IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

The Global Techniques for Search Engine Optimization and Marketing

SHARAD CHOUDHARY
ASSISTANT PROFESSOR IN MANAGEMENT DEPARTMENT
VIVEK COLLEGE OF MANAGEMENT & TECHNOLOGY
BIJNOR

ABSTRACT

This paper defines what a **search engine** is and explains how various **search engines** work. There is a **search engine** glossary which provides an explanation of the various **search** methods and an evaluation of some of the bigger **search engines** on the Internet. With this paper we can analyzes the transition of B2C versus B2B to the internet. It provides recommendations for the brand manager to establish a brand presence online, including: **search engine optimization**, paid **search**, emails, rich media, sponsorship and banner advertisements. Each of these items is then digressed upon to show how a brand manager would measure the results and associated success of the online branding tools. Paper also describes the process by which Internet **search engines**, such as "Google" and "Yahoo", search and retrieves information requested by users. The paper explains that the process works on the basis of created indexes and then explain some elements of the indexing system, which are known as a document processor, query processor, **search** and matching function, and ranking ability.

This paper will examine the **Search Engine** and how it has significantly affected previous human activities, pointing out in addition some of the advantages and disadvantages of the **Search Engine**.

Search engines have forever changed the way and provides the way to people access and discover knowledge, allowing information about almost any subject to be quickly and easily retrieved within seconds. As increasingly more material becomes available electronically the influence of search engines on our lives will continue to grow. This presents the problem of how to find what information is contained in each search engine, and how Web is applicable for social navigation. What bias a search engine may have, and how to select the best search engine for a particular information need. This research introduces a new method and algorithm for social application. We further present features and algorithms that facilitate online communication and collaboration towards common searching targets. The utility of our system is established by experimental studies. The extensions we present can be easily adopted in a typical web browser.

Keywords
Speculative search engines, Google Scholar, ranking algorithm, search engine optimization, SEO.

1. INTRODUCTION

Researchers should have an interest in ensuring that their paper are indexed by speculative search engines such as Google Scholar, IEEE Xplore, PubMed, and SciPlore.org, which greatly improves their ability to make their articles available to the academic community. Not only should authors take an interest in seeing *that* their articles are indexed, they also should be interesting in *where* the articles are displayed in the results list. Like any other type of ranked search results, articles displayed in top positions are more likely to be read.

This paper presents the concept of *speculative search engine optimization* (2SEO) to optimize scholarly literature for academic search engines. The first part of the article covers related work that has been done mostly in the field of general search engine optimization for Web pages. The second part defines SEO and compares it to search engine optimization for Web pages. The third part provides an overview of ranking algorithms of speculative search engines in general, followed by an overview of Google Scholar's ranking algorithm. Finally, guidelines are provided on how authors can optimize their articles for speculative search engines. This article does not cover how publishers or providers of academic repositories can optimize their Web sites and repositories for speculative search engines. The guidelines are based on three studies we have recently conducted [1] and on our experience in developing the academic search engine SciPlore.org.

2. GLOBAL SPECULATIVE SEO PROCESS

Companies have many options for communicating with prospective customers. The Internet has become the primary medium for reaching potential clients anywhere across the globe. With advent of search and popular brands like Google and Yahoo, the Internet is an excellent source to obtain new clients and build your bottom line as well as build your brand beyond your wildest dreams. Let us begin by outlining the primary components of Online Marketing:

3. 2.1. SEO / SEARCH ENGINE OPTIMIZATION:

SEO is short for search engine optimization, the process of increasing the amount of visitors to a Web site by ranking high in the search results of a search engine. The higher a Web site ranks in the results of a search, the greater the chance that that site will be visited by a user. It is common practice for Internet users to not click through pages and pages of search results, so where a site ranks in a search is essential for directing more traffic toward the site. SEO helps to ensure that a site is accessible to a search engine and improves the chances that the site will be found by the search engine.

