Evaluating the Value of Lensing Wikipedia During the Information Seeking Process

ABSTRACT

While Wikipedia is an excellent source of information about entities, discovering relationships among them is not well-supported by its search features. Lensing Wikipedia was designed as an alternate search and summarization interface, providing a set of filtering, visualization, and exploration tools that enable searching among the connections between people, organizations, and locations based on the descriptions of activities in the text. In this paper, we present the results of a user study on how these features are used in each of Vakkari's stages of information seeking (pre-focus, focus formulation, and post-focus), and the participants perceptions of the utility of these features to their overall information-seeking goals. Participants primarily used the input and control features during the pre-focus stage, the informational features during the focus formulation stage, and the personalization features during the post-focus stage. We also found differences between the assigned search activities and those chosen by the participants. Findings from this study contribute to understanding how people use advanced search, summarization, and visualization tools to aid their information seeking tasks.

CCS Concepts

•Information systems → Search interfaces; •Human-centered computing → Empirical studies in HCI;

Keywords

Wikipedia search, User study, Information seeking

1. INTRODUCTION

Discovering relationships between entities within complex document collections such as Wikipedia can be a difficult and cognitively taxing activity. For example, suppose a student wishes to study the relationship between Charlemagne and Pope Adrian I for a class report. Doing so within a simple search interface, such as that provided by Wikipedia, requires that the searcher formulate their information needs and devise queries, evaluate the relevance of documents, view and evaluate the information within the selected documents, synthesize what they have learned, and prepare an assessment of what has been discovered. Lensing Wikipedia provides a series of search and summarization features that are designed to enable searchers to discover and explore among such relationships [28]. Following the taxonomy proposed by Wilson [36] these features can be classified as being input, control, informational, or personalization. The goal of this study is to assess how these different types of features are used during each of Vakkari's stages of information seeking [33] (pre-focus, focus formulation, and post-focus), how the participants perceive the usefulness of the features, and their overall satisfaction with Lensing Wikipedia.

Knowing which features are used during each of the stages of the information seeking process will not only contribute to a better understanding of how to support search within digital encyclopedias, but will also allow us to refine Lensing Wikipedia. In particular, features that are not useful may be removed in order to avoid interface complexity. Those features that are useful may then be re-ordered to more closely follow the information seeking stages. If we are able to detect the searcher changing from one stage to another, Lensing Wikipedia may be able to dynamically adjust the availability of features based on the type of task the searcher is expected to perform.

The remainder of this paper is organized as follows. Section 2 provides a review of the recent literature on searching within Wikipedia, as well as an overview of the information seeking literature that informed the design of this study. Section 3 provides brief explanation of the core features of Lensing Wikipedia. The user study methodology, design, analysis methods, and results are provided in Section 4. The paper concludes with a summary of the primary contributions and future work in Section 5.

2. LITERATURE REVIEW

2.1 Wikipedia Search and Exploration

The default search interface in Wikipedia matches the user-supplied query against the titles of the articles. For instance, a search for "Napoleon" will directly lead to the wiki/Napoleon page, but does not show alternatives such as wiki/Napoleon_III. The information in the text of the page itself is ignored in such a search interface. Although there is an advanced search feature that matches the contents of the page, it follows the traditional query box and search re-

CHIIR '17 Oslo, Norway

© 2017 ACM. ISBN 978-1-4503-2138-9…$15.00

DOI: 10.1145/1235

sults list paradigm, requiring the searchers to consider each search result one-by-one.

A visual approach to displaying the full text search results was adopted in [32], in which a visual clustering of Wikipedia articles was created using a vector-space model over the bag of words in the Wikipedia articles. While such an approach may help a searcher to identify groups of related documents, it does not help the searcher to identify the relationships among the entities in these clusters.

Apart from the titles of pages, there is a wealth of structured information in most Wikipedia pages in the form of infoboxes and Wikipedia Categories. The structured information in Wikipedia has been an important topic of research in the semantic web community and in the knowledge-base construction community. Two major projects in this area are DBPedia [4, 24, 34] and YAGO [30, 22, 16, 31], both of which create entity based knowledge bases from Wikipedia.

The information in infoboxes extracted in DBPedia lends itself to a faceted search interface such as the one that was explored by Hahn et al. [12]. A similar approach has leveraged user queries to support the facet construction [38]. Others have sought to take advantage of how information is structured within Wikipedia articles to enhance the search process. For example, Vispedia [8] allows searchers to select a table of data, apply a visualization technique to the data (e.g., timeline), and subsequently use that to enable interactive querying to supplement the visualization.

The geo-location of pages can also be exploited and combined with the YAGO database to provide map based browsing of Wikipedia articles [15]. It also provides graph-based visualizations based on typed links between entities in the YAGO knowledge base.

In all of these approaches, the information encoded within the text of the Wikipedia articles themselves are largely ignored. Our goal with Lensing Wikipedia is to use natural language processing to extract information from the textual descriptions about entities, which are grounded in a certain time and location, and expose this information to searchers through an interactive visual interface.

2.2 Information Seeking

Although some have been advocating for the importance of considering searchers' interactions with information retrieval systems as a central concern for many years [2], support for the study of interactive information retrieval has only recently made its way to the forefront of information retrieval research [3]. The goal among such research is to study how search systems can support people in fulfilling their information seeking tasks and goals. What has become clear is the importance that the search interface has in allowing searchers to be effective in the information seeking processes.

The importance of the search interface is exemplified by the inclusion of chapters on this topic within current information retrieval textbooks [14, 36], and books dedicated solely to the design and study of search interfaces [13]. These works generally focus on discussing the accepted models of information seeking, surveying current trends in search interface design, providing best-practices advice, and advocating for careful evaluations of new interfaces.

