SearchScape: From Ranked Lists to Interactive Landscapes

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

Whether planning a trip or researching medical symptoms, many complex and exploratory search tasks involve intense cross referencing of multiple sources. One has to determine the popularity and trustworthiness of competing options and avoid missing valuable but less common options. Traditional search systems leave this task to users, while automatic result clustering approaches often fail to produce structures coherent enough to meet these needs. We introduce a novel approach that builds coherent topical landscapes from search results. By combining the rich but sparse representations of knowledge bases with word vector models that are shallow but dense in coverage, we produce structures more coherent than alternative approaches. We instantiate this in a prototype interactive system and explore which elements users find useful. Across three information seeking scenarios, we found that participants valued both the structure and distribution information provided by our system for complex exploratory searches and, surprisingly, simple searches as well.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

Author Keywords

Exploratory Search; Sensemaking

INTRODUCTION

Foraging and consuming information online has became how many people make sense of the world today, using search engines to plan a trip to a new city, figure out which university to attend, or to learn the important topics in a new domain. Unlike simple searches such as finding the address of a restaurant or checking the weather, these more complex and exploratory searches require significant amounts of learning and synthesis across multiple websites. Planning a trip or researching medical symptoms often deals with scenarios where the user is unfamiliar with the domain of interest, unsure about what

information is needed to achieve their goals, and have difficulties figuring out the right queries to find such information [33].

Consider an example of an exploratory search such as looking for the best day trips from Barcelona. Searching for this online returns a set of web pages primarily composed of lists of day trip destinations. These range from the popular beach town of Sitges to the less well known market town of Vic. However, looking at any single page raises a number of potential issues. A reader might want to know the relative popularity of each option: while Sitges appears many different sources, Vic appears in only a few, despite being given approximately equal weight on a top search result.[1] Conversely, a page may be missing a popularly cited option such as Sitges and the reader would have no indication they might be missing out. Providing an overview of the range of options could also expose interesting options in the long tail that could be relevant to certain readers, despite not appearing in the top few results (e.g., a Utopian textile mill's church designed by Gaudi, Colònia Güell).

To understand the prevalence of this problem, we conducted a pilot survey with 103 participants from Amazon Mechanical Turk (age between 20 and 63, M=36, SD=12, 60% male and 40% female, mostly from the US), focusing on their experiences when conducting complex exploratory searches. Around half of the participants reported having conducted more than four exploratory search tasks in the past two weeks, in which 17% reported having conducted more than ten tasks. Our results reaffirm that when conducting exploratory searches, people read from multiple information sources to make sure they do not miss something important (84%), look for more information about a topic from other webpages in the same search result (84%), verify previously seen but uncertain information using multiple sources (88%), and get a sense of what's popular or important in the search results (76%). At the same time, many also found conducting complex searches can be stressful (60%), and that they often end up with more browser tabs than they can manage efficiently (67%). These findings are consistent with previous findings that shows that up to 33% of this time is related to complex exploratory search tasks [33, 24, 43].

Despite their importance and ubiquity, the activities described above are poorly supported by modern approaches. Recent advances in search interfaces have optimized for providing information as quickly as possible, such as knowledge cards
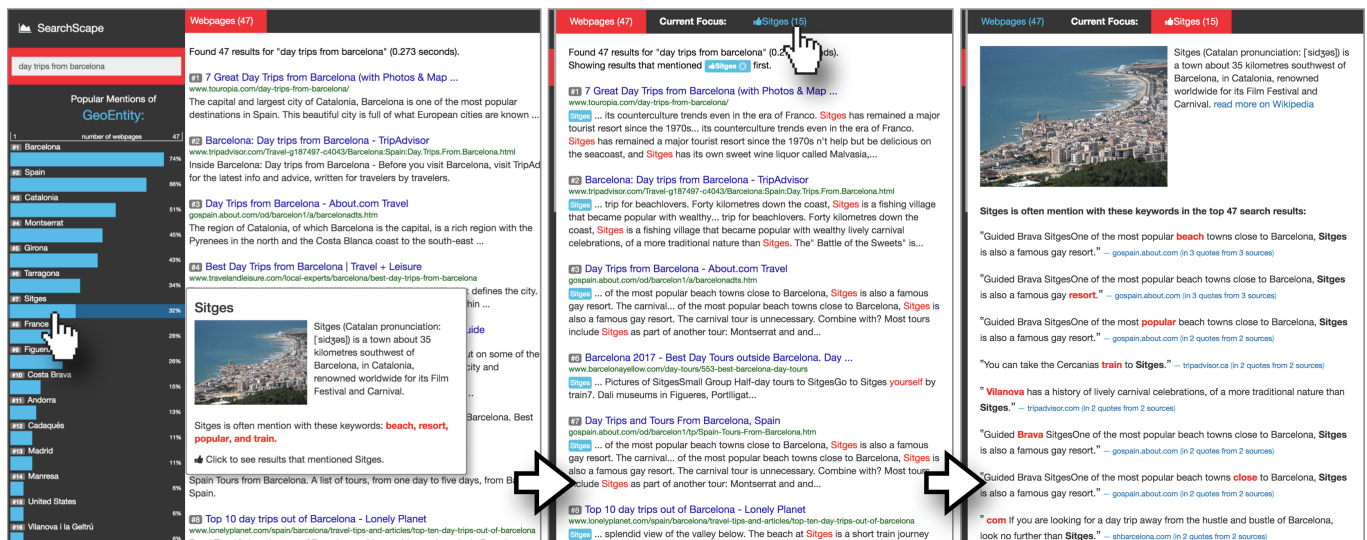
---

[1] LonelyPlanet - Top 10 day trips out of Barcelona, https://www.lonelyplanet.com/spain/barcelona/travel-tips-and-articles/top-ten-day-trips-out-of-barcelona

Figure 1. An overview of the SearchScape system. The Document Panel on the right shows a ranked list of search results, while the Structure Panel on the left shows the distribution of automatically identified key concepts in the search results. Exploratory searchers can click on an item to filter the search results and see how the item of interest is mentioned across sources. Alternatively, they can also grouped these mentions using keywords were frequently mentioned together with the item of interest.

in which an answer is surface directly on a search results page or visual carousels of possible options that a user can quickly click through. These approaches excel at finding specific information, but notably fail to address the activities involved in complex search such as cross-referencing information across different webpages, identifying and learning important subtopics, planning out subsequent queries, and synthesizing information gathered from multiple sources.

