NET Institute*

www.NETinst.org

Working Paper #04-13

October 2004

MEDQUAL: Improving Medical Web Search over Time with Dynamic Credibility Heuristics

Mark Ginsburg

University of Arizona

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

MEDQUAL: Improving Medical Web Search over Time with Dynamic Credibility Heuristics

Mark Ginsburg

mginsbur@eller.arizona.edu

MIS Department, Eller College of Business, University of Arizona

1130 E Helen St., #430 Tucson, AZ 85721

***Abstract*—Performing a search on the World Wide Web (WWW) and traversing the resulting links is an adventure in which one encounters both credible and incredible web pages. Search engines, such as Google, rely on macroscopic Web topology patterns and even highly ranked 'authoritative' web sites may be a mixture of informed and uninformed opinions. Without credibility heuristics to guide the user in a maze of facts, assertions, and inferences, the Web remains an ineffective knowledge delivery platform. This report presents the design and implementation of a modular extension to the popular Google search engine, MEDQUAL, which provisions both URL and content-based heuristic credibility rules to reorder raw Google rankings in the medical domain.**

MEDQUAL, a software system written in Java, starts with a bootstrap configuration file which loads in basic heuristics in XML format. It then provides a subscription mechanism so users can join birds of feather specialty groups, for example Pediatrics, in order to load specialized heuristics as well. The platform features a coordination mechanism whereby information seekers can effectively become secondary authors, contributing by consensus vote additional credibility heuristics. MEDQUAL uses standard XML namespace conventions to divide opinion groups so that competing groups can be supported simultaneously. The net effect is a merger of basic and supplied heuristics so that the system continues to adapt and improve itself over time to changing web content, changing opinions, and new opinion groups. The key goal of leveraging the intelligence of a large-scale and diffuse WWW user community is met and we conclude by discussing our plans to develop MEDQUAL further and evaluate it.

***Index Terms*—Credibility, Web Credibility, Heuristics, Medical Informatics, Authoritative, Opinions, XML, Java**

Word Count: 6440

I. INTRODUCTION

Since its inception in the early 1990s, the World Wide Web (WWW) has seen a lowering of its barriers of entry to information providers and seekers alike. The ease of publishing, coupled with the commercialization of the NSFNet backbone in 1995, has led to an intermingling of facts (such as dictionary entries) and assertions of varying degrees of credibility. Individuals constantly access and try to make sense of the Web's vast content using a variety of tools, such as search engines and digital libraries. The explosion of Web content (both in sheer volume of pages and in supported format types, such as streaming media), coupled with the increasing ease of access to high speed bandwidth, confronts the information seeker with a mixture of high quality and dubious information sources, crippling the Web's potential to be an effective large-scale knowledge transfer platforms. As Pattie Maes points out, " […] the computer is a window into a world of information, people, software […] and this world is vast, unstructured, and completely dynamic" (Maes 1997). Maes, who hails from the Artificial Intelligence (AI) community, focuses on software agents that ideally would "know the user, know what the user's interests, habits, and goals are" (Maes 1997). Work along agent lines includes early attempts to mitigate search difficulties with a personalized software 'robot', or 'softbot' assistant (Etzioni 1994) to traverse Web resources intelligently on the user's behalf. However, over time, attention has shifted to indexing as many web pages as possible and leaving it up to the ad-hoc information seeker to sort through the resulting ranked list of results. He or she would then need to use personal heuristics to weed out undesirable content, such as off-topic items, stale information that has since been superceded, or assertions that are outright erroneous. Alternatively, one could consult only a set of trusted portals to gain information. These situations, with isolated ad-hoc behavior, hinder the ability of the Web to be an effective knowledge delivery platform.

A more dynamic approach involves converting the information flow from one-way to two-way, by allowing the information seekers to be secondary contributors of information. This has been seen in electronic marketplaces of expertise such as Answer Garden (Ackerman 1994) (Ackerman 1998) and the Annotate! system, which allowed organizational workgroup-level document annotation to augment search engine results (Ginsburg 1999). In situations where all participants are potential information donors, coordination mechanisms are critical between the primary content providers (authors, in the case of a digital library), and the secondary content providers (readers).

For example, notification mechanisms could alert authors to contributed reader activity, and aggregation measures can populate software recommender systems (Resnick and Varian 1997) to benefit the future readers. It is also important to note that individuals often form part of a specialized 'birds of a feather' group, for example Pediatrics or Internal Medicine among MDs, or Intensive Care Nurses among nurses.

Given a specific domain of interest and its audience pool, there are two important aspects of a networked knowledge transfer platform. We have (a) knowledge delivery, where the system is able to answer a broad range of questions within the domain to the satisfaction of a broad range of the audience pool, and (b) knowledge acquisition, where the audience can contribute ideas to the system's knowledge base for the subsequent benefit of all. In (b) we must be careful because ideas can be any combination of fact, assertion, and inference. For certain birds of a feather groups, a set of assertions may be axiomatic while for another, the same assertions are highly questionable. The current paper focuses on a practical example, the medical domain and its universe of Web pages. In our knowledge acquisition, we will implement a subscription model to allow users to join birds of a feather groups. The knowledge acquisition portion will take the form of the users supplying credibility heuristics (rules of thumb) which are of two types for a given query: (a) pattern matching versus a Google-returned URL system and (b) content matching on a phrase or set of phrases contained in the web page returned by Google.

We thus have the idea of using the Internet audience to build a large-scale information resource of interest via a two-way flow of information, in this case a set of medical information quality heuristics. There are prior examples of this type of approach. WordNet, OpenCYC, and Wikipedia, while adopting differing implementation philosophies (Lenat, Miller et al. 1995) (Wagner 2004) all leverage large numbers of users to build the resource; a dictionary in the case of WordNet and a freely available encyclopedia in the case of OpenCYC and Wikipedia. Our idea is more ambitious because medical opinions are often more debatable than dictionary or encyclopedia entries.

To realize this idea in concrete terms, we discuss in this paper the MEDQUAL[1] system, layered on top of Google, which is designed to provide a more effective knowledge transfer platform.

In the remainder of the paper we model a typical web text-based query session and some visualization extensions in Section II. We discuss pre-processing and post-processing coordination approaches to improve the situation in Section III and move on to our MEDQUAL system description in Section IV. Section V presents the concluding

[1] The website http://louvain.bpa.arizona.edu/medqual is the home page for this effort, to facilitate documentation and software distribution download.

comments and our immediate plans for further development and validation of this platform.

II. Modeling the Web Search Session

A typical text-based web query session has been modeled previously (Ginsburg 1998; Ginsburg 1999) and consists of a (usually stark) search interface, where a few keywords are inputted, a retrieval interface, where a list of results are rank-ordered (this typically includes a document title, a hyperlink to the base document, and a short summary of its contents), and finally the document interface, where the base document can be viewed by following a link in the retrieval interface. The basic situation with the three major interfaces is shown in Figure 1.

