A Perspective-Aware Approach to Search: Visualizing Perspectives in News Search Results

Muhammad Atif Qureshi†‡, Colm O’Riordan†, Gabriella Pasi‡
†CIRG, NUIG, Ireland
‡IRLAB, DISCO, UNIMIB, Italy
muhammad.qureshi@nuigalway.ie

ABSTRACT
The result set from a search engine for any user’s query may exhibit an inherent perspective due to issues with the search engine or issues with the underlying collection. This demonstration paper presents a system that allows users to specify at query time a perspective together with their query. The system then presents results from well-known search engines with a visualization of the results which allows the users to quickly surmise the presence of the perspective in the returned set.

Categories and Subject Descriptors
H.3.3 [Information Search and Retrieval]: Search process; H.5.2 [User Interfaces]: Natural language

Keywords
Perspective, Wikipedia, News Search

1. INTRODUCTION
There have been notable efforts in the information retrieval research community to provide users with an insight into the relationship between the query and the result set [1]. Recent research into exploratory search systems [2] recommends the integration of cognitive aspects within the information retrieval process. However, current information retrieval systems do not support means to investigate “potential bias” (an essential cognitive element) towards a certain perspective introduced during the search process. According to the Oxford Dictionary, the definition of perspective is as follows: “Perspective is a particular attitude towards or way of regarding something.” In line with this we argue for incorporating the essential cognitive element of “perspective” within the search engine interface thereby introducing “perspective-aware” search in this demonstration paper.

The proposed system allows the user to specify an additional input perspective terms/phrase along with the query through a standard type-keywords-in-entry-form interface (see Figure 1). Note that this is not equivalent to appending the query with the perspective terms/phrase as the perspective may not necessarily be a part of the user’s search intent. However, there may be a bias in the result set towards a certain “perspective”; and hence, we propose perspective-aware search as a means to investigate and analyze a leaning towards an agenda. We explain through the following motivating example: consider a case in which a user wishes to find information about a certain event (say, a bomb blast in a certain region). The search results returned may be polarized instead of focusing on factual aspects i.e., relating to a certain race, ethnicity, or political movement which caused violence. This can prompt a user to explicitly evaluate a move from objective factual reporting to subjective reporting within the top results.

Our system utilizes knowledge from Wikipedia to make conceptual sense of the perspective terms/phrase. This knowledge does not modify the query (as would an additional query term) but is instead used to highlight the presence of a perspective in the result set.

2. SYSTEM DESCRIPTION
Fig. 2 shows the architecture of our system. The user enters the query together with the perspective phrase and the query is fed to the underlying information retrieval system which generates a ranked list of documents. The document extractor then forwards the content of the documents to the tokenizer and the extracted tokens along with the input perspective are fed into our perspective scoring system which...

Figure 1: Perspective-Aware Search Entry Form

---

1The “potential bias” may be introduced due to issues with the search engine itself or with the underlying collection.

2An external and collaboratively created knowledge resource which is less likely to be biased in a given direction.
uses our custom-built WikiMadeEasy API\(^3\). The perspective scoring system scores each token with respect to the perspective entered by the user and the score of each token is aggregated to produce a perspective score for a document in the ranked list returned by the information retrieval system(s). Finally, the ranked list returned by the information retrieval system(s) and the perspective scores of tokens & documents returned by the scoring system is returned as output to generate the HTML result page.

![Figure 2: Perspective-Aware Search Overall Architecture](image)

The perspective scoring system uses the Wikipedia category-article structure to score the amount of content present inside a document with respect to the input perspective. The underlying perspective computation algorithm makes use of the title (or redirect) of a Wikipedia article for matching perspective terms/phrase, then utilizes the categories\(^4\) and sub-categories\(^5\) to a depth count of two of the matched Wikipedia article. We refer to these categories related to the perspective under investigation as \(RP\) (i.e., it contains all related categories of the perspective from depth count zero to two). Next, we retrieve the set of all articles within the Wikipedia category set \(RP\) (we refer to this set as \(Articles_{RP}\)). Finally, all categories associated with these articles are retrieved. The extracted tokens from a document which are contained in \(Articles_{RP}\) are called matched phrases (phrases defining perspective). We use these matched phrases to calculate the perspective score. The following summarizes important factors which contribute in calculating the perspective score for a token using Wikipedia category-article structure: 1) The significance of category depth at which a matched phrase occurs; the deeper the match occurs in the taxonomy the less its significance to the perspective under investigation. 2) The significance of a matched phrase’s categories corresponding to the perspective under investigation. The more categories of matched phrase in \(RP\), the higher the significance related to the perspective, and 3) The significance of the phrase itself which is a combination of phrase word length and frequency of the phrase within the document.

\(^3\)http://bit.ly/1eMADG9, we aim to release the API as an open source Wikipedia tool to facilitate easy access to Wikipedia data and graph structure.

\(^4\)These are basically the categories of a perspective phrase’s Wikipedia article i.e., categories at the depth zero.

\(^5\)Each Wikipedia category can have an arbitrary number of subcategories.

Finally, for user convenience, we have provided a perspective autocompletion feature within the system which simplifies the perspective input process for the user.

3. DEMONSTRATION FOR NEWS DOMAIN

The perspective-aware search prototype we describe in this demonstration paper is tailored for news search and fetches news search results from three search engines (i.e., Yahoo!, Google and Bing). The video of the demonstration can be accessed at [http://youtu.be/mPO763z6H4Y](http://youtu.be/mPO763z6H4Y). The system provides additional information within snippets of returned news search results where the perspective score for each result together with its perspective rank is displayed as the snippet in Figure 3 shows. Furthermore, the search results of each search engine are displayed in a side-by-side manner. The system also employs visualization techniques (i.e., bar charts and line charts) to display the comparative perspective scores. Figure 4 shows the visualizations incorporated in the system as they appear on the search results page giving the user further insights into the result sets of major search engines together with the web sites from which the result sets are retrieved. A few screenshots of the system are available at [http://bit.ly/Mpetmc](http://bit.ly/Mpetmc). An interface such as the proposed one can be particularly useful in exploratory tasks such as those commonly encountered in the news domain by journalists, media studies researchers or by end-users.

4. REFERENCES


\(^6\)It is the ordering by perspective scores for the top ten results.