Collaborative Information Seeking with Ant Colony Ranking in Real-Time

Tommaso Turchi[1(✉)], Alessio Malizia[1], Paola Castellucci[2], and Kai Olsen[3]

[1] Department of Computer Science, Brunel University London, Uxbridge, UK
{tommaso.turchi,alessio.malizia}@brunel.ac.uk
[2] Department of Arts and Humanities, Sapienza University of Rome, Roma, Italy
paola.castellucci@uniroma1.it
[3] Molde College, Molde, Norway
kai.a.olsen@himolde.no

Abstract. In this paper we propose a new ranking algorithm based on Swarm Intelligence, more specifically on the Ant Colony Optimization technique, to improve search engines' performances and reduce the information overload by exploiting users' collective behavior. We designed an online evaluation involving end users to test our algorithm in a real-world scenario dealing with informational queries. The development of a fully working prototype – based on the Wikipedia search engine – demonstrated promising preliminary results.

1 Introduction

Going back to the bibliographical databases in the seventies, text retrieval was based on a Boolean search query on the keywords of a document; predetermined attributes such as date and file size were used to rank results based on their relevance. With faster computers and more direct access storage it became possible to search the document's full text: relevance was then determined by the number of occurrences of each search term in the document, for example in relation to document size. The first search engines on the Web used this approach.

However, search term occurrences are not a good indicator for relevance: for example, when trying to find information on "VW Golf" search engines directed the user to online advertisements for used cars as these had the keywords, combined with a high occurrence to length ratio. Google's PageRank algorithm saved the day: now the number of links to a site determined relevance, higher if the sites with the links also had a high ranking. Then the "VW Golf" query would direct the user to an official site for Volkswagen, which most users would find more relevant than the car-for-sale advertisements.

While the PageRank algorithm functions well and offers a notion of "relevance" that is shared by many users it has the disadvantage of being static. It may be enhanced by other data and other techniques, but is in principle based on the current structure of the Web. The algorithm will even freeze this picture, as the high ranking will make the important sites more important. New interesting sites may get a low PageRank value and be presented further down on

D. Calvanese et al. (Eds.): IRCDL 2015, CCIS 612, pp. 104–115, 2016.
DOI: 10.1007/978-3-319-41938-1_12

the search engine results page, and may therefore be noticed only by the most persistent users; thus, the algorithm may be self-fulfilling.

To make Web searching more dynamic we could try to exploit the experience of these "persistent" users: this could be done as easy as presenting an other-users-found-this list. For example, we have been looking for a small lightweight camera with GPS to take on hiking trips; our search terms are "compact camera GPS". After some effort, trying a site here, another there, we have found what we have been looking for. To indicate that this site is interesting we could use a feature recently introduced by Google, the +1[1] service. However, it will also be possible to extract the "interesting site" information indirectly, for example by evaluating the time used on each site, how many links on the site that were explored, if the user printed anything from the site or if the "buy" button was clicked. The next user giving the same or similar search terms could then go directly to this site, following the "other users found this" link.

In this paper we shall explore this concept and present a model that describes a trend towards new Web-searching paradigms, which are both social and dynamic. The idea is taken from biology, from the way ants forage for food: the image in itself refers also to the seminal work of Norbert Wiener, the pioneer both of Cybernetics and of the political approach to technology. Particularly in [31], Wiener uses the figurative speech of ants as a dystopian one. Ants are perceived as a meek colony instead of an unpredictable cluster of *single* ants, each with its own identity. Therefore, the most appreciated qualities of ants turn out as dangerous disadvantages: ants can be easily controlled by a totalitarian ideology just because of their well organized and collaborative way of behaving.

Our approach to Web searching and ranking relies on the positive meaning of the metaphor of the ants. But, at the same time, it is important not to undervalue the possible risks in a collaborative approach. Any single "ant" (i.e. any *user*) has his own unique and value-added perspective and knowledge, and should then be helped to enhance his sense of awareness about his "singularity". Seen under contemporary eyes, "ants" can easily be exploited by a single-minded, market-oriented society, or by a too generalist and massive search engine and ranking algorithm. Instead, it is time to break monopolistic ways to access the Web. The information need of any single user must be taken into consideration, interpreted, and fulfilled. To that aim, it is necessary to make use of different methodologies of analysis. A variety of distinctive research communities can cooperate like "virtual ants": interdisciplinary approaches will surely prove helpful to find new paths.