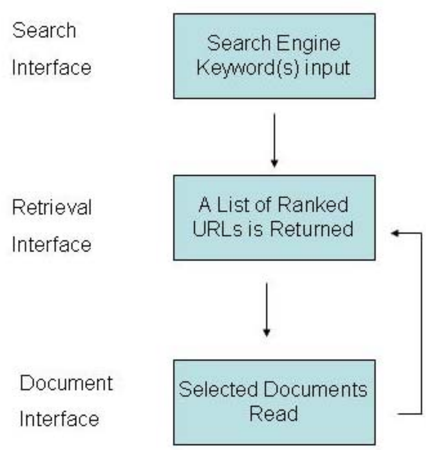

Figure 1: The Basic Search Sequence of Events: Three Interfaces

As can be seen, the situation is manually intensive. After typing in one or more keywords, the user must check the retrieval interface which offers a meager set of metadata clues: a ranked list of URLs along with document titles (if available) or summary snippets (if available). Then, selected documents are scanned for interesting content and the 'back' button is used on the browser to return to the Retrieval Interface to continue the process (Ginsburg 1998; Ginsburg 1999).

The first question revolves around the search interface and the relative merits of a full text search versus a controlled vocabulary search. One stream of work in the 1990s has focused on building a controlled vocabulary

[4] A key component of the system, which is in progress, is a simple logon whereby the individual can identify (subscribe) himself or herself with

system, similar to the Library of Congress Subject Headers classification scheme, which can be adapted to specific domains and help users with specific goals. (Ferguson, Allen et al. 1996; Ferguson and Allen 1998) These approaches can lead to maintenance difficulty and semantic drift as vocabularies change and evolve (Pejtersen 1998). There has been significant debate in the Information Retrieval Community between the controlled vocabulary approach (Blair and Maron 1985; Blair and Maron 1990) and the alternative of full text (Salton 1986; Salton, Allan et al. 1994) search. The debate focuses on the tradeoffs between ‘precision’ (the ratio of results on topic divided by the total number of results) and ‘recall’ (the ratio of relevant results divided by the total number of results in the document corpus universe). In our study, we do not constrain the user to a controlled vocabulary and use open-ended full text queries, following in the footsteps of today’s popular search engines, such as Google, and Alta Vista.

Let’s consider the specific case of a Google search. When an individual executes a Google Search, a query phrase is submitted to the system. The order of words in this phrase matters, and stopwords such as “`a`” and “`the`” are discarded, Furthermore, users can supply special characters such as + to ensure that the term actually appears in the document (and not just in sites pointing to that document), or “`<some phrase>`” to search for a phrase rather than its component words.

We can write the Google query as a vector or words: Q: (w_1, w_2, w_3, w_n). We do not pursue mathematical formalism in this paper, but it is interesting to note that the words themselves are combinations of alphabet letters and the set of all words forms a mathematical semigroup where the required associative operation is concatenation (Truss 1992) . Since Google drops stopwords, its allowable query set is a sub-semigroup of the query universe. The result from Google is a vector of URLs (pointers to web pages), together with titles (if they exist) and summary snippets (if they exist). Google operates using a variant of Jon Kleinberg’s algorithm (Kleinberg 1998) (Chakrabarti, Dom et al. 1998), which identifies macroscopic Web topology properties: ‘hubs’ and ‘authorities’. ‘Hub’ web pages point to ‘authority’ web pages – for example, many sites point to Ford Motor Company’s web site as a good resource to learn about the new Ford automobile models. By ranking highly the ‘authority’ web sites and the hubs that point to them, the Google search hopes to avoid isolated (‘non-authoritative’) opinions and assertions showing up highly in their rank order. For example, in the medical arena, many sites point to DrGreene.com as a trusted source of medical opinions (since Dr. Greene, a pediatrician, personally screens all content on that site) so

Google elevates DrGreene.com content. Similarly, many sites refer to Medline.Gov as a trusted source of clinical data so MedLine is elevated. In the retrieval interface, document data is present (the summary snippet) and a few metadata elements, such as the title, and the Google rank order. It is up to the user to divine which links are best to follow, given the usual constraints of time and energy, to satisfy the given information need. To accomplish this guesswork, the user peruses the data and metadata clues in the retrieval interface. The situation is inefficient due to limited clues, and limited screen real estate in the interface.

Numerous avenues for improvement suggest themselves. One line of research improves the retrieval interface by representing visually the concepts present in the result document set. In a review of visualization techniques, Ginsburg writes “if the search space and the potential audience are both broad, algorithms such as a Kohonen self-organizing map (SOM) technique may be used to generate two dimension concept maps. SOM maps have been widely used in conjunction with various visualization strategies; for example the comparison of fisheye and fractal views as described in a broad Internet search task by Yang et al. (Yang, Chen et al. 2003). The authors report that the concept subspaces may overload the user (excessive ‘visual load’) and visualization strategy is an important consideration. In addition, concept maps may be generated with Latent Semantic Indexing (Deerwester, Dumais et al. 1990); this technique has been applied in numerous specialized domains, for example the biomedical area (Chute, Yang et al. 1991). The Alexandria Digital Earth project visualizes geospatial data (Hill, Janee et al. 1999) using the ADEPT digital library architecture (Janee and Frew 2002) which provides logical “buckets” for metadata collection descriptions. Finally, the GenNav project (Bodenreider 2002) provides an interface to link a large glossary of genetic terms to the visualization of gene ontology pathways for biomedical researchers” (Ginsburg 2004).

If the document topics can be neatly categorized into an *a priori* classification scheme, then a hyperbolic tree is appealing. Figure 2 shows a situation where Information Systems research documents from the Cornell arXiv e-Print Technical Report archive were classified according to Association of Computing Machinery (ACM) and the joint ACM/IEEE (“CC2001”) scheme, with the reader able to switch between the two tree views. The interface supports fast ‘berrypicking’ (Bates 1989), i.e. locating rapidly items of interest, and supports user learning. The user, while interacting with the interface over time, can learn more about the underlying structure of the document collection. When possible, the tree interface is an effective navigation tool, offering “Focus + Context” information visualization (Green, Marchionini et al. 1997) (Leung and Apperley 1994) (Plaisant, Carr et al. 1995). This type of visual organization has been shown to lead to faster and more intuitive user navigation (Pirolli, Card et al. 2000; Pirolli, Card et al. 2001; Börner and Chen 2002). The Focus + Context frames give the users clues in both the top and bottom window (cf. Figure 2) to help locate information more quickly, i.e. "reduce the cost structure of information" (Card, Mackinlay et al. 1999).

