Improving Search Result Comprehension by Topic-Relevance Map Visualization

Jaakko Peltonen
Faculty of Natural Sciences,
University of Tampere, Finland
Department of Computer
Science, Aalto University, Espoo,
Finland
jaakko.peltonen@uta.fi

Kseniia Belorustceva
Department of Computer
Science, Aalto University,
Espoo, Finland

Tuukka Ruotsalo
Department of Computer
Science, Aalto University, Espoo,
Finland
Department of Computer
Science, University of Helsinki,
Helsinki, Finland
tuukka.ruotsalo@aalto.fi

IUI'17 Companion, March 13-16, 2017, Limassol, Cyprus
ACM 978-1-4503-4893-5/17/03.
http://dx.doi.org/10.1145/3030024.3038263

Abstract
We introduce topic-relevance map, an interactive search result visualization that assists rapid information comprehension across a large ranked set of results. The topic-relevance map visualizes a topical overview of the search result space as keywords with respect to two essential information retrieval measures: relevance and topical similarity. Non-linear dimensionality reduction is used to embed high-dimensional keyword representations of search result data into angles on a radial layout. Relevance of keywords is estimated by a ranking method and visualized as radiuses on the layout. Similar keywords are modeled by nearby points and more relevant keywords are closer to the center of the radial display. We evaluated the effect of the topic-relevance map in a search result comprehension task where 24 participants were summarizing search results and produced a conceptualization of the result space. Topic-relevance map significantly improves participants' comprehension capability compared to a ranked list.

Author Keywords
Sense-making; Visualization; Exploratory Search

ACM Classification Keywords
H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous; H.3. [Information Storage and Retrieval]

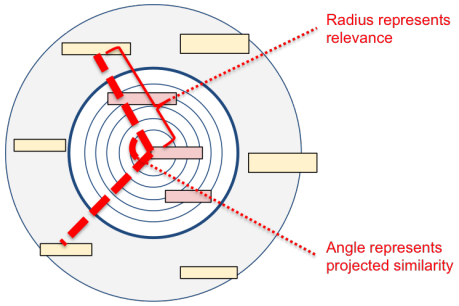

Figure 1: An illustration of the Topic-Relevance Map. The distance from the center represents relevance for a keyword. The angle between keywords represents their topical similarity.

Introduction

Search systems often present the retrieved information as a relevance-ranked list of documents [6], forcing the user to scan through the ranked list and pick sufficiently relevant documents from the top. In many cases users are not looking for a single highest-ranked document, but the initial query is issued to broadly comprehend the information space [8]. This requires proficiency in information literacy based on the result list and poses problems detecting topical subspaces within the list [1]. A promising solution is to visualize the search result space for the user [2]. Unlike a ranked list, which uses one dimension (ranking), visualization methods allow additional dimensions to represent search results, with two main types: *visualizing topics or terms within retrieval results* (i.e. content of a result document) and *visualizing an overview of results* [3].

We present *topic-relevance map*, a novel overview visualization approach considering two dimensions essential for search results: topical similarity and relevance. Topic-relevance map organizes keywords representing the search results onto a radial layout. Non-linear dimensionality reduction is used to embed high-dimensional keyword representations into angles on the radial layout. Relevance of keywords is estimated by a ranking method and visualized as radiuses on the radial layout. As a result, similar keywords are modeled by nearby points, dissimilar keywords are modeled by distant points, more relevant keywords are closer to the center of the radial display, and less relevant keywords are distant from the center of the radial display. Figure 1 illustrates the underlying principle.

We study the effectiveness and user behavior of topic-relevance map in a search result comprehension task; skimming and understanding the structure and conceptualization of the search results by creating a conceptual structure organized as main topics and their subtopics. In experiments 24 users comprehended the search results retrieved in response to eight pre-defined queries. The results suggest the topic-relevance map yields a more comprehensive understanding of search results than a comparison system that visualized topics within the search result list. This demo complements our long paper ([4]; see also [5]) and lets IUI attendees use the topic-relevance map to comprehend search results from a huge scientific article database.