4. 2.2. PPC / PAY PER CLICK:

Pay per click (PPC) is an advertising technique used on websites, advertising networks, and search engines. Advertisers bid on "keywords" that they believe their target market (people they think would be interested in their offer) would type in the search bar when they are looking for their type of product or service. For example, if an advertiser sells red widgets, he/she would bid on the keyword "red widgets", hoping a user would type those words in the search bar, see their ad, click on it and buy. These ads are called "sponsored links" or "sponsored ads" and appear next to and sometimes above the natural or organic results on the page. The advertiser pays only when the user clicks on the ad.

While many companies exist in this space, Google Adwords and yahoo search marketing, which was formerly Overture, are the largest network operators as of 2006. In the spring of 2006, MSN started beta testing their own PPC service, MSN. Depending on the search engine, minimum prices per click start at US$0.01 (up to US$0.50). Very popular search terms can cost much more on popular engines. Abuse of the pay per click model can result in click fraud.

5. 2.3. PAY PER INCLUSION

Paid inclusion is a search engine marketing product where the search engine company charges fees related to inclusion of websites in their search index. Paid inclusion products are provided by most search engine companies, the most notable exception being Google.

The fee structure is both a filter against superfluous submissions and a revenue generator. Typically, the fee covers an annual subscription for one webpage, which will automatically be cataloged on a *regular* basis. A per-click fee may also apply. Each search engine is different. Some sites allow only paid inclusion, although these have had little success. More frequently, many search engines, like Yahoo!, mix paid inclusion (per-page and per-click fee) with results from web crawling. Others like Google (and a little recently, Ask.com), do not let webmasters pay to be in their search engine listing (advertisements are shown separately and labeled as such). Some detractors of paid inclusion allege that it causes searches to return results based more on the economic standing of the interests of a web site, and less on the relevancy of that site to end-users. Often the line between pay per click advertising and paid inclusion is debatable. Some have lobbied for any paid listings to be labeled as an advertisement, while defenders insist they are not actually ads since the webmasters do not control the content of the listing, its ranking, or even whether it is shown to any users. Another advantage of paid inclusion is that it allows site owners to specify particular schedules for crawling pages. In the general case, one has no control as to when their page will be crawled or added to a search engine index. Paid inclusion proves to be particularly useful for cases where pages are dynamically generated and frequently modified. Paid inclusion is a search engine marketing method in itself, but also a tool of search engine optimization, since experts and firms can test out different approaches to improving ranking, and see the results often within a couple of days, instead of waiting weeks or months. Knowledge gained this way can be used to optimize other web pages, without paying the search engine company.

2.4. BANNER ADVERTISING PROGRAMS

Cost Per Impression (CPI) / Cost Per Thousand (CPM), Cost per Action (CPA) / Cost Per Click (CPC)

The web banner or banner ad form of online advertising entails embedding an advertisement into a web page. It is intended to attract traffic to a website by linking them to the web site of the advertiser. Banner Ad programs can be structured into any of the following types of programs. Cost Per Impression is a phrase often used in online advertising and marketing related to web traffic. It is used for measuring the worth and cost of a specific e-marketing campaign. This technique is applied with web banners, text links, e-mail spam, and opt-in e-mail advertising, although opt-in e-mail advertising is more commonly charged on a Cost Per Action (CPA) basis. The Cost Per Impression is often measured using the CPM (Cost Per Thousand; "M" is the Roman numeral for 1,000) metric. (A CPM is the cost of one thousand (1,000) impressions.) CPM is considered the optimal form of selling online advertising from the

publisher's point of view. A publisher gets paid for each ad that is shown. This type of advertising arrangement closely resembles Television and Print Advertising Methods for speculating the cost of an Advertisement. Often, industry agreed approximates are used. With Television the Nielsen Ratings are used and Print is based on the circulation a publication has. For Online Advertising, the numbers of views can be a lot more precise. When a user requests a Web Page, the originating server creates a log entry. Also, a third party tracker can be placed in the web page to verify how many accesses that page had. CPM and/or Flat rate advertising deals are preferred by the Publisher/Webmaster because they will get paid regardless of any action taken. For Advertisers a Performance Based system is preferred. There are two methods for Paying for Performance: 1) CPA - Cost per Action/Acquisition and 2) CPC - Cost per Click Through.