When designing new search interfaces, it is helpful to consider the type of interaction that the user needs to perform, and only add new interaction methods that support specific information seeking tasks. Wilson provided a useful taxonomy in which interactive features may be classified: *input*, *control*, *informational*, and *personalization* [36]. Input features are those that allow a searcher to provide input into the search process (e.g., query box, facets). Control features are those that allow a searcher to manipulate how the search results are shown (e.g., custom sort, filters). Informational features are those that present information to the searcher (e.g., search results list, visualizations). Personalization features are those that allow a searcher to build a view that is unique to their specific needs (e.g., rejecting irrelevant search results, saving results to evaluate later).

Building upon an exhaustive list of information seeking tasks that one may undertake when seeking to resolve an information seeking problem (initiation, selection, exploration, formulation, collection, and presentation [20]), Vakkari developed a three-stage model of information seeking [33]. The *pre-focus* stage deals with the initiation of the search and undertaking a broad exploration of the topic. In the *focus formulation* stage, searchers refine their goals and explore these in detail. During the *post-focus* stage, searchers collect, synthesize, and present the results of their search. While this model was developed in the context of fulfilling information seeking goals that are complex in nature and are performed over multiple search sessions, it has been used to study single-session search activities as well [17].

In this paper, we have drawn inspiration for the organization of our study from Huurdeman et al. [17]. They developed a custom search interface that included a number of lightweight interactive features designed to aid the searcher, and evaluated it in the context of situated work tasks that guide the searchers through Vakkari's three stages of information seeking. Measuring both active use (interface logging) and passive use (eye tracking), they found that certain types of interface features are more useful during certain stages, and less useful during others (e.g., control features are more useful during the early stages, whereas personalization features are more useful during the late stages).

3. LENSING WIKIPEDIA

We employ a "machine reading" approach to the creation of a knowledge base from Wikipedia natural language text. Previous work has focused on extracting semantic relations between entities [25] observed in a large number of instances by parsing the text. Some construct a rich ontology (e.g., NELL [7] and PROSPERA [26]), or simply deal with token-level facts (e.g. TextRunner [1, 23]). In all of these approaches, each semantic relation extracted has to occur frequently in the collection.

In our approach we extract information from every clause in every sentence in the data set. We aggregate information from multiple sentences, using a natural language processing approach commonly called *Semantic Role Labeling* (SRL) [10]. SRL is used to extract predicate argument structures from text to automatically populate a predicate-centric ontology for a domain. A predicate argument structure looks for a predicate (an action or a state commonly instantiated by a verb) and the arguments of the predicate (the entities, typically noun phrases). We use supervised machine learning to train a SRL parser.

We choose a subset of Wikipedia: events about human history because it contains facts about events, very few opinions, and almost no sentiment. The sentences describe who

did what to whom, when, and where, as represented in the predicate-argument structure. At the time of the crawl, English Wikipedia contained about 2600 URLs (from 1500 BCE to 2016 CE), which are natural language summaries of important events in each year or decade in human history. We crawled all these Wikipedia URLs and obtained about 41,000 events, with each event described by one to several sentences. The result is 83,000 predicate argument structures discovered through semantic role labeling. Using methods from [37] we combine the information extracted from the natural language text with other structured information in Wikipedia such as dates and geo-location information.

3.1 Semantic Roles

For SRL, we used supervised machine learning to train a parser capable of taking natural language sentences and returning predicate-argument structures [10, 21], using the human annotated corpus called the Proposition Bank (PropBank) [27].

For example, in the sentence *The House of York defeats the House of Lancaster* we identify *defeat* as the predicate with arguments *The House of York* (*arg0*) and *the House of Lancaster* (*arg1*), where (*arg0* and *arg1*) are the abstract semantic roles. These abstract roles are then converted into human readable types (*arg0:'entity victorious'* and *arg1:'entity defeated'*) using information in PropBank. This is done using machine learning of the mapping between abstract semantic role labels and verbose descriptions such as *'entity victorious'*. The verbose label depends on identifying the sense of the verb. For instance, 'arg0' for 'get' might be labeled as 'receiver', but the right label could also be 'instigator' for another sense of the verb (get across). We explored many machine learning approaches to solve this task, and our best result could find verbose labels with 92% accuracy [28] on unseen test data.

These verbose labels are crucial in our visualization because it provides a lightweight ontology (or useful types). For instance, we could search for all entities that match the "buyer" or "entity defeated" types. These predicate-centric types are often more useful than the types from Freebase [5] which need further disambiguation to be applied correctly to the entities and events in our data.

3.2 Entity Extraction

We also perform named entity extraction from text data. We have done this in two ways: using Wikipedia hyperlinks and named entity recognition (NER). Lensing Wikipedia focuses on *Person* and *Location* as named entity types. We used Stanford NER [9] to detect candidate entities along with their types. Each candidate entity is then verified by linking that entity mention to an article in Wikipedia. In addition to entities recognized by NER, we take advantage of hyperlinks embedded in the event's texts. Using NER and Wikipedia hyperlinks we identify mentions in the text with Wikipedia URLs and then verify the entity type (person, geographic location, or NIL). For this we use the Wikipedia categories associated with the URL. Wikipedia categories are organized as a hierarchical ontology but they are not anchored in general conceptual classes like entity types (e.g., person). We use the category labels and infoboxes to verify entity mentions as the correct entity type (person, location) (e.g., categories like *Category:y_births* where y is a year, are associated with entity type *person*).

3.3 Temporal and Geographical Identification

From the SRL output and the entity linked with Wikipedia URLs, we extract temporal information such as the date when the event occurred. For event geo-location, we used the information in the Wikipedia URLs to obtain latitude and longitude information. Having geo-location information, we are able to extract the current country(ies) where the event happened by reverse geocoding using the Google geocoding API [11].

3.4 Search Interface

To show the effectiveness of this knowledge base to represent the underlying data, we created an interactive search interface called *Lensing Wikipedia*. The system consists of five core interface elements, organized on separate tabs: facets, storyline, timeline, comparison, and map (see Figure 1). In addition, as the queries are dynamically generated using these features, a search results list is created consisting of information extracted from the Wikipedia articles. Clicking on any links within this list will open the associated Wikipedia page in a new browser tab, allowing the searcher to keep relevant documents open while returning to the Lensing Wikipedia interface. A free-form note-taking feature is provided to allow searchers to document information relevant to their search.

