Exploring Market Mechanisms for Community Web Search

Gordon Rios
gparker@gmail.com

June 15, 2004

Abstract

Community based web search should in principle improve the quality of search engine results via taking advantage of large user communities. However, there are many challenges towards realizing the goal of leveraging user endorsements for improving search engine relevance. This paper explores some of the challenges and presents a candidate approach based on creating a market for trust among identified users able to form highly relevant query-url associations.

1 Introduction

This paper is concerned with using community or social network effects to enhance the quality of user experience with web search. Web search is viewed as a simple application whereby users expend time in the expectation of satisfying information needs. Further, the concerns are narrowed to mechanisms that can enhance the search engine's ability to bring documents into their top ten results to help the user satisfy the information need associated with the query. Earlier studies, for example [4], have identified that most searchers rarely look beyond the first ten results. In general, the framework for web search and internet users assumed for this paper follows that described in works like [1, 2].

1.1 User Click Feedback and Web Page Annotations

There have been many attempts to leverage the communities of search engine users and web page authors to increase the relevance of search result ranking. Two well explored approaches are the analysis of user behavior on search results and web page link annotations known as *anchortext*. The primary mechanism for capturing user feedback is via measuring various transforms of clicks on urls for a given query.[1] Currently, certain engines look closely at the last url that the user clicks on for a given query session.[2] It's not surprising that if many users end their session on the same url for a given query then that url is often relevant for that query. Another important feature is *anchortext* which proven to be quite helpful in relevance ranking[3] and this effect is analyzed in [3].

[1]DirectHit*(TM)* and Inktomi*(TM)* were early adopters of this approach.

[2]While at Inktomi*(TM)* the author referred to these as *end clicks* and found them to be highly predictive of url relevance.

[3]Google*(TM)* was probably the first engine to make widespread use of this mechanism – to quite good effect.

For user feedback and annotations the most dramatic results are seen on content for popular queries. Importantly, both approaches have severe limitations due to ongoing efforts by search engine optimizers (SEO) and outright spammers who subvert feedback and annotation efforts to increase the ranking of their submitted urls.[4] In addition, there are subtle modeling issues for both approaches. In user click analysis one must measure the amount of information in click behavior and control for the effects of ranking on the prior probability of click given the rank (i.e. control for the *rank effect*.) However, since the quality of the url at any given rank varies by query, the *rank effect* is nearly impossible quantify.[5] For *anchortext* the problem of sparseness is often an issue and attempts to use anchortext for all but exact query matches is error prone and often low precision.[6]

1.2 New Content Discovery and Promotion

Another aspect of relevance is the discovery of novel content. Most search engines have some type of *submit url* mechanism which presumably would allow the popular engines to include important new content. Again, due to the economic incentive provided by affiliate revenues and techniques of the SEO community, url submissions from the user community are almost entirely discounted.[7] Despite these difficulties it is critical that producers of good (if perhaps obscure) content have a way to introduce it into the index. In the world of paper based content, many publishers develop expertise in discovering unknown producers and bringing their work to larger audiences.[8]

1.3 Community Feedback on Query-Url Associations

Without too much difficulty one can imagine a community of search users who endorse or introduce urls for specific queries. For example, an informed user might select a url at rank eight to be highly relevant for a given query. Thereafter, other users issuing the same query would see the result boosted and annotated with the user-id of the endorser. Of course this kind of mechanism is already in widespread use in many large sites on the web.[9] For web search we try to satisfy a user's *information needs* [10] which can be assumed to drive query generation. Thus, the content returned for a query and it's ranking should reflect the user *information needs* most likely to have generated the query (i.e the focus is on the *maximum likelihood* information need.)

[4] This is such a widespread phenomenon that it hardly needs mentioning.

[5] Technically, we'd like to control for $p(click|rank)$ but instead we're forced to consider $p(click|rank, query)$ which is sparse for all but the most popular of queries.

[6] For example extracting *Linux Documentation* from a link annotation *Deprecated Linux Documentation* or *Humorous Linux Documentation* would probably reduce precision for most searches for *Linux Documentation*.

[7] During his tenure at Inktomi*(TM)* the author was told that editor sampling on the *submit url* revealed more it to be overwhelmingly made up of spam. For *paid inclusion* submissions associated with keywords editor sampling reveals that over 99% of it is unfit for use.