2.5. AFFILIATE MARKETING

Affiliate marketing is a method of promoting web businesses in which an affiliate is rewarded for every visitor, subscriber, customer, and/or sale provided through his/her efforts. It is a modern variation of the practice of paying a finder's fee for the introduction of new clients to a business. Compensation may be made based on a certain value for each visit (Pay per click), registrant (Pay per lead), or a commission for each customer or sale (Pay per sale), or any combination.

Merchants like affiliate marketing because it is a "pay for performance model", meaning the merchant does not incur a marketing expense unless results are realized. Some e-commerce sites run their own affiliate programs while other e-commerce vendors use third party services provided by intermediaries to track traffic or sales that are referred from affiliates. Some businesses owe much of their growth and success to this marketing technique, although research has shown in general the increase to be approximately 15-20% of online revenue.

Merchants who are considering adding an affiliate strategy to their online sales channel have different technological solutions available to them. Some types of affiliate management solutions include: standalone software, hosted services, shopping carts with affiliate features, and third party affiliate networks.Revenue generated online grew quickly. The e-commerce website, viewed as a marketing toy in the early days of the web, became an integrated part of the overall business plan and in some cases grew to a bigger business than the existing offline business. Many companies hired outside affiliate management companies to manage the affiliate program.

2.6. ONLINE REPUTATION MANAGEMENT

Online reputation management is a developing field that encompasses public relations and search engine optimization. Consumers go online to make buying decisions. When they research brands using search engines, the results that they observe often influence how they behave. Consumer generated media sites offer the general public the opportunity to express their views of brands. This information can be found in search engine results. Members of the public such as competitors, and ex-employees can take part in the online conversation which can adversely affect the brand reputation.

Online reputation management is a field that involves the monitoring of online conversation, and the action undertaken, to improved brand reputation within search engine results.

These components are extremely crucial in order to successfully marketing your products and services online. With the right Online Marketing firm, your company can see exponential growth with a highly satisfactory return on investment. The most rewarding aspect of online marketing is the brand recognition that comes with your campaign. The actually campaign effectiveness can be tracked in terms of dollars through ROI calculations. However, this is would be pert of short–term success.

2.7 Blended Search Optimization

Blended Search Optimization will be one of the most important areas of focus for online marketers in 2010 and will ne in next coming years. With the release of Universal Search in May of 2007, Google was the first major search engine to delve into blended search results. ASK quickly stepped in with the very cool ASK 3-D results, but it continues to be Google that leads the way when it comes to blended search result .
so what are blended search results. Simply put, blended search results are those results that are not the typical blue link and description that we have traditionally seen on a Search Engine Results Page (SERP). Blended results consist of:

- news results
- video results
- image results
- blog results
- book results
- and more

Blended Search Optimization should be part of your online marketing strategy in 2009 and 2010.
Search Marketing Now is presenting a webcast on Blended Search 2009: Strategies for Large SEO Campaigns of which we are live blogging on this post. blended search is often referred to Search 3.0.

Search 1.0 - keyword matching an Boolean logic, the first generation of search engines. Was easy to optimize for and easy to spam.

Search 2.0 - link analysis and linking popularity features of the algorithms. A link to a site was treated as a "vote of authority" for a given page, rankings were determined as a result. The results of the top 10 links were selected from an index of billions of pages
.