Within the facets tab, the information regarding the extracted entities and their roles in the history of events are provided as the primary mechanism for constructing a query. The faceted classification allows users to explore among a collection of information by interactively applying multiple filters. While faceted browsing has been applied for closed domains with structured data (e.g., Nobel prize winners, and recipes [29]), to our knowledge this is the first time faceted browsing has been implemented for an open domain data set where the facets are generated from unstructured data.

The storyline tab provides a representation of the relationships between a selected entity and other entities within the knowledge base. For example, if Napoleon is selected, all other entities of this same type (person) are depicted within a flow-based visualization. The interaction between two or more entities are identified using vertical bars.

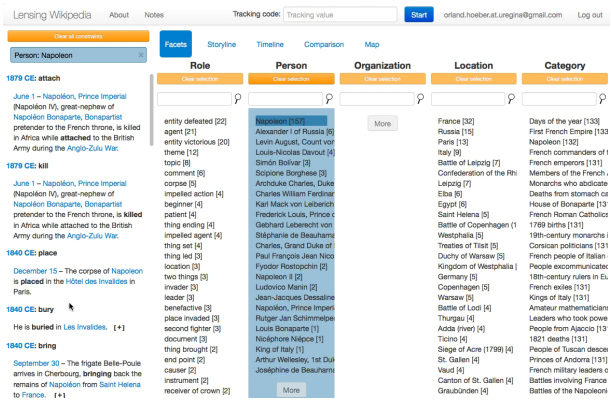
The timeline tab shows when in the full history of events the currently selected events have occurred. Using stacked views, the searcher can zoom into a specific temporal range, yet remain aware of when this exists in the full history of events in Wikipedia.

The comparison tab illustrates the number of events over time for the selected entity and other significant related entities. This can be used to show how the interaction between sets of entities overlap in time.

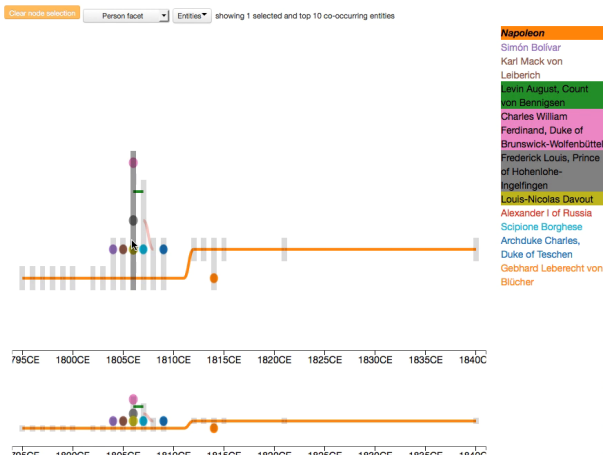
The map tab leverages the extracted geo-location data, situating the events on an interactive map. Variably sized discs are used to represent the number of events in the specific location. In order to show where the entities' events have occurred, numerical values are provided on the map.

4. USER STUDY

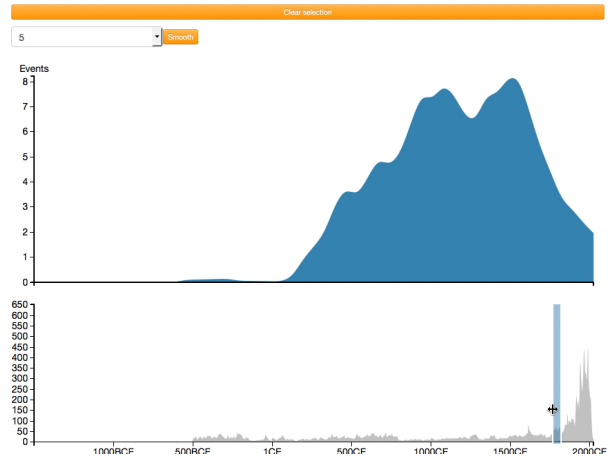
While a researcher may be well-intentioned in the design and development of novel features to support interactive information retrieval, the actual usefulness of such features remains unknown until the interface is properly evaluated.



