SearchLens: Composing and Capturing Complex User Interests for Exploratory Search

Joseph Chee Chang
Carnegie Mellon University
Pittsburgh, PA
josephcc@cs.cmu.edu

Nathan Hahn
Carnegie Mellon University
Pittsburgh, PA
nhahn@cs.cmu.edu

Adam Perer
Carnegie Mellon University
Pittsburgh, PA
adamperer@cs.cmu.edu

Aniket Kittur
Carnegie Mellon University
Pittsburgh, PA
nkittur@cs.cmu.edu

ABSTRACT

Whether figuring out where to eat in an unfamiliar city or deciding which apartment to live in, consumer generated data (i.e. reviews and forum posts) are often an important influence in online decision making. To make sense of these rich repositories of diverse opinions, searchers need to sift through a large number of reviews to characterize each item based on aspects that they care about. We introduce a novel system, SearchLens, where searchers build up a collection of "Lenses" that reflect their different latent interests, and compose the Lenses to find relevant items across different contexts. Based on the Lenses, SearchLens generates personalized interfaces with visual explanations that promotes transparency and enables deeper exploration. While prior work found searchers may not wish to put in effort specifying their goals without immediate and sufficient benefits, results from a controlled lab study suggest that our approach incentivized participants to express their interests more richly than in a baseline condition, and a field study showed that participants found benefits in SearchLens while conducting their own tasks.

CCS CONCEPTS

• **Information systems** → **Query representation**; **Search interfaces**; • **Human-centered computing** → *Graphical user interfaces*; *Web-based interaction.*

KEYWORDS

Sensemaking; Exploratory Search Interfaces; Results Visualization

ACM Reference Format:
Joseph Chee Chang, Nathan Hahn, Adam Perer, and Aniket Kittur. 2019. SearchLens: Composing and Capturing Complex User Interests for Exploratory Search. In *24th International Conference on Intelligent User Interfaces (IUI '19), March 17–20, 2019, Marina del Rey, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3301275.3302321

IUI '19, March 17–20, 2019, Marina del Rey, CA, USA

ACM ISBN 978-1-4503-6272-6/19/03…$15.00
https://doi.org/10.1145/3301275.3302321

1 INTRODUCTION

Whether figuring out where to eat in an unfamiliar city or deciding which apartment to live in, people often rely on reading online reviews and forum posts to make predictions about how well different options might match their personal interests and needs. With the proliferation of online reviews, people now have instant access to millions of online reviews from people with varying perspectives and interests. It was estimated that in 2013 Amazon provided shoppers access to more than one million reviews for just their electronics section [34], and in 2016 Yelp provided around 250,000 reviews for over 6,000 restaurants for the city of Toronto alone [29]. Having access to this rich repository of diverse perspectives based on the past experiences of others has the potential to empower consumers to understand their choices thoroughly and make better decisions for themselves without being overly influenced by marketing and branding [15].

Unfortunately, it is often difficult for users to be able to quickly and efficiently match their personal interests to the large amount of information available for each potential option. One problem is that simple star ratings are often not sufficient, and recent research has shown reviews often play an important role in users online purchase decisions [21, 36]. For example, restaurants might receive negative reviews for its simple decor and lack of good ambiance, but some searchers might value more the authenticity of the food or whether vegan options were available on the menu. Subsequently, finding, reading, and evaluating relevant reviews is time-consuming and challenging. Users have to manually parse through the reviews for each restaurant and match them to their personal interests (e.g., kid friendly, authentic Indian cuisine). They then have to track which restaurant meets which criteria, and if they discover and add any additional criteria, they must back-fill that information and re-evaluate previously seen restaurants. Furthermore, once a user has finished searching, the work performed discovering and evaluating factors is lost, resulting in having to start from scratch even if a similar need arises in the future. For example, a traveler who has spent a lot of time choosing between ramen restaurants in Los Angeles must start from scratch evaluating ramen restaurants in Toronto, despite having discovered several important factors (e.g., thickness and chewiness of noodle,

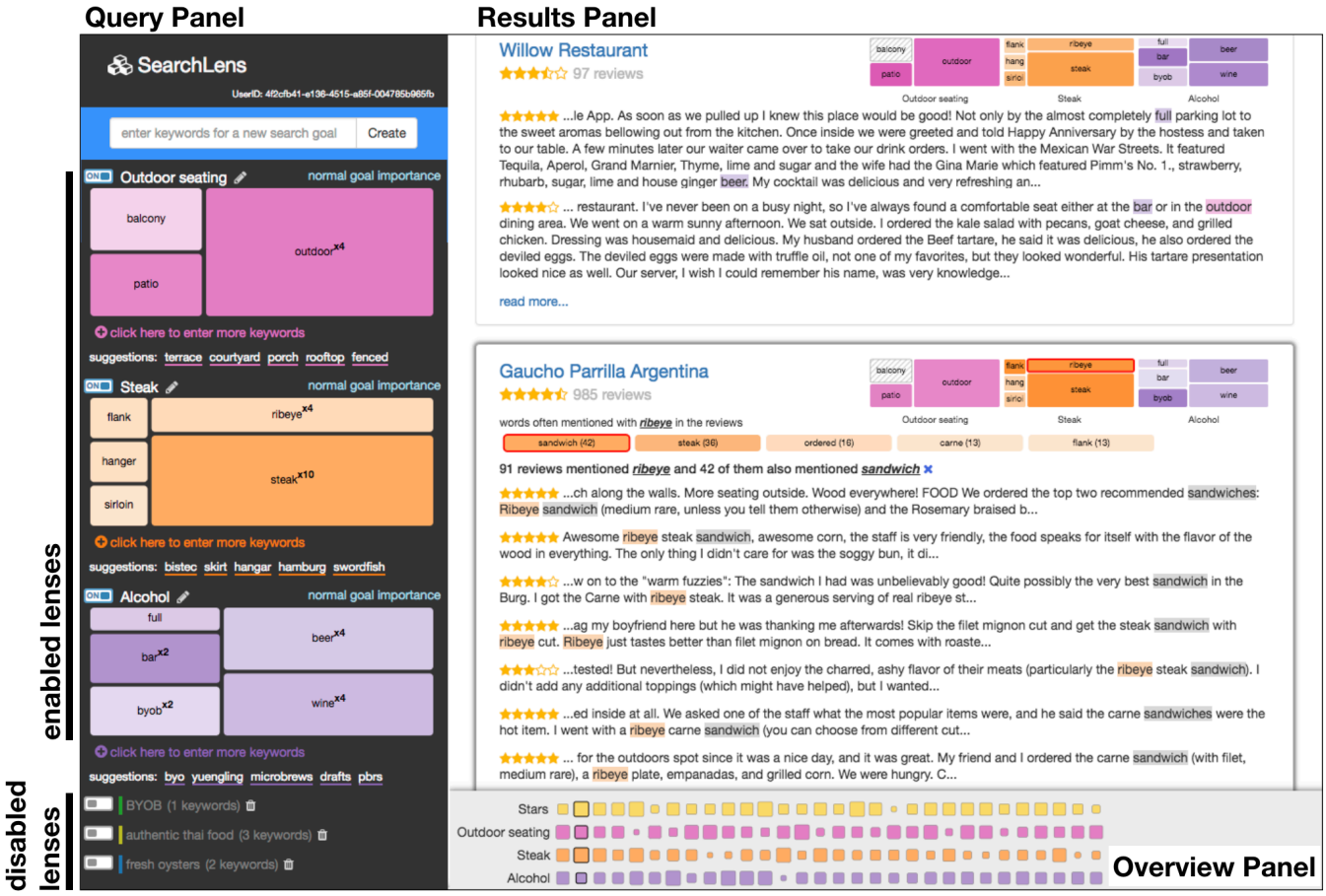

Figure 1: An overview of the SearchLens system. The Query Panel on the left allows users to specify search topics, or Lenses, by specifying multiple keywords. The keywords for a given Lens are show in colored cells sized by importance (weight). Lenses can be freely disabled or enabled for different scenarios. The Results Panel on the right shows a ranked list of search results that best match the enabled Lenses from the searcher. The same visualization for specifying queries are then used for explaining how each result matches with user's interests and mental model, and also serve as an interactive navigation for filtering mentions of specific keywords. The Overview Panel at the bottom shows a collapsed version of the cells that allows for quick comparison between results.

whether the broth is simmered for a long time with pork bones) that will be similarly utilized in their decision making.