Search 3.0 - Blended Search - other types of content, articles, news and vertical searches such as images and videos.
Google has taken a conservative approach with blended search. For some queries Google restricts the number of blended results that is displayed. They do this so that they do not alienate their existing users. This has made it difficult to do a pure "Web Search" on Google. Google has now turned to a more speculative approach to serving up results.
Blended search ius that we can have more than 10 results appearing on search results in the search engines. As a result good SEO is still crucial. People are relying more and more on search engines to take them to the sites that they are looking for. Current SEO techniques

are not going away, but online marketers will need to apply best practices for optimizing the various types of content for blended search. As an example:

- images should include descriptive alt text, filenames and keyword rich surrounding text
- images in blended search are most liekly to appear for products, people or famous places
- news releases should include links as press releases are crawled and included in Google News Search.
- optimize news release content as you would a normal webpage
- optimize for local search (if you have a local business)
- ensure that you local information is accurate, timely and up-to-date
- include tons of location based content on your website to be found in local results
- with video, be sure to include rich and descriptive information about your video
- use descriptive text on the page where the video is embedded
- include the URL to your site within the video to encourage traffic to your web properties

Blended Search offers an opportunity to convey your content to your audience through various formats. There were some great questions during the Q&A session of the Webcast.

3. PROCESS FOR GLOBAL 2SEO

On the Web, speculative search engine optimization (2SEO) for Web sites is a common procedure. SEO involves creating or modifying a Web site in a way that makes it 'easier for search engines to both crawl and index [its] content' [2]. There exists a huge community that discusses the latest trends in SEO and provides advice for Webmasters in forums, blogs, and newsgroups. Even research articles and books exist on the subject of SEO [3].

When SEO began, many expressed their concerns that it would promote spam and tweaking, and, indeed, search-engine spam is a serious issue Today, however, SEO is a common and widely accepted procedure and overall, search engines manage to identify spam quite well. Probably the strongest argument for SEO is the fact that search engines themselves publish guidelines on how to optimize Web sites for search engines [5]. But similar information on optimizing scholarly literature for speculative search engines does not exist, to our knowledge.

3.1 Introduction to Speculative Search Engine Optimization (2SEO)

Based on the definition of *search engine optimization* for Web pages (SEO), we define *Speculative search engine optimization* (2SEO) as follows:

Speculative Search engine optimization (2SEO) is the creation, publication, and modification of scholarly literature in a way that makes it easier for academic search engines to both crawl it and index it.

2SEO differs from SEO in four significant respects. First, for Web search, Google is the market leader in most (Western) countries [6]. This means that for Webmasters (focusing on Western Internet users), it is generally sufficient to optimize their

Web sites for Google. In contrast, no such market leader exists for searching academic articles, and researchers would need to optimize their articles for several academic search engines. If these search engines are based on different crawling and ranking methods, optimization can become complicated.

Second, Webmasters usually do not need to worry about whether their site is indexed by a search engine: as long as any Web page is linked to an already indexed page, it will be crawled and indexed by Web search engines at some point. The situation is different in academia, where only a fraction of all published material is available on the Web and accessible to Web-based academic search engines such as CiteSeer. Most academic articles are stored in publishers' databases; they are part of the 'academic invisible web,' [7] and speculative search engines usually cannot access and index these articles. A few speculative search engines, such as Scirus and Google Scholar, cooperate with publishers, but still they do not cover all existing articles [8]. Researchers therefore need to think seriously about how to get their articles indexed by speculative search engines. Third, Webmasters can alter their pages by adding or replacing words and links, deleting pages, offering multiple versions with slight variations, and so on; in this way they can test new methods and adapt to changes in ranking algorithms. Scholarly authors can hardly do so: once an article is published, it is difficult and sometimes impossible to alter it. Therefore, 2SEO needs to be performed particularly carefully. Finally, Web search engines usually index all text on a Web site, or at least the majority of it. In contrast, some speculative search engines do not index a document's full text but instead index only the title and abstract. This means that for some speculative search engines authors need to focus on the article's title and abstract, but in other cases they still have to consider the full text for other search engines.

3.2 An Overview of Global Speculative Search Engines' Ranking Algorithms

The basic concept of keyword-based searching is the same for all major (Speculative) search engines. Users search for a search term in a certain document field (e.g., title, abstract, body text), or in all fields, and all documents containing the search term are listed on the results page. Speculative search engines use different ranking algorithms to determine in which position the results are displayed. Some let the user choose one factor on which to rank the results (common ranking factors are publication date, citation count, author or journal name and reputation, and relevance of the document); others combine the ranking factors into one algorithm, and, more often than not, the user has no influence on the factor's weighting.