Meanwhile, research on supporting complex information seeking has attempted to address some of these activities through faceted navigation [16] and semantic web interfaces [50]. While useful in an environment with well-structured data, it's prohibitive to expand these approaches to support the unstructured Web [3, 47]. One promising direction has been using unsupervised clustering algorithms to automatically extract structures from search results, which have been found useful for separating relevant sources from irrelevant sources (e.g., differentiating fruit related from technology related results when querying "apple") [55]. However, these automatically generated structures are rarely coherent enough to provide an overview of the space of useful information. Consider the top five clusters that Clusty (now yippy.com) shows for "Day trips from Barcelona": Spain, Europe, Cruise, Photos, Reviews. Even within the Spain cluster the subclusters provide no structure or coherence: Blog, Tours from Barcelona, Flights, Guided, Cordoba, Cultural, etc. The topics for "Guacamole recipes" similarly do not provide any concepts of the same class: Mexican (country), Chef (profession), Health (topic), Blog (format), and Free (cost). While these structures may be useful for navigating to sources in different dimensions (e.g., websites for guided tour services, for photo albums, or for blog posts), they provide little utilities for comparing between options under the same dimensions. For example, comparing the popularity and the types of activities you can do in different locations when searching for day trip destinations, or compare the reviews and popularity of different guided tour services. Fundamentally, automatic structures often provide no structure to define a space – there is no base for comparison and no understanding of relationships between concepts in their structure. At a higher level, faceted search and clustering approaches make better filers for people to zero in on what they know they are searching for (e.g., Apple the tech company), whereas we are instead trying to provide an atlas for exploratory scenarios (e.g., how popular are the different destinations, and how different sources describe them): a representative description of the space to help searchers learn about their options in the information space and how key concepts are distributed among the search results.

In this paper we explore a novel approach for extracting coherent structures from arbitrary unstructured web content (i.e., in a set of web search results), and also how to present such structures to the users to provide an overview of the information landscape. We instantiate this approach in a prototype system called SearchScape, an interactive search result exploration interface. SearchScape dynamically identifies useful structures by combining external knowledge bases (i.e., Wikipedia[2] and NELL[6]), and machine learned unsupervised semantic models (i.e., Word2Vec[36] and GloVe[41]). The intuition behind this approach is that on the one hand, knowledge bases provide high quality entities and ontological frameworks, but often lack good coverage over arbitrary search results. On the other hand, unsupervised models provide substantial coverage, but can often be overly generalized and difficult to understand due to its lack of structure and incoherence. SearchScape combines the two to identify high quality categories from knowledge bases that are easy to understand (e.g., Geographic Locations, or Agricultural Products), while expanding the coverage of these categories using automatically learned semantic models. In our first study, we find that structures produced by SearchScape for 50 different exploratory search queries

was found to be significantly more coherent than structures produced by a search result clustering system.

Extracting coherent structure enabled us to further expose the distribution of key items of the same classes in the search results, in hopes that it can allow exploratory searchers to better understand all the options in the information landscape. For example, exposing the popularity of different *location* items when searching for day trip destinations, or how common different *ingredient* items are when searching for recipes. In our second study, the overall system was found to be useful for three diverse scenarios including both exploratory and fact finding search. Further examination using a control condition hiding distribution information found that in exploratory situations showing distribution was significantly more beneficial to searchers.

RELATED WORK

Exploratory Search
Traditional information retrieval systems make the assumptions that users have clear ideas about their goals when searching for online documents, and search engines should therefore optimize for high precision, recall, and efficiency [5, 32, 46]. However, these assumptions rarely hold for searchers seeking information in an unfamiliar domain. The term exploratory search emerged from the field of Information Retrieval to describe scenarios where the users were uncertain about what they were looking for [33].

Several exploratory search interfaces have been developed in order to help searchers orient themselves in the information space, review and explore the different subtopics, and keep track of their overall progress [17, 35, 39, 48, 37]. Two closely related studies include Topic-Relevance Map and Exploration Wall, which explored ways to provide overviews of search results of academic papers using document keywords and entities and easily choose keywords to build up subsequent queries [40, 26]. SearchScape also makes use of key entity mentions in the document for overview and drill-down purposes, but we automatically extract key items from the unstructured web using combined hand-crafted knowledge bases and machine learned unsupervised semantic models.

Diversity and Distribution of Information
Initially motivated to reduce redundant information in top search results produced by traditional relevance-oriented ranking algorithms, Carbonell et. al first introduced the idea of ranking search results to increase the diversity of information in the top results for scenarios where a large amount of potential relevant document exists in the corpus [5]. More recent research has also explored how diversifying search results can benefit exploratory searchers (e.g., [23]). Our perspective differs from these pure diversity-oriented ranking approaches, which optimize for showing as many different topics to the user as possible. We instead argue that the distribution of different topics and key entities can actually be a useful signal for exploratory searchers to judge the importance of different subtopics, and prioritizing their subsequent exploration. In this case, showing a uniformly diversified list of search results

could actually give people an incorrect sense of the information distribution. For example, showing as many restaurant choices as possible can at the same time obscure a sense of how frequently each restaurant is mentioned. A similar idea was pointed out by Dang et. al, who proposed instead to diversify search results while accounting for the proportionality of information in the search result [12]. In SearchScape we take a more direct approach than re-ranking search results and present the distribution of extracted entities on the interface so that exploratory searchers can easily compare and evaluate the different subtopics in the information space.

Faceted Search for Structured Corpus
Faceted search is perhaps the most commonly used search interface, in which users can filter results by selecting or deselecting categories or "facets" (e.g., products on an online shopping site). Although originally designed to help searchers efficiently narrow down on relevant sources and exclude irrelevant sources from their search results (e.g., showing only digital cameras of specific price range and manufactured by a specific company on an online shopping site), researchers have also found faceted search interfaces to benefit exploratory scenarios, where searchers are less certain about their information needs [18, 52]. Facets can provide an overview of the different topics in the information space in addition to supporting drill-down on those topics [52, 4, 31] and have been widely explored for corpora where metadata are readily available, such as images [52], music [50], social media [38], library catalogs [34, 31], and scholarly articles [22, 13, 26]. However, in cases where metadata is not readily available, computationally inferring fine-grained facets (e.g., day trip destinations) remains a challenge, inhibiting the faceted interface from being expanded to support exploratory search on the open Web [3, 47]. Instead of requiring documents to be annotated with metadata, SearchScape uses a novel technique for automatically producing structures by inferring topics from unstructured web content by combining readily available knowledge bases and ontologies. Furthermore, we also explore the benefits of visually depicting the distribution of content across structures in helping the user better understand the information space.

Automatically Identifying Structure
More generally, numerous challenges have been identified [47] in developing structured search systems to better support exploratory scenarios without pre-compiled metadata. Researchers have long explored ways to extract structure by clustering search results using machine learning [54, 55, 20], lexical and HTML patterns [29], crowdsourcing [15, 7, 10], or interaction techniques [20, 21, 21]. However, a number of papers [19, 11, 7] point to the fact that automatically techniques, such as unsupervised clustering, can often produce incoherent structures that are difficult to comprehend by users unfamiliar with the domain of interest. Further, providing a quick overview using coherent structures can be critical in supporting early stages of the exploratory search process. This helps users get a sense of the information space, and further investigate subtopics that are interesting or important to the them [33]. Conversely, techniques that utilize human judgments via crowdsourcing or user interaction can produce structures that

are more coherent, but do so at the expense of either additional time and monetary cost or user effort[7, 15].

Perhaps most closely related to our work in intent, MetaSpider extracts noun phrases from webpages to use as features for clustering webpages in search results into visualized topic maps using the Kohonen SOM algorithm [9, 28]. Instead of trying to identify structures solely based on information presented in the search results, SearchScape utilizes rich and readily available knowledge bases and ontologies to automatically identify dynamic but coherent structures that best fit the search results. In addition, instead of producing document clusters, SearchScape presents distributions of entities over documents in the search results, as an overview of the information space (e.g., list of day trip destinations and how often they are mentioned in the search results.)