Getting users to specify these nuanced interests and preferences has been a long standing challenge. Several decades of research have explored ways of getting users to externalize their interests [3, 30], for example by: using prompt and text field designs that promote longer query terms [4, 19], asking for relevance feedback on the results provided [38, 41, 42], or explicitly asking users to build up sets of query terms of different topics [27, 28]. There are two primary challenges brought up by this work. First, users have trouble specifying their interests, which includes challenges with identifying query terms that were neither too general nor too specific; providing more than a few terms (even when longer queries were more likely to lead to useful results); and learning terms from the content, rather than knowing them all beforehand [4, 42]. The other main issue found is that it is very difficult to get users to put in the work to externalize their interests, either as query terms or as explicit feedback, due to perceptions that the work will not be sufficiently paid off in the future or not understanding how their work will affect their results.

To tackle this issue of capturing, leveraging and exposing user interests, we introduce SearchLense, where users construct externalized representations of their interests as "Lenses". Lenses are leveraged as an explanatory tool, providing users with a way to quickly parse, understand and make judgments based on the vast amount of review data instantaneously. Additionally, Lenses can be reused in different contexts and combined in different configurations. In the example above, imagine a system which could capture the factors that the traveler found important for ramen in Los Angeles and reuse them to quickly make a confident, personalized decision about ramen in Toronto. If traveling to Toronto with kids, a "kids" Lens might also be added with factors such as whether

the restaurant typically has long lines and how many seats it has. These persistent Lenses could be useful in a variety of situations beyond reviews, ranging from academics keeping track of interesting research topics; travelers deciding which places to visit in an unfamiliar city; consumers deciding between products; lawyers doing case discovery; or voters tracking important issues. We explore this problem in the context of restaurant reviews, conducting a controlled lab study with 29 participants to examine if our visual interface for explanation and exploration is effective in providing immediate benefits to elicit rich interest expressions from the users. Additionally, we performed a three day field deployment study with 5 participants to explore the benefits of Lenses when users were conducting their own tasks. Results suggest that our prototype system SearchLens was able to learn richer representations of its users' interests when compared to a baseline system by allowing users to fluidly capture, build, and refine Lenses to reflect their interests and needs, and that the user-generated interfaces can be reused over time and transfer across contexts.

2 RELATED WORK

Past research has proposed a variety of approaches to collecting, modeling and leveraging users' interests and intents through both interface design and computation. Our work builds on this diverse of literature by allowing the system to learn the personal interests of the users through interaction to retrieve relevant data, and present data based on its understanding of the different users. This allows us to elicit structures that can be reused across different contexts and tasks and are more nuanced and personalized to each users when compared to traditional search structures such as search results clustering or pre-compiled facets.

2.1 Eliciting and Modeling Interests and Intents

A significant topic of research has been interfaces that can collect, explicitly or implicitly, the personal goals and interests of users as they search for information and modify their viewing of content correspondingly. While there is extensive literature on doing so in the context of personalized search and re-ranking of search results (e.g., [7, 8, 44, 45]), we focus here on work that enables more interactivity and transparency of users' interests to support more complex searching. One such thread lies in the collection of users' interests through keywords or interest vectors into an agent or user interest or intent model. This includes seminal work such as WebMate [11], which built up an agent composed of sets of TF-IDF [49] vectors to represent the user's different interests. Similar to WebMate, we aim to build collections of terms that represent the user's interests, but focus on explicit user selection of those sets, and making them explainable and composable. Interestingly, WebMate's "Trigger Pair Model" which looked at co-occurrence of words within a sliding window across a set of documents can be seen as a precursor to the word vector model that we use for keyword suggestions. More recent work in this vein includes user modeling of concepts, such as AdaptiveVIBE [1] and Intent Radar [38], which include two dimensional visualizations of documents and their relation to the user's inferred interests. Our work builds upon these but aims at increasing the richness of the structure, nuance, and specificity of the user's expression of interests. Specifically, our Lenses, composed of multiple keywords that can capture multiple levels of specificity, can be themselves composed into more complex expressions and reused across different contexts and tasks. We also focus on supporting users in the discovery process of building good terms that are discriminatory and explanatory.

2.2 Faceted Search Interfaces

In cases were item metadata is available, faceted search is perhaps the most commonly used search interface [23], 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 relevant sources and exclude irrelevant sources from their search results, researchers have also found faceted search interfaces beneficial in exploratory scenarios, where searchers are less certain about their information needs [24, 53]. Although these expert structures can cover many general and objective aspects of the data for navigation or filtering (such as price range and location of restaurants, or authors and publishing year of books), it can be difficult to scale to nuanced and subjective criteria that vary between users. For example, different users may have different ideas about what makes a good date night restaurants. SearchLens instead uses a novel interactive interface that allowed searchers to specify and refine their personal and idiosyncratic search topics in a composable and dynamic way that enables personalized visual explanation and deeper exploration of search results.

2.3 Automatically Identifying Structure

In cases where metadata is not available, computationally inferring task-specific fine-grained facets (e.g., day trip destinations) remains a challenge [6, 46]. Researchers have long explored ways to extract structure by clustering search results using machine learning [26, 54, 55], lexical and HTML patterns [32], crowdsourcing [9, 10, 13, 22], or interaction techniques [26, 27, 27]. The Scatter/Gather system in particular, allowed users to navigate and explore large collections of documents using an interactive hierarchical clustering paradigm [26]. However, a number of papers [10, 14, 25] point to the fact that automatically techniques can often produce incoherent structures that are difficult to comprehend by users. Further, the automatically generated structures were designed to reflect the characteristics of data for exploration, and does not take into account the interests of the users when trying to infer structures. While they may be effective for exploration and navigation, they typically provide little support for allowing users to externalize idiosyncratic search goals, and do not present data in ways that reflect the interests of the users. In SearchLens, we allow the user to explicitly express their idiosyncratic search goals using structured queries we call Lenses. The system provides a novel interactive visualization that allows users to both refine their query structures, and help them interpret the results based on their interests. This draws from degree of interest (DOI) functions used in the visualization literature, which drive human attention to areas of the information that have high expected utility for a user's expressed or inferred goals [20, 47], and suggest a bottom-up and iterative,

user-driven process of searching in which the user is continually updating the expected utility function.