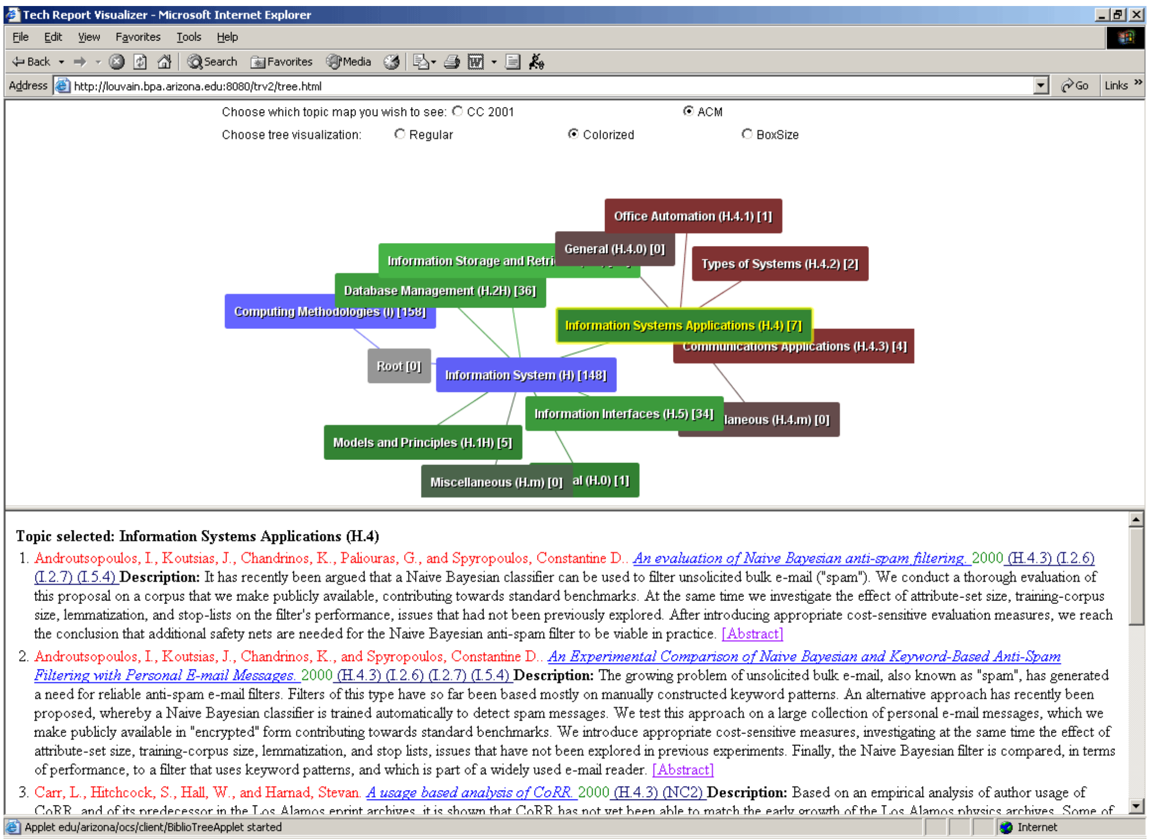

Figure 2. The Hyperbolic Tree Approach of Query Result Visualization (Ginsburg 2004)

However, if we unconstrain the document universe and consider all of the billions of items in the Google index, we cannot easily make use of classification structures. If we have to fall back on a text search interface (think of the typical Google experience), we must instead focus on providing coordination mechanisms to leverage the intelligence of the user community, and provide support to birds-of-a-feather common interest groups as well. The ultimate goal is to improve the Web search as a knowledge transfer vehicle for both users and groups, and this means improving either the performance of the front-end (the search interface), or the data and metadata clues made available by the back-end (the retrieval interface), or both. This is the focus of the next section.

III. IMPROVING THE WEB SEARCH SESSION

A. Query Transformation and Reduction

Keeping in mind the Search, Retrieval and Document interfaces that are part of a generic web search session, the first potential improvement is in the behavior of the front-end, or Search interface. Let's consider the medical domain as an example. If we could filter the query input and reduce the query set to a set of root queries, the system would have more consistent behavior. One way to do this is by analyzing the query vector, token by token, and find the most general match to the set of root queries. The technology to do this has already been developed: ALICE, or the Artificial Linguistic Internet Chat Entity, has a pattern matching engine. (Wallace 2004). It encodes patterns in AIML (Artificial Intelligence Markup Language), an XML dialect. Its standard distribution (there are various ALICE distributions freely available via http://www.alicebot.org) comes with approximately 24,000 patterns covering assorted geography, nature, and human interest facts. In addition, ALICE has mechanisms to engage the participant in conversational small talk and it supports access to third party networked resources via a scripting language. A test implementation linking ALICE to a set of Web services in a prototype portal was demonstrated in (Ginsburg 2002). The distribution also provides automated dialog logging, archiving, and visualization using XML and XSL. Transforming the query to a known root case has the additional advantage of cleaning up potential typographical errors, or ungrammatical sentences, from e.g. a non-native English language speaker. The pros of this approach are forming a robust taxonomy of root queries, presumably with the assistance of an expert panel. Another appealing aspect is that coordination mechanisms can inform and evolve the taxonomy over time, as the system is used. In addition, an advantage is that the core ALICE distribution is geared toward dialog. Thus, queries judged to be incomplete can trigger dialog to elucidate the full query or clarify an ambiguous query. Thus, for example, the query `“pain in my arm”` can trigger a follow-up dialog to find out which part of the arm hurts. A more specific location will result in a better-tuned search query. The cons are the time and effort to construct the taxonomy, and the possibility that a new query is not reducible to one of the root queries. In that case, the new query is passed to the Search engine untransformed and potentially flagged for addition to the taxonomy. Maintenance costs might be onerous in this approach, however. Another potential difficulty is that different birds-of-a-feather interest groups may have different reduction and transformation rules that are optimal for them. The system would have to maintain multiple transformation rule namespaces for group support; this consideration applies as well to the MEDQUAL system which we describe in Section IV.