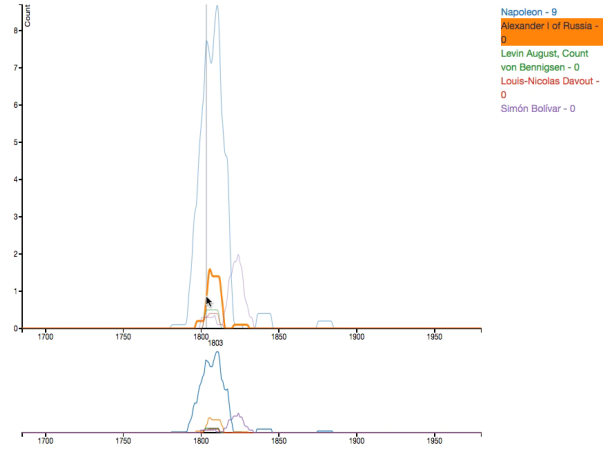
(a) Facet view



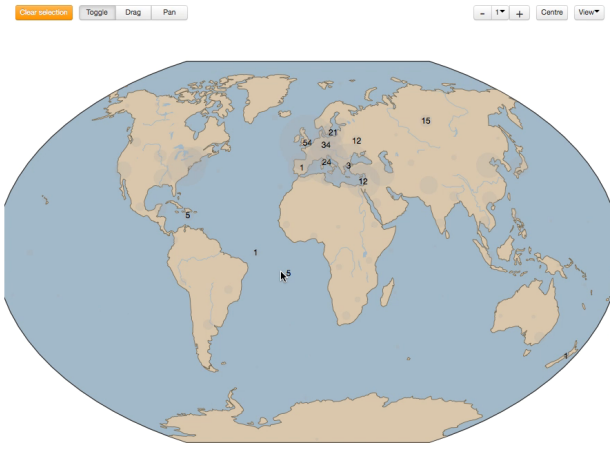
(b) Storyline view



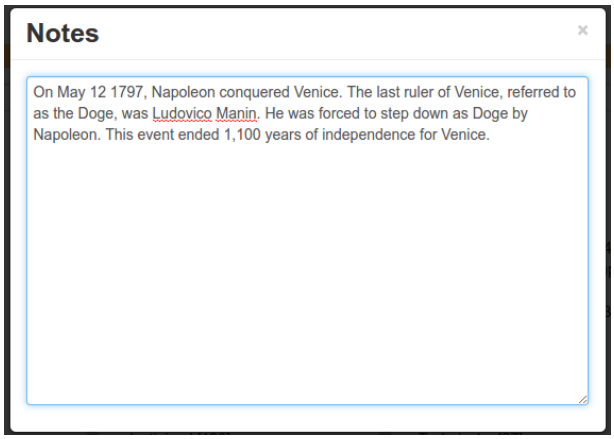
(c) Timeline view



(d) Comparison view



(e) Map view



(f) Notes input

Figure 1: Views available in Lensing Wikipedia.

Such evaluations will often take the form of user studies, where the goal is to study searchers' behaviours and experiences as they use the software, with particular focus on how they interact with the interface and the information provided [18]. Furthermore, the searchers' impressions of the usefulness of a search interface is a fundamentally important measure for evaluating information retrieval performance [3].

Studying a search interface in the context of the breadth of strategies, behaviours, and complex search activities [19] is necessary when the search interface is designed for general use. However, when the search interface is designed for a specific search context, the evaluation can be more focused.

In the case of Lensing Wikipedia, the interface was designed to support exploring among the relationships between entities, which is used as the motivating information seeking activities in the study design. By evaluating the software in the context for which it was designed, we can answer the question: does it support the intended tasks?

Rather than evaluating each feature of Lensing Wikipedia in isolation, we take a holistic view in the evaluation.Following the IIR evaluation model [6], specific situated work tasks in the context of defined information seeking activities were provided to each participant, which focus on the issue of discovering information about connections between entities. In particular, we have prepared these situated work tasks in the context of a student writing an outline for a history course report centred on specified historical figures.

To guide the analysis of Lensing Wikipedia and the support it provides for the information seeking activities, we have classified its features according to Wilson's taxonomy [36] (see Table 1). This provides an organizational structure for the discussion regarding which types of features are useful during which stage of the information seeking process.

4.1 Methodology

The specific methodology for the evaluation was designed to answer the following four research questions:

RQ1: How does the search stage influence the use of the different types of Lensing Wikipedia features?

RQ2: How much knowledge was gained as a result of undertaking the information seeking activities using Lensing Wikipedia?

RQ3: How useful do the searchers perceive the Lensing Wikipedia features for fulfilling their overall information seeking goal?

RQ4: How satisfied were the searchers in using Lensing Wikipedia?

Because of the importance of being able to effectively read and write in English, the participant recruitment process included an online screening component where we collected self-reported English language proficiency details, along with history knowledge and Wikipedia search expertise. All participants with good or excellent English skills, reasonably good knowledge of history, and some experience with using Wikipedia were invited to continue participation in the study (n=14).

Each study session began with an overview of the study procedures and a confirmation of informed consent to continue with the study. A 10-minute training video was show to the participant to introduce each of the core features of the software, after which the researcher answered any questions the participants may have had. The participant was then introduced to the software setup and specific study procedures. On a 27-inch display, Lensing Wikipedia was provided in a window that took 2/3 of the screen width. The remaining 1/3 of the screen was filled with an online questionnaire that lead the participant through the procedures of the study, the specific steps of the information seeking tasks, and the post-study questionnaire. Two printed pages were provides: a brief description of the Lensing Wikipedia features, and an overview of the study procedures.

Table 1: A summary of the feature interaction types in Lensing Wikipedia.

Interaction Type	Lensing Wikipedia Features
Input	facets tab
Control	remove constraints
Informational	storyline tab timeline tab comparison tab map tab follow link
Personalization	return to document edit notes

Three information seeking activities were assigned to the participant, focusing on using Lensing Wikipedia to explore and study the relationships between a historical figure and other important people in history that they discover. The first two historical figures were randomly assigned between "John, King of England" and "Franklin D. Roosevelt". The third activity allowed the participant to choose their own historical figure upon which to base their search. Note that the training video focused on a similar activity centred around "Napoleon". Participants were given specific instructions that guided them through Vakkari's stages of information seeking [33] (see Figure 2).

At the beginning of each of the three information seeking activities, participants were asked to rate their level of prior knowledge on the topic. While using Lensing Wikipedia, all interactions with the software were logged. As the participant completed each stage of the process, they were asked to notify the researcher and explicitly indicate within the interface movement to the next stage of the process. At the end of each activity, participants were asked to rate how much they learned about the topic.

After all activities were complete, a post-study questionnaire was administered. A set of questions were provided regarding the perceived usefulness of each core feature of Lensing Wikipedia (facets, storyline, timeline, comparison, and map tabs), along with other supporting features (modify constraints, follow link, return to document, and edit notes). The participants were then asked three questions regarding the overall usefulness of Lensing Wikipedia (based on the Technology Acceptance Model instrument [35]) and one question regarding their satisfaction with Lensing Wikipedia for finding information on the relationships between historical figures.

The study took approximately 1.5 hours, and each participant was given a $10 gift-card for their time and trouble. In order to inspire the participants to take the assigned activities and tasks seriously and do their best work, a further incentive of a $50 gift card was given to the participant who provided the best essay outline for each of the three activities. These incentives were indicated in the informed consent procedure.

