

Users' Adjustments to Unsuccessful Queries in Biomedical Search

G. Craig Murray, Jimmy Lin
College of Information Studies
University of Maryland
{gcraigm,jimmylin}@umd.edu

John Wilbur, Zhiyong Lu
National Center for Biotechnology Information
U.S. National Library of Medicine
{murraycr,wilbur,luzh}@ncbi.nlm.nih.gov

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *query formulation*. H.3.7 [Information Storage and Retrieval]: Digital Libraries – *user issues*.

General Terms

Human Factors, Measurement

Keywords

Medical search, query reformulation, user modeling, PubMed

1. INTRODUCTION

Biomedical researchers depend on on-line databases and digital libraries for up to date information. We introduce a pilot project aimed at characterizing adjustments made to biomedical queries that improve search results. Specifically we focus on queries submitted to PubMed[®], a large sophisticated search engine that facilitates Web access to abstracts of articles in over 5,200 biomedical journals. On average 2 million users search PubMed each day. During their search, nearly 20% will experience a result page from one of their queries that has zero results. In some cases there really is no document or abstract that will satisfy a particular query. However, in analyzing one month of queries submitted to PubMed, we find that more often than not, queries that retrieved no results are queries that *would* retrieve something relevant if they were constructed differently. This paper describes a new effort to identify some of the characteristics of a query that produces zero results, and the changes that users most often apply in constructing new, “corrected” queries. Zero-result queries afford us an opportunity to examine changes made to queries that we know did not return relevant data, because they did not return *any* data. An investigation of the changes users make under these circumstances can yield insight into users’ search processes.

2. METHOD AND EARLY RESULTS

We took a collection of queries submitted to PubMed in one month and analyzed differences between *zero-result* queries and the queries that followed these queries. We randomly sampled pairs of consecutive queries where one or both queries produced zero result. Our dataset contained just over 125,000 query pairs. We analyzed typographic changes made to queries as well as conceptual changes.

Typographic Changes—We focused first on a set of possible typographical changes that can be made to a query. We counted the number of times the change was an addition, removal, or replacement of a term or a character, and whether the change was made to the interior of the query or to the left or right ends. We compared the distribution of instances where these types of changes were “successful”—i.e. retrieved *some* results—and “unsuccessful”—where the revised query *also* retrieved zero results. The distribution of typographic changes differs for successful and unsuccessful query changes, as shown in Figure 1.

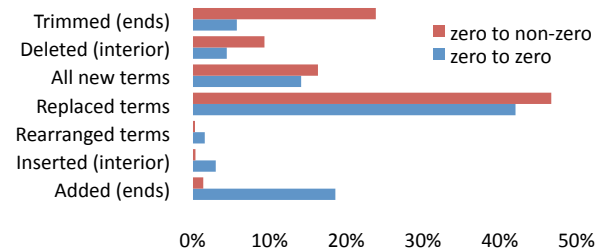


Figure 1. Typographical changes made to zero-result queries.

Conceptual Changes—Following these analyses, we took a smaller sub-sample of the query pairs where terms or characters were removed. Four annotators classified the changes made to these queries using 15 categories in three separate facets.

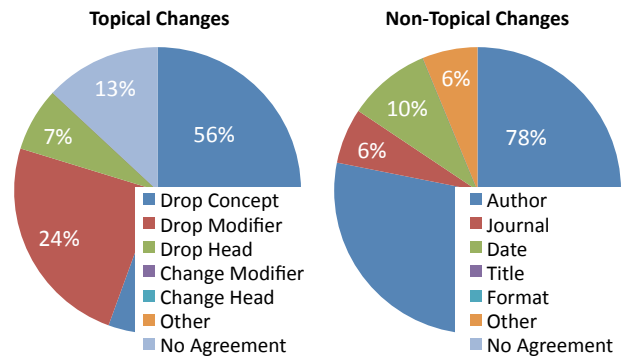


Figure 2. Conceptual changes made to zero-result queries.

These summative statistics are results from research in progress. We present the distributions of types of changes that are made to queries when they produce zero results. This exploratory analysis is a valuable first step toward building new tools to support users with “unsuccessful” queries.

ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine. This work is in the public domain in the United States.