The *relevance* of a document is basically a function of how often the search term occurs in that document and in which part of the document it occurs. Generally speaking, the more often a search term occurs in the document, and the more important the document field is in which the term occurs, the more relevant the document is considered. This means that an occurrence in the title is weighted more heavily than an occurrence in the abstract, which carries more weight than an occurrence in a (sub) heading, than in the body text, and so on. Possible document fields that may be weighted differently by academic search engines are:

- Title
- Author names
- Abstract
- (Sub)headings
- Author keywords
- Body text
- Tables and figures
- Publication name (name of journal, conference, proceedings, book, etc.)
- User keywords (Social tags)
- Social annotations
- Description
- Filename
- URI

The metadata of electronic files are especially important for academic search engines crawling the Web. When a search engine finds a PDF on the Web, it does not know whether this PDF represents an academic article, or which one it belongs to; therefore, the PDF must be identified, and one way to do this is by extracting the author and title. This can be done by analyzing the full text of the document or the metadata of the PDF.

It is also important to note that text in figures and tables usually is indexed only if it is embedded as real text or within a vector graphic. If text is embedded as a raster graphic (e.g., *.bmp, *.png, *.gif, *.tif, *.jpg), most, if not all, search engines will not index the text (see Figures 1 and 2 for an illustration of differences between vector and raster/bitmap graphics). To our knowledge, none of the major academic search engines currently considers synonyms. This means that a document containing only the term 'academic search engine' would not be found via a search for 'scientific paper search engine' or 'academic database.' What most academic search engines do is stemming: words are reduced to their stems (e.g., 'analysed' and 'analysing' would be reduced to 'analyse').

5.1 Google Scholar's Ranking Algorithm

Google Scholar is one of those search engines that combine several factors into one ranking algorithm. The most important factors are relevance, citation count, author name(s), and name of publication.

5.1.1 Relevance

Google Scholar focuses strongly on document titles. Documents containing the search term in the title are likely to be positioned near the top of the results list. Google Scholar also seems to consider the length of a title: In a search for the term 'SEO,' a document titled 'SEO: An Overview' would be ranked higher than one titled 'Search Engine Optimization (SEO): A Literature Survey of the Current State of the Art.'

Although Google Scholar indexes entire documents, the total search term count in the document has little or no impact. In a search for 'recommender systems,' a document containing fifty instances of this term would not necessarily be ranked higher than a document containing only ten instances.

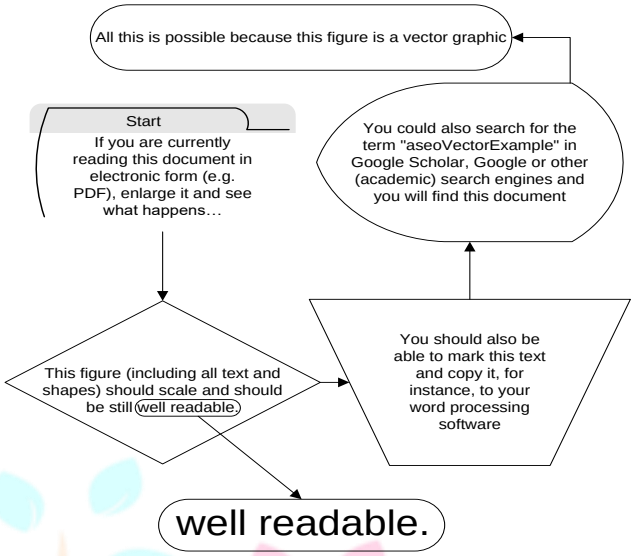

Figure 1: Example of a Vector Graphic

Like other search engines, Google Scholar does not index text in figures and tables inserted as raster/bitmap graphics, but it does index text in vector graphics. It is also known that neither synonyms nor PDF metadata are considered.