SEARCHSCAPE
The SearchScape system augments standard web search results with a list of dynamically identified concepts useful for the current query. This gives exploratory searchers the ability to first gain a quick overview of the information space and explore different subtopics. The system consists of two main components. First, SearchScape analyzes the content of all webpages in the search results to identify the latent structures in the information space leveraging existing knowledge base structures (which typically have higher accuracy but lower coverage) and expands them with unsupervised word vector models (lower accuracy but higher coverage). Then, the search results along with the identified structures are presented to the user through an interactive search result interface which visually presents the distribution of structures. In the following subsections, we will describe in detail our data sources, the back-end components, and the front-end interface.

Data Source
In an ideal setting, the back-end components of SearchScape would be integrated as a part of a general web search engine, so that content fetching, indexing, and analysis can be performed before query time. However, since it would require enormous time and resources to download and index the entire web to build a full-scale web search engine, we used a readily available API from a commercial search engine as our data source.[2] In the front-end, SearchScape focused on augmenting traditional web search results with automatically identified, task-specific structures using an interactive interface to better support exploratory scenarios. Similar to search engine backends, we also pre-fetch and pre-index webpages for a fixed set of queries by first retrieving a list of top 50 relevant webpages from the API, then download and extract the main content of each webpage to remove menus and advertisements using an readily available API[3] and build an inverted-index file for each webpages [1]. Finally, we perform further analysis to identify task-specific structures as described in the next subsection.[4]
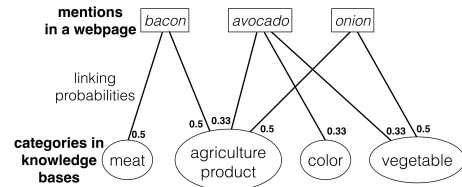
Figure 2. When a common noun can be associated with multiple categories due to word-sense ambiguity, SearchScape linked it to all potential categories with uniform linking probabilities.

Identifying Structures
After downloading webpages in the search results, SearchScape analyzes the content to identify potentially useful items and categories. SearchScape generates structures using both content from the webpages in the search results, two external knowledge bases, and two word semantic models. As an overview, words and phrases in the webpages are first linked to concepts in the knowledge bases. We then use the linked concepts and the ontological framework of knowledge bases to infer and rank topical categories in the knowledge bases (such as *GeographicLocations* or *AgricultureProducts*). Finally, we expand the top categories using the word semantic models to ensure good coverage over the search results.

Linking Mentions to Knowledge Base Concepts
Linking entity mentions in free text to concepts in knowledge bases and ontologies is a well researched problem in the field of natural language processing known as entity link (for named entities) and word sense disambiguation (for common nouns). SearchScape uses the open source AIDA algorithm to link named entity mentions in the search results to concepts in the Freebase knowledge base [53, 2], covering concepts such as *Banff National Park* and *Yosemite National Park*. For each named entity mention, the AIDA algorithm reports the linking probability of possible concepts in Freebase. To link common nouns such as *avocados* or *bacon*, we link each word in the search results to all of their possible concepts in NELL with a uniform probability distribution. For example, the term "avocado" can refer to three NELL[6] entities of categories *AricultureProducts*, *colors*, and *vegetables*, respectively (Figure 2), and SearchScape links "avocado" to each of the three concepts with 0.33 probability. In the next subsection, we describe how we identify the most useful categories based on these linking probabilities using a class-based word-sense disambiguation approach similar to [51, 8].

Topic Inference and Category Ranking
To identify the most commonly used categories in the search results, SearchScape used a probabilistic model based on the linking probabilities of both named entities and common nouns. The fundamental assumption here is that webpages from the same search result will likely mention concepts of the same categories. Similar assumptions were found to be effective in past work for word-sense disambiguation [51]. For example, when multiple results for "guacamole recipes" mentioned "avocado", SearchScape uses mentions of different *AgricultureProducts* in the search results to determine its likely that all avocado mentions were referring to cooking ingredients (*AgricultureProducts*) rather than *colors* or a mix

of the two (Figure 2). More specifically, it used the linking probabilities to vote on the categories in the knowledge bases. To account for the relevance of a search result based on its position in the ranked list, we weighted the linking probabilities by the reciprocal rank of their sources when voting.

$$Votes_{category,results} = \sum_{\substack{entity \in category \\ webpage \in results \\ mention \in webpage}} \frac{linkP_{mention,entity}}{rank_{webpage,results}}$$

For example, if "avocado" is mentioned in the fourth document, and was linked to three categories each with a probability of $linkP = 0.33$, each of the three categories receives $linkP/rank = 0.33/4 = 0.083$ votes. The total number of votes for a category given a ranked search result list is therefore the sum of linking probabilities to all entities in the category, weighted by the reciprocal ranks of the webpages that mentioned the entities.

Expanding Coverage
Knowledge bases can provide high quality ontological information. For example, the Freebase knowledge base defines *Sitges* as a *MunicipalCitiesInBarcelona* and a *GeographicLocations*. However, the coverage of categories over its entities can still be sparse, especially for general categories that can contain up to millions of entities such as *GeographicLocations* or *Person*. On the other hand, unsupervised word vector models that automatically project words and phrases onto semantic vector spaces to measure semantic similarity can yield high coverage over common words and phrases due to their capacity to train on large, unlabeled corpora [36, 41]. To take advantage of both the quality of structures in knowledge bases and the coverage of semantic word vector models, SearchScape first use the knowledge bases to identify useful categories as described above. Then, for each category extracted, SearchScape searches the word vector models with the identified key items for semantically similar phrases to add to the category and expand its coverage. For example, when a searcher queried SearchScape with "*day trips from Barcelona*", SearchScape first identified a list of *GeographicLocations* including *Sitges*, *Girona*, and *Tarragona* from Freebase. To expand the coverage of the *GeographicLocations* category, SearchScape searches the semantic vector space for phrases similar to the already identified key items, and add these similar phrases to the *GeographicLocations* category to expand its coverage. To expand NELL categories that contain mostly common nouns, we used a GloVe model trained on the Common Crawl corpus with 42 billion words and a vocabulary size of 1.9 millions.[5] To expand Freebase categories that contain mostly proper nouns, we used a Word2Vec model trained on a news corpus with 100 billion words and a vocabulary size of 1.4 million Freebase entities.[6]

Hand-off to the Front-end Interface
Finally, the back-end delivers the following data objects to the front-end interface (Figure 3):

---

[5]Pre-trained GloVe model downloaded from https://nlp.stanford.edu/projects/glove/
[6]Pre-trained Word2Vec model downloaded from https://code.google.com/archive/p/word2vec/
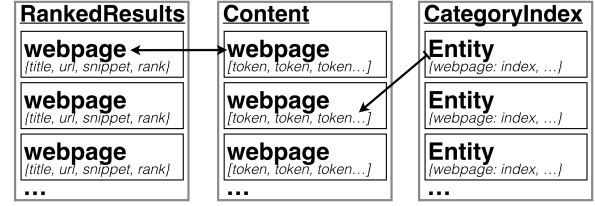

Figure 3. The three data objects produced in the SearchScape back-end component, and passed to its front-end component to enable interactive information exploration. The arrows show the mapping relations between elements in different object.