2 Related Work

2.1 Search Engines

Although more than half (59 %) of Internet users in the US use a Web search engine during a typical day[2], in general the users' degree of satisfaction with

[1] http://www.google.com/+/learnmore/+1/.
[2] http://www.pewinternet.org/2012/03/09/search-engine-use-2012/.

major search engines is - to the best of our knowledge - largely unsettled, and can only be investigated thanks to rather small studies conducted in experimental environments.

According to Silverstein [27] (1) a maximum of two queries is needed to solve users' information needs (67 %) and, usually, (2) users scan only the first page of results (58 %). However, Hawking et al. [16] state that 50 % of proposed search results are irrelevant, thus there are complex informational needs most likely receiving irrelevant results.

Other studies pointed out the low degree of satisfaction with search engine: Fox et al. [13] devised a machine learning approach which employs users' actions (e.g. the time spent on a page, the scrolling usage, the page visits, etc.) concluding that users consider 28 % of search sessions unsatisfactory and 30 % only partially satisfactory. Xu and Mease [34] have measured the average duration of a search session: typically, users end a session – even without satisfying their informational need – after 3 min.

Summarizing, many users employ search engines to satisfy their informational needs and as a starting point of their Web browsing [5]; nevertheless, the search experience is far from being perfect, in fact a substantial amount of searches end up unsatisfied.

In this paper, we deal with the problem of improving search engines' performance by exploiting the actions performed by the users; the problem that we try to address is the information overload, i.e. the inability to take a decision due to the huge quantity of information obtained by the users. As a matter of fact, search engines are tools designed to help people solving their own informational needs and – as we have discussed before – there is much room for improvements.

Persistent users must be taken in due consideration too. On the average, they are also highly specialized users, often with highly specific information needs [12]. The old and dismissed library catalog, the traditional Online Public Access Catalog (OPAC), or even the generalist search engine are any use for them [8,35]. As it has been appropriately underlined by the LIS research community, a Next Generation Catalog (NGC) is tremendously needed, and can cooperate with web-scale discovery service [17].

2.2 Collaborative Filtering

Large scale searching on the Web can be applied only after a careful qualitative and quantitative analysis of user's satisfaction [6,22]. Here the problem of information overload if faced from a personalization perspective, without exploiting users' collective behavior: they focus on personalizing search engines' results rather than improving their performances. LIS research community shares the same critical approach [29]. Therefore, it must be held in due consideration, looking for a synergistic cooperation [24].

The first system employing a real Collaborative Filtering (CR) was the Tapestry mail system at Xerox PARC, described by Goldberg et al. [15] as "people collaborate to help one another perform filtering by recording their reactions to documents they read", which is thus considered orthogonal to content based filtering used by all the previously summarized systems. Following this study, several systems applied CR to face information overload, like GroupLens [23] or Ringo [26], always asking users' feedback on suggested resources.

Rucker and Polanco [25] devised a different and simpler approach in their Siteseer system: it collects users' bookmarks and use them both to find similar users and recommend bookmarks to other users that are unaware of them. In a similar way, the PHOAKX system [28] inspects Usenet groups in order to find posts containing URLs, ranking them according to the number of posts.

2.3 Information Foraging on the Web

One of the theories trying to explain users' behavior while searching for information in complex systems (e.g. the Web) is the Information Foraging [21]; it is inspired by the optimal foraging theory, and one of its key concepts is the so-called "information scent", i.e. the perception the user has of the cost, value and ease to access of a resource, given some available clues (e.g. link, snippet, tag, comments, etc.). Applying the information foraging theory to Web information seeking seems quite natural and many approaches have been developed to improve this process, which can be traced back to this theory [1,7].

Nevertheless, many studies have chosen to extend the information provided by hyperlinks suggesting to users the most promising paths to follow in order to achieve their goals, usually working within a single website [2,3]. ScentTrails [20] continuously allows users to supply keywords and enriches hyperlinks providing a path that achieves the goal described by them. Finally, Wu and Aberer's method [32] operates within a single website, enriching the information provided by hyperlinks with a technique inspired by the ant foraging behavior (i.e. heavily clicked links are recommended in favor of less visited links).

2.4 Other Approaches

There are many other ranking mechanisms exploiting users' behavior, but with a different goal than personalizing the searching/browsing experience. A significant number of contributions comes from a "critical" analysis of PageRank. This "post-monopolistic" approach is currently pushing users beyond Google's way of evaluation, and towards new kind of "metrics" [9].