2.4 Concept Discovery and Evolution

Research in interactive machine learning has also explored techniques to support data annotators or searchers in discovering and externalizing useful concepts when working in unfamiliar domains. For example, Alloy used a *sample-and-search* technique to categorize textual datasets with novice crowdworkers where they first explore the space of information through sampling items in the dataset to discover useful categories, then externalize each category using a set of query terms and search for other relevant items [10]. Past work has further suggested that the working concepts of an annotator may change over time as new items were examined [33]. Different techniques that can better support this concept evolution process were proposed, such as structured labeling [33], crowd collaboration [9], and interactive visualization [12]. These point to the importance of providing mechanisms that allow users to not only discover and define concepts based on data, but also to easily evolve their concept representations during the process of exploring an unfamiliar domain. In a study more closely related to our work, CueFlik allowed image searchers to define conceptual filters (e.g., listing only *action shots* when searching for *baseball* images) by labeling items in a search result list as positive or negative training examples [18]. Previously defined filters are persisted and can be applied to future searches (e.g., applying the same *action shots* filter when searching for *football* images), but evolving existing conceptual filters would require recreating filters from scratch or re-labeling items in existing filters. Our work builds this past work to allow exploratory searchers in unfamiliar domains to discover concepts of interests from data and externalize these concepts in the form of "Lenses" that can be continually refined. Finally, the Lenses are persisted across different search sessions similar to [18], and can be modified and composed for different scenarios and goals.

3 SYSTEM DESIGN

The key motivating concept behind SearchLens was providing users with a way to externalize their complex interest profiles in a way that could be useful for ranking, explanation, and transference to different contexts. We aimed to make the interface simple and transparent but also powerful enough to express higher level, abstract concepts and differing levels of specificity. To do this, we introduce the idea of "Lenses": reusable collections of weighted keywords that contain "honest signals" of a user's interests that can be composed in different configurations to match a user's current needs. The Lenses that are enabled in a particular configuration drive various visualization and explanation elements to help the user understand how the information space meets their needs, and also whether they need to fix or reformulate their Lenses.

A key challenge here is incentivizing users to create rich Lenses by providing sufficient and immediate benefits. For this, SearchLens provides visual explanation of items in the search results based on users' Lenses, which also serves as an interface for deeper exploration. When a new Lens is created or enabled, its visual representation appears on the interface for each item, allowing users to understand how well each item matches with the Lens, and how frequently each keyword is mentioned in its reviews. To further explore each item, users can click on keywords in each Lens to see relevant reviews.

A typical use case is as follows. A user just moved to Pittsburgh and wants to go out to eat ramen. She starts by pulling up a restaurant she knows she likes from Toronto and goes through some of the reviews, noticing that the reviews of her favorite tonkatsu ramen mention interesting signals such as "bone" and "umami" and adds them to her ramen Lens along with other useful words such as "tonkatsu", "ramen", "bowl", etc. After checking to see that her Lens is bringing up other restaurants that serve ramen she likes in Toronto and adding a few of their terms to her Lens, she switches to Pittsburgh and looks for how her Lens is being used. She also activates her drinks Lens, which she's built up over time to incorporate her particular interests in unfiltered sakes as well as hoppy beers. Using the Lenses, she quickly see which ramen restaurants in the results list serve unfiltered sakes and/or hoppy beers. To further explore her different options, she can click on each keyword in her Lenses to filter relevant reviews. For example, "tonkatsu" might be often mentioned with "spicy" in one restaurant, and "creamy" in another, allowing her to further differentiate her options based on aspects that she cares about.

The following subsections describe the designs of the SearchLens system. We will first present our concept of "Lenses," and how users can use SearchLens to fluidly express and refine their different nuanced interests, and freely compose their Lenses for different contexts. We will also describe how search Lenses can provide immediate benefits once specified, providing users visual explanation of each item in the search results, and also an interface for deeper exploration.

To test our prototype system in a realistic and manageable setting, we focused on the domain of restaurant reviews where personalization and searching with multiple goals is especially important. We used a subset of the dataset from the Yelp challenge [29] that included local business in 11 metropolitan areas.[1] Restaurants and reviews were selected by string matching on the *city* field of each restaurant available as metadata in the Yelp challenge dataset, resulting a subset of 48,485 restaurants and 2,577,298 reviews. This allows us to explore how user-specified Lenses can be composed and reused for different scenarios, as well as for the same scenario across different cities. In addition, we also use the same data to train a Word2Vec model [35] for generating Lens-specific query term suggestions.

3.1 Capturing User Interests with Lenses

Our goal was to develop a way to elicit users' interests which is both highly expressive and immediately beneficial. To explore the natural discovery and collection of users' interests we conducted a preliminary study in which we asked people to read reviews of their favorite restaurants on Yelp and see if they could identify terms that were good indications of their interests. We discovered that people found it intuitive to identify many different terms that

[1] Pittsburgh, Charlotte, Phoenix, Las Vegas, Toronto, Montréal, Mesa, Mississauga, Cleveland, Scottsdale, and Edinburgh.

matched their interests. Many of these terms were not simply general descriptors (e.g., "good", "tasty") but instead terms they considered indicative of matching their personal interests (e.g., an authentic ramen restaurant would include terms talking about the thickness of the noodles; a popular restaurant might be less favored if it also had very long lines). Terms also fell into different classes of factors users were interested in (e.g., service vs. food quality vs. parking). Users seemed to focus on finding reviews that mentioned these terms and use them in their decision making.

Based on these initial findings we developed a system for users to easily collect terms from reviews into "Lenses" and to use those terms to identify and summarize reviews that mentioned those terms. Similar to [27], we enable users to search with multiple Lenses at the same time. However, our Lenses differ from traditional search queries or faceted metadata in several important ways.

First, our system encourages the iterative development of Lenses as the user explores. A common activity in online exploratory search involves discovering new and interesting aspects from data. SearchLens aims to make it easy for users to add new Lenses and improve existing ones throughout their searching process. Users can create a new Lens by specifying a set of keywords using the text field in the Query Panel on the left (Figure 1). As users browse the results on the right, they might find some keywords in their Lenses were too general to be useful (e.g., "tasty broth"), and find discover more indicative keywords either from prior knowledge or from the reviews (e.g., "rich and thick broth"). In this case, users can refine their Lenses by adding new keywords using three different interactions, each for a different scenario. First, users can click on the plus icon under each Lenses to enter new keywords in a Lens specific text field. Second, as users discover more indicative keywords or new topics of interests from the reviews, they can highlight the keywords and use a context menu to add them to an existing Lens. In addition, a list of keyword suggestions are also listed under each Lens based on current keywords (Figure 2). Users can hover over each suggestion to see example mentions, and click on the keyword include it. This allows users to assess the usefulness of the suggestions, such as to avoid ambiguous terms. The Lens-specific suggestions were computed based a word semantic model described in the below subsection. To remove a keyword, users can click on its cell and select remove keyword in the context menu.

Once constructed, Lenses can then be used to visually inspect and adjust their "projections" onto the data. Lenses are represented visually as boxes subdivided into cells, one for each term the user added. Initially, all keywords in the same Lens have equal importance (as reflected by being the same size), but users can click on each cell to select different importance in a context menu (x1, x2, x4, x10, exclude) to better reflect their personal preferences. The size of the cells will adjust accordingly to reflect the importance of each keyword (excluded keywords are represented using fixed size cells with a unique pattern fill). The shade of each cell shows the overall frequency of each keyword in the top 30 search results (Figure 1, Query Panel). This allows the user to get a sense of how items in the corpus reflect their mental representation of each topic. For example, a large cell with very light shade represents a concept that the users deemed as an important feature of the topic, but was rarely found in the results. Surfacing this information ensures user

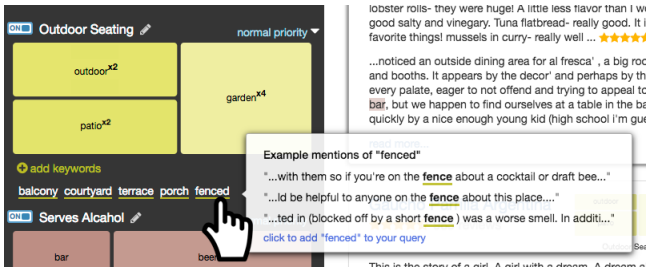

Figure 2: SearchLens provides keywords suggestions based on currently Lenses. Hovering shows a preview panel with mentions of the suggested keyword, allowing users to better understand the effect of adding the suggested keyword. In this case, SearchLens suggested balcony, terrace, fenced, and other keywords for the "Outdoor Seating" Lens. However, further inspection showed that fenced may not be a indicative keyword for the purpose of this Lens.

are aware of how useful each of their keywords are, and can refine their Lenses to include more indicative keywords.