1. **RankedResults**. The ranked search results retrieved from the Microsoft Bing Web Search API, containing the top 50 results for the given query. Each result contains the rank, URL, title, and a short text snippet from the webpage.
2. **WebpageContent**. An index that contains an array of tokenized main content of each webpage in the search results.
3. **CategoryIndex**. The identified category and its list of entities as described in the previous subsections. Each entity has an inverted-index file[1] of webpages and indices of where the entity was mentioned in WebpageContent.

Figure 3 shows the mapping relations between the three data objects. The addition of CategoryIndex and WebpageContent to traditional web search engines enabled the SearchScape front-end to perform the following complex associations in real-time:

- Filter the RankedResults by using the CategoryIndex to find the list of webpages that mentioned a specific entity of interest, and calculate its proportion of mentions.
- List sentences that mentioned a specific entity of interest by using the CategoryIndex to find webpages and token indices where the entity was mentioned, and extract sentences from WebpageContent.
- Count word frequencies in the extracted sentences to find keywords that frequently co-occur with the entity of interest.

We will describe in detail how the SearchScape interface utilize these operations to enable interactive data exploration in the next subsection.

Interactive Search Landscape
Using the structures identified by its back-end components, the SearchScape interactive front-end interface augments traditional ranked list search results both to provide a quick overview of the information landscape, and to enable deep user interaction for exploring important subtopics specific to the current query.

SearchScape supports many key tasks defined in Shneiderman's task taxonomy of information visualization [44], such as *overview*, *zoom*, *filter*, and *details-on-demand*. The interface consists of a Structure Panel on the left, and a Document Panel on the right (Figure 1, left). When a new search results is loaded, the Structure Panel on the left displays an *overview* of the current information space using the distribution information of items in the identified category, and the Document Panel on the right lists webpages in the search results. Similar to many commercial search engines, the Document Panel shows the title, URL, and a short snippet related to the query
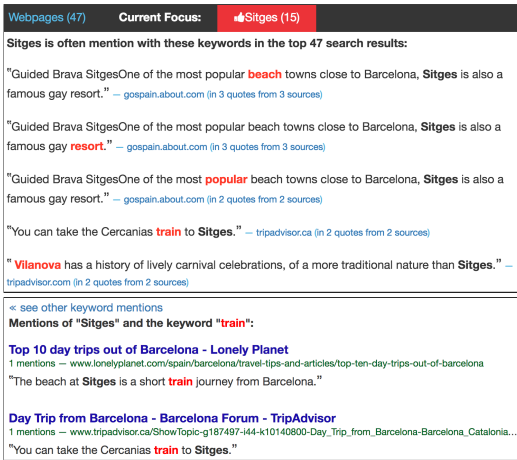
Figure 4. The searcher clicked on "Sitges" from the result page of "*day trips from Barcelona*" to explore its mentions grouped by frequently co-occurring keywords, such as *beach*, *resort*, and *train*.
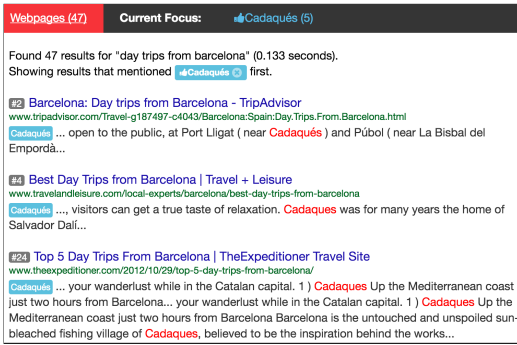
Figure 5. The searcher filtered the search results using "Cadaqués" from the result page of "*day trips from Barcelona*". Notice the snippets are updated to show relevant parts of each webpage.

terms for each webpage in the search results. Even though SearchScape can automatically identify key concepts from the search results, exploratory searchers might not be familiar with the extracted concepts, especially in early stages of the exploratory search process. Therefore, SearchScape allows searchers to *demand-for-details* when needed by hovering over the items to see short descriptions of the concepts from external knowledge bases (Figure 1, left). In addition, SearchScape also extracts a list of keywords that frequently co-occur with the concept of interest in the search results using the inverted-index file. When the searchers discovered an interesting item, they can *filter* the Document Panel by clicking the item of interest (Figure 5, left). A focus tab will appear on the top of the Document Panel to indicate the the current items of interest (Figure 5, center), and the Document Panel will re-rank webpages that mentioned the item to the top of the list, as well as update the short descriptions to show parts of each webpage that mentioned the item. Instead of looking through the filtered search result list, searchers also have the option to *zoom-in* to the current focused item by clicking on the focus tab to see sentences in the search results that mentioned the item of interest, grouped by frequently co-occurred keywords (Figure 4).
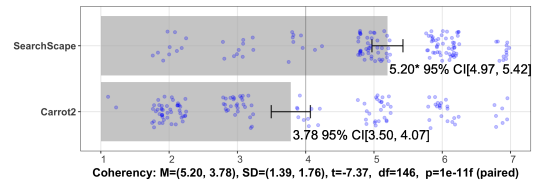
Figure 6. Using 7-point Likert scales, SearchScape structures were found to be strongly coherent (M=5.20*, 95% CI[4.97, 5.42]), and significantly more coherent compared to Carrot2 (M=3.78*, 95% CI[3.50, 4.07]) based on paired t-test (N=147, t=-7.37, df=146, p<0.001). The error bars indicate confidence intervals.

To give a sense of how an exploratory searcher might use SearchScape, we will use Figure 1 as an example scenario. In our scenario, the user searched for "*day trips from Barcelona*", the SearchScape back-end analyzed sources in the search results and identified the Freebase category *GeographicLocations* as the most useful category for the current task. The list of ranked search results are presented to the user in the Document Panel, and a list of different destinations and how frequently each of them are mentioned by different sources are presented in the Structure Panel. By examining the items in the *GeographicLocations* category, the user quickly gained an overview of how many different destinations are in the search results, and how popular each of the choices are (Figure 1, left). The user then proceeded to explore the different choices in the Structure Panel by hovering over each of the items. The user noticed that the destination "Sitges" is mentioned in a third of the sources, and is also frequently mentioned with keywords that matched with his or her interests, such as "*beach*" and "*resort*". To access more information about Sitges, the user clicked on Sitges to surface sources that mentioned Sitges (Figure 1, center. Figure 5), and also directly explored sentences from different sources that mentioned both "Sitges" and "beach" (Figure 1, right. Figure 4).

EVALUATION

We evaluated SearchScape through two studies aimed at assessing the coherence of the identified structures and the usefulness of the interface, respectively. For the first study, we compared structures generated for 50 different search queries from an exploratory search database [42, 14] using both SearchScape and Carrot2 [45], a search result clustering system. For the second study, we recruited 210 unique participants from Amazon Mechanical Turk for a user study [25]. Three diverse search scenarios were designed as between-subject conditions in Study 2 to evaluate the usefulness and generalizability of SearchScape. To further examine whether providing distribution information is actually beneficial to searchers, we grouped participants under each of search scenario into two front-end conditions to test the effects of hiding and exposing distribution information of the extracted items (Figure 7). In the following subsections, we will describe the two studies in detail.