Topic-Relevance Map

The design principles behind the topic-relevance map are: **1.** The map should illustrate, in a compact manner, the full result space of topical information across all search results. **2.** The map should efficiently show the estimated relevance of each topical content. **3.** The map should compactly show relationships among the topical content. Figure 2 illustrates the topic-relevance map. It represents topical content of search results based on keywords in the result documents, and organizes keywords as directions on a radial layout.

Visualization method. We optimize a data-driven layout for keywords, by nonlinear dimensionality reduction of their topical content. We quantify goodness of a layout by analyzing neighborhood relationships in the high-dimensional topical-features space and the low-dimensional angles on the topic-relevance map. Two keywords can be called *neighbors* if their topical features are similar: we define an input neighbor probability as a Gaussian falloff with respect to distance between topical features of keywords. On the outer ring, two keywords seem neighbors if they have close-by angles: we define an output neighbor probability as a Gaussian falloff of angular distance. We optimize angles by minimizing Kullback-Leibler divergence of input and output neighborhoods. We create topical features of keywords from lookup searches, emphazising each top keyword in turn and estimating relevance of other keywords by Bayesian linear regression.

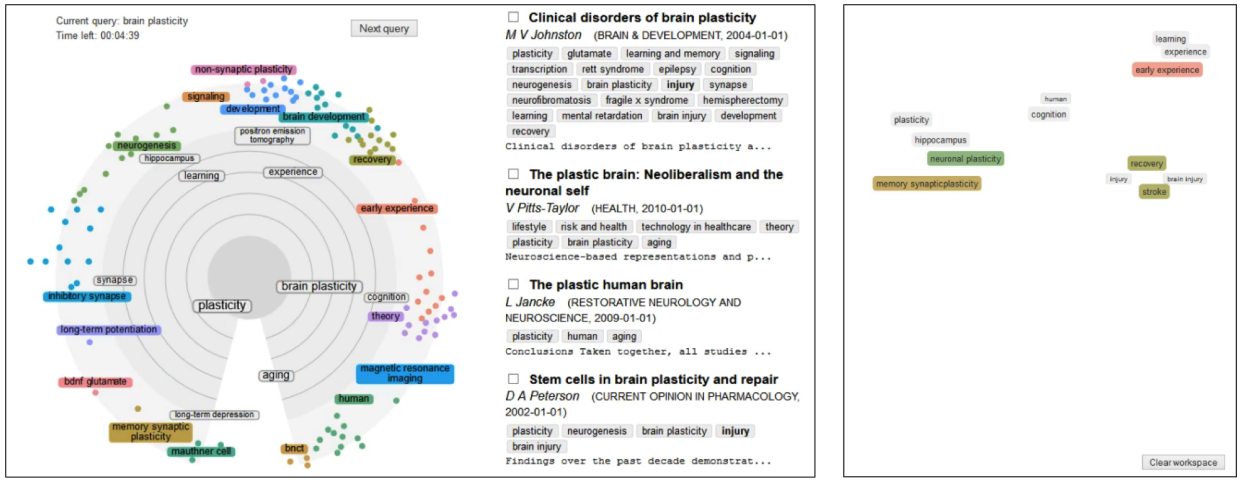

Figure 2: A screenshot of the interface in response to a search query "Brain Plasticity". *Left:* The visualization and the ranked list of search results. In the visualization, the *inner area* (gray background) shows the top ten keywords related to the search query; the *outer ring* around it shows a set of other keywords clustered and the most important in each cluster highlighted. *Right:* the workspace where particants could drag keywords. The workspace was placed under the visualization as a floating element to ensure equal screen estate with the baseline system.