Probably the first search engine to take into account users' behavior in its ranking computation was the no-longer available DirectHit, devised by Gary Culliss in 1995 and later bought by Ask Jeeves, which combined it with the Teoma search engine; it employed a ranking formula composed by three main factors: (1) content based ranking, (2) Link Analysis Ranking (LAR), and (3) usage based ranking; the latter takes into account all those clicks issued by users on a result in relation to a specific query, besides their own time of access and the time spent on the page.

Baeza-Yates et al. [4] devised a ranking algorithm in which the relevance of each document is boosted in relation to previous users' preferences; their method includes a preliminary phase of clustering, when similar queries are grouped. Then, the URLs are extracted and ordered by the number of clicks.

Summarizing, it's important to notice how none of the techniques just mentioned can be used to adapt a search engine to users in real time; in fact, all of them need to be periodically retrained to adapt the search engine's responses according to the last recorded usage behavior.

3 Ant Colony Ranking

We summarized some techniques to improve search engines' performance, highlighting a few key concepts: (1) the relationship between users seeking information and the optimal foraging theory; (2) the need of a search engine to adapt itself to users' behavior; and (3) the need to perform such adaptation in real time. Almost none of the aforementioned approaches take all those three aspects into account – especially the latter – and those that do might also benefit from a deeper implementation of some of the Information Foraging's key concepts [10].

As stated by Wu and Aberer [32] and by Olston and Chi [20], a swarm-based approach – thus one employing some Swarm Intelligence (SI) ideas – can beyond a doubt take into account all those three key factors, being nonetheless a much more elegant and simple method than all the others "ad-hoc" ones. Swarm Intelligence (SI) refers to the emergence of "intelligent" behaviors from a group of simple and/or loosely organized agents. Ants are a typical example of SI and their use of stigmergic processes[3] inspired the famous family of Ant Colony Optimization (ACO)[4] algorithms. Thus, we'll now introduce a simple ranking algorithm based on SI that can be used to improve search engine's performance, adapting themselves to users' behavior.

Each day ants leave the colony in search of food and building materials; they will exploit the surroundings in all directions in a somewhat random fashion. If an ant finds anything of interest, it will return to the colony depositing pheromone, a chemical substance that ants are able to detect. Thus they create trails to signal the path between the colony and the food. The quantity of pheromone deposited, which may depend on the quantity and quality of the food, will guide other ants to the food source. That is, other ants in the colony may now use the pheromone as trail markers to reach the food. These markers evaporate over time, so that uninteresting trails disappear. Shorter trails will get a higher level of pheromone, thus shorter trails will endure longer, inserting a notion of optimization. This form of organization may be used to characterize social behavior on the Web.

As we have previously stated, we will adapt the strategies employed in food searching by ant colonies in the building of ranking algorithms employing users' behavior: it's pretty intuitive to find a parallelism between the way ants forage for food and the way users employ search engines to satisfy their informational

[3] Introduced in 1950s by P. Grasse during his research on termites, it denotes a method of communication whereby individuals modify their surrounding environment.

[4] The ACO is a bio-inspired (ant colony) probabilistic meta-heuristic for solving computational problems related to searching for an optimal path in a graph.

needs [18]; yet the latter don't leave any trace, so they can't provide any clues to users with their same informational needs, and – as about 30–40 % of queries issued have already been submitted [33] – that's a pretty common scenario.

We propose a simple algorithm that implements the model we just proposed; we called it NaïveRank. We will assume that interactions between users and search engine are available in the form of query-sessions (briefly sessions) that are formed by the query and by the different (possibly none) documents selected by the user among the results related to it; so, a session $s \in S$, where S is the set of all the available sessions, is defined by $s = (q, c)$, where $q \in Q$ is the query and $c = (d_1, \ldots, d_{|c|}) \in D_q^{|c|}$ is the ordered sequence of $|c|$ results selected by the user. Moreover, $D_q \subseteq D$ is the set of relevant documents (selected by the search engine) in relation to the query q, Q being the set of all the known queries and D the one of all the available documents.

Given a query $q \in Q$ and a document $d \in D_q$, the pheromone $w \in \mathbb{R}^+$ associated to the couple (q, d) is denoted by $\phi_W(q, d)$, where the function ϕ is defined by

$$\phi : Q \to D \to \mathbb{R}^+ \times T,$$

while the last time the document was clicked among the results of the query is denoted by $\phi_T(q, d) \in T$.