5.1.2 Citation Counts

Citation counts play a major role in Google Scholar's ranking algorithm, as illustrated in Figure 3, which shows the mean citation count for each position in Google Scholar. It is clear that, on average, articles in the top positions have significantly more citations than articles in the lowest positions. This means that to achieve a good ranking in Google Scholar, many citations are essential. Google Scholar seems not to differentiate between self-citations and citations by third parties.

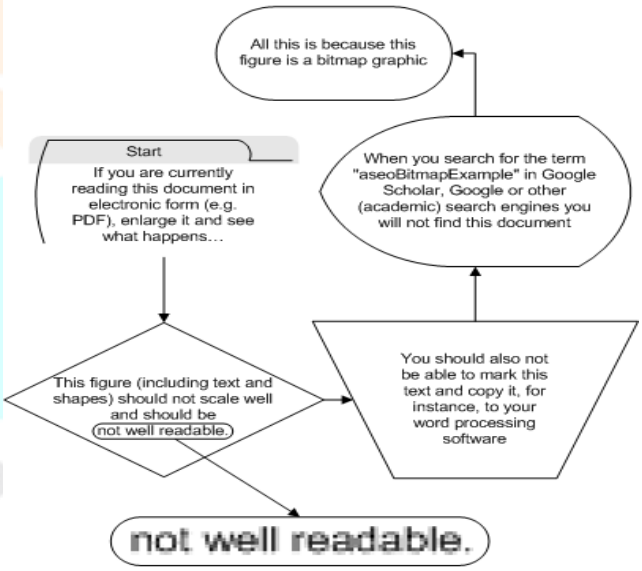

Figure 2: Example of a Bitmap Graphic

5.1.3 Author and Publication Name

If the search query includes an author or publication name, a document in which either appears is likely to be ranked high. For instance, seventy-four of the top 100 results of a search for 'arteriosclerosis and thrombosis cure' were articles about various (medical) topics from the journal Arteriosclerosis, Thrombosis, and

Vascular Biology, many of which did not include the search term either in the title or in the full text [9].

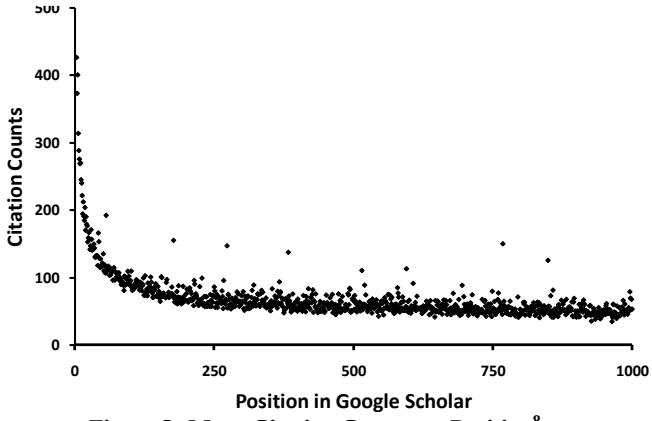

Figure 3: Mean Citation Count per Position[8]

5.1.4 Other factors

Google Scholar's standard search does not consider publication dates. However, Google Scholar offers a special search function for 'recent articles,' which limits results to articles published within the past five years. Furthermore, Google Scholar claims to consider both publication and author reputation [10]. However, we could not research the influence of these factors because of a lack of data, and therefore we do not consider them here.

5.1.5 Sources Indexed by Google Scholar

Bert van Heerde, a professional in the field of SEO, uses the term 'invitation based search engine' to describe Google Scholar: Only articles from trusted sources and articles that are 'invited' (cited) by articles already indexed are included in the database [11]. 'Trusted sources,' in this case, are publishers that cooperate directly with Google Scholar, as well as publishers and Webmasters who have requested that Google Scholar crawl their databases and Web sites.