B. Augmenting the Query Results

Another approach, which is the first thing we undertook in the MEDQUAL system, is to augment the search engine's standard results with additional metadata. There are numerous potential metadata sources that come to mind. In fact, the Yahoo search engine relies on Yahoo workers to canvass manually many popular web sites and to incorporate the workers' opinions when fine-tuning the rank ordering of the results. Another notion is to filter the result set on two major considerations: (i) the originating site, as given in the result URL and (ii) patterns in the content of the webpage, (which may or may not be read by the user in a typical web search; recall the opportunity cost of traversing repeatedly from the Retrieval Interface to the Document Interface). If URL and content filtering rules can be successfully applied to accomplish a reordering of the standard result set, then the opportunity cost of finding the appropriate information should be reduced. To put it another way, the credibility (credibility is roughly synonymous with 'believability' (Tseng and Fogg 1999)) of the top results is increased and the overall effectiveness of the web search as a knowledge transfer mechanism is increased. It is worth noting that even web sites deemed to be authoritative by macroscopic structure (e.g. the Google search engine using the Kleinberg algorithm), such as www.wedmd.com, are in fact a mish-mosh of professional and lay opinions. In additions, lay questions, potentially unanswered, may also match the user's query and show up highly ranked in the result set. Thus, we are concerned with individual web pages as they appear in the search results, not with the overall credibility of a web site (which was the focus of earlier work, for example (Fogg, Soohoo et al. 2002) (Tseng and Fogg 1999). It is also worth noting that web pages offer a diverse assortment of facts, opinions, and assertions. At any given time, for a given query and a given user, documents in the result set will range from the extremely useful to the extremely useless.

C. Grouping the User Population

An avenue not typically pursued in web search is to group the users into birds of a feather shared-interest groups. Shared interest groups will often harbor similar opinions and will make similar inferences given a set of basic facts.

For example, consider the Pediatric specialty in medicine. Mainstream pediatricians believe there is no link between the Measles-Mumps-Rubella (MMR) immunization and childhood autism, despite a few studies in the United Kingdom which suggest this link. Therefore, web pages promulgating this hypothesized link would not be credible to mainstream pediatricians but would be highly credible to those who believe in the soundness of the studies mentioned above. It would be advantageous for a search session to support subscription into one or more interest groups in order to make use of 'common wisdom' – it would also be advantageous to make the system flexible so that the individual can decouple from group opinions as need be. The system should also have the flexibility to adapt to scientific advance – in applied science, beliefs are often updated by advances in core disciplines (for example, biochemical and genetic advances informing medicine).

In Section IV we present MEDQUAL, a system that operates in conjunction with the popular search engine Google to augment the credibility of a rank-ordered list of results given a particular medical web query. From the above discussion, our basic approach will be a combination of (b), augmenting the results (implemented), and (c), grouping the user population (in development). To accomplish (b), MEDQUAL makes use of URL and content rules that come from two sources: (i) bootstrapped at system initialization, and (ii) user supplied via two-way author-reader coordination mechanisms. These rules are applicable at both the general and the specialized (common interest group) level. MEDQUAL does not, as of this writing, incorporate query transformation as discussed in Section IIIa. However, the two approaches are not exclusive and may be combined. In the next section, the prototype implementation is presented that demonstrates (b), augmenting Google results.

IV. MEDQUAL: Improving Web Query Result Credibility

The MEDQUAL system interposes itself between the user and the standard Google search engine, making use of the standard Google Application Programming Interface (API) (Calishain and Dornfest 2003). At startup, the MEDQUAL system reads in a simple XML configuration file with promotion and demotion rules[4].

The first rule level is the domain name of the result site (a match versus the URL pattern). Some sites, according to a certain set of birds of a feather groups, have built-in credibility (such as NIH.GOV). Depending on the user's group subscription elections, we apply the URL pattern rules. After the domain name is checked, we also check content. The system follows the result link to retrieve the contents internally and it scans for certain patterns. Some content rules are followed on a domain-by-domain basis (dependent on the given domain) and some are global, applied to all domains. The XML tree is a natural way to nest Domain and Content promotion and demotion rules. The system runs on the client computer as a Java applet (for more information on the system details, see Appendix A). For demonstration purposes, we initialized the configuration file by consulting medical professionals who ran sample medical queries and then were able to identify a few simple URL and Content rules.

The introductory point of the system is shown in Figure 3. For simplicity and demonstration purposes, we pre-loaded the system with some sample queries. The system will be developed to support both pre-written and new (ad-hoc) queries.

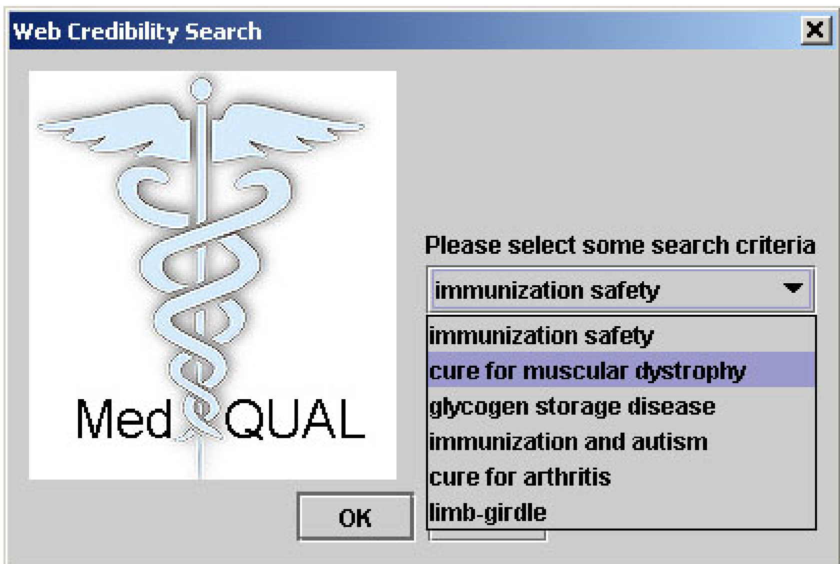

Figure 3. MEDQUAL Start Point

After a query is selected, the Google Web Service is invoked and the Google results are loaded into memory. The URL and content rules are then consulted to reorder the list. An example of the reordered list, with a promotion rule activated, is shown in Figure 4.