As Lenses and terms are collected a user can over time build up a repository that reflects her personal interests. Each Lens can be disabled and re-enabled and are persisted across different visits to the SearchLens interface, with disabled Lenses are listed at the bottom of the Query Panel (Figure 1). Various combinations of Lenses can be activated depending on the goal and context. For example, for a date night a user might enable their personalized Lenses for "cozy and intimate", and "vegan", or for a weekday lunch activate their Lenses for "fast casual", "vegan", and "easy parking". Although our main thrust in this paper is exploring the viability of this approach, further work will likely be needed to understand as Lenses accumulate how to scale them. For example, in the current prototype all disabled Lenses are shown, but future systems could further contextualize Lenses by inferring the task context (e.g., what type of item someone is searching for).

3.1.1 Keyword Suggestions. While creating a new Lens, listing all keywords from prior knowledge can be mentally taxing and have poor recall. To further reduce the required effort for building expressive Lenses, SearchLens generates Lens-specific keyword suggestions. As an example, when a user created an "Outdoor Seating" Lens with only three keywords ("outdoor", "patio", and "garden"), SearchLens automatically suggested relevant keywords including "balcony", "courtyard", and "terrace" (Figure 2). To do so, we trained a Word2Vec model [35] with 300 dimensions using the entire Yelp dataset of 2,577,298 reviews. The trained word model can project words onto a semantically meaningful vector space, which in turn allows for measuring semantic similarity between words. Alternatively, it can also be used to find a set of words that are semantically similar to a given term by searching in the vector space of nearby words. To generate Lens-specific keyword suggestions, we first project all its keywords in a Lens onto the vector space and calculate the average vector to obtain a list of similar terms around the average vector. To further increase the chance of presenting useful and discriminatory search terms, we only used

terms that appeared more than 50 times in the corpus, were mentioned in reviews of more than three restaurants, and were mentioned in less than 40% of all restaurants.

3.2 Interest-driven Explanation

Persistent, decoupled user interest models would be beneficial to users the long run by providing separate reusable and recomposable interests across multiple search sessions. However, without immediate and perceivable benefits, users typically are not willing to spent extra effort expressing their separate interests for future tasks. For this, SearchLens uses each user's Lenses to provide visual explanation of each item in the search results. This is based on our approach of allowing users to express their multiple topics of interest separately, which enables SearchLens to distinguish between keywords of different topics and opens the possibility of visualizing each result according to users' interests in easy-to-interpret ways. Explanation is especially important for supporting searching with multiple interests, as it can be difficult for the users to understand which interests and keywords were associated with each result. Consider traditional search interfaces that only offer a short snippet for each result as explanation. These short summaries provide little support for personalized interpretation beyond a few highlighted query terms and their context. Even if users listed keywords of many different topics at once, the linear result list also provides little information about each result beyond their overall relevance ranking.

One obvious approach to explaining items in the search results is to surface mentions and statistical information, such as mention frequencies, at the topic level. For example, [28] visualized the overall frequency of different search terms in different topics for each search result, and [27] visualized the mention locations of different topics within each document. Visualizing at the topic level allowed these systems to provide mechanisms for specifying many topics and keywords, while at the same time visualized deeper information about each result in a way that matches the mental model of the searchers. However, visualizing at the topic level can be prohibitive for keyword-level operations, such as query reformulation and assigning importance levels to different keywords based on their frequencies.

SearchLens supports rich explanation at the topic and keyword level through its user-specified Lenses. Explanation occurs by showing the each Lens visualization from the Query Panel (Figure 1) on each result and adjusting the term shading to correspond to the frequency of the term within that search result (Figure 3). By using identical colors and layouts of each Lenses, and showing result-specific keyword frequencies, users can quickly interpret how each result matches with their different interests at both the topic and at the keyword level using a familiar visualization. As an example, Figure 3 shows how a user might examine two restaurants in a search result list using her Lenses for "Steak", "Alcahol", and "Outdoor Seating". At the topic level, both restaurants matched well with her Steak Lens rendered in dark shades that incorporated her stronger preference for "ribeye" steak, and also also her other interests such as "flank" steaks. She can also see that the first restaurant matched her Outdoor Seating Lens better than the second one. Looking at the same Alcohol Lens at the keyword level, she can

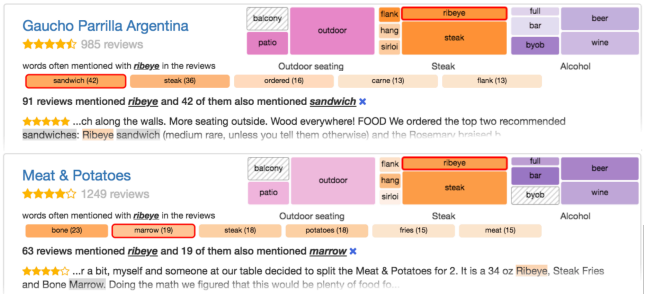

Figure 3: The visual explanation and exploration feature allows comparison of results at different levels of granularity using a familiar interface used for specifying queries - at the levels of Lenses, keywords, co-occurring terms, and mentions, allowing users to query with multiple Lenses at the same time, while still being able to comprehend how each result matches their different Lenses.

easily see that the two restaurants matched differently with her "Alcahol" Lens where the first one has many mentions of "byob" in the reviews and the second one with many mentions of "beer" and "bar" instead.

Finally, to provide a more compact, higher-level, topic-centric overview of all restaurants in the search results, SearchLens collapses the colored cells for each Lens into a single cell similar to [28]. The size of each cell to shows the overall frequencies of keywords in different Lenses for each result (Figure 1). This allows users to get a quick overview of restaurants in the search results, and compare different options at the topic level using the Overview Panel at the bottom.

3.3 Supporting Deeper Exploration of Items

In addition to acting as a visual explanation for each result, the cells in the visualization also act as a navigation tool for deep exploration at the keyword level. Users can explore mentions of different keywords by clicking on its corresponding cell and the summary will update in real-time to show a list of its mentions. In addition, the Lens also shows the top co-occurring words that were frequently mentioned near the selected keyword as overview and deeper navigation, a strategy found useful in exploratory scenarios by prior work [16, 17, 37]. As an example, Figure 3 shows the how the Lenses allow users to explore and compare options at different levels of granularity. At the highest level, users can use the shading of different cells to see that the *Outdoor Seating* Lens has more mentions in the first restaurant (Figure 3). Searchers can use the shading of individual cells to compare options at the keyword level. For example, the term "BYOB" was frequently mentioned in reviews for the first restaurant, but did not show up in reviews for the second restaurant. Finally, clicking on the individual cells allows users to explore mentions of its corresponding keywords and words that were frequently mentioned together. For example, when exploring mention of the work "ribeye" for both restaurants, SearchLens shows that there were many mentions of "sandwich"

near the word "ribeye" for the first restaurant, and many mentions of "bone marrow" near "ribeye" for the second restaurant (Figure 3).