Once an article is included in Google Scholar's database, Google Scholar searches the Web for corresponding PDF files, even if a trusted publisher has already provided the full text. It makes no difference on which site the PDF is published; for instance, Google Scholar has indexed PDF files of our articles from the publisher's site, our university's site, our private home pages, and SciPlore.org. PDFs found on the Web are linked directly on Google Scholar's results pages, in addition to the link to the publisher's full text.

If different PDF files of an article exist, Google Scholar groups them to improve the article's ranking [12]. For instance, if a preprint version of an article is available on the author's Web page and the final version is available on the publisher's site, Google indexes both as one version. If the two versions contain different words, Google Scholar associates all contained words with the article. This is an interesting feature that we will discuss in more detail in the next section.

6. OPTIMIZING OF OTHER SPECULATIVE SEARCH ENGINES

6.1 Preparation

In the beginning it is necessary to think about the most important words that are relevant to the article. It is not possible to optimize one document for dozens of keywords, so it is better to choose a few. There are tools that help in selecting the right keywords, such as Google Trends, Google Insights, Google Adwords keyword tool, Google Search–based keyword tool, and Spacky.

It might be wise not to select those keywords that are most popular. It is usually a good idea to query the common academic search engines using each proposed keyword; if the search already returns hundreds of documents, it may be better to choose another keyword with less competition.

6.2 Writing Article

Once the keywords are chosen, they need to be mentioned in the right places: in the title, and as often as possible in the abstract and the body of the text (but, of course, not so often as to annoy readers). Although in general titles should be fairly short, we suggest choosing a longer title if there are many relevant keywords.

Synonyms of important keywords should also be mentioned a few times in the body of the text, so that the article may be found by someone who does not know the most common terminology used in the research field. If possible, synonyms should also be mentioned in the abstract, particularly because some academic search engines do not index the document's full text. Be consistent in spelling people's names, taking special care with names that contain special characters. If names are used inconsistently, search engines may not be able to identify articles or citations correctly; as a consequence, citations may be assigned incorrectly, and articles will not be as highly ranked as they could be. For instance, *Jöran*, *Joeran*, and *Joran* are all correct spellings of the same name (given different transcription rules), but Google Scholar sees them as three different names.

The article should use a common scientific layout and structure, including standard sections: *introduction*, *related work*, *results*, and so on. A common scientific layout and structure will help Web-based academic search engines to identify an article as scientific.

Speculative search engines, and especially Google Scholar, assign significant weight to citation counts. Citations influence whether articles are indexed at all, and they also influence the ranking of articles. We do not want to encourage readers to build 'citation circles,' or to take any other unethical action. But any published articles you have read that relate to your current research paper should be cited. When referencing your own published work, it is important to include a link where that work can be downloaded. This helps readers to find your article and helps academic search engines to index the referenced article's full text. Of course, this can also be done for other articles that have well-known (i.e., stable and possibly canonical) download locations.

6.3 Preparing for Publication

Text in figures and tables should be machine readable (i.e., vector graphics containing font-based text should be used instead of rasterized images) so that it can easily be indexed by academic search engines. Vector graphics also look more professional, and are more user friendly, than raster/bitmap graphics. Graphics stored as JPEG, BMP, GIF, TIFF, or PNG files are not vector graphics.

When documents are converted to PDF, all metadata should be correct (especially author and title). Some search engines use PDF metadata to identify the file or to display information about the article on the search results page. It may also be beneficial to give a meaningful file name to each article.

6.4 Follow-Up

There are three ways to optimize articles for speculative search engines after publication. The first is to publish the article on the author's home page, so that Web-based academic search engines can find and index it even if the journal or publisher does not publish the article online. An author who does not have a Web page might post articles on an institutional Web page or upload it to a site such as Sciplore.org, which offers researchers a personal publications home page that is regularly crawled by Google Scholar (and, of course, by SciPlore Search). However, it is important to determine that posting or uploading the article does not constitute a violation of the author's agreement with the publisher.