4.2 Analysis

Because of the subtle differences between each of the assigned information seeking activities, we analyzed and present each separately. The average amount of time and average number of interactions were calculated and presented in graphical format using standard error about the mean as a measure of variability in the data. These are organized by the information seeking stage, allowing the reader to observe the pattern of feature use across the stages. Statistical analyses of the data were performed, using ANOVA to de-

Introduction: For a history course, you have been given an assignment to write a short 2-page essay about the relationship between [primary person] and some other important person in history. Today, you have given yourself the goals of (1) finding a small set of important people related to [primary person], (2) narrowing this down to one person and studying the relationship between these two people that will be the focus of your paper, and (3) producing a list of sources to support the writing of your essay.

Stage 1 (pre-focus): Using Lensing Wikipedia, search within Wikipedia to find three people that had significant interactions or relationships with [primary person]. Feel free to use any feature of Lensing Wikipedia and the browser that you feel will help you with this activity. A list of features and their primary purpose is provided to help you learn what each feature is for. Document the names of these individuals in the space provided, as well as your ideas about the essay you will write.

Stage 2 (focus formulation): Select one of the people you have identified in the previous step, and use Lensing Wikipedia to learn about the relationship between these two individuals. Feel free to use any feature of Lensing Wikipedia and the browser that you feel will help you with this activity. Document the name of the individual that will be the focus of your paper, along your notes on the relationship, when it took place, and any other information you feel is relevant.

Stage 3 (post-focus): Your goal now is to collect and organize the source material that will help you with writing the short essay. Find five articles that you think will be useful, copy their URLs in the space provided, and then write a short summary of what is in each that is relevant to your paper. Finally, write a short (200 word) outline for your essay.

Figure 2: Instructions given to participants.

termine if there is a statistically significant difference in the feature use between each stage.

In order to compare the prior knowledge and the knowledge gained during each activity, the responses were aggregated by task and stacked. Similarly, the usefulness of each feature, as reported by the participants at the end of the study, was aggregated and presented in graphical format using a stacked bar chart. Likewise, the overall perceived usefulness measures and level of satisfaction are presented in a similar manner. These provide a summary of the perceptions of the participants regarding their knowledge gained and the value of Lensing Wikipedia

4.3 Results

4.3.1 Feature Use in Each Information Seeking Stage

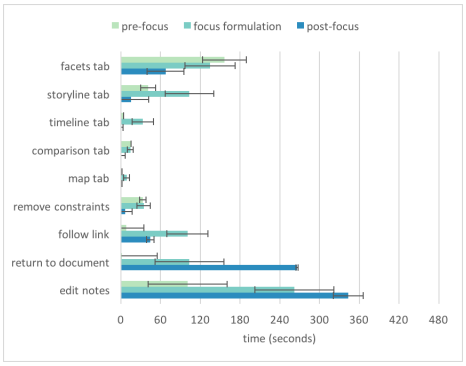
For each of the three activities, the amount of time and the number of interactions with the core interface features were analyzed across each of the three information seeking stages. Figures 3, 4, and 5 show the average values of these measures, with error bars representing the error about the mean. One can see a general pattern emerging: during the pre-focus stage (light colour in the figures), participants spent time using the facets tab (input) to explore among the relationships and removed constraints (control) that were added during the interactive exploration process. There was some use of the storyline and comparison tabs (informational), but these were minimal in comparison to the facets tab. Some time was also spent with the notes, but these consisted of many short interactions as the searchers entered brief information relevant to the task.

Things changed somewhat in the focus formulation stage (mid-weight colour in the figures). Here we can see that participants spent less time with the facets tab (input) than in the previous stage, and made more use of the storyline tab (informational). The other tabs that were similar in informational nature (timeline, comparison, and map tabs) were used somewhat, in what we characterize as experimentation (many interactions, but little time spent using the features). In this stage, we notice that participants followed the most links (informational) and spent time with these documents judging their relevance. More time was spent editing the notes, as participants added more information to aid their decision-making process.

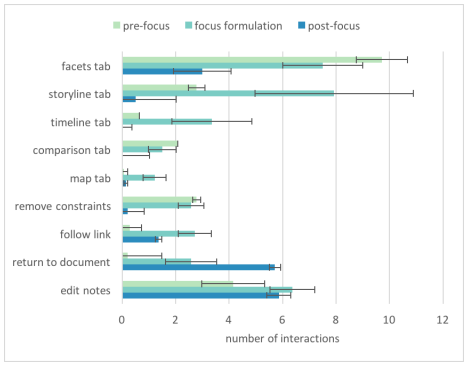
In the post-focus stage (dark colour in the figures), it is no surprise that the participants spent the majority of their time editing the notes (personalization) and returning to previously opened documents (personalization). The high number of interactions with returning to documents and editing the notes was an indication of the participants flipping back and forth between these to two views. Some supplemental exploration of the data was also performed (facets and storyline tabs, removing constraints, and following links). Post-hoc analysis of the logs revealed that this was often done near the end of the stage, presumably as a method to verify what was written and to ensure completeness.

While the pattern is generally consistent between the first two activities, things were somewhat different with the third activity. Since this activity was self-selected by the participants, they had a strong interest in fulfilling the information seeking goals. They appeared to make quicker decisions regarding relevance during the pre-focus stage, evidence of which can be seen in similar number of pre-focus interactions with the input (facets tab), control (remove constraints), and informational (storyline tab) features as in the first two activities, yet much less time spent with these feature. The focus formulation stage took much less time than with the assigned activities, with much attention given to following links (informational). The post-focus stage was consistent with the other activities, following a pattern of returning to previously opened documents (personalization) and editing the notes (personalization). However, less time was spent with the previously opened documents than in the two assigned activities.

