ICTE 2016, December 2016, Riga, Latvia

Search Intention Analysis for Task- and User-Centered Visualization in Big Data Applications

Dirk Burkhardt[a,b,*], Sachin Pattan[a], Kawa Nazemi[a,b], Arjan Kuijper[b]

ᵃFraunhofer-Institute for Computer Graphics Research (IGD), Fraunhoferstr. 5, 64283, Darmstadt, Germany
ᵇTU Darmstadt, Fraunhoferstr. 5, 64283, Darmstadt, Germany

Abstract

A new approach for classifying users' search intentions is described in this paper. The approach uses the parameters: word frequency, query length and entity matching for distinguishing the user's query into exploratory, targeted and analysis search. The approach focuses mainly on word frequency analysis, where different sources for word frequency data are considered such as the Wortschatz frequency service by the University of Leipzig and the Microsoft Ngram service (now part of the Microsoft Cognitive Services). The model is evaluated with the help of a survey tool and few machine learning techniques. The survey was conducted with more than one hundred users and on evaluating the model with the collected data, the results are satisfactory. In big data applications the search intention analysis can be used to identify the purpose of a performed search, to provide an optimal initially set of visualizations that respects the intended task of the user to work with the result data.

Keywords: User-centered-design; Human-computer-interaction; Information retrieval; Predictive analysis; Interactive search

1. Introduction

Nowadays search engines are designed to search for clear objectives, such as concrete resources, products or information, which results are almost presented in list representation ordered by complex algorithms that majorly focus on keyword matching. Advanced search systems that use even graphical visualizations are able to visualize

---

* Corresponding author. Tel.: +49-6151-155-578; fax: +49-6151-155-139.
  E-mail address: dirk.burkhardt@igd.fraunhofer.de

results beyond result listings. These advanced visualization opportunities allow showing structures or in general specific aspects such as geographical or temporal properties visually. In fact that means, searches are not limited on clear target searches and therewith objectives. Modern information retrieval systems can cover also exploratory searches, e.g. to gather an overview about unknown topics or complex linking of entities, resources and objects. But also analysis is a typical tasks for modern information retrieval systems, which aim lays on analyzing certain properties or complex numeric values.

The challenge of modern information retrieval system is to show the results with appropriate visualizations or visualization algorithms. If users perform searches about persons, e.g. the biography, a simple result listing ordered by keyword matching is sufficient. But what if users aim to compare different cameras by their properties to find the best option to buy? Or what if users aim to search for German renowned search institutions that focus on renewable energy? The list presentation is not that helpful to start the data analysis or exploration.

To be able to show the results in a beneficial set of visualizations, to know the search intention would help to better identify what visualization might be useful. Indeed, the idea of identifying the search intention is not new. The current existing approaches can be classified in semantically analysis[1] and generic/situational query analysis[2,3,4]. The semantical query analysis approaches is language depending and requires predefined rules. The generic or situational query analysis using only basic information, such as query length, or additional behavioral information like the geographical location or details about the used (mobile) device. However, all of the approaches are not generic and therefore have deficits in particular in normal search contexts.

Therefore, this papers describe a novel approach to identify the intention of a performed search just on the basis of the entered search query, and majorly on the basis of the word frequency that are used in the search phrase.

2. Search intention analysis approach

In this section we describe the principle approach to calculate the search intention majorly based on the word frequency. Before the method can be described, we explain the preliminary study we performed before, to identify relevant influencing factors to calculate the user's search intention.

2.1. Preliminary study

Since there is no existing work to identify search task majorly based on word frequency, we had to perform a preliminary study to identify rules or patterns. Therefore we asked more than 50 people, almost with a computer science background, to check their search history on Bing, google etc. to send us real performed search phrases. Together with the search phrases, we ask them to classify the searches, if they would identify them more likely as target, exploratory or analysis search. Even though the people could provide some notes, e.g. if they were not sure what the goal was most likely.

On the basis of this data we tried to identify rules or patters on which basis we could classify a search phrase. Even more we developed a metric to be able to calculate the search intention into the task categories: targeted, exploratory and analysis search.

2.2. Rules and patterns to identify search intentions

With this classification information in hand, we tried to recognize some patterns, which would help us in forming the hypotheses. Based on these found hypotheses, six were considered for classifying a query as exploratory, targeted or analytical search query. As there were only few analytical search queries, only few patterns/hypotheses were recognized for them (see Fig. 1):