[8] For example, J.K. Rowling, the author the popular *Harry Potter* series was a struggling single mother before she was discovered by a small Scottish publishing house.

[9] At Amazon, for example one can often see reviews from very experienced users annotated as *Top 500 Reviewer*. For movies, one can visit *movies.yahoo.com* to see individual reviews sorted by numbers of people who found the review helpful, etc.

[10] While working for Inktomi*(TM)* the author used the phrase *query Intention* to describe this idea. The name *information need* is more widely prevalent. The author's first exposure to the term was in a talk given by Prof. John Canny from UC Berkeley in 1999.

2 Community Web Search

There's no reason to believe that explicit user endorsements for web search results would fare any better than the passive collection of user click behavior or *anchortext*. A naive approach to collecting endorsements would in all likelihood draw the same significant volume of spam submissions and feedback. However, there are social networking aspects that might be exploited. For example, if we consider the case of drawing endorsements only from a pool of registered users we can annotate the search results with the user-id. Then, not only can we track the user click behavior on annotated urls we also provide high information cues for system users to seek out or avoid results of endorsers they trust or distrust. Thus, we are exploiting past user history and can hope to sharply improve the efficiency with which untrustworthy endorsers are rooted out of the system. We'll say more about this later, but the idea that reputation can improve the effectiveness of exchange mechanisms is certainly gaining support. An interesting study demonstrating the effects of reputation on an exchange mechanism can be found in [5].

Community endorsements for relevance ranking also provide possible avenues for personalizing web search. The user interface (UI) could be enhanced to allow users to block the ranking effects of endorsers they distrust or boost those from trusted sources.[11] It might also be beneficial to provide the user with the option of applying these block or boosting effects on a query specific basis or globally across all future searches. Most importantly, we empower the user to personalize results and simultaneously create a rich feedback signal by aggregating the independent trust or distrust signals we get for registered endorsers.

Our discussion so far begs the question of what impact should user endorsements have on ranking? For example, a widely trusted source that endorses a specific url for a query certainly should improve the rank but by how much? For example, any direct quantitative assessment could require us to condition on the features or attributes of all the other urls in the result set to fully capture the correct ranking of a url. This is important because the ranking of one type of url depends on the information need driving the query and how well the results satisfy that need. The point is that the best result set mix is strongly query dependent. For example, if one is buying a product there might exist common combinations and rankings of transactional sites, review sites, or auction sites. Where good urls are missing, the correct rewards (whether intrinsic or extrinsic) could induce informed users to introduce content for queries generated by that need. Some methods require that the graph of recommenders be analyzed directly using techniques from large scale graph analysis.[12] In general, search ranking is an open ended question and may never yield a complete answer because of it's connections with the AI Problem [10]. For this paper, we'll leave this important topic and assume that given trustworthy endorsements from real people we can improve the ranking and content of our search engine.

[11] We could also leverage mechanisms from social networking and allow the user to extend his judgments to those endorsers who have endorsed his target. For now, we'll confine our attention to a simple rating system.

[12] A good summary of graph techniques for the search problem is contained in [6].

3 Economic Mechanisms for Community Web Search

The problem with current approaches to imputing or aggregating user feedback (via click behavior, annotation) is that promoting or submitting an item has essentially zero cost. The volume of endorsements an individual can produce is limited only by his or her time, energy, and scripting skills. As long as there is no friction on the endorsement process there will be abuses of the system that largely wipe out potential gains. Making endorsements explicit can be expected to somewhat mitigate these effects but are unlikely to provide a full solution. Instead, we must find a way to extract a cost or bid from the endorser so that in a sense he or she is putting something at risk by making the endorsement. Ideally, the right mechanism would find an equilibrium where people only give endorsements when they are comfortably sure that the result should have higher ranking or inclusion.

3.1 An Early Economic Mechanism for Web Search

For completeness we should note that *Overture(TM)*, a company that allows Vendors or SEO's to explicitly bid for ranking, solved part of this problem. By pricing the cost per click of a result for a specific query via auction mechanism Overture*(TM)* was able to produce better rankings for popular commercial queries. The better ranking occurred since presumably only those who were willing to pay for the high cost of the click would bid up the price. However, there were shortcomings that limited the success of this approach. First, vendors or SEO's had a strong tendency to be overly aggressive in forming associations. Essentially, bidding for obscure or low priced queries was an effective way to gain high ranked emphimpressions of their urls and descriptions at very low costs. Other paid inclusion type programs face the same limitations including even degenerate cases like the SEO practice of associating results with viable misspellings of popular terms or queries.[13] In addition, a more subtle point is that the vendors or SEO's are not actually experiencing the *information need* that driving query generation. Endorser's drawn from the set of users could be expected to generate queries that more accurately reflect the set of information needs related to the query.