3.4 Indexing and Ranking

Traditionally, faceted search systems typically combine factors from multiple facets for ranking using disjunctions (factors within facets, such as brands selected by the user on a shopping website) and conjuctions (factors between facets, such as brands and price ranges). In an early iteration of SearchLens, we tested using the Boolean OR operator between keywords within the same Lens, treating keywords within the same Lens as synonyms while ranking. However, users reported this approach lead them to restaurants that poorly reflected their Lenses, as some restaurants may have many mentions of few keywords in a Lens, but very few mentions of other keywords. Fundamentally, unlike faceted search systems, different keywords in the Lenses typically describe a criteria as a whole. For example, an authentic ramen Lens might contained keywords describing creamy bone broth and freshly made noodles. In this case, the different keywords combined represented what the user considered good ramen restaurants, instead of as alternate options in a facet (such as a set of preferred brands). In a later iteration, we switched to Okapi BM25 for ranking that used inverse document frequencies to weight keywords instead of eliciting importance rating from the users. However, users reported unable to construct Lenses that reflect their priorities and unable to construct expressive Lenses that lead to useful results. This lead to the current iteration where we used a modified version of the standard Okapi BM25 ranking function to combine keywords across Lenses [40], which by default considers both term frequency and document frequency to rank documents similar to TF-IDF ranking function, but also adjust for the length of each documents.

We modify the Okapi BM25 ranking function to account for the importance levels specified by users in the following ways. By default, Okapi BM25 uses the inverse document frequencies to weight each keywords, with the motivation that words appearing in many documents tend to be less important. Since in SearchLens users can specify keyword importance using the interactive visual explanation, we instead weight each keyword according to their user-specified importance level. By default, SearchLens assume each Lens is equally important, and normalizes the weights of keyword q in a Lenses ℓ in proportion to the user-specified importance level of all keywords $\acute{q}$ in search Lens ℓ:

$$weight(q) = \frac{importance(q)}{\sum_{\acute{q} \in \ell} importance(\acute{q})}$$

SearchLens then uses the normalized keyword weights in place of the inverse document frequency term in the Okapi BM25 ranking function, and the score of each document d in the corpus for a set of Lenses L is therefore:

$$score(d, L) = \sum_{\substack{\ell \in L \\ q \in \ell}} \frac{weight(q) * tf(d,q) * (k+1)}{tf(d,q) + k * (1 - b + b * |d|/avgDL)}$$

where ℓ is the different user-specified Lenses, q is the different keywords in each Lens ℓ, $tf(d,q)$ is the term frequency of keyword q in document d, $|d|$ is length of the document d, and the constant $avgDL$ is the average document length in the corpus. We used the default parameters $k = 1.2, b = 0.75$ for Okapi BM25. Finally, we sum up the score of each Lens weighted by a coordination factor, which is the proportion of keywords in a Lens that has a non-zero document frequency. This modified version of the Okapi BM25 function can be easily translated to SQL queries for standard relational databases, or as a custom ranking function for the popular open sourced document retrieval engine Apache Lucene. This allows the SearchLens interface to be easily implemented using readily available tools that were already optimized for scaling and computational efficiency. Admittedly, more sophisticated ranking approaches may further improve the quality of results, but this simple method allowed us to explore the costs and benefits of providing reusable, re-composable, explanation-centric Lenses to users.

3.5 Implementation Notes

The backend of SearchLens was implemented in Python, using NLTK [5] and gensim [39] for indexing and word semantic model, respectively. In the indexing phase, text in each review is lowercased, tokenized, and stemmed using the Word Punkt Tokenizer [31] and Porter Stemmer [48]. Stop words are filtered out. An inverted index that records the document and the offsets of the mentions of each word stems is computed and stored in a PostgreSQL relational database. The Flask Python framework was used for our HTTP server. We implemented front-end of the SearchLens prototype as a web-based system using Javascript (ES6) and the ReactJS GUI framework, and the interactive visualizations are implemented using the D3.js library. User-specified Lenses were stored on client-side using browser cookies, so that they are persistent for the searchers between multiple visits.

4 EVALUATION

We evaluated SearchLens in two studies. First, we conducted a usability study in a controlled lab environment. Using predefined tasks, we tested the usefulness and usability of the system, as well as whether the visual explanation and exploration features provide enough benefit to encourage participants to express their rich and multifarious interests. Second, we conducted a field deployment study where participants use SearchLens for their own tasks. This allowed us to explore the benefits and limitations of our reusable and re-composable Lenses in real-life scenarios.

4.1 Usability Study

The main goal of the usability study was to verify in a controlled lab environment the usability of the interface and whether the visual explanation and exploration features can provide benefits to encourage users to express their nuanced and multifarious interests. We considered these the preconditions for conducting a field deployment study to test the real-life benefits of reusable and re-composable Lenses. Therefore, we focused on the following:

- whether the interface encouraged participants to externalize multiple interests and structure them using Lenses

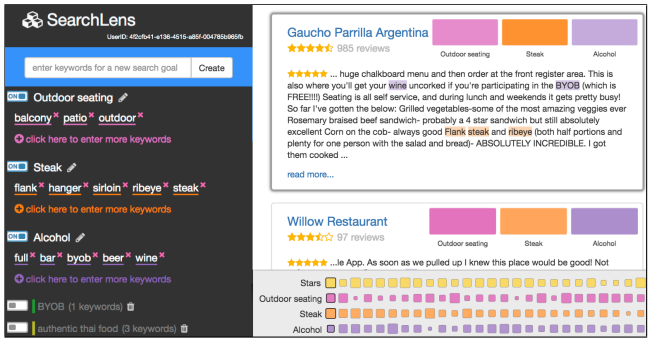

Figure 4: A Baseline system with topic-level visual explanation by collapsing the colored cells in each Lens and visualizing results only at the topic level.

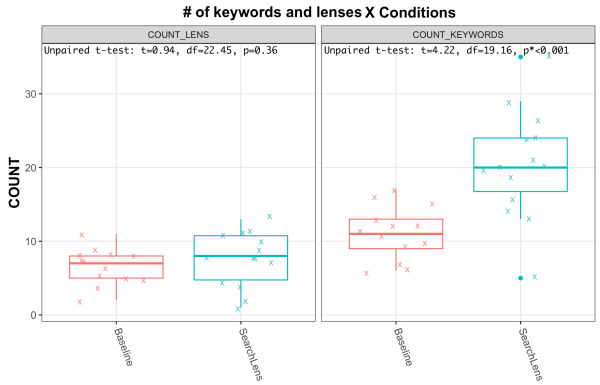

Figure 5: Number of Lenses and keywords saved by each participants at the end of the study. Participants in both conditions created comparable number of search Lenses, but participants in the SearchLens condition collected significantly more keywords in their Lenses.

- whether participants found the visual explanation and exploration feature to be useful
- whether the added benefits of visual explanation and exploration encouraged participants to spend more effort to express, iterate, and refine their Lenses

To test the above, we compared SearchLens to a baseline interface as a between subject condition, where the detailed visual explanation and exploration features were removed by collapsing the colored cells in each Lens and visualizing results only at the topic level (Figure 4), resulting an interface similar to the TileBars and the HotMap systems [27, 28]. Unlike in the SearchLens condition, users can only explore each restaurants at the topic-level, but not at the individual keyword level. Since searchers can not assign importance levels for each keyword in the baseline interface, we used the standard Okapi BM25 ranking function that weights keywords based on inverted document frequencies [40]. We chose this baseline as a more conservative test of the interactive explanation features than, for example, a comparison to Yelp or other search query-driven site (which are the implicit comparisons for the field study below).