Every time a result $d \in D_q$ is picked among the results of query $q \in Q$, or – carrying on the similarity with ACO – the path $q \to d$ is covered by one user, a certain quantity of pheromone is deposited on it. The straightforward implementation of ACO's principles, also described by Gayo-Avello and Brenes in their paper [14], is to employ the simplest incrementing function ever, namely the successor. Thus, the upgrade is issued applying the rule

$$\phi_W(q, d) = \phi_W(q, d) + 1.$$

Evaporation is obtained by an exponential decay, using the rule

$$\phi_W(q, d) = \phi_W(q, d) \left(\frac{1}{2}\right)^{\frac{t - \phi_T(q,d)}{\delta}},$$

where $t \in T$ is the current time-stamp and $\delta \in \mathbb{R}^+$ is the time required for the pheromone to half its value.

Pheromone evaporation will be performed periodically for each pair query-document, always before the upgrade; the frequency of upgrading is related to the users' perceived relevancy of considered documents: evaporation is a useful mechanism to forget registered behaviors, thus issuing it frequently causes the increase of new registered behaviors' importance.

Finally, we consider the quantity of pheromone deposited on each pair query-document and rank results based on it; thus

$$R_{\phi(q)} = \{(d_i, d_j) : d_i, d_j \in D_q \wedge \phi_W(q, d_i) \geqslant \phi_W(q, d_j)\}$$

defines the actual documents' ranking: for all known query $q \in Q$, the results' rank will be the one given by the chain $(D_q, R_{\phi(q)})$.

Summarizing, the ranking algorithm based on ACO proposed in this section employs pheromone's traces for each pair query-document; the pheromone increases every time a user selects a page among the results related to a query and, also, vaporizes itself in time, being a simple mechanism to take into account the gradual loss of interest by users. By doing this, once a user performs a known query the search engine is able to present a new ranking based on the behavior shown by users with the same informational need – i.e. users who previously issued the same (or a *similar*) query – by exploiting the pheromone's traces.

Thus, it's important to establish whether it's really possible to improve search engines' performance by employing this new approach; therefore, we devised an experiment involving online participants. In the following section we will present details about the setup and results.

4 Evaluation

We devised an experiment to test the hypothesis that a search engine employing our new approach based on ACO provides a new ranking based on users' behaviors, and that this new ranking somehow improves the users' degree of satisfaction in performing a search.

In fact, given that a search can be considered satisfactory if it's successful, i.e. if the user can easily (and quickly) identify the content he was looking for, we devised the experiment involving 8 participants (for the most part students of the faculty), asking them to find some contents and describing them only the informational need they have to solve (trying, as much as possible, not to give clues about how to formulate the query).

Although we are aware of the reduced size of the sample, its characteristics result compliant with demographics of search engine users, whom are more likely to employ our system in future. However, as a preliminary evaluation 8 users are enough, in fact, according to Nielsen and Landauer [19] conducting a usability testing with a single user a third of the usability problems will be discovered; with five or more users a little more can be gained. Nevertheless, if the goal is to run a controlled experiment, from which a statistical analysis has to be performed, at least twice this number is necessary [11]. Evaluations of user's satisfaction carried out by other disciplinary communities can also offer a proper methodology [30].

For sake of simplicity, we chose an existing search engine, adjusting the default ranking and using the proposed algorithm to compute the new one; the choice fell on the MediaWiki[5] search engine – made available by Wikimedia Foundation to search contents among the famous online encyclopedia Wikipedia. Consequently, we devised a search engine that works over the Wikipedia's contents and uses MediaWiki to fetch them; besides, it records the users' behavior and exploits it to improve the provided results' ranking by using our algorithm.

Moreover, given the small number of participants available, we chose to ignore the evaporation mechanism, since there's no need to adapt to the shifting of users' interests, which we reasonably assumed static.

[5] http://www.mediawiki.org/wiki/MediaWiki.

We proposed six informational needs to be solved (summarized in Table 1) to the 8 participants, providing them with the devised search engine that, even if it doesn't display the same Wikipedia graphic, offers the same information about the proposed results (the page title and a brief snippet); besides, the users were only allowed to use the provided search engine, without any time limitation or being in any way controlled by the examiners.