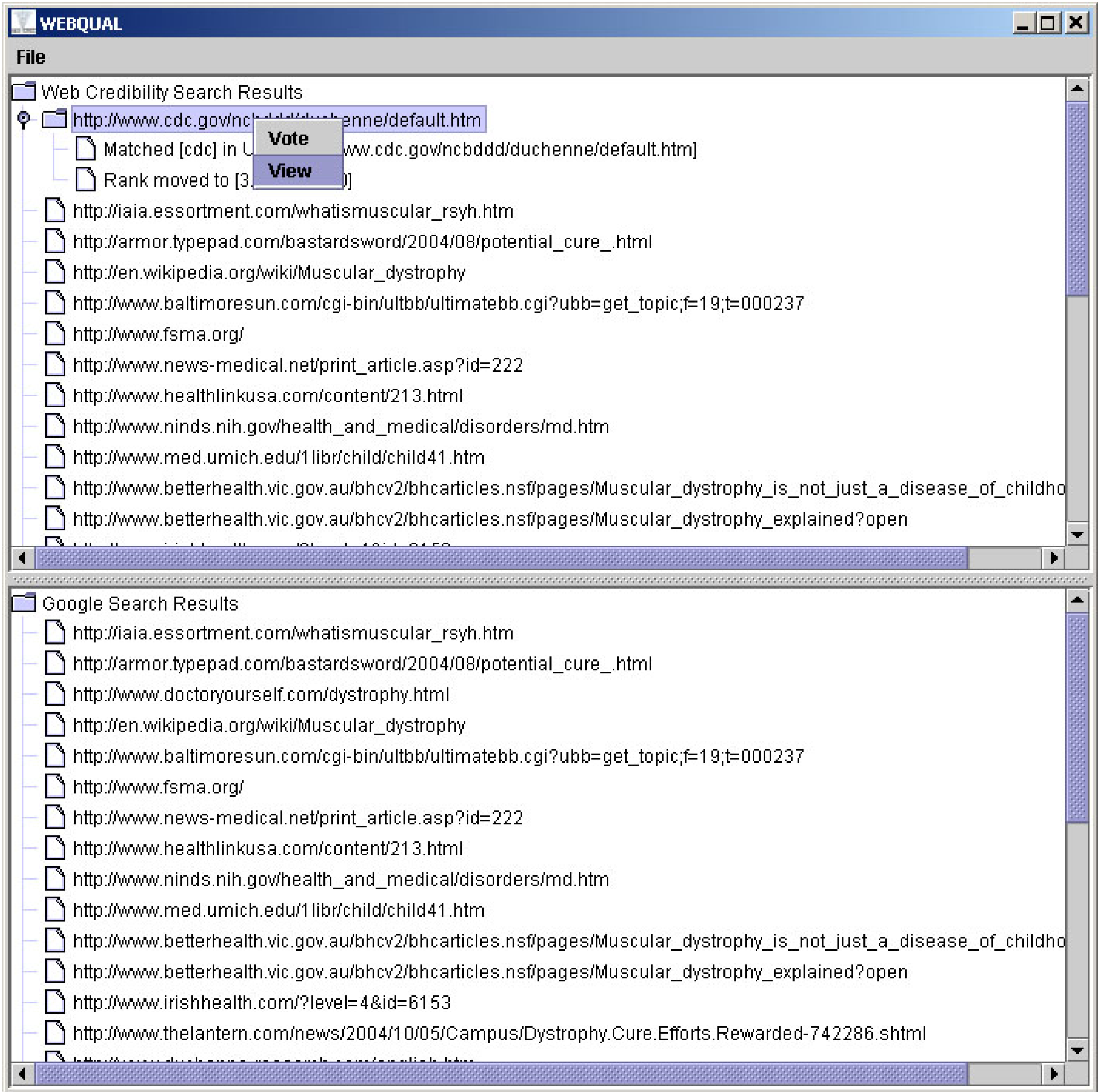

Figure 4. CDC.GOV site is Promoted.

As Figure 4 shows, a match of the pattern `cdc.gov` in the URL has promoted that entry to the top. The URL and Content rules, when applied, appear in the interface with an "expand-here" icon. If the user expands the selection, the exact rules that were triggered are shown. Figure 4 shows the entry already expanded. The user can similarly elect to compact the view, hiding the rule activations, and show only the result URLs. In the bottom frame, the raw (unordered) Google set is shown for comparison. Our promotion rules are simple: in a given result set, all promotions are moved to the top in weight-order. We assigned, arbitrarily, a weight of +3 to `cdc.gov`. In practical use, an expert committee will be able to assign weights using consensus opinion. Or, specialty groups can contribute rules over time to bring the system to the same level it gains by the input of an expert committee. Similarly, demotions move to the bottom of the result set again in weight-order. We actually pull out the top thirty Google results and hence demoted appears can disappear entirely from the reordered view. Referring to Figure 4, another action the user can take is to right-click on an entry. This will pop up the "Vote/View" dialog box. The user can elect to View the page (which pops up a new web browser) or to Vote on its applicability and credibility to the task at hand. The Vote mechanism will contain a simple Likert scale, from 1 to 7 as shown in Figure 5 . Likert scales are advantageous for statistical analysis of aggregate voting patterns. The vote module as shown is incomplete; it is currently under development and will also include other user feedback. To make the voting mechanism consistent with prior work (Ginsburg 1998; Ginsburg 1999) we plan to have a list of reasons (e.g. rationales for the given vote), for example "information out of date", "information gave me new ideas", "information superceded by", and so on – represented as a set of intuitive icons. There will also be an optional text box so the user can input another URL in the case, for example, of new information (a web page potentially not on Google's list) superceding the web page returned by Google. The important design consideration is to make the vote mechanism lightweight and easy to use; to minimize the opportunity cost for the user to actively participate and contribute.

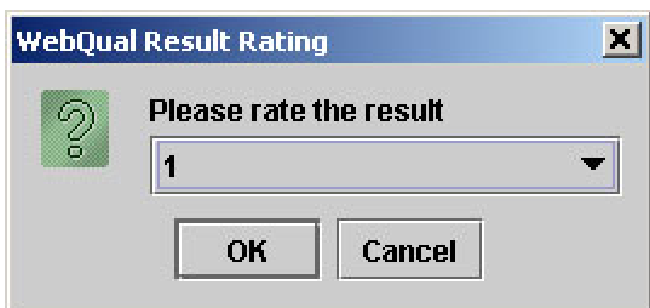

Figure 5. Prototype Voting Mechanism (under development).

The flip side of Figure 4 in shown in Figure 6. Here, multiple demotion rules are in effect, both on URL and on content.