Study 1: Automatic Structuring

To evaluate the quality of the structures produced by SearchScape, we compare it against the open source search results clustering system Carrot2 [45], which produces interpretable names for each of its clusters. To cover a wide variety of

Statements	Barcelona DIST	Barcelona RANK	Guacamole DIST	Guacamole RANK	Guggenheim DIST	Guggenheim RANK
list of items (S1)	**4.35** [4.11, 4.60]	4.12 [3.82, 4.43]	**3.97** [3.99, 4.66]	3.86 [3.69, 4.42]	4.22 [3.88, 4.57]	**4.36** [4.08, 4.64]
distribution/rank (S2)	**4.32*** [4.02, 4.63]	3.82 [3.45, 4.19]	3.89 [3.45, 4.33]	3.89 [3.53, 4.26]	4.28 [3.93, 4.63]	**4.44** [4.20, 4.69]
webpage mentions (S3)	**4.19** [3.89, 4.50]	4.09 [3.78, 4.40]	**4.11** [3.70, 4.51]	4.08 [3.69, 4.47]	4.28 [3.96, 4.60]	4.28 [3.99, 4.57]
keyword mentions (S4)	**4.39** [4.12, 4.65]	4.09 [3.74, 4.44]	**3.89** [3.42, 4.37]	3.76 [3.35, 4.17]	4.03 [3.65, 4.40]	4.03 [3.71, 4.35]
improvement (S5)	**4.19**** [3.89, 4.50]	3.52 [3.15, 3.88]	4.03 [3.62, 4.44]	4.03 [3.65, 4.41]	3.94 [3.54, 4.35]	**4.17** [3.86, 4.47]
prefer (S6)	**4.19*** [3.87, 4.51]	3.58 [3.20, 3.95]	**3.76** [3.32, 4.20]	3.46 [3.02, 3.89]	**3.83** [3.42, 4.24]	3.78 [3.43, 4.12]
find better (S7)	**3.21*** [2.84, 3.58]	2.60 [2.26, 2.95]	**3.00** [2.59, 3.41]	2.70 [2.32, 3.09]	3.03 [2.61, 3.45]	3.03 [2.67, 3.38]
Aggregated	**4.15***** [4.03, 4.27]	3.69 [3.55, 3.83]	**3.81** [3.65, 3.97]	3.68 [3.53, 3.84]	3.95 [3.80, 4.09]	**4.01** [3.89, 4.14]
distraction&stress (S8)	**2.45** [1.92, 2.99]	2.79 [2.35, 3.23]	**2.19** [1.77, 2.61]	2.49 [2.03, 2.94]	**2.28** [1.86, 2.70]	2.39 [1.94, 2.83]
N	31	33	37	37	36	36

Table 1. (Full statements in Table 2) Survey responses (mean and 95% confidence interval) under different tasks and conditions using a 5-point Likert scale. A score of 5 indicates strong agreement with the statement, and a score of 1 indicates strong disagreement. Confidence intervals above 3 indicates significant agreement. Asterisks indicate significant differences between conditions under independent t-test. *p $\leq$ 0.05,**p $\leq$ 0.01, ***p $\leq$ 0.001

Statements
S1. Seeing the **list of items** on the left is useful
S2. The **distribution/rank** each item is on the left is useful
S3. Clicking on an item to see **webpages** that **mentioned** it is useful
S4. Seeing **keywords** often **mentioned** together with the items is useful
S5. For complex searches, this is an **improvement** to current interfaces
S6. I would **prefer** it if my search engine have these extra features
S7. I think the extra features would help me **find better** information
S8. The interface is a **distraction** or makes searching more **stressful**

Table 2. Full statements rated by the participants. Statements were not stylised in the survey.

Figure 7. The DIST condition (left) shows the proportion of results mentioned each items in the category. The RANK condition (right) hides the distribution information, but still shows the ranking of each item.
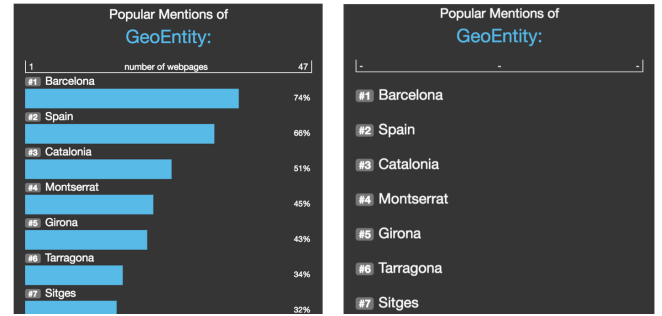
search topics and scenarios, we randomly selected 49 queries from different topics found in a exploratory search database that contains a total of 150 search tasks derived from the TREC Web track topics of the years 2009 to 2011 [14]. The queries are collected from participants conducting exploratory search for the 150 search tasks in a lab setting [42]. We submit each of the 49 queries to both SearchScape and Carrot2 and obtained a total of 98 search result structures. Finally, we present each of the task descriptions, queries, and the pair of structures to 3 crowdworkers. For each structure, we asked the crowdworkers to rate the statement "*Items in the list are of the same type*" using a 7-point Likert-scale ranging from Strongly Disagree (1) to Strongly Agree (7), and a score of 4 indicates neutral.

Results

As shown in Figure 6, participants agreed strongly that SearchScape produced structures that were coherent (M=5.20*, 95% CI[4.97, 5.42]). When compared side-by-side against structures produced by Carrot2, crowdworkers rated SearchScape as more coherent (M=(3.78 vs 5.20), SD=(1.76, 1.39)). This difference was significant based on a paired t-test (N=147, t=-7.37, df=146, p=1e-11f < 0.001). The results suggest that for a wide variety of exploratory search scenarios, SearchScape can produce structures more coherent than traditional search result clustering approaches.

Study 2: Exploratory Search

Automatically identifying important structures and displaying distribution information are the two key components that differentiate SearchScape from traditional Web search interfaces and previous work. In order to evaluate the usefulness of the SearchScape interface, and more specifically, the benefits of exposing structural and/or distribution information to searchers, we conducted a user study with 210 unique participants recruited from Amazon Mechanical Turk (age between 18 and 68, M=35, SD=11, 57% male and 43% female, mostly from the US) [25].

The experiment was a 3x2 between subjects design, with participants randomized into one of the three different search scenarios and one of the two interface conditions (rank+distribution or rank only) as shown in Figure 7. The **DIST** condition consisted of all features described in the previous section, which showed both the ranks and proportions of webpage mentions of the items using text labels and horizontal blue bars of different lengths. The **RANK** condition consisted of all the same interactive features, but hid the distribution of items by only showing their ranking. During the study, participants first filled out a pre-survey about their demographic information, we then presented each participant with their assigned search tasks. They then watched a 70 second video tutorial introducing the features of SearchScape according to their front-end conditions. They were then asked to conduct the assigned search tasks using SearchScape. As they interact with the interface, we recorded their actions for further analysis. Finally, we collected their opinions about SearchScape in a post-survey (Table 1).