Table 1. The six tasks submitted to the participants of the experiment; the title of the page satisfying the task and the brief description of the informational need given to the participants are reported.

Japan	Colostrum	John Von Neumann
The country with the highest life expectancy	The first milk a mother produces after giving birth	The first computer virus's theorist.
Californium	Saturn	Beagle
The chemical element which takes its name from one of the United States	The last planet of the Solar System which can be seen by naked eye	The breed of dog which shares its name with the ship on which Darwin did his explorations

The search engine, in addition to recording the users' behavior, shows the content of the fetched pages next to the proposed results; this way one can also record the actions carried out inside each page, allowing us to build the entire click graph. These information were included in the ranking computation too, i.e. considering not only the unit length's paths, but also the longer ones; thus we can operate also among those pages considered less relevant by MediaWiki, but considered relevant by the users, even if they reached them with a higher number of clicks than the ones displayed in the results.

4.1 Results

If the proposed approach somehow improves over the default ranking we should witness some changes in a way that could better satisfy new users facing the same informational needs than their predecessors; moreover – given that our approach closely follows the ACO approach's suggestions - in order for the system to exploit the collective behavior of its users, the queries submitted to the search engine shouldn't be too heterogeneous, thus – even with such a limited number of users' interactions – our algorithm could still be effective. Naturally, given the uncontrolled and limited nature of the experiment, the results will be taken as preliminary and only have informative value, but we can still rely on them to better understand some underlying dynamics driving the users during a search, and to analyze how the algorithm performs in a real confined environment.

Once the experiment ended, we gathered the query-click logs and analyzed them to extract the evolution of the pheromones deposited on the key pages solving each task; the chart depicted in Fig. 1 summarizes the results of our experiment. The figure shows a timeline of the interactions that occurred between participants and our system on the x-axis. On the y-axis all the different queries issued to the system are shown, grouped by the corresponding task. Every interaction corresponds to one point on the graph, whose size represents the quantity of pheromone deposited on the document that solved that particular task, related to the corresponding query issued by the user; the shape represents the ranking status in which the interaction occurred. The circle indicates that the page was not found in the first result for the query issued by the user; thus, it was a disadvantageous situation for users seeking the correct page among the other results. The diamond indicates that the page appeared first among the results, being an advantageous situation for users who could immediately spot the answer to their informational need among the list of results.

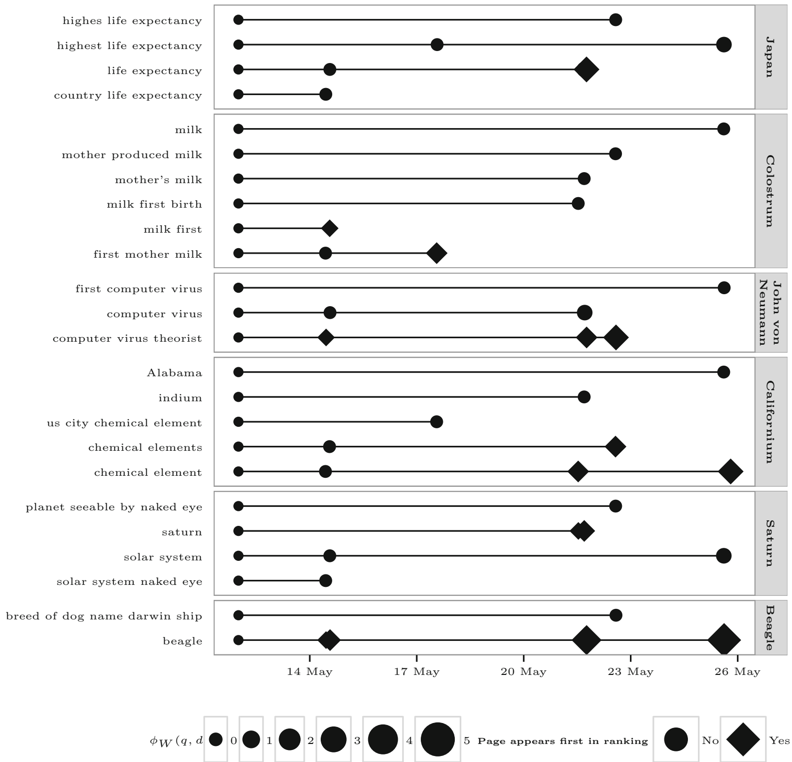

Fig. 1. Results obtained from the experiment.