We focused the statistical analysis on the features that were used regularly by all participants in the study, and excluded the timeline, comparison, and map tabs because of insufficient data on their use. Table 2 summarizes this analysis, with statistically significant results in bold. While there was some variability regarding the input (facets tab) and control (remove constraints) features, there are three

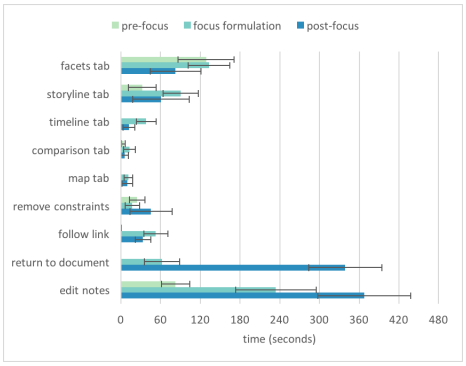


(a) Time spent using each feature.

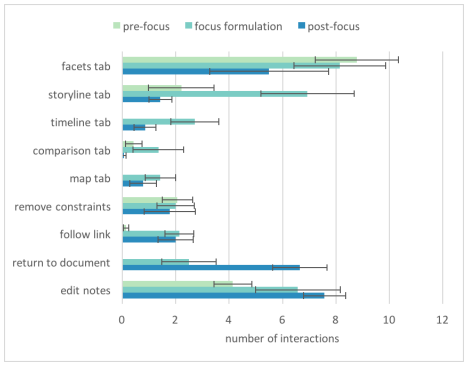


(b) Frequency of interaction with each feature.

Figure 3: The use of the core features during the "John, King of England" activity.

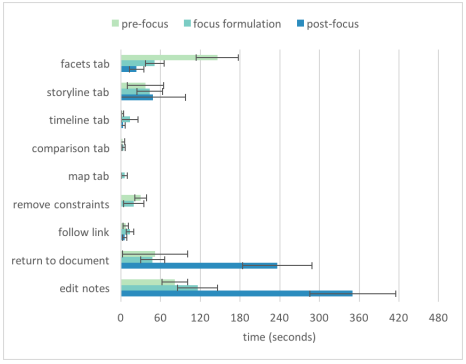


(a) Time spent using each feature.

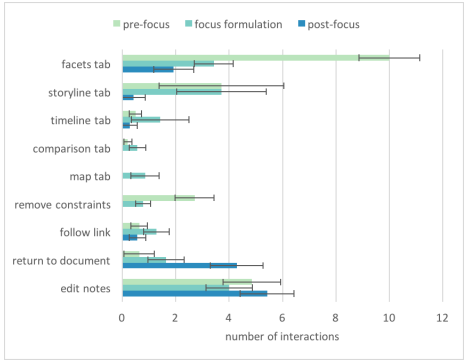


(b) Frequency of interaction with each feature.

Figure 4: The use of the core features during the "Franklin D. Roosevelt" activity.



(a) Time spent using each feature.



(b) Frequency of interaction with each feature.

Figure 5: The use of the core features during the participant-selected activity.

noteworthy aspects of this statistical analysis.

Contrary to the two assigned tasks, for the participant-selected tasks we found statistically significant differences in which stage participants used the facets tab (input), and no statistically significant differences in when links were followed (informational). For these tasks, participants used the facets tab much more in the pre-focus stage than the other two, and followed links throughout all three stages. What this tells us is that when searchers seek information that is of personal interest, they may be able to quickly and easily narrow down the search results to the information they need for the later stages, follow the relevant links to load the required documents, and not find the need to continue to explore during the focus formulation stage or verify completeness post-focus stage.

For the assigned tasks, we found a consistent pattern of participants following links (informational) during the focus formulation stage. During the pre-focus stage, participants were able to discover the information they needed from within the Lensing Wikipedia interface, and did not follow many links. During the post-focus stage, few links were followed and instead participants returned to documents found

Table 2: Statistical analysis (ANOVA) of the time and frequency of feature use across the stages of the information seeking process.

Feature	Task	Time Spent Using the Feature	Frequency of Interaction
facets tab	John, King of England	F(2, 39) = 1.944, p = 0.157	**F(2, 39) = 8.190, p <0.005**
	Franklin D. Roosevelt	F(2, 39) = 0.564, p = 0.573	F(2, 39) = 0.887, p = 0.420
	Participant's Selection	**F(2, 39) = 9.361, p <0.005**	**F(2, 39) = 23.167, p <0.001**
storyline tab	John, King of England	F(2, 39) = 2.811, p = 0.072	**F(2, 39) = 3.900, p <0.05**
	Franklin D. Roosevelt	F(2, 39) = 0.867, p = 0.428	**F(2, 39) = 5.600, p <0.05**
	Participant's Selection	F(2, 39) = 0.030, p = 0.970	F(2, 39) = 1.281, p = 0.239
remove constraints	John, King of England	F(2, 39) = 3.215, p = 0.051	**F(2, 39) = 9.911, p <0.005**
	Franklin D. Roosevelt	F(2, 39) = 0.501, p = 0.610	F(2, 39) = 0.038, p = 0.963
	Participant's Selection	F(2, 39) = 2.239, p = 0.121	**F(2, 39) = 9.752, p <0.005**
follow link	John, King of England	**F(2, 39) = 3.760, p <0.05**	**F(2, 39) = 7.563, p <0.005**
	Franklin D. Roosevelt	**F(2, 39) = 4.632, p <0.05**	**F(2, 39) = 5.172, p <0.05**
	Participant's Selection	F(2, 39) = 0.846, p = 0.437	F(2, 39) = 1.090, p = 0.346
return to document	John, King of England	**F(2, 39) = 9.593, p <0.001**	**F(2, 39) = 8.909, p <0.001**
	Franklin D. Roosevelt	**F(2, 39) = 25.976, p <0.001**	**F(2, 39) = 16.490, p <0.001**
	Participant's Selection	**F(2, 39) = 6.353, p <0.005**	**F(2, 39) = 6.020, p <0.05**
edit notes	John, King of England	**F(2, 39) = 5.971, p <0.05**	F(2, 39) = 1.767, p = 0.184
	Franklin D. Roosevelt	**F(2, 39) = 6.747, p <0.005**	F(2, 39) = 2.577, p = 0.089
	Participant's Selection	**F(2, 39) = 11.720, p <0.001**	F(2, 39) = 0.531, p = 0.592