3.2 Single Endorsement Model

At first, let's focus on the case of endorsing a result that is already in the result set for a given query (though perhaps not in the coveted top ten.) The simplest form of endorsement would be that the result is boosted by one rank. Rank effects on click probabilities have been widely identified – though these effects are query specific we can certainly expect them to be monotonically non-increasing as we move down the ranked list of results.[14] Presumably, even a single rank will

[13]While at Inktomi*(TM)* the author was instrumental in rooting out this practice in the paid inclusion program. One measure revealed that over 15% of the query stream contained misspellings and many paid inclusion submissions capitalized on this by including reasonable looking results to get top rankings while including the correct spelling so that the casual searcher would be fooled. Almost certainly, the misspelling effect caused damage to the perceived relevance of the engine. Previously, the managers of the program had looked on this practice as being a "free service" performed by the SEO's.

[14]In fact, directly using click feedback to rank results usually results in ranking damage to the effects of *information cascades* or *herding* since many users simply click on the top ranked results reflecting the effects of earlier click

slightly increase the chance that a result is seen and possibly further endorsed. In this simple model we reorder the results each time the query is issued by tallying the current endorsements and applying them to the score for each result prior to ranking. Further, we accept only one endorsement per url per user for any given query. If the pool of endorsers was much larger than the number of results endorsed and their input were the sole factor in relevance this mechanism would correctly sort the search results.[15]

3.3 Estimating the Value of an Endorser

In practice we will quickly remove the assumption that the endorsements are the only input to the ranking. Instead, for each result set we can reorder taking into account the value of each endorser's input in some general ranking function. If we relax the assumption that each endorser's input is treated equally then we need a method for valuing the information content of each endorser. Referring back to the UI mechanism described earlier allowing registered users to explicitly assign *trust* or *distrust* ratings to an endorser we have a way to price the endorser's input. Perhaps the simplest mechanism is for any given user to assign 0.0 for untrusted endorsers, a 1.0 for trusted endorsers, and the current price v for unjudged endorsers. The current price v for each endorser is simply the average trust level of that endorser and assumed to be 0.5 for endorsers with no exposure to users.[16] Further, let's assume that for urls with multiple endorsements that the endorser with the highest value is displayed (perhaps along with the total number of endorsers.)[17] This is important because it would quickly highlight endorsers with high value that make bad endorsements and speed up the opportunity for the community to provide corrective feedback.

For endorsers seeking wide influence, status, or reputation, the cost of endorsing a result spuriously or to capture affiliate revenue will quickly become prohibitive. For endorser's only serving a small number of relatively obscure queries scores would be unlikely to deviate much from the uniform or default trust value. One of the most important tasks for endorsers is to introduce content into the index or promote a very low ranking result (e.g. 300th, etc.) For these cases one might consider a separate area in the results akin to Google*(TM) sponsored links* but instead called *endorsed links* and ranked by endorser trust rating. While giving endorsers considerable power to introduce new content endorsing new content quickly exposes their degree of trustworthiness. Best of all, the speed with which an endorser's trust is computed is directly proportional to the popularity of the queries they seek to influence.

feedback. Some of the research on *information cascades* or *herding* with respect to group decision making is summarized in [9] which was published while I was preparing this paper.

[15]The problem is the same as that of ordering a list by the request frequency of each element using the *move one step ahead rule* and time reversible Markov chains [7]. Since this is a transactional system we also assume that the endorsements are maintained in time-stamp order.

[16]With time, we could try to estimate conditionals like $p(x|Trust)$ for the general user, where x is the vector of trust ratings with elements drawn from $\{0, 1\}$. Using the class conditional probabilities, and with the reasonable assumption of *conditional independence*, we could estimate $p(Trust|x)$ using a *Naive Bayes* classifier.

[17]Alternatively, we could display the user-id of the most recent endorser.