In general the tasks were solved by the majority of participants, since all of them were solved by at least 6 out of 8 participants, with the "Japan" task solved by 7 and the "Californium" one by everyone; this variation may be caused by several factors, e.g. difficulties encountered by users in performing effective queries to solve specific tasks, such as the "Beagle" one, which was apparently solved indirectly by 5 out of 6 participants, who then searched for the right page.

As we expected, queries performed by a few users – even though they managed to solve the task – don't really affect the ranking in favor of the goal page; indeed users solving tasks through non-specific queries need a higher number of clicks, given that the results aren't enough specific and thus a more accurate inspection is needed. This way, some pheromone has been deposited even onto documents not solving the particular user's informational need, thus voiding the effect of finally reaching the right page among the results.

Furthermore, another beneficial effect noticed through the experiment to be pointed out is the stability yielded by the optimal ranking, since the goal page keeps appearing in the first position of the ranking once earned; this is caused by the algorithm's underlying approach employing user's behaviors, but could also reveal some degree of self-reinforcement in the ranking algorithm: in fact, when a page reaches the first position among the results, it will be prone to be selected by the majority of users, causing the progressive increase of its pheromone and a convergence to a sub-optimal solution. Although in our case the domain makes this effect negligible (it's unlikely that the computed ranking will stop being the optimal one, since an encyclopedia is a rather static collection), in reality here is when the evaporation mechanism takes place, making the ranking more flexible thus preventing the system's convergence to a non-optimal solution.

In conclusion, from our results we could argue that our algorithm could offer a considerable improvement to the online search experience, in the case of informational queries; thus, it could be helpful to consider further relevance measures in our ranking computation, such as Link Analysis Ranking-based ones.

5 Conclusion

Recently, Google introduced their Social Search service declaring: "with these changes, we want to help you finding the most relevant information from the people who matter to you". That is, in a way, our definition of a colony. The mechanism is the Google+1 button, which let users share interesting pages with their contacts - a way of releasing pheromone. This case shows the paradigm shift in Web searching that we are experiencing today, and we hope that our approach can be a first step in modeling and describing this trend, considering even implicit ways of releasing the pheromone.

We designed an algorithm employing an ACO strategy to provide implicit collaborative seeking features in real-time to search engines. It seems particularly adequate for informational queries for retrieving results about relatively static information on the Web, such as looking for products in a catalog or encyclopedia entries. We evaluated the algorithm with a preliminary online experiment with 8 participants employing the MediaWiki search engine augmented with our NaïveRankygorithm. It proved to be effective for the sample set of participants and was relevant to get a first insight about our approach with real users.

In future we plan to extend the online experiment to a more extended sample of participants and test slightly different ACO-based algorithms, in order to validate our approach against different situations (other kinds of queries and search sessions); we are also devising an offline experiment to test our algorithms with publicly available datasets of query-click logs, released by some real-world search engines. We hope to prove that also in an online environment real-time relevant results can be obtained by users employing an implicit collaborative approach for information seeking and selecting the right algorithm depending by the type of query.

