

Democratisation of Digital Search by Decentralisation

Judith Winter

University of Applied Sciences
Frankfurt

winter@fb3.fh-frankfurt.de

EXTENDED ABSTRACT

Today's knowledge society is based on using individual and especially collective information. As the amount of digital knowledge available on the web is still growing exponentially, it is essential to access and process huge collections of complex data [4]. In order to provide relevant information to the user, modern web search engines successfully apply information retrieval techniques [1]. The usefulness and necessity of web search engines is essentially beyond dispute, more than 90% of all users applied them regularly in 2010.

However, with their spreading use, web search engines have met with growing skepticism and criticism not regarding technical, but social issues [11]. In particular the market leader Google hits the headlines regularly. Not only data protectionists and privacy advocates criticize the big amount of user information collected and exploited by Google, but also politicians like the German minister of consumer protection see a need for action. Furthermore, the possibility to censor information is a controversial topic. Whereas Google recently shut down their activities in China due to excessively strict censorship regulations, other companies such as Bing or Yahoo have decided to continue subordinating themselves under these policies [3].

It is a *centralized* architecture that enables web search engines to exploit private information about their searcher or to censor search results. Collection and search information is stored in a central index, making manipulation of that information easy [8]. The same applies to the processing of user requests that are hence easily logged and exploited [2]. Also, centralized systems can be attacked or forced to be shut down, e.g. by authorities in countries with little respect for freedom of opinion [3].

This paper thus raises awareness of the social issues of centralized search engines.

It furthermore acknowledges the growing amount of users who reject centralized search engines, but who are looking for alternatives. Their reasons range from institutions and authors looking to keep full control and autonomy in terms of access to their provided documents, to users who appreciate privacy, or users suspicious of systems susceptible to manipulation and censorship. Also, the quasi-monopoly of commercial web search engines is of concern to people, especially when those might be biased to paying advertisers [7].

Third, the paper proposes the use of decentralized web search architectures that are based on networks of autonomous and equal computers, so called peers. Search engines based on such Peer-to-Peer (P2P) systems are less susceptible to abuse, as all

information (documents, index, statistics) is distributed among a huge amount of participating peers and thus cannot be centrally evaluated or manipulated [9]. Due to their self-organization methods, such distributed search engines can process search requests without any central control, without any central servers and without any central indexes. P2P-based search engines are hence subject to current research as they are promising architectures that can develop into real alternatives to classical client/server-based search engines for the growing amount of privacy-aware users [6].

Even where no concerns are raised against centralized control, there are reasons to use P2P-applications. For example, loosely coupled independent users or organizations that plan to exchange information may have neither the personnel, nor the technical and financial means to build up a centralized search systems. P2P-networks can be a solution here by using the „edge of the internet“ – the unused potential of individual computers that can be pooled together into a self-organizing system. The load of answering search request, as well as the storage consumption or the document collection itself, can be distributed among the participating peers. Thus, new web applications can be realized with low starting costs and hence serve the general public or disadvantaged groups. P2P-systems can empower financially weak developing countries to realize digital projects, even if there are no financial means to build a centralized system. Also, non-profit organizations can take advantage of low cost P2P-applications in order to serving the public good.

P2P systems are no cure-all for problems arising with centralized search systems, with a lot of research still ongoing that regards socio-technical issues such as trust in the peers as well as security issues regarding the distributed data. However, with all the described benefits, web search engines based on P2P-architectures provide possible solutions for two demands of the 21. Century. First, they are a means to access the basic resource of our society and pool together collective but often fragmented knowledge. At the same time, they achieve this aim in a democratic manner. P2P-systems allow for collaborative use of private computing resources. In open P2P-systems, individual users provide their resources to the public for common use. Thus, they contribute to the networking and the cohesion of society. Even P2P-systems used by closed groups on the web are a contribution to democracy. As they are powerful and can be realized with low starting costs, they can empower disadvantaged groups.

The last part of this paper presents a fully implemented P2P-based search system that has been developed in order to promote research on decentralized search [12]. This system can process

queries for structured information with decentralized techniques. It dispenses of any use of central control or central information storage but distributes queries, documents as well as index information over participating nodes, using a distributed hash table. We especially focus on searching annotated information in the form of XML-documents, as XML has developed to be the standard for exchanging structured information on the web [5]. Experiments have shown that our system can provide search results not only effectively but also efficiently. It is thus a contribution to the promotion of P2P-search engines as democratic, privacy-aware means to provide and access knowledge in our society, means that are far less susceptible to manipulation and censorship than centralized systems.

REFERENCES

- [1] Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier: *Modern Information Retrieval*. ACM Press Series / Addison Wesley, New York, USA, 2010.
- [2] Barbaro, Michael; Zeller, Tom Jr.: *A Face Is Exposed for AOL Searcher No. 4417749*. In: The New York Times, 9. August 2006.
- [3] Briegleb, Volker: *Google wartet auf China-Lizenz*. heise online 01.07.2010.
- [4] Chute, Christopher; Manfrediz, Alex; Minton, Stephen et al.: *The Diverse and Exploding Digital Universe*. IDC White Paper, 2008.
- [5] Harold, Elliotte R.: *The future of XML - How will you use XML in years to come?* IBM Technical Library, 2008.
- [6] Risson, J.; Moors, T.: *Survey of research towards robust peer-to-peer networks – search methods*. Technical Report UNSW-EE-P2P-1-1, University of New South Wales, Australia, 2004.
- [7] RP ONLINE GmbH, Düsseldorf: *Verdacht der Manipulation von Suchergebnissen - EU eröffnet Kartellverfahren gegen Google*. http://www.rp-online.de/digitale/internet/EU-eroeffnet-Kartellverfahren-gegen-Google_aid_936617.html, Artikel vom 20.11.2010.
- [8] Sander-Beuerman, Wolfgang: *Schürfrechte im Informationszeitalter*. Magazin für Innovation Technology (MIT), Feb 2005.
- [9] Steinmetz, R.; Wehrle, K. (eds.): *Peer-to-Peer Systems and Applications*. LNCS No. 3485, Springer-Verlag, Berlin-Heidelberg, 2005.
- [10] Steinmetz, R.; Wehrle, K. (eds.): *Peer-to-Peer Systems and Applications*. LNCS No. 3485, Springer-Verlag, Berlin-Heidelberg, 2005.
- [11] SuMa-eV: Gemeinnütziger Verein zur Förderung der Suchmaschinen-Technologie und des freien Wissenszugangs e.V. (SuMa-eV). Hannover, Deutschland, <http://www.suma-ev.de>, 25.03.2011.
- [12] Winter, Judith; Kühne, Gerold: *Achieving High Precision With Peer-to-Peer Is Possible!* In: Focused Retrieval. Lecture Notes in Computer Science (LNCS), Springer-Verlag, Berlin-Heidelberg, 2010.