It has two main areas: an *inner area* shows top-ten keywords representing the most relevant topics, and an *outer ring* shows all other keywords representing other topics in the search results. The *inner area* represents the search: the closer a keyword is to the center the more relevant it is. The *outer ring* shows other keywords: topically similar keywords are arranged to similar angles; keywords form angular clusters representing directions in the information space. Keywords are colored according to the clustering; the most relevant keyword of each cluster is given, the rest shown as dots for clarity. The *inner area* is also arranged along angular directions, and with radial positions representing relevances. Top keywords act as signposts along angular directions towards clusters of keywords along the directions. Users can inspect clusters with a *fish-eye lens*.

Our interface directly supports key tasks in Shneiderman's [7] taxonomy of tasks for information visualizations, including the visual information seeking mantra *overview first, zoom and filter, details on demand, and relate*. **Overview:** the map shows an at-a-glance overview of keyword content in the search results: top keywords, keyword relevances, groups of keywords as directions. **Zoom and filter:** the fisheye lens lets the user zoom into a parts of the visualization. **Details on demand:** hovering over a keyword highlights it on the radar and in documents where it appears. **Relate:** the map shows data-driven relationships of keywords by co-location along similar directions in the visualization, as well as direct co-occurrences in documents.

Experiments

We measure the effect of the visualization component on information comprehension performance. We compared the proposed system to one that is otherwise the same but does not have the visualization component, only the conventional result list. Both systems listed keywords under each result document. 24 participants from two universities performed a task in which they had to comprehend and summarize search results; the scenario was "You are searching information about a pre-defined topic using an information retrieval system. Your task is to comprehend the topic by describing, at least two, main concepts related to the overall topic and describe as many as possible subconcepts related to each main concept." Participants were asked to use a two-level hierarchical conceptualization: **1.** Find as many *main topic keywords*, but at least two, that you find important to cover the overall topic. **2.** Find as many *subtopic keywords* under each main keyword that you find important to cover the main keyword. Participants stored their conceptualization by dragging keywords to a *workspace* component (Figure 2 right) during the task, then used it to compose a written answer. Eight topics were used in the experiment: Human Memory, Web Design, Cognition, Distributed Systems, Language Processing, Kernel Function, Wearable Sensors, and Compiler Design. Two assessors assessed the relevance of main topic and subtopic keywords and the subtopic keywords on a Likert scale.

Results in brief. *Participants inspected the search results using the topic-relevance map.* Recordings of mouse movements showed that over one third of the time the users inspected the search results using the map. *Participants selected significantly more and more comprehensive set of keywords from the map than from the result list. The topic-relevance map improved the comprehension outcome for the main topic keywords*, but no overall difference was found in case of the subtopic keywords. This suggests that the visualization enabled participants to obtain a broader view on the search results, but did not help gather better information under an individual subtopic.

Acknowledgments. Authors belong to HIIT, JP&TR to COIN; work funded in part by TEKES & Academy of Finland decisions 252845, 256233, 295694, 305739, 278090.

REFERENCES

1. C. Bruce. 1999. Workplace experiences of information literacy. *Int J Inform Manage* 19, 1 (1999), 33 – 47.
2. C. Chen and Y. Yu. 2000. Empirical Studies of Information Visualization. *Int. J. Hum.-Comput. Stud.* 53, 5 (Nov. 2000), 851–866.
3. M. A. Hearst. 2009. *Search User Interfaces* (1st ed.). Cambridge University Press, New York, NY, USA.
4. J. Peltonen, K. Belorustceva, and T. Ruotsalo. 2017a. Topic-Relevance Map: Visualization for Improving Search Result Comprehension. In *Proc. IUI*. In press.
5. J. Peltonen, J. Strahl, and P. Floreen. 2017b. Negative Relevance Feedback for Exploratory Search with Visual Interactive Intent Modeling. In *Proc. IUI*. In press.
6. S. E. Robertson. 1997. Readings in Information Retrieval. Morgan Kaufmann, Chapter The Probability Ranking Principle in IR, 281–286.
7. B. Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. IEEE Symposium on Visual Languages*. IEEE.
8. J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. 2004. The Perfect Search Engine is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proc. CHI'04*. 415–422.