References

1. Aggarwal, C.C.: Collaborative crawling: mining user experiences for topical resource discovery. In: IBM Research Report, pp. 423–428. ACM (2002)
2. Ali, K., Ketchpel, S.P.: Golden path analyzer: using divide-and-conquer to cluster web clickstreams. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 349–358. ACM, New York (2003)
3. Armstrong, R., Freitag, D., Joachims, T., Mitchell, T.: WebWatcher: a learning apprentice for the world wide web. In: AAAI Spring Symposium on Information Gathering, pp. 6–12 (1995)
4. Baeza-Yates, R., Hurtado, C.A., Mendoza, M.: Query clustering for boosting web page ranking. In: Favela, J., Menasalvas, E., Chávez, E. (eds.) AWIC 2004. LNCS (LNAI), vol. 3034, pp. 164–175. Springer, Heidelberg (2004)
5. Broder, A.: A taxonomy of web search. ACM Sigir Forum **36**(2), 3–10 (2002)
6. Calhoun, K.: The changing nature of the catalog and its integration with other discovery tools (2006)
7. De Roure, D.C., Hall, W., Reich, S., Hill, G.L., Pikrakis, A., Stairmand, M.A.: MEMOIR - an open framework for enhanced navigation of distributed information. Inf. Proces. Manage. **37**, 53–74 (2001)
8. Dempsey, L.: Thirteen ways of looking at libraries, discovery and the catalogue: scale, workflow, attention (2013)
9. Devine, J., Egger-Sider, F.: Going beyond Google again (2014)
10. Ding, C., Chi, C.H.: Towards an adaptive and task-specific ranking mechanism in Web searching. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 375–376. ACM, New York (2000)
11. Dix, A.: Human-computer interaction: a stable discipline, a nascent science, and the growth of the long tail. Interact. Comput. **22**(1), 13–27 (2010)
12. Exner, N.: Research information literacy: addressing original researchers' needs. J. Acad. Libr. **40**(5), 460–466 (2014)
13. Fox, S., Karnawat, K., Mydland, M., Dumais, S., White, T.: Evaluating implicit measures to improve web search. ACM Trans. Inf. Syst. **23**(2), 147–168 (2005)
14. Gayo-Avello, D., Brenes, D.J.: Making the road by searching - a search engine based on swarm information foraging, November 2009. arXiv.org
15. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Commun. ACM **35**(12), 61–70 (1992)
16. Hawking, D., Craswell, N., Bailey, P., Griffihs, K.: Measuring search engine quality. Inform. Retrieval **4**(1), 33–59 (2001)
17. Hull, D., Pettifer, S.R., Kell, D.B.: Defrosting the digital library: bibliographic tools for the next generation web. PLoS Comput. Biol. **4**(10), e1000204 (2008)

18. Malizia, A., Olsen, K.: Toward a new search paradigm-can we learn from ants? Computer **45**(5), 89–91 (2012)
19. Nielsen, J., Landauer, T.K.: A mathematical model of the finding of usability problems. In: CHI 1993: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, pp. 206–213. ACM Request Permissions, New York, May 1993
20. Olston, C., Chi, E.H.: ScentTrails: integrating browsing and searching on the Web. Trans. Comput. Hum. Interact. (TOCHI) **10**(3), 177–197 (2003)
21. Pirolli, P., Card, S.: Information foraging in information access environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 51–58. ACM Press/Addison-Wesley Publishing Co., New York (1995)
22. Quint, B.: Attacking our problems. Inf. Today **31**(2), 8 (2014)
23. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, pp. 175–186. ACM, New York (1994)
24. Richardson, H.: Revelations from the literature: how web-scale discovery has already changed us. Computers in Libraries (2013)
25. Rucker, J., Polanco, M.J.: Siteseer: personalized navigation for the Web. Commun. ACM **40**(3), 73–76 (1997)
26. Shardanand, U., Maes, P.: Social information filtering: algorithms for automating "Word of Mouth". In: Proceedings of ACM CHI 1995 Conference on Human Factors in Computing Systems, pp. 210–217 (1995)
27. Silverstein, C., Marais, H., Henzinger, M., Moricz, M.: Analysis of a very large web search engine query log. SIGIR Forum **33**(1), 6–12 (1999)
28. Terveen, L., Hill, W., Amento, B., McDonald, D., Creter, J.: PHOAKS: a system for sharing recommendations. Commun. ACM **40**(3), 59–62 (1997)
29. Thomsett-Scott, B., Reese, P.E.: Academic libraries and discovery tools: a survey of the literature. Coll. Undergraduate Libr. **19**(2–4), 123–143 (2012). dx.doi.org
30. Vaughan, J.: Web Scale Discovery Services (2011)
31. Wiener, N.: The Human Use of Human Beings: Cybernetics and Society. A Da Capo paperback (Da Capo Press), New York (1954)
32. Wu, J., Aberer, K.: Swarm intelligent surfing in the Web. In: Cueva Lovelle, J.M., Rodríguez, B.M.G., Gayo, J.E.L., Ruiz, M.P.P., Aguilar, L.J. (eds.) ICWE 2003. LNCS, vol. 2722. Springer, Heidelberg (2003)
33. Xie, Y., O'Hallaron, D.: Locality in search engine queries and its implications for caching. In: INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, pp. 1238–1247. IEEE (2002)
34. Xu, Y., Mease, D.: Evaluating web search using task completion time. In: SIGIR 2009: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 676–677. ACM, New York, July 2009
35. Young, M., Yu, H.: The impact of web search engines on subject searching in OPAC. Inf. Technol. Libr. **23**(4), 168–180 (2004)