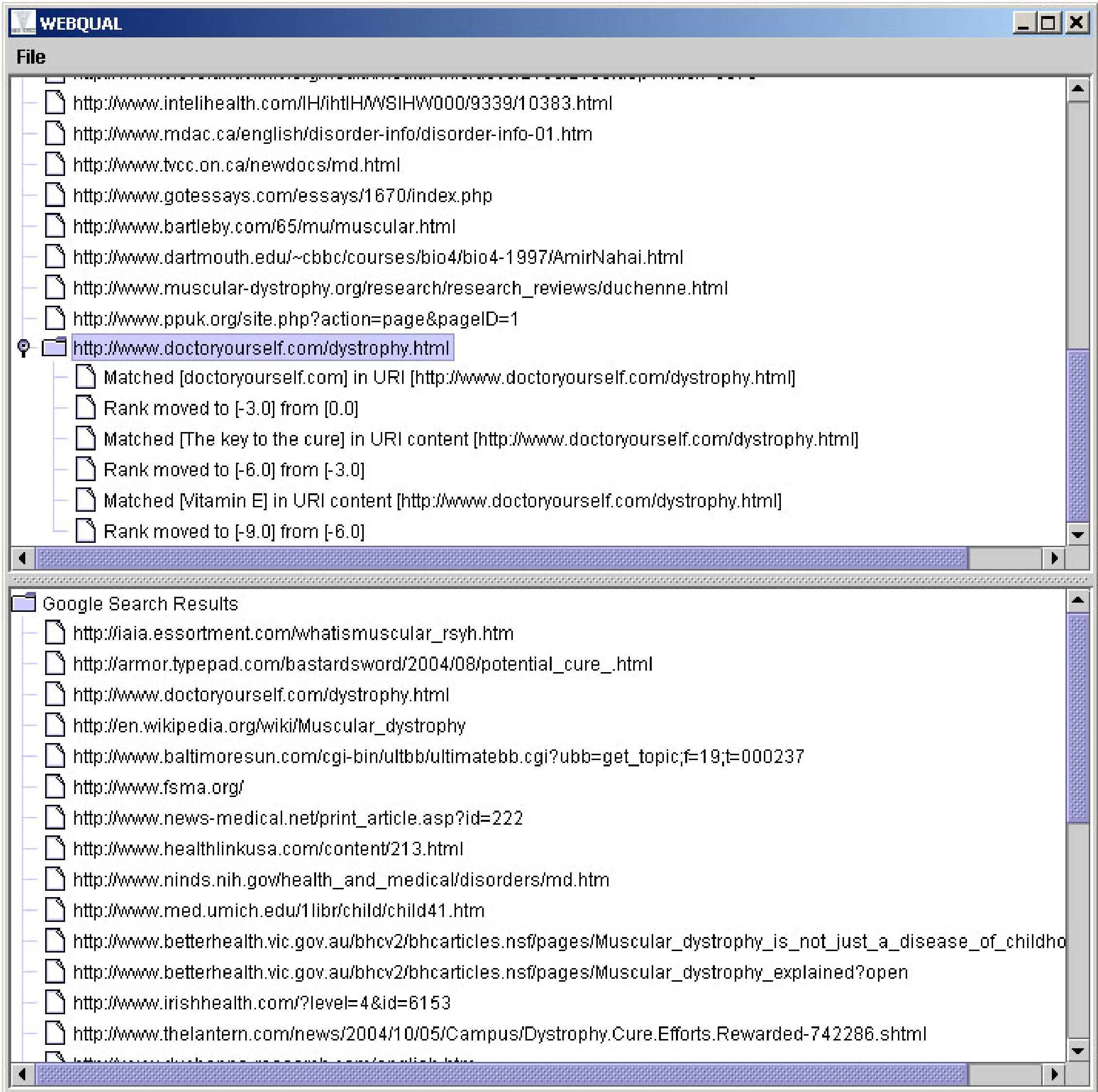

Figure 6. DoctorYourself.Com is demoted on both URL and Content.

In Figure 6, a user (let's assume this user has subscribed to the mainstream Pediatric specialty group, although the group subscription module is under development) has run a query on "Muscular Dystrophy". One of the highly ranked Google results is a website, http://www.doctoryourself.com. According to the expert committee we canvassed, this web site contains highly questionable assertions about the causes of Muscular Dystrophy and purports that Vitamin E is linked to a cure; this claim is labeled as false by the expert group. Thus, the URL is demoted and chained content rules are invoked. Catch phrases "key to the cure" and "Vitamin E" are detected and further demotion occurs. As can be seen from the bottom half of Figure 6, this web site was ranked highly (#3) in the unordered Google results.

V. Concluding Comments and future plans

The MEDQUAL system was conceived to address one of the Holy Grails of networked computing: leveraging the intelligence of the user community. It will do so by permitting two-way information flow – authors contribute documents to the web, and readers become secondary authors by assigning credibility weightings to the documents' origins (URLs) and contents. These weights in turn become part of the heuristic rule base of the system. MEDQUAL, when fully implemented, will support individuals and birds of a feather specialty groups and will be an excellent experimental platform for further studies on the related issues of web credibility and what constitutes an authoritative web page. It is designed to be flexible, extensible, and simple which is particularly important in networked computing and groupware. Thus, people should be able to subscribe and unsubscribe from multiple opinion groups as need be, and the voting should be as painless as possible.

The most important goals we are trying to attain is that of self-sustainability and growth – the system should bootstrap with a core set of heuristics to make it immediately useful , and ongoing use of the system should strengthen it for the benefit of all.

We are focusing now on completing the Voting interface, and the Group subscription logon functionality. Then we will be ready to conduct system trials in both laboratory and field settings. Validation of the system will be an interesting exercise. It will involve various users, belonging to various groups, running specialized queries in a given domain (for example, the medical field). For the task at hand, is the system's reordering of search results effective? Is the subscription model (merging opinion groups' heuristics) effective? Effectiveness must be gauged in terms of user constraints; typically the user is constrained on both time and effort. We feel the approach is very exciting and offers much potential to a wide range of specialized opinion groups. We look forward to evaluating and hopefully demonstrating its worth in improving basic search engine functionality.

Acknowledgments

The seed development work described here was supported by a summer 2004 grant by the NET Institute. The author thanks Prof. Nicholas Economides for bringing the NET Institute to his attention, and John C. Dale, a MS-MIS student at the University of Arizona, for prototype coding.

References

Ackerman, M. (1994). Answer Garden: A Tool for Growing Organizational Memory, Massachusetts Institute of Technology.

Ackerman, M. S. (1998). "Augmenting Organizational Memory: A Field Study of Answer Garden." ACM Transactions on Information Systems **16**(3): 203-224.

Bates, M. J. (1989). "The design of browsing and berrypicking techniques for the on-line search interface." Online Review **13**(5): 407-431.

Blair, D. C. and M. E. Maron (1985). "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System." Communications of the ACM **28**: 289-299.

Blair, D. C. and M. E. Maron (1990). "Full-Text Information Retrieval: Further Analysis and Clarification." Information Processing and Management **26**: 437-447.

Bodenreider, O. (2002). Experiences in visualizing and navigating biomedical ontologies and knowledge bases. Intelligent Systems for Molecular BIology 2002, Edmonton, Canada.

Börner, K. and C. Chen (2002). Visual interfaces to digital libraries. Proceeding of the second ACM/IEEE-CS joint conference on Digital libraries, ACM Press.

Calishain, T. and R. Dornfest (2003). Google Hacks: 100 Industrial-Strength Tips and Tools. Sebastopol, CA, O'Reilly.

Card, S. K., J. D. Mackinlay, et al. (1999). Information Visualization. San Francisco, CA, Morgan Kaufmann.

Chakrabarti, S., B. E. Dom, et al. (1998). Experiments in Topic Distillation. San Jose, CA, {IBM} Alamaden Research Center.

Chute, G. C., Y. Yang, et al. (1991). Latent semantic indexing of medical diagnoses using UMLS semantic structures. Proc., 15th SCAMC, Washington, DC.

Deerwester, S., S. T. Dumais, et al. (1990). "Indexing by latent semantic analysis." Journal of the American Society for Information Science **41**(6): 391-407.