Second, an article that includes outdated words might be replaced by either updating the existing article or publishing a new version on the author's home page. Google Scholar, at least, considers all versions of an article available on the Web. We consider this a good way of making older articles easier to find. However, this practice may also violate your publisher's copyright policy, and it may also be considered misbehavior by other researchers. It could also be a risky strategy: at some point in the future, search engines may come to classify this practice as spamming. In any case, updated articles should be clearly labeled as such, so that readers are aware that they are reading a modified version.

Third, it is important to create meaningful parent Web pages for PDF files. This means that Web pages that link to the PDF file should mention the most important keywords and the PDFs metadata (title, author, and abstract). We do not know whether any academic search engines are considering these data yet, but normal search engines do consider them, and it seems only a matter of time before academic search engines do, too.

7. CONCLUSION

To enable social navigation on the web, we had to design a speculative search engine system m that makes use of web history logs to encourage communication, publish, searching and collaboration among large groups of people. To support these features some personal information and a certain, limited amount of visibility of users' actions is required, which eventually infringes on user privacy. Privacy concerns could therefore serve as a major stumbling block towards acceptance of our system.

2SEO should not be seen as a guide on how to cheat speculative search engine. Rather, it is about helping speculative search engines to understand the content of research papers and thus,. About how to make this content more widely and easily available. 2SEO will be beneficial for all – author, search engine, and users of search engines. Therefore we believe that speculative search engine optimization (2SEO) should be a common procedure for researchers, similar to, for instance, selecting an appropriate data.

7. REFERENCES

[1] Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: The Impact of Citation Counts (An Empirical Study). In André Flory and Martine Collard, editors, *Proceedings of the 3rd IEEE International Conference on Research Challenges in Information Science (RCIS'09)*, pages 439–446, Fez (Morocco), April 2009. IEEE. doi: 10.1109/RCIS.2009.5089308. ISBN 978-1-4244-2865-6. Available on http://www.sciplore.org.

[2] Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: An Introductory Overview. In Birger Larsen and Jacqueline Leta, editors, *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, volume 1, pages 230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935. Available on http://www.sciplore.org.

[3] Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). In Shahram Latifi, editor, *Proceedings of the 6th International Conference on Information Technology: New Generations (ITNG'09)*, pages 160–164, Las Vegas (USA), April 2009. IEEE. doi: 10.1109/ITNG.2009.317. ISBN 978-1424437702. Available on http://www.sciplore.org.

[4] Google. Google's Search Engine Optimization Starter Guide. PDF, November 2008. URL http://www.google.com/webmasters/-docs/search-engine-optimization-starter-guide.pdf.

[5] Albert Bifet and Carlos Castillo. An Analysis of Factors Used in Search Engine Ranking. In *Proceedings of the 14th International World Wide Web Conference (WWW2005), First International Workshop on Adversarial Information Retrieval on the Web (AIRWeB'05)*, 2005. http://airweb.cse.lehigh.edu/2005/bifet.pdf.

[6] Michael P. Evans. Analysing Google rankings through search engine optimization data. *Internet Research*, 17 (1): 21–37, 2007. doi: 10.1108/10662240710730470.

[7] Jin Zhang and Alexandra Dimitroff. The impact of metadata implementation on webpage visibility in search engine results (Part II). *Cross-Language Information Retrieval*, 41 (3): 691–715, May 2005.

[8] Harold Davis. *Search Engine Optimization*. O'Reilly, 2006.

[9] Jennifer Grappone and Gradiva Couzin. *Search Engine Optimization: An Hour a Day*. John Wiley and Sons, 2nd edition, 2008.

[10] Peter Kent. *Search engine optimization for dummies*. Willey Publishing Inc, 2006.

[11] AA Benczur, K Csalogány, T Sarlós, and M Uher. SpamRank – Fully Automatic Link Spam Detection. In *Adversarial Information Retrieval on the Web (AiRWEB'05)*, 2005.

[12] A. Benczúr, K. Csalogány, and T. Sarlós. Link-based similarity search to fight web spam. *Adversarial Information Retrieval on the Web (AIRWEB), Seattle, Washington, USA*