hurley. To matrix, network, or hierarchy: That is the question. *Cognitive Psychology*, 42(2):158–216, 2001.

[24] S. Pace. A grounded theory of the flow experiences of web users. *International journal of human-computer studies*, 60(3):327–363, 2004.

[25] P. Pirolli and S. Card. Information foraging. *Psychological review*, 106(4):643, 1999.

[26] R. M. Ratwani, J. G. Trafton, and D. A. Boehm-Davis. Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied*, 14(1):36, 2008.

[27] D. E. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, pages 13–19. ACM, 2004.

[28] D. Russell, M. Stefik, P. Pirolli, and S. Card. The cost structure of sensemaking. In *Proc. of INTERACT'93 and CHI'93*, pages 269–276. ACM, 1993.

[29] F. Sanchez-Zamora and M. Llamas-Nistal. Adaptive concept maps: Issues on design and navigation. In *Concept Mapping: Connecting Educators. Proceedings of the Third International Conference on Concept Mapping. Tallinn, Estonia & Helsinki, Finland*, 2008.

[30] B. Sarrafzadeh, R. Guttikonda, K. Suleman, J. Thomas, and O. Vechtomova. Automatic discovery of related concepts. Technical report, March 2013.

[31] B. Sarrafzadeh, O. Vechtomova, and V. Jokic. Exploring knowledge graphs for exploratory search. In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 135–144. ACM, 2014.

[32] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[33] A. Singhal. Introducing the knowledge graph: things, not strings, 2012. *Official Blog (of Google).*

[34] C. Ware. *Information visualization: perception for design*. Elsevier, 2012.

[35] R. W. White and R. A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.

[36] B. M. Wildemuth and L. Freund. Assigning search tasks designed to elicit exploratory search behaviors. In *Proc. of HCIR 2012*, page 4. ACM, 2012.

[37] M. L. Wilson, B. Kules, B. Shneiderman, et al. From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends in Web Science*, 2(1):1–97, 2010.

[38] T. D. Wilson. Models in information behaviour research. *Journal of documentation*, 55(3):249–270, 1999.