Search Scenarios

Creating tasks to effectively evaluate exploratory search systems is itself a challenging task [30, 49]. Challenges include finding topics that are unfamiliar yet still relatable for the participants and motivating participants to explore and evaluate different options. We designed two exploratory tasks of different complexity, following the task design suggestions outlined in [30], such as providing ample imaginative context, and be-
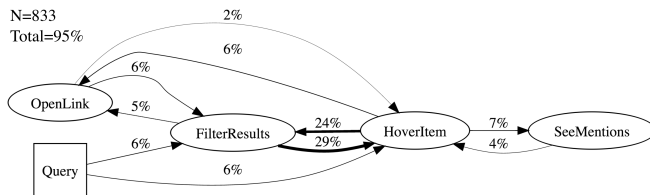
Figure 8. Top transitions between different features of participants across all conditions. Around half of the participants started by hovering items, while the other half by filtering the search results.

ing specific about the information necessary for their assigned search task. In addition, we also included one fact-finding task to see if the additional features of our interface would interfere or cause annoyance in non-exploratory scenarios. For each task, participants were given a fixed query that we have fetched and indexed in advance using the back-end of Search-Scape. Below are the three tasks we utilized in this study and their corresponding queries:

- **Barcelona**. "*You have a friend who's planning a day trip from Barcelona, and you want to help her figure out 3 destinations that you considered the best options. Give a short explanation about why you picked each of the destinations.*" **Query**: "*day trips from Barcelona*"

- **Guacamole**: "*You are in charge of making guacamole for a party you and your friends are hosting. You've decided to do some searching online to find 3 recipes you like the most. Give a short explanation about why you picked each of the recipes.*" **Query**: "*guacamole recipes*"

- **Guggenheim**: "*You need to find out who designed The Guggenheim Museum in New York City for a school report, and you decided to search the Internet for the answer. Give a short explanation for how you found the answer.*" **Query**: "*who designed the Guggenheim Museum?*"

The three tasks were designed with the following factors in mind. The **Barcelona** scenario promotes objective decision making based on the information presented, since we did not specify a persona for the person participants were gathering information for. Conversely, the **Guacamole** scenario promotes subjective decision making, since the participants are searching for recipes for themselves. In addition, the relationship between the identified structures and information needs also differs. SearchScape identified a list of *geographic locations* for the **Barcelona** query, which were mostly competing *places* for day trip destinations, whereas for the **Guacamole** query, SearchScape identified *agriculture products*, which contained mostly composable *ingredients* of different recipes. Finally, we included the simple fact-finding **Guggenheim** task to see if participants find the extra features overwhelming or detrimental for finding a single correct answer.

RESULTS AND DISCUSSION

Behavior
SearchScape augments traditional search interfaces with a interactive information landscape. First, we examine whether participants were actually using the added features. Figure 8 shows the top 10 transitions between different SearchScape

features, and participants almost always explored the overview information in the Structure Panel first, instead of opening links from the list of search results directly. Similar findings have been reported in previous work on using a faceted interface to explore a library catalog [31].

SearchScape vs Traditional Web Search Interface
SearchScape augments traditional web search interfaces with structure and distribution information, which enables interactive information exploration. The results of our studies suggested the structure and distribution information helped participants process and navigate large amounts of information efficiently. Participants were asked to compare SearchScape to their current search engine interfaces using a 5-point Likert scale (Table 1, DIST condition). Across three search tasks, participants generally agreed that SearchScape is an improvement to current interfaces (average: 4.2*, 4.0*, 4.0*, respectively, 95% CI > 3.0), and would prefer it if their search engines also had the extra features (average: 4.2*, 3.8*, 3.8*, respectively, 95% CI > 3.0). Further, they also agreed that the added features did not cause distraction or stress, even for the fact-finding Guggenheim task (average: 2.5*, 2.2*, 2.3*, respectively, 95% CI < 3.0). In the general feedback, participants also expressed how SearchScape is an improvement to the current interfaces, and how they felt more organized and/or efficient with the extra features (12.9%, 5.4%, and 15.2% of the participants in the three search tasks, respectively):

> "*I found the interface to be very effective and efficient even while hosting a large amount of information it was easy to wade through the results. When I first watched the video I was skeptical but it really did make finding information much easier to process and understand.*"

We also asked participants about the different SearchScape features (first four questions in Table 1), and participants generally agreed that the added features were useful during their search sessions. Of which, using the Structure Panel to quickly drill-down to the different subtopics was the most commonly mentioned feature in the feedback (19.8% of participants under DIST and 14.9% of participants under RANK):

> "*I like the information on the left. It has all the information i needed without making new searches all of the time. You just have to click on each link to see if it has the new information that you are looking for.*"

These results combined with the behavior logs suggest that participants in our studies were actively using the features provided by SearchScape to complete their tasks, and seemed to consider them improvements to the traditional web search interfaces. In the next subsections, we will examine more closely on how SearchScape performed under different type of search tasks using the three search scenarios we designed.

The Barcelona Scenario
Of the search scenarios, we consider the Barcelona scenario to be the most complex task of the three. First participants were faced with competing choices that were generally unfamiliar to our participants (i.e., places near Barcelona). In addition, unlike ingredients in a recipe, destinations often have multiple

important attributes relevant to the task that participants need to keep track of, such as common activities, or distance from Barcelona. For this scenario, SearchScape identified 46 *geographic locations* mentioned in the search results of the query *"day trips from Barcelona"*, and many of which were presented as day trip destinations from Barcelona in the webpages.

In general, the participants answered favorably on post-survey questions with participants in the DIST condition (the full SearchScape system) responding with an average rating of 4.15* (95% CI[4.03, 4.27]) across questions regarding the usefulness of SearchScape (Table 1). In addition, based on independent t-tests, participants in the DIST condition also responded to the survey questions significantly more favorably when compared to participants in the RANK condition, both aggregated (t=4.90, p=1.3e-06) and on four individual questions, suggesting that in this complex scenario, showing the distribution of the items was perceived as significantly more useful (S2, t=2.13, p=0.037), more improvement to the current interface (S5, t=2.78, p=0.007), was more preferred by the participants (S6, t=2.53, p=0.01), and was believed more strongly that they can find better information (S7, t=2.41, p=0.02). Qualitative results also showed similar pattern, with more participants mentioning they found the distribution being useful in the DIST condition then the number of participants in the RANK condition (12.9% vs 3.0% of participants):

"I enjoyed using the interface because for whatever topic you're searching for, it gives you great "summary" type results on the left. ... To see percentages helps a lot with pinpointing what you want to find."

Some participants in the RANK condition raised concerns about the interface being cluttered and the amount of information was too overwhelming (15.2% of participants):

"I feel like there is almost TOO much information, making it harder for me to narrow down choices. I could almost see myself spending a few hours just diving deeper into what the best spots really would be. ... with so much info, I'd begin to wonder if I had made the wrong choice."

Interestingly, less participants (3.2%) in the DIST condition (the full SearchScape) mentioned *clutter* or *overwhelming* in their feedback, even though their interface contained more information. This may suggest the counterintuitive potential for extra information to reduce perceived complexity if it promotes a sense of efficacy and understanding of the information space, and may be worth following up in future work.

The Guacamole Scenario
We designed the Guacamole scenario to be simpler and more subjective than the Barcelona task, but to still have potential for exploration and discovery. For example, searchers might want to find the most authentic recipes or discover unusual ingredients that might not be in the first few search results. For this scenario, SearchScape identified 59 *agricultural products* mentioned in the search results of the query *"guacamole recipes"*, and many of which were presented as ingredients for making guacamole in the webpages.

