

Reference EXTRACT



Using Reference Authoring to Provide Credible Search Tools

“Imagine Borrowing the Brain of a Librarian to search the web”

At the [Internet Credibility and the User Symposium](#) the concept of a search engine built with Google’s simplicity and librarians’ credibility was proposed. Reference Extract (and it’s research twin Bibliomancer) was created to explore this type of service.

Reference Extract uses modern web search engine technology (NUTCH <http://lucene.apache.org/nutch/>) to search the output of digital reference authoring services; either in a library or a credible AskA service, like AskA Scientist and Ask Dr. Math. In the first search users can look for answers in digital reference archives. Users can then expand this search to look for information in web sites cited by these high-quality digital reference services. Reference Extract does not search the whole web, just the part created and cited by authoritative experts.

Reference Extract is founded in the concept of “reference authoring.” Reference authoring is the use of digital reference artifacts like transcripts and question/answer pairs to create new products. There are examples of creating training materials, search archives, and even text books from the process of answering questions.

This exploration makes the following assumptions:

- If an authoritative group builds tools it can be at least surmised such tools will be seen as credible (at least initially)

This exploration also has several research questions associated with it:

- How much of the original context of a digital reference transaction are needed to ensure the credibility of a tool? Do people searching Reference Extract need to know how often the URL was cited by experts? Do they need to know who cited the resources? Do they need to see the answers in which the URL’s were cited...the question that prompted the citation to web resources?
- Does weighting the results of a search towards more cited resources improve the system’s effectiveness?
- How can semantic web technologies enhance user’s determination of credibility? Imagine clicking on a search engine hit and finding who cited the resource, then click on that person to find their biography and background. Will users take the time to do this? Under what circumstances?

This is an ongoing research project, but centered on a real service. The intention is to run Reference Extract as a reliable service, and conduct research around real use. The following links provide more information on the project:

[Reference Extract](#): The actual service

[Reference Extract Blog](#): Blog for ongoing development and discussion of Reference Extract

[Bibliomancer Information](#): The research underpinnings of this project began their life as Bibliomancer. You can read that information and a more extensive (but drafty) [white paper on this work here](#).

R. David Lankes

This overview prepared by the director of the Information Institute of Syracuse at Syracuse University’s School of Information Studies. Special thanks to Michael Eisenberg, Blythe Bennett, Lisa Pawlewicz, Stuart Sutton, Scott Nicholson, and the DIG_REF community for their contributions to the ideas and operations of Reference Extract

<http://www.digref.org>