The three scenarios for the usability study are listed below. The first scenario was designed to have both clear criteria (nice decor and good atmosphere and serves beer or wine), and an exploratory aspect (find a specific type of Japanese restaurant based on your own preferences). Scenarios 2 and 3 were designed to explore whether users would be able to reuse their Lenses for different contexts and find value in doing so. Scenario 2 had overlapping criteria to Scenario 1 (serves beer, cocktails, or wine), and Scenario 3 involved performing an identical search to Scenario 1 but in a different city.

- **Scenario 1**: Stanley is in Pittsburgh, USA visiting some friends and he is in charge of finding a few good restaurants for the group. They are interested in Japanese restaurants. They're not familiar with Japanese food or the different types of Japanese restaurants, so it is up to you to find Japanese restaurants based on reading the reviews and your personal preferences. The restaurants should have a nice decor and good atmosphere. Some of his friends like to have a few drinks with their meal, so if the place has a bar that serves beer or wine it would also be great. Since its pretty nice out, it would also be nice if the restaurants has outdoor seating or a patio, too.
- **Scenario 2**: John is looking for good seafood restaurants in Pittsburgh, USA, particularly places that serves fresh oysters and has a bar that serves beer, cocktails or wine. Decor or atmosphere are not important, but big plus if they offer outdoor seating, for example, a patio. Some of his friends are allergic to seafood, so the place must also have non-seafood options, preferably steak.
- **Scenario 3**: (Same as Scenario 1 but for finding restaurants in Montreal, Canada instead of in Pittsburgh, USA.)

A total 29 participants were recruited from a local participant pool, where 14 participants were randomly assigned the SearchLens interface with three predefined search tasks (N=14, Age=18-61, M=28.1, SD=12.7, 7 male, 6 female, and 1 other/not listed), and 15 participants assigned the baseline interface with the same search tasks (N=15, Age=18-54, M=28.1, SD=10.7, 7 male, 7 female, and 1 other/not listed). Each participant was given 60 minutes to complete the study and was compensated 10 USD. Before conducting the three tasks, participants watched a five minute introduction video that described the features in their given interfaces, which is followed a step-by-step training where participants created two pre-defined Lenses, report the name of the third restaurant in their search results, and report which keyword is missing from its reviews. Participants finished the training steps using an average of 5.9 minutes (N=29, SD=3.8). For the main task, participants were told to spend 10 to 15 minutes on each of the three tasks listed above in order. Finally, participants answered a short post-survey where we collected their subjective opinions about the systems using 7-point Likert scales and free-form responses.

4.1.1 Results for the Usability Study. One of our key hypotheses was that the immediate visual explanation provided by Lenses would encourage participants to express their interests and continually collect and refine those interests throughout the search process. This hypothesis appears to have been validated by the data. On

Action	Lab Baseline		Lab SearchLens		Field SearchLens	
add terms by typing	3.67	σ=2.82	5.50	σ=4.86	7.00	σ=5.39
add from suggestions	n/a		1.57	σ=1.99	3.20	σ=1.79
add from reviews	n/a		0.29	σ=0.73	0.40	σ=0.55
total add actions	3.67	σ=2.82	7.36	σ=6.10	10.60	σ=3.71
remove a keyword	4.67	σ=4.27	3.50	σ=2.79	4.20	σ=2.68
adjust weights	n/a		8.93	σ=7.54	12.80	σ=7.89
	N=15		N=14		N=5	

Table 1: Mean statistics for number of Lens editing actions performed by participants. Participants used SearchLens in the lab study more frequently add keywords to refine Lenses compared to baseline (t(27)=2.12, p<0.05). Participants in the field study conducted their own tasks.

Figure 6: Participants in the SearchLens condition were less likely to read through unfiltered lists of reviews than the baseline condition, which was accompanied by increased use of the SearchLens-specific ability to filter reviews relevant to different keywords.

average, participants in the SearchLens condition saved 20.43 keywords across their Lenses (N=14, SD=7.33), significantly more than participants in the baseline condition who saved 11.15 keywords (N=15, SD=3.58; t(27)=4.12, p<0.001). Importantly, this difference is likely not attributable to different perceptions of the task across conditions, as in both the SearchLens and baseline conditions participants generally created one Lens for each task criteria and combined multiple Lenses for each task (e.g., decor, drinks) and there was no difference between the total number of Lenses created between conditions (SearchLens: 7.6, baseline: 6.5; t(27)=0.92, p=0.36). In other words, the term-based interactive visual affordances supported by SearchLens seemed to encourage people to collect more terms indicative of their interests.

This pattern appeared to hold true throughout the search process for the iterative refinement of Lenses as well (Table 1). On average, participants using SearchLens added keywords to existing Lenses 7.4 times (N=14, SD=6.1) while those in the baseline condition did so 3.7 times (N=15, SD=2.8), which was found to be a significant difference (t(27)=2.12, p<0.05). This suggests that the added benefits from the visual explanation and exploration feature encouraged participants to iteratively refine their Lenses and allowed them to discover useful keywords more often.

We also examined whether participants found the added visual exploration features to be useful, and how the added benefits affected their behavior. By examining the behavior logs, we found participants using SearchLens frequently use the visual exploration feature. On average, each participant clicked on 25.86 (SD=29.19) keywords to filter reviews that mention a specific keyword instead of sifting through reviews to find ones that mentioned it (Figure 6). In both conditions, participants can also click on the name of a restaurant to see a list of reviews ranked by all active Lenses. While there is suggestive evidence that the filtering of reviews led to less use of the generic review lists, the result was not significant based on the number of participants in the study (M=6.33, 3.07; SD=5.78, 5.92; t(27)=1.50; p=0.15).

These results suggest SearchLens allowed participants to maintain a broader search goal with multiple interests, while at the same time explore and compare different options at a finer-grain level interactively instead of sifting through the reviews of each restaurant.

4.2 Field Study

Our field deployment study aimed to test our idea of reusable and re-composable Lenses in real-world settings. Five participants were recruited from the first study based on their high self-reported interest in researching restaurants online and in participating in a follow up study (N=5, Age=18, 20, 22, 23, and 25, 4 male, and 1 others/not listed). The participants were given access to the SearchLens system via the internet, and were asked to use the system for at least 60 minutes in total over a three day period. Although they were free to choose from any of the 11 cities in the dataset for this study, all five participants conducted tasks for their current city. Afterwards, they return to the lab and were given 45 minutes to finish a survey with primarily free-form questions, and were interviewed for another 15 minutes. Each participant was compensated with 40 USD for finishing the study.

Participants created more Lens keywords when conducting their own tasks comparing to participants in the lab study (Figure 7). On average, participants in the field study created 13.40 (SD=3.65) Lenses, significantly more than participants in the lab study that created 7.64 Lenses (SD=6.54; t(17)=2.46, p<0.05). They also saved significantly more keywords than participants in the lab study (lab: 20.4, field: 30.0, t(17)=2.50, p<0.05). Admittedly, it can be difficult to measure how much time participants actually spent using SearchLens in the field, nevertheless, results suggest that participants were able to accumulate more interests Lenses over a three day period than participants who spent 60 minutes in the lab study.

All five participants conducted multiple tasks during the study. Many explored different types of restaurants that they liked in the city using multiple Lenses, using SearchLens to build *"an overview interface for restaurants in the city that I might like"* (P1, P3, P4, P5). Participants also had more specific goals, including to check if there are vegan restaurants she has not discovered yet (P5), restaurants that serve bubble tea (P2), pizza places that offer Chicago deep dish-styled pizza (P3), and Mexican restaurants that has vegan options on the menu (P2).