Participants still responded favorably overall to the Search-Scape system with a score of M=3.81* (95% CI[3.65, 3.97]) averaged across all seven questions on a 5-point Likert scale. However, unlike the Barcelona scenario, participants did not find the DIST condition to be significantly more useful than the RANK condition (Table 1). Indeed, many participants considered the Guacamole task to be too simple for the SearchScape interface to be useful (14% of participants in the Guacamole task, and none in the other two tasks):

"I'd definitely use it for more complex searches. For guacamole, no. HAHA. I smoosh avocados, add salt and lemon. No Internet search needed, but I would definitely use it for something more complex."

Further analysis of the qualitative feedback, as well as the short explanations for their answers revealed that many were finding recipes that contained uncommon but interesting ingredients, such as *bacon* (only mentioned in 6% of web pages), *mango* (11%), or *cumin* (15%). Their feedback mentioned they discovered these interesting choices from the Structure Panel (19% of participants in the Guacamole task, and around 12% in the other two):

"I like this search engine interface. I like the options that I was able to get in a quick amount of time through those keywords listed on the left. They allowed me to rethink the kind of guacamole recipes I wanted, and exposed me to new types of guacamole recipes. That added to my creativity and curiosity."

This could also have contributed to the less substantial difference between the survey responses from the two front-end conditions, since many participants were using the Structure Panel to pick out recipes that contained ingredients that they personally preferred, regardless of how frequently they were mentioned in the sources. Finally, many participants expressed needs for filtering with multiple ingredients from the Structure Panel, in order to find recipes that contained multiple ingredients or to exclude a particular ingredient (14% of participants in this task, 0% in the other two tasks):

"...(it) lacked functionality that would have made it useful, like searching for multiple things at once (yes to garlic and avocado, no to any animal products)."

These comments provided valuable insights to the further development of SearchScape to better support scenarios where key items (such as ingredients) can be composed to form larger options (such as recipes).

The Guggenheim Scenario
Differing from the two previous search scenarios, the Guggenheim scenario was about finding the correct answer to a factual question (i.e., *who designed the Guggenheim museum in New York City*). We intentionally included this scenario to see if the additional features provided by SearchScape would be detrimental for simple fact-finding searches, since SearchScape was designed for exploratory search scenarios (e.g., comparing different choices, or diving into different subtopics). For this scenario, SearchScape identified 65 *Persons* mentioned in the search results of the query *"who designed the Guggenheim*

museum", where the most frequently mentioned name *"Frank Lloyd Wright"* was the correct answer to the task.

To our surprise, participants still responded to the post-survey questions favorably to SearchScape when compared to traditional interfaces with an score of around 3.95* (95% CI[3.80, 4.09]) averaged across the seven questions on a 5-point Likert scale (Table 1). Examining the qualitative responses we found participants were using SearchScape to fact-check their prior findings, mentioning a combination of different usages, including validating information by looking at multiple sources (9% of participants), filtering search results using specific item (15% of participants), and that seeing distribution was useful (18% of participants):

> *"If I had no idea who did design the museum, while I can find it on Google Search, I would have to click several web pages to ensure that I had the correct answer...because let's face it, there's a LOT of misinformation on the internet."*

These responses highlight a key issue for supporting exploratory search on the general Web. Since sources online can be unreliable, people often rely on source credibility or cross-referencing the same information on multiple sources to make certain a previously seen information is correct. In this case, even though the correct answer is in the top first search result (a Wikipedia page), some still used SearchScape to verify their findings by examining how different sources in the search results mentioned *Frank Lloyd Wright*. This suggests even for tasks with a single correct answer, the uncertainty from showing only a single source might be addressed by surfacing information about how the returned result is cited across multiple sources.

Discussion

Developing search interfaces to better support exploration for the open web possess unique challenges that differs from search engines for well-structured corpora such as products on ecommerce sites or library catalogs. One fundamental challenge is the subjective nature of the Web. Whereas many past work on exploratory search interfaces assumed information in the corpora are objective and trusted by searchers, such as faceted search systems that showed price listings on ecommerce sites or author listing on library websites, when conducting exploratory search on the web in unfamiliar domains, information is often subjective and sources often do not agree. For example, searching for day trip destinations from Barcelona yielded many partially overlapping lists of top destinations in the search results. Therefore, exploratory searchers often rely on cross-referencing multiple sources just to validate a new piece of information is accurate or fits well with their personal preferences. For example, choosing a destination that was frequently mentioned favorably in multiple sources trusted by the user, or paying more attention to a frequently discussed topic in an online forum. Our results suggest that in scenarios where the users are trying to make sense of subjective information or contradicting sources, exposing distribution of key concepts can be beneficial in quickly identifying promising subtopics or options while not overwhelming users.

One interesting extension of this approach could be to show correlations between key items. For example, showing ingredients commonly used in the same recipes to help searchers learn which ingredients fit well together and develop their own recipes. Alternatively, correlations between key items of different categories could be useful, for example, by showing how different day trip *destinations* were often mentioned together with different types of *transportation methods* such as *buses* or *trains*. How to accurately calculate correlation between items, and to visualize them in useful ways that does not overwhelm users remains an interesting future direction.

Some participants noted cases where SearchScape did not perform well. Overall, 5% of participants had concerns about quality, some citing SearchScape picked up ingredients in the wrong recipes from webpages with multiple recipes:

> *"I think searchscape could be a good idea, however with the indexing, it's picking up words and parts of recipes in this situation that don't belong to the correct recipe."*

On the other hand, even though many participants used SearchScape to discover less popular items in the search results that they might otherwise overlooked, some still raised concerns about how the distribution information can potentially cause fixation on the popular items (5% of participants under the Barcelona task, none in the other two):

> *"Honestly I think it makes people focus too much on popularity and not enough on what's actually good. It'd be good for data research and statistics, maybe, but that's not really the sort of thing most people are looking for."*

How to balance using distribution information to show the prominence of options while avoiding over fixation on the top key items remains an interesting future design challenge. A deeper semantic understanding of why certain options are cited in a particular way might support landscapes that could be organized by context instead of popularity. For example, by clustering destinations popular for first-time visitors or explaining how a choice is popular because it has a famous church (vs. a sunny beach). Alternatively, different means of depicting the popularity of items (e.g., bucketing items with similar popularity while hiding absolute popularity within a bucket, or visualizing the uncertainty of distributions) might ameliorate or make people more aware of such effects. Overcoming these challenges will be important to approaches helping people aggregate and make decisions from multiple sources.

Conclusion

In this paper we introduced SearchScape, a novel approach to building information landscapes. SearchScape automatically generates coherent topic structures from a set of web pages and surfaces the distribution of key items to help users understand and explore those pages. Our approach demonstrates the value of combining the precision of curated but sparse knowledge bases with the recall of large scale but low accuracy word vectors to ensure both coherency and coverage. We believe SearchScape represents a first step towards aggregating and interacting with the web in a way that addresses the subjective and fragmented nature of information today.

REFERENCES

1. Ricardo Baeza-Yates, Berthier Ribeiro-Neto, and others. 1999. *Modern information retrieval*. Vol. 463. Page 513. ACM press New York.

2. Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, 1247–1250.

3. Christine Susan Bruce. 1999. Workplace experiences of information literacy. *International journal of information management* 19, 1 (1999), 33–47.

4. Robert G Capra and Gary Marchionini. 2008. The relation browser tool for faceted exploratory search. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 420–420.

5. Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 335–336.

6. Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. 2010. Toward an Architecture for Never-Ending Language Learning.. In *AAAI*, Vol. 5. 3.

7. Joseph Chee Chang, Aniket Kittur, and Nathan Hahn. 2016. Alloy: Clustering with crowds and computation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3180–3191.

8. Joseph Z Chang, Richard Tzong-Han Tsai, and Jason S Chang. 2009. WikiSense: Supersense Tagging of Wikipedia Named Entities Based WordNet.. In *PACLIC*. 72–81.

9. Hsinchun Chen, Haiyan Fan, Michael Chau, and Daniel Zeng. 2001. MetaSpider: Meta-searching and categorization on the Web. *Journal of the American Society for Information Science and Technology* 52, 13 (2001), 1134–1147.

10. Lydia B Chilton, Greg Little, Darren Edge, Daniel S Weld, and James A Landay. 2013. Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1999–2008.

11. Jason Chuang, Daniel Ramage, Christopher Manning, and Jeffrey Heer. 2012. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 443–452.

12. Van Dang and W Bruce Croft. 2012. Diversity by proportionality: an election-based approach to search result diversification. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 65–74.

13. Marian Dörk, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. 2012. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2709–2718.

14. Matthias Hagen, Martin Potthast, Michael Völske, Jakob Gomoll, and Benno Stein. 2016. How Writers Search: Analyzing the Search and Writing Logs of Non-fictional Essays. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16)*. ACM, New York, NY, USA, 193–202. DOI: http://dx.doi.org/10.1145/2854946.2854969

15. Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The Knowledge Accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2258–2270.

16. Marti Hearst. 2006a. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search*. Seattle, WA, 1–5.

17. Marti Hearst. 2009. *Search user interfaces*. Cambridge University Press.

18. Marti Hearst, Ame Elliott, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ka-Ping Yee. 2002. Finding the flow in web site search. *Commun. ACM* 45, 9 (2002), 42–49.

19. Marti A Hearst. 2006b. Clustering versus faceted categories for information exploration. *Commun. ACM* 49, 4 (2006), 59–61.

20. Marti A Hearst and Jan O Pedersen. 1996a. Reexamining the cluster hypothesis: scatter/gather on retrieval results. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 76–84.

21. Marti A Hearst and Jan O Pedersen. 1996b. Visualizing information retrieval results: a demonstration of the TileBar interface. In *Conference Companion on Human Factors in Computing Systems*. ACM, 394–395.

22. Marti A Hearst and Emilia Stoica. 2009. NLP support for faceted navigation in scholarly collections. In *Proceedings of the 2009 workshop on text and citation analysis for scholarly digital libraries*. Association for Computational Linguistics, 62–70.

23. Xiaoran Jin, Marc Sloan, and Jun Wang. 2013. Interactive exploratory search for multi page search results. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 655–666.

24. Melanie Kellar, Carolyn Watters, and Michael Shepherd. 2007. A field study characterizing Web-based information-seeking tasks. *Journal of the Association for Information Science and Technology* 58, 7 (2007), 999–1018.

25. Aniket Kittur, Ed H Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–456.

26. Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. 2015. Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4189–4198.

27. Christian Kohlschütter, Peter Fankhauser, and Wolfgang Nejdl. 2010. Boilerplate detection using shallow text features. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 441–450.

28. Teuvo Kohonen. 1998. The self-organizing map. *Neurocomputing* 21, 1 (1998), 1–6.

29. Weize Kong and James Allan. 2014. Extending faceted search to the general web. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 839–848.

30. Bill Kules and Robert Capra. 2008. Creating exploratory tasks for a faceted search interface. *Proc. of HCIR 2008* (2008), 18–21.

31. Bill Kules, Robert Capra, Matthew Banta, and Tito Sierra. 2009. What do exploratory searchers look at in a faceted search interface?. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 313–322.

32. Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, and others. 2008. *Introduction to information retrieval*. Vol. 1. Cambridge university press Cambridge.

33. Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

34. Gary Marchionini and Ben Brunk. 2006. Towards a general relation browser: A GUI for information architects. *Journal of Digital Information* 4, 1 (2006).

35. Gary J Marchionini, Gary Geisler, and Ben Brunk. 2000. Agileviews. In *Proceedings of the ASIST Annual Meeting*, Vol. 37. 271–280.

36. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

37. Dan Morris, Meredith Ringel Morris, and Gina Venolia. 2008. SearchBar: a search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1207–1216.

38. Brendan O'Connor, Michel Krieger, and David Ahn. 2010. TweetMotif: Exploratory Search and Topic Summarization for Twitter.. In *ICWSM*. 384–385.

39. Emily S Patterson, Emilie M Roth, and David D Woods. 2001. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition, Technology & Work* 3, 4 (2001), 224–237.

40. Jaakko Peltonen, Kseniia Belorustceva, and Tuukka Ruotsalo. 2017. Topic-Relevance Map: Visualization for Improving Search Result Comprehension. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 611–622.

41. Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation.. In *EMNLP*, Vol. 14. 1532–1543.

42. Martin Potthast, Matthias Hagen, Michael Völske, and Benno Stein. 2013. Exploratory Search Missions for TREC Topics. *EuroHCIR* 1033 (2013), 7.

43. Daniel E Rose and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*. ACM, 13–19.

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

45. Jerzy Stefanowski and Dawid Weiss. 2003. Carrotˆ2 and Language Properties in Web Search Results Clustering. In *AWIC*. Springer, 240–249.

46. Jaime Teevan, Kevyn Collins-Thompson, Ryen W White, Susan T Dumais, and Yubin Kim. 2013. Slow search: Information retrieval without time constraints. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*. ACM, 1.

47. Jaime Teevan, Susan T Dumais, and Zachary Gutt. 2008. Challenges for supporting faceted search in large, heterogeneous corpora like the web. *Proceedings of HCIR* 2008 (2008), 87.

48. Simon Tretter, Gene Golovchinsky, and Pernilla Qvarfordt. 2013. SearchPanel: A Browser Extension for Managing Search Activity.. In *EuroHCIR*. 51–54.

49. Ryen W White, Gheorghe Muresan, and Gary Marchionini. 2006. Report on ACM SIGIR 2006 workshop on evaluating exploratory search systems. In *ACM SIGIR Forum*, Vol. 40. ACM, 52–60.

50. Max Wilson, Alistair Russell, Daniel A Smith, and others. 2006. mSpace: improving information access to multimedia domains with multimodal exploratory search. *Commun. ACM* 49, 4 (2006), 47–49.

51. David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 189–196.

52. Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 401–408.

53. Mohamed Amir Yosef, Johannes Hoffart, Ilaria Bordino, Marc Spaniol, and Gerhard Weikum. 2011. Aida: An online tool for accurate disambiguation of named entities in text and tables. *Proceedings of the VLDB Endowment* 4, 12 (2011), 1450–1453.

54. Oren Zamir and Oren Etzioni. 1999. Grouper: a dynamic clustering interface to Web search results. *Computer Networks* 31, 11 (1999), 1361–1374.

55. Hua-Jun Zeng, Qi-Cai He, Zheng Chen, Wei-Ying Ma, and Jinwen Ma. 2004. Learning to cluster web search results. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 210–217.