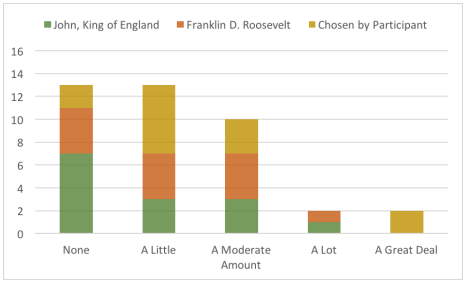
in the focus formulation stage. This finding validates some important features of Lensing Wikipedia: showing snippets that describe the extracted relationships from the Wikipedia articles helps searchers to make decisions during the pre-focus stage; providing links to the source documents helps searcher to find specific information that is needed in the focus formulation stage, and opening links in browser tabs allows searchers to return to previously found documents during the post-focus stage.

The personalization features of returning to documents and editing notes followed a consistent pattern across all tasks. Participants spent significantly more time viewing documents they previously clicked on during the post-focus stage, and undertook this action significantly more frequently. Similarly, they spent significantly more time editing their notes during this stage, although the number of times they opened the notes was not significantly different across the stages. This finding highlights the need to make it easy for searchers who are in the post-focus stage to easily change between reading previously found documents and writing their notes about the information they discover.

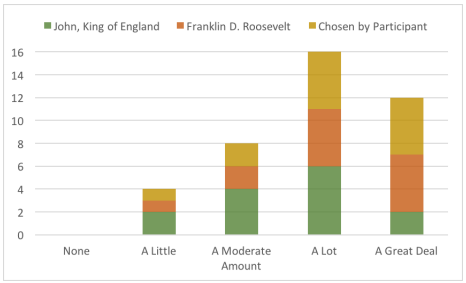
4.3.2 *Perceived Knowledge Gained*

Before and after each activity, participants self-reported how much prior knowledge they had on the historical figure in question, and how much knowledge they gained as a result of using Lensing Wikipedia to fulfil the information seeking goal. Figure 6a shows that most participants had a low to moderate amount of prior knowledge, and Figure 6b shows that most participants reported having learned a moderate amount, a lot, or a great deal. Of special note is the amount of knowledge the participants gained in their self-selected activity. Even though this may be attributed to their growing confidence with using Lensing Wikipedia (this activity was always performed last), we consider it an indicator of the success that searchers experienced using Lensing Wikipedia.

4.3.3 *Perceived Usefulness of Features*



(a) Prior knowledge.



(b) Knowledge gained.

Figure 6: Amount of knowledge participants reported to have had before and gained during the three activities.

While there were a core set of features that were consistently reported as being useful by the participants, there was also a set that had mixed reactions (see Figure 7a). Virtually all participants reported positively regarding the usefulness of the facets tab, the ability to remove constraints, following links, returning to documents, and editing notes. This supports the time spent with these features and the frequency of interaction.

The informational features (storyline, timeline, comparison, and map tabs) were perceived by some participants as

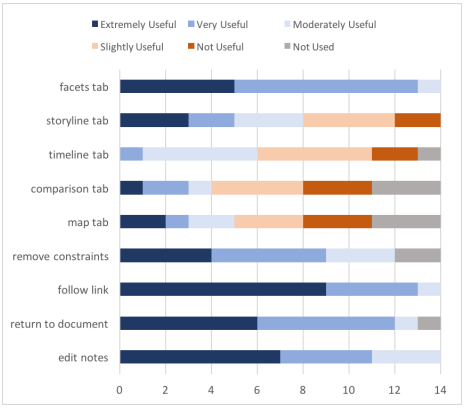
useful, and by others as not useful. Some participants did not bother to use these features, others used them briefly and then reverted to the more familiar facets tab, and the remainder made an honest effort to use these features to learn about the information presented in them. Of these, the storyline, comparison, and map tabs showed promise and could prove to be more useful with additional refinement and training.

4.3.4 *Perceived Overall Usefulness & Satisfaction*

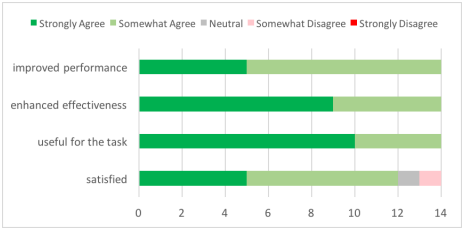
Overall, participants in this study reported positive perceptions of usefulness of Lensing Wikipedia, and were satisfied with using it to fulfill their information seeking goals (see Figure 7b). In fact, over half of the participants strongly agreed that Lensing Wikipedia enhanced their effectiveness and was useful for the task. There was general agreement of being satisfied, with the caveat of one participant being neutral and another disagreeing. We observed within the usage logs the difficulties these two participants experienced in finding information about their self-selected historical figures, which was due to limited information in Wikipedia, rather than the features of Lensing Wikipedia.

5. CONCLUSIONS & FUTURE WORK

This study illustrated the utility of various different features of Lensing Wikipedia during three information seeking stages. The findings were consistent with that of other similar studies [17]. Of note was the difference we found between the assigned and self-selected search activities. Participants self-reported positive knowledge gain, and generally high degrees of usefulness and satisfaction.



(a) Feature-level usefulness.



(b) Overall usefulness and satisfaction.

Figure 7: Responses to questionnaires regarding the feature-level usefulness and overall usefulness and satisfaction with using Lensing Wikipedia.

Findings from this study contribute to understanding how people use advanced search tools to aid their information seeking tasks, and will be used to further refine the features of Lensing Wikipedia before releasing it as a public service. In particular, we plan to group features that are used together within the three separate information seeking stages, and to re-evaluate the utility of the less-used visualization features. Efforts will be made to combine the storyline, timeline, comparison, and map tabs, thereby providing a unified view of the information that may be helpful for a searcher to "see" the story.