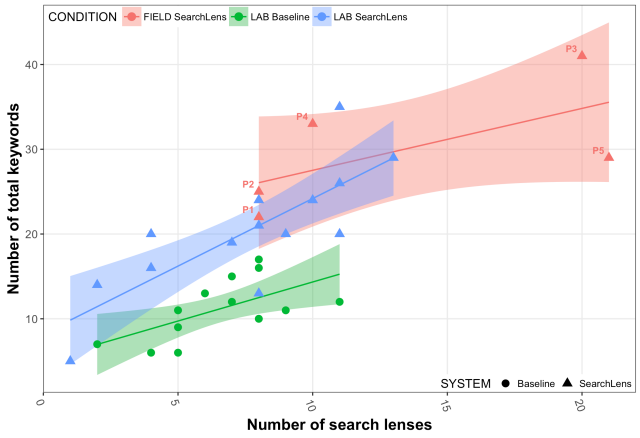

Figure 7: Number of Lenses and keywords specified by participants under different conditions. In the lab study with predefined search tasks, participants using SearchLens (blue) created a similar number of Lenses but used more keywords than the baseline condition (green). Participants in the field study (red) conducted their own tasks.

4.2.1 Refining Lenses. While participants reported creating Lenses based primarily on prior knowledge, all five participants also reported refining their Lenses throughout the process. Several cited that the shaded cells of the visual explanation helped them quickly noticed some keywords were too uncommon, and that an important concept of interest was missing from the search results (P1, P2, P5). One also mentioned noticing and removing ambiguous keywords when using the mention filtering features (P4). Participants also learned about new keywords which they added to their Lenses, sometimes replacing existing keywords, from both the suggestions (P1, P2, P3, P5) and from the reviews (P1, P2). Interestingly, the behavioral logs (Table 1) suggest they frequently discovered them from the systems' suggestions, indicating the value of the word2vec approach which we initially were concerned about for being noisy. This also points to potential future work in auto-suggesting Lenses which we intentionally avoided here due to concerns about agency and explainability.

4.2.2 Breadth and Depth. Participants created both general, breadth-oriented Lenses and more specific, depth-oriented Lenses. P4 specifically mentioned that it was useful being able to search for different genre (i.e., American, Mexican, or Indian restaurants) and at the same time pay attention to very specific dishes (i.e., cheese steak sandwich made with chicken), while still being able to see how each result match with different things, citing that "*more specific things are hard to search for on Yelp.*" Alternatively, P3 presented an interesting use case for deeper exploration of a specific genre, by first creating an more general Indian Food Lens, and then creating multiple more specific Lenses describing specific dishes from different regions of India, generating an overview of different styles of Indian restaurants in the city. This suggests that some users may want to create higher level groups of Lenses

4.2.3 Reusing Lenses: Combinations and Task Resumption. Participants reported their strategies for how they reused their Lenses, which can be broken down into two non-exclusive categories. The first use case we observed was task resumption between multiple search sessions (P1, P3, P4). Participants described having the ability to switch to a different sets of Lenses yet still keep the original Lenses for the future being useful (P3). One participant (P1) searched with a single Lens most of the time, but still cited that being able to re-enable Lenses from past sessions and to continue work on previous tasks and refined restaurants being useful. For the second use case, participants mentioned reusing Lenses in combination with other Lenses (P2, P3, P5). When asked about which of their Lenses were used in combination with different other Lenses, participants reported Lenses that concerned style and environment (*Cute and Quirky* (P5), *Atmosphere and Vibe* (P2, P5), *Friendly Staff* (P3)), price (*Inexpensive* (P2, P3), *Large Portion* (P3)), and some food-related but not for a general genre (*Fresh* (P2), *Fast Casual* (P2), *Vegan Options* (P2, P5), *Strong Beer* (P3)).

4.3 Overall Usefulness and Other Usecases

Through the lab and the field studies, we found evidence that using user-generated Lenses to provide visual explanation for deeper exploration was beneficial and effective in incentivizing users to externalize and iteratively refine their interests using Lenses. This occurred throughout the search process almost twice as frequently when compared to participants in the baseline condition which did not include the visual explanation and exploration features (Figure 4). As a result, participants using SearchLens created richer Lenses with nearly double the number of keywords on average compared to participants in the baseline condition. Participants also frequently used the visual explanation feature to explore the individual items in their search results, filtering reviews using different keywords in their Lenses 25.9 times on average. To test SearchLens in real-world settings, participants in the field study conducted their own tasks, and provided insights into their strategies in building and refining Lenses, as well as their strategies of composing and reusing Lenses across context and across search sessions over a three day period.

From the field study interviews, three out of the five participants said that they actually found and saved interesting restaurants during the study, and intend to visit those restaurant in the near future (P1, P3, P4). P1 in particular went to one of the restaurants he discovered using SearchLens and was happy about the visit, and P3 used SearchLens to complete a previous task, saying "*I wanted to try deep dish pizza for some time since I moved to US. Finally found one near the city. Kudos!*" All participant expressed that they would be interested in using SearchLens in the future if available, many also cited other scenario that might benefit from SearchLens. P2 pointed to scenarios where he needed to "*find a place for many people that may want different things*", and mentioned that SearchLens would be useful when her family visits her soon for his graduation. These results suggest that SearchLens was effective at helping users effectively find items that matched their specific interests.

5 LIMITATIONS AND FUTURE WORK

One limitation of the current implementation of SearchLens is its lack of ability to filter restaurants using their metadata, such as geographic location. We intentionally did not expose this information to our participants so we can focus our studies on allowing them to build personalized Lenses. However, practical systems would likely combine both paradigms to maximize efficiency. Utilizing metadata can also augment user-defined Lenses, for example, taking into account whether the a review that matched a specific Lens was positive or negative and whether the review poster's interests matched with the user's personal interests. However, the interactions between the two paradigms would require further studies. On the other hand, utilizing existing techniques for query term generalization beyond stemming or lemmatization, such as synonyms, semantic word models, or query expansion, can potentially improve recall, but their effects on the visual explanations would also require further studies.

Another obvious limitation of SearchLens it that it required more user effort upfront in order to receive the benefits provided by the system, such as reuse, explanation, and exploration. On a 7-point Likert scale, most participants from our lab study responded favorably in the post-survey to this trade-off with 64% agreed or strongly agreed that SearchLens is an improvement to the traditional search interfaces, and another 21% somewhat agreed with the statement, however, the long-term effect remained to be seen. One way to extend SearchLens is to combine machine learning and information retrieval approaches to reduce the effort of building Lenses, such as building interest profiles automatically, or using collaborative filtering and query expansion for expanding or inferring Lenses automatically [2, 43, 50], or word-sense disambiguation techniques for resolving ambiguous keywords [52].

Alternatively, we could also explore ways to allow users to share their Lenses with each other through explicit or implicit collaborations. For example, one participant mentioned "*It would be nice if I can see what Lenses a local person would use if I'm traveling, because I always try to ask the locals about where I should eat.*" Allowing access to Lenses created by previous users or expert users could potentially enable expertise transfer and accumulation through continuing refinement of a set of Lenses. For example, locals and past travelers could iteratively curate a set of Lenses that leads to an interactive and explorable list of local specialties for future travelers.