Etzioni, O. (1994). "A Softbot-Based Interface to the Internet." Communications of the ACM **37**(7): 72-76.

Ferguson, G. and J. F. Allen (1998). TRIPS: An Integrated Intelligent Problem-Solving Assistant. American Association for Artificial Intelligence**:** 567-572.

Ferguson, G., J. F. Allen, et al. (1996). The Design and Implementation of the TRAINS-96 System: A Prototype Mixed-Initiative Planning Assistant. Proceedings of the Third Conference on Artificial Intelligence Planning Systems.

Fogg, B. J., C. Soohoo, et al. (2002). How do People Evaluate a Web Site's Credibility? Results from a Large Study, Stanford University Persuasive Technology Lab.

Ginsburg, M. (1998). Annotate: A Knowledge management support system for Intranet Document Collections. IS. New York City, NYU.

Ginsburg, M. (2002). The Catacomb Project: Building a User-Centered Portal the Conversational Way. WIDM 2002 (Fourth International Workshop on Web Information and Data Management), McLean, VA, ACM.

Ginsburg, M. (2004). "Visualizing Digital Libraries with Open Standards." Communications of the AIS.

Ginsburg, M., Ajit Kambil (1999). Annotate: A Knowledge Management Support System. HICSS-32, Hawaii, IEEE.

Green, S., G. Marchionini, et al. (1997). Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. College Park, MD, University of Maryland CS.

Han, S. and Y. Kim (2001). Intelligent Dialogue System for Plane Euclidean Geometry Learning. International Conference on Computers in Education, Seoul, Korea.

Hill, L. L., G. Janee, et al. (1999). "Collection metadata solutions for digital library applications." Journal of the American Society for Information Science **50**(13): 1169-1181.

Janee, G. and J. Frew (2002). The ADEPT DIgital Library Architecture. JCDL'02, Portland, OR, ACM.

Kleinberg, J. (1998). Authoritative Sources in a Hyperlinked environment. Proceedings, 9th SIAM Symposium on Discrete Algorithms, ACM.

Lenat, D., G. Miller, et al. (1995). "CYC, WordNet, and EDR: critiques and responses." Communications of the ACM **38**(11): 45-48.

Leung, Y. K. and M. D. Apperley (1994). "A review and taxonomy of distortion-oriented presentation techniques." ACM Transactions on Computer-Human Interaction **1**(2): 126-160.

Maes, P. (1997). "Pattie Maes on Software Agents: Humanizing the Global Computer." IEEE Internet Computing: 10-19.

Moore, R. and G. Gibbs (2002). Emile: using a chatbot conversation to enhance the learning of social theory. Huddersfield, England, Univ. of Huddersfield.

Pejtersen, A. M. (1998). "Semantic Information Retrieval." Communications of the ACM **41**: 90-92.

Pirolli, P., S. K. Card, et al. (2000). The effect of information scent on searching information: visualizations of large tree structures. Proceedings of the Working Conference on Advanced Visual Interfaces, ACM Press.

Pirolli, P., S. K. Card, et al. (2001). "Visual Information Foraging in a Focus + Context Visualization." ACM Conference on Human Factors in Computing Systems, CHI Letters **citeseer.nj.nec.com/pirolli01visual.html**
3: 506-513.

Plaisant, C., D. Carr, et al. (1995). "Image-browser taxonomy and guidelines for designers." IEEE Software **12**(2): 21-32.

Resnick, P. and H. R. Varian (1997). "Recommender Systems." Communications of the ACM **40**: 56-58.

Salton, G. (1986). "Another Look at Automatic Text-Retrieval Systems." Communications of the ACM **29**: 648-656.

Salton, G., J. Allan, et al. (1994). "Automatic Structuring and Retrieval of Large Text Files." Communications of the ACM **37**: 97-108.

Truss, J. K. (1992). Discrete Mathematics for Computer Scientists, Addison-Wesley.

Tseng, S. and B. J. Fogg (1999). "Credibility and Computing Technology." Communications of the ACM **42**(5): 39-44.

Wagner, C. (2004). "WIKI: A Technology for Conversational Knowledge Management and Group Collaboration." Communications of the AIS **13**: 265-289.

Wallace, R. S. (2004). The Anatomy of A.L.I.C.E. A.L.I.C.E. Artificial Intelligence Foundation, Inc.

Yang, C. C., H. Chen, et al. (2003). "Visualization of large category map for internet browsing." Decision Support Systems **35**: 89-102.

VI. Appendix A. MEDQUAL System Details

The MedQUAL system was built with modern technology tools: Java, JDOM (a Java XML parser), J2EE components, and Java 'Swing' Graphical User Interface Components. The system runs as a Java applet on the client desktop. These components were selected for portability, maintainability, and extensibility. XML is used to encode heuristic patterns so that we can make use of an XML standard, XML Namespaces, to support multiple opinion groups simultaneously. XML Namespaces allow us to easily merge (take the union of) several opinion sets.

The MEDQUAL web site is http://louvain.bpa.arizona.edu/medqual and will contain links so that interested parties can download the software distribution. We will also provide current development comments and to-do lists on that website, as well as accept bug reports.

The Primary Components of the System

- *MedQUAL Server* **–** The MedQUAL server brokers user query requests concurrently. In addition, it exposes a secure interface wherein the users of the system can execute the implemented functionality. It is composed of several logical components.
    - Vote Controller – The vote controller manages the associations between the Root Queries, the root query results, and the result rankings. It is responsible for keeping track of user ranking data by associating a ranking to a user and to to the particular result that is being ranked.
    - Heuristic Controller – The Heuristic Controller is responsible for reading and summing the votes for a query's results so that they can be ordered based on expert opinion within a namespace.
    - Query Resolver Controller – This is as yet unimplemented. When we pursue the ALICE stategy, referenced in Section IIIa, the query resolver controller will take in a user's query string and resolves it to some root query in the system (if one exists).

- *MedQUAL Client*

    - Bootstrap Heuristic Controller – The Bootstrap configuration is specific to the namespace and is loaded when the client starts up. The Bootstrap heuristics are used in conjunction with specialized group heuristics, contributed by system users or by expert committee.
    - Misc – There are standard presentation framework components that will be used to display and sort the results that are returned. In addition, a simple network communication layer component will wrap a network communication protocol for communication with both the search engine of choice and the MedQUAL server. For the time being, MEDQUAL is hooked into the Google search engine only.