6. REFERENCES

[1] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni. Open information extraction from the web. In *Proceedings of the International Joint Conference on Artifical Intelligence*, pages 2670–2676, 2007.

[2] N. Belkin. Intelligent information retrieval: Whose intelligence? In *Proceedings of the International Symposium for Information Science*, pages 25–31, 1996.

[3] N. J. Belkin. People, interacting with information. *ACM SIGIR Forum*, 49(2):13–27, 2015.

[4] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, 2009.

[5] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 1247–1250, 2008.

[6] P. Borland. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), 2003.

[7] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr., and T. M. Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the Conference on Artificial Intelligence*, pages 1306–1313, 2010.

[8] B. Chan, L. Wu, J. Talbot, M. Cammarano, and P. Hanrahan. Vispedia: Interactive visual exploration of Wikipedia data via search-based integration. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1213–1220, 2008.

[9] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the Annual Meeting on Association for Computational Linguistics*, pages 363–370, 2005.

[10] D. Gildea and D. Jurafsky. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288, 2002.

[11] Google Developers. Google geocoding. https://developers.google.com/maps/documentation/geocoding/.

[12] R. Hahn, C. Bizer, C. Sahnwaldt, C. Herta, S. Robinson, M. Bürgle, H. Düwiger, and U. Scheel.

Faceted Wikipedia search. In *Proceedings of the International Conference on Business Information Systems*, pages 1–11, 2010.

[13] M. Hearst. *Search User Interfaces*. Cambridge University Press, Cambridge, UK, 2009.

[14] M. Hearst. User interfaces for search. In R. Baeza-Yates and B. Ribeiro-Neto, editors, *Modern Information Retrieval: The Concepts and Technology Behind Search*, pages 21–55. Addison-Wesley Professional, 2nd edition, 2011.

[15] J. Hoffart, F. M. Suchanek, K. Berberich, E. Lewis-Kelham, G. de Melo, and G. Weikum. YAGO2: Exploring and querying world knowledge in time, space, context, and many languages. In *Proceedings of the International World Wide Web Conference*, pages 229–232, 2011.

[16] J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. *Artificial. Intelligence*, 194:28–61, 2013.

[17] H. C. Huurdeman, M. L. Wilson, and J. Kamps. Active and passive utility of search interface features in different information seeking task stages. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval*, pages 3–12, 2016.

[18] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1-2):1–224, 2009.

[19] D. Kelly, S. Dumais, and J. O. Pedersen. Evaulation challenges and directions for information-seeking support systems. *IEEE Computer*, 42(3):60–66, 2009.

[20] C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5):361–371, 1991.

[21] Y. Liu and A. Sarkar. Experimental evaluation of LTAG-based features for semantic role labeling. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing and Computational Natural Language Learning*, pages 590–599, 2007.

[22] F. Mahdisoltani, J. Biega, and F. M. Suchanek. YAGO3: A knowledge base from multilingual wikipedias. In *Proceedings of the Biennial Conference on Innovative Data Systems*, 2015.

[23] Mausam, M. Schmitz, S. Soderland, R. Bart, and O. Etzioni. Open language learning for information extraction. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534, 2012.

[24] P. N. Mendes, M. Jakob, and C. Bizer. DBpedia: A multilingual cross-domain knowledge base. In *Proceedings of the International Conference on Language Resources and Evaluation*, pages 1813–1817, 2012.

[25] T. M. Mitchell, J. Betteridge, A. Carlson, E. R. Hruschka Jr., and R. C. Wang. Populating the semantic web by macro-reading Internet text. In *Proceedings of the International Semantic Web Conference*, pages 998–1002, 2009.

[26] N. Nakashole, M. Theobald, and G. Weikum. Scalable knowledge harvesting with high precision and high recall. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 227–236, 2011.

[27] M. Palmer, D. Gildea, and P. Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, 2005.

[28] M. Siahbani, R. Vadlapudi, M. Whitney, and A. Sarkar. Knowledge base population and visualization using an ontology based on semantic roles. In *Proceedings of the Workshop on Automated Knowledge Base Construction*, pages 85–90, 2013.

[29] E. Stoica, M. A. Hearst, and M. Richardson. Automating creation of hierarchical faceted metadata structures. In *In Proceedings of the Human Language Technology Conference*, pages 244–251, 2007.

[30] F. M. Suchanek, G. Kasneci, and G. Weikum. YAGO: A core of semantic knowledge. In *Proceedings of the International World Wide Web Conference*, pages 697–706, 2007.

[31] F. M. Suchanek, G. Kasneci, and G. Weikum. YAGO: a large ontology from Wikipedia and WordNet. *J. Web Sem.*, 6(3):203–217, 2008.

[32] J. Szymański. Self–organizing map representation for clustering Wikipedia search results. In *Proceedings of the International Conference on Intelligent Information and Database Systems*, pages 140–149, 2011.

[33] P. Vakkari. Task-based information searching. *Annual Review of Information Science and Technology*, 37(1):413–464, 2003.

[34] M. van Erp, P. N. Mendes, H. Paulheim, F. Ilievski, J. Plu, G. Rizzo, and J. Waitelonis. Evaluating entity linking: An analysis of current benchmark datasets and a roadmap for doing a better job. In *Proceedings of the International Conference on Language Resources and Evaluation*, pages 4373–4379, 2016.

[35] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204, 2000.

[36] M. L. Wilson. Interfaces for information retrieval. In I. Ruthvan and D. Kelly, editors, *Interactive Information Seeking, Behaviour and Retrieval*. Facet Publishing, 2011.

[37] F. Wu and D. S. Weld. Open information extraction using Wikipedia. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 118–127, 2010.

[38] N. Yan, C. Li, S. B. Roy, R. Ramegowda, and G. Das. Facetedpedia: Enabling query-dependent faceted search for Wikipedia. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1927–1928, 2010.