Another promising direction is to more deeply explore the idea of user-generated interest profiles and how they could dynamically influence the different interfaces accessible to the user or interacting with users in more proactive ways. Since we asked the field study participants to use SearchLens for their own tasks, most participants searched for restaurants in the city they lived in. Some participants that conducted more targeted search tasks (P2, P3, P5) mentioned that they were already familiar with most of the options in the city that fits their goals, but would still occasionally search online to see if there were new restaurants that match their interests (P2, P5). As users continue to use SearchLens, the system will accumulate more understanding of what the users is interested in, and can potentially detect and notify the users of new information that might be of interests with high accuracy [51]. Alternatively, existing users may use their repository of Lenses to explore or curate the restaurants in an unfamiliar city. Participants in the field study also pointed to the potential of Lenses being useful for other types of information and domain, including shopping (P2, P3), trip planning (P2, P5), buying a house (P2), and job hunting (P4).

6 CONCLUSION

In this paper we introduced SearchLens, a novel approach that allows users to specify and maintain their profile of multifarious and idiosyncratic interests. This enabled them to reuse and re-compose their different interests across scenarios, as well as maintaining context across multiple search sessions. To encourage users to put in the up-front effort of curating Lenses, we explored ways of using Lenses to provide immediate benefits of visual explanation and deeper exploration of search results. Across a lab and field study we observed that participants expressed their interests with significantly more query terms, and found benefits in the SearchLens approach, including being able to transfer and reuse their Lenses across contexts, being able to interpret new information that reflects their own personal interests with transparency, and working at multiple levels of specificity and hierarchy. More fundamentally, being able to visualize and explore new information in ways that promote transparency can potentially empower users to be more aware of their online information diet. For example, as a way to manipulate their own social media feeds, and being more aware of how posts were selected or hidden. We believe SearchLens represents a first step towards a transparent and user-centered approach to addressing subjective and fragmented nature of information today.

7 ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (grants PFI-1701005 and SHF-1814826), Google, and the Carnegie Mellon University Center for Knowledge Acceleration.

REFERENCES

[1] Jae-wook Ahn and Peter Brusilovsky. 2009. Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization* 8, 3 (2009), 167–179.

[2] Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. 2007. Open user profiles for adaptive news systems: help or harm?. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 11–20.

[3] Nicholas J Belkin, Colleen Cool, Judy Jeng, Amymarie Keller, Diane Kelly, Ja-Young Kim, Hyuk-Jin Lee, Muh-Chyun (Morris) Tang, and Xiao-Jun Yuan. 2001. Rutgers' TREC 2001 interactive track experience. In *TREC*.

[4] Nicholas J Belkin, Diane Kelly, G Kim, J-Y Kim, H-J Lee, Gheorghe Muresan, M-C Tang, X-J Yuan, and Colleen Cool. 2003. Query length in interactive information retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 205–212.

[5] Steven Bird and Edward Loper. 2004. NLTK: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*. Association for Computational Linguistics, 31.

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

[7] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. 2005. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*. ACM, 89–96.

[8] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on Machine learning*. ACM, 129–136.

[9] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. In *Proceedings of*

the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA. https://doi.org/10.1145/3025453.3026044

[10] 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.

[11] Liren Chen and Katia Sycara. 1998. WebMate: A personal agent for browsing and searching. In *Proceedings of the second international conference on Autonomous agents*. ACM, 132–139.

[12] Nan-Chen Chen, Jina Suh, Johan Verwey, Gonzalo Ramos, Steven Drucker, and Patrice Simard. 2018. AnchorViz: Facilitating Classifier Error Discovery through Interactive Semantic Data Exploration. In *23rd International Conference on Intelligent User Interfaces*. ACM, 269–280.

[13] 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.

[14] Jason Chuang, Daniel Ramage, Christopher Manning, and Jeffrey Heer. [n. d.]. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proc. CHI 2012*. ACM, 443–452.

[15] Bart De Langhe, Philip M Fernbach, and Donald R Lichtenstein. 2015. Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research* 42, 6 (2015), 817–833.

[16] Cecilia di Sciascio, Peter Brusilovsky, and Eduardo Veas. 2018. A Study on User-Controllable Social Exploratory Search. In *23rd International Conference on Intelligent User Interfaces*. ACM, 353–364.

[17] Cecilia di Sciascio, Vedran Sabol, and Eduardo E Veas. 2016. Rank as you go: User-driven exploration of search results. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 118–129.

[18] James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. 2008. CueFlik: interactive concept learning in image search. In *Proceedings of the sigchi conference on human factors in computing systems*. ACM, 29–38.

[19] Kristofer Franzen and Jussi Karlgren. 2000. Verbosity and interface design. *SICS Research Report* (2000).

[20] G. W. Furnas. 1986. Generalized Fisheye Views. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '86)*. ACM, New York, NY, USA, 16–23. https://doi.org/10.1145/22627.22342

[21] Qiwei Gan, Qing Cao, and Donald Jones. 2012. Helpfulness of online user reviews: More is less. (2012).

[22] 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.

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

[24] 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.

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

[26] Marti A Hearst and Jan O Pedersen. 1996. 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.

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

[28] Orland Hoeber and Xue Dong Yang. 2006. A comparative user study of web search interfaces: HotMap, Concept Highlighter, and Google. In *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on*. IEEE, 866–874.

[29] Yelp Inc. 2016. The Yelp Dataset Challenge: Discover what insights lie hidden in our data. https://www.yelp.com/dataset/challenge. Accessed: 2017-09-10.

[30] Bernard J Jansen, Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management* 36, 2 (2000), 207–227.

[31] Bryan Jurish and Kay-Michael Würzner. 2013. Word and Sentence Tokenization with Hidden Markov Models. *JLCL* 28, 2 (2013), 61–83.

[32] 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.

[33] Todd Kulesza, Saleema Amershi, Rich Caruana, Danyel Fisher, and Denis Charles. 2014. Structured labeling for facilitating concept evolution in machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3075–3084.

[34] Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 165–172.

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

[36] Susan M Mudambi and David Schuff. 2010. Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly* (2010), 185–200.

[37] 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.

[38] Jaakko Peltonen, Jonathan Strahl, and Patrik Floréen. 2017. Negative Relevance Feedback for Exploratory Search with Visual Interactive Intent Modeling. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 149–159.

[39] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.

[40] Stephen Robertson, Hugo Zaragoza, et al. 2009. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval* 3, 4 (2009), 333–389.

[41] Yong Rui, Thomas S Huang, Michael Ortega, and Sharad Mehrotra. 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on circuits and systems for video technology* 8, 5 (1998), 644–655.

[42] Gerard Salton and Chris Buckley. 1990. Improving retrieval performance by relevance feedback. *Journal of the American society for information science* 41, 4 (1990), 288–297.

[43] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. ACM, 285–295.

[44] Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 824–831.

[45] Mirco Speretta and Susan Gauch. 2005. Personalized search based on user search histories. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*. IEEE, 622–628.

[46] 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.

[47] Frank Van Ham and Adam Perer. 2009. "Search, show context, expand on demand": supporting large graph exploration with degree-of-interest. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009).

[48] Cornelis J Van Rijsbergen, Stephen Edward Robertson, and Martin F Porter. 1980. *New models in probabilistic information retrieval*. British Library Research and Development Department London.

[49] Ho Chung Wu, Robert Wing Pong Luk, Kam Fai Wong, and Kui Lam Kwok. 2008. Interpreting tf-idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)* 26, 3 (2008), 13.

[50] Jinxi Xu and W Bruce Croft. 1996. Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 4–11.

[51] Beverly Yang and Glen Jeh. 2006. Retroactive answering of search queries. In *Proceedings of the 15th international conference on World Wide Web*. ACM, 457–466.

[52] 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.

[53] 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.

[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.