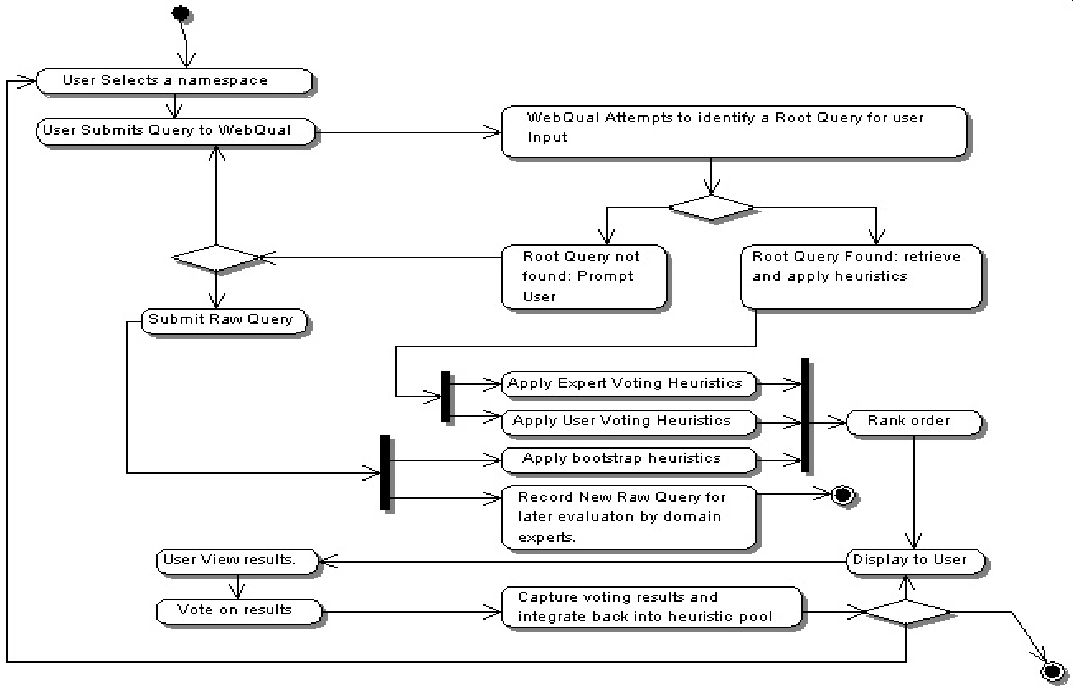

Figure 7. MEDQUAL Activity Diagram (Part 1 of 2)

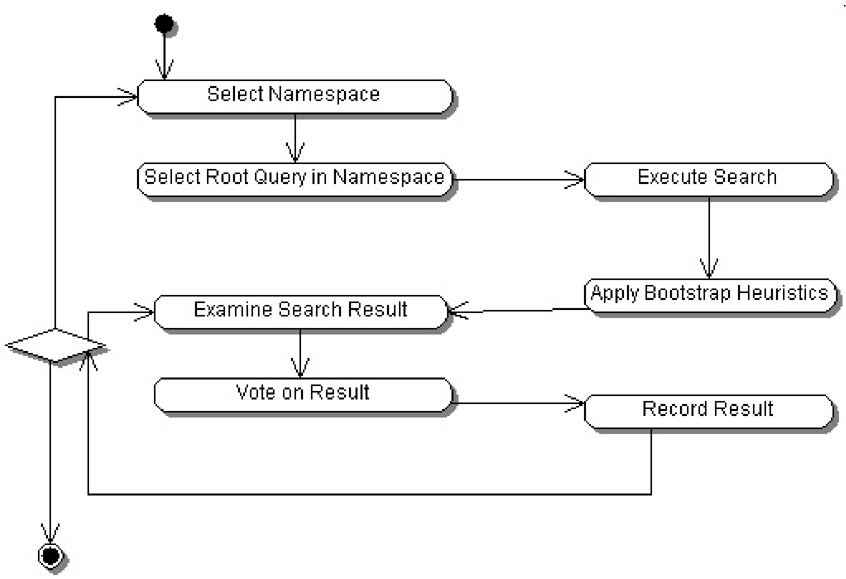

Figure 8. MEDQUAL Activity Diagram (Part 2 of 2)

Figure 7 and Figure 8 shows the activity diagram that represents the end user's primary experience with the system. Bootstrap heuristics and specialized heuristics are combined and the MEDQUAL server reorders the Google results appropriately. We are developing the vote mechanism to coordinate readers and authors, by allowing readers' votes to flow into the contributed heuristics rule base. We are also developing the organization of the readers into specialized groups. Implied in these figures is a mechanism for a system administrator(s) to make changes to the bootstrap heuristics which are applied globally for all users.

The maintenance of the system will be on the shoulders of the expert groups within the various namespaces. This activity diagram details the process of examining and ranking search engine results for root queries. With the exception of the last step, all activity occurs on the client side as queries are made to a search engine, the results examined, and rankings determined. For a given root query, the results are then stored in the database in a structure that will allow the Heuristic Controller to read and tabulate the ranking heuristics for a given namespace.

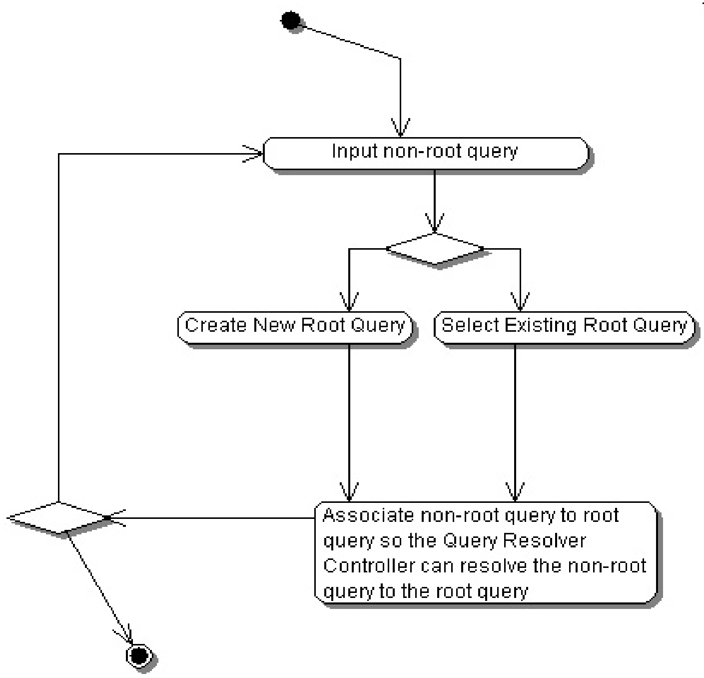

Figure 9. System Training to Learn New Root Queries

Figure 9 represents the training of the system to resolve non-root queries to root queries. This would be used in, e.g., adapting the ALICE engine to pre-search query processing. Pattern matching techniques will be used to resolve user queries to well known system root queries. Then, under normal usage scenarios, on the end user's behalf, the root query will be submitted to the MedQUAL system, heuristics applied, and results sorted accordingly.