

ABSTRACT

The growth in the past decade of both the infrastructure and the number of users of the Internet has enabled a corresponding growth in the number of users of digital reference services on the Internet. This increase has led to an increase in the number of questions received by these services, putting a strain on the human intermediaries employed therein. The ability of a digital reference service to “scale up” to handle an increasingly large number of questions is directly affected by the amount of automation employed by that service: the more processes that are automated, the more of human intermediaries’ time and effort can be dedicated to tasks that cannot yet be automated. There is, now more than ever, an increased and immediate need in digital reference for automation.

This study identifies (1) the types of questions that are received by digital reference services, according to several taxonomies of questions at different levels of linguistic analysis, and (2) the rules by which questions are triaged within and between services (triage being the process of routing and assigning questions to expert digital reference question answerers and other services). Taxonomies of questions are identified through an extensive review of literature that deals with questions, from several fields: desk and digital reference, question answering, and linguistics. The rules by which questions are triaged are identified through a think-aloud study of digital reference triagers performing the task of triage. The goal of this study is to develop specifications according to which an automated triage system can be built. These specifications will take into account the question type, as well as other attributes of questions that affect triage decisions. These taxonomies of questions may also prove to be useful as the basis for algorithms for automating other steps in the digital reference process.

Question Taxonomies for Digital Reference

By

Jeffrey Pomerantz

B.A. University of Massachusetts at Amherst, 1993

M.S.(L.I.S.) Simmons College, 1997

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Information Transfer
of the Graduate School of Syracuse University

May 2003

Approved by

Dr. Elizabeth D. Liddy

Date

23 July 2003

ACKNOWLEDGEMENTS

First and most importantly, I want to thank my wife, Yvonne. She deserves an honorary Ph.D. in Information Science for listening to me at every step, and another one in Psychiatry for talking me through this process.

I want to thank my advisor, Liz Liddy, for being a real joy to work with, for always making time regardless of how busy she was, and for letting me proceed with this study at my own pace.

I also want to thank my dissertation committee members: Barbara Kwaśnik, David Lankes, and Scott Nicholson, for their feedback, time, ideas, and support.

Thanks to the Information Institute of Syracuse, for supporting me by means of a Research Assistantship.

Thanks to Beta Phi Mu for assistance in the form of a Eugene Garfield Doctoral Dissertation Fellowship.

“The really great promise of automation is to be sought in... the opportunity it affords to analyze the reference process and re-define reference service. ... The fullest utilization of the potential of automation does necessitate a thorough study of the total reference process.”

Jesse Shera, 1964