3.4 A Market for Endorsers

Perhaps the best approach is to form a market for endorsers. Each registered user of a given endorser's information, distributed over the endorser's query recommendations, is holding partial information about that endorser's value. The presence of independent and distributed observers of the endorser's performance means that a *Distributed Information Market* approach, like the one discussed more generally in [8], might be employed. The results of such a market based approach could be made to converge to the best available estimate of the probability that an endorser is actually trustworthy. In this approach, trusting an endorser is analogous to bidding at the current value of his trust rating. Over time, a user assigning ratings will generate a set of transactions whose impact on his portfolio can be tracked. Thus, trusting an endorser whose value is on the rise or distrusting one whose value declines will add value to the user's portfolio. This mechanism not only provides strong incentive for endorsers to be trustworthy it also provides continuous measurement and assessment of an endorsers trustworthiness.

3.5 Community Search Refinements and Sharing the Wealth

Although a community feedback mechanism may have a dramatic impact on search quality there are also many possible refinements to consider. For example, there might be value in compartmentalizing endorser trust by type and popularity of query. Along a similar line, one might establish limits on the ranking power for endorsers below some high trust level (e.g. for endorsers with trust levels below say 0.9 they are incapable of boosting a urls rank under 10, etc.) More radically, financial incentives like prizes or affiliate revenue sharing could be considered. One can imagine the right financial incentives helping to create an entire ecology of query-url association producers[18] and consumers adding a powerful community assist to the relevance and robustness of search engine results. At some stage perhaps the search engine provides only baseline search and clearing house functions as it matches vendors, skilled revenue sharing endorsers, and search engine consumers, in a powerful market for query-url associations.

4 Conclusions and Further Research

This paper proposes a simple community search mechanism where registered users of a system can endorse urls as results for specific queries. The endorsement process is then subjected to community feedback where the individual and independent decision of community members to trust or distrust an endorser are used to establish the value of that endorser's information. Even a cursory examination of the ongoing research in information retrieval reveals a growing interest in community feedback enhancements for these kinds of systems. And, exploitation of community feedback follows the ongoing research to explore and understand the potential of markets and other group decision mechanisms for performing a broad set of information aggregation tasks. Clearly, this paper only begins the process of trying to understand how community feedback might be used to enhance search engine relevance. Further research is needed to understand what aggregating mechanisms best exploit the information in users' trust ratings, how to best to exploit

[18] Producers could perhaps be required to establish *PayPal(TM)* accounts.

the information once aggregated, and how to create incentives that encourage broad and active use of the endorsement system.

Acknowledgment

I wish to thank Luke Lu from Proofpoint Inc. (formerly of Inktomi Corp.) for many insightful discussions on the topic of personalizing search engine through client side mechanisms. And, of course, a big thanks to the open source software community for producing and maintaining such invaluable tools as R, XEmacs, and Python used in much of the research I did to formulate these ideas over the past several years.

References

[1] Huberman, B. A. *The Laws of the Web: Patterns in the Ecology of Information* The MIT Press, 2001

[2] Huberman, B. A., P. L. T. Pirolli, J. E. Pitkow, and R. M. Lukose Strong Regularities in World Wide Web Surfing *Science* 280 , 5360, 95–97

[3] Nadav Eiron, Kevin S. McCurley Analysis of Anchor Text for Web Search *citeseer.ist.psu.edu/eiron03analysis.html*

[4] Craig Silverstein, Monika Henzinger, Hannes Marais, and Michael Moricz Analysis of a Very Large AltaVista Query Log SRC Technical note #1998-14

[5] Chen, K., Hogg, T., Wozny, N. Experimental Study of Reputation Mechanisms in an Exchange Economy HP Labs *www.hpl.hp.com/research/idl/papers/reputationExpt*

[6] David Andrew Gibson Communities and Reputation on the Web PhD thesis, Computer Science Division, UC Berkeley, Fall 2002

[7] Sheldon M. Ross *Stochastic Processes* 2nd Edition John Wiley & Sons, Inc 1996

[8] Joan Feigenbaum, Lance Fortnow, David M. Pennock, Rahul Sami Computation in a Distributed Information Market Preprint submitted to Elsevier Science

[9] James Surowiecki *The Wisdom of Crowds* Doubleday, June 2004

[10] Hubert L. Dreyfus *On The Internet (Thinking in Action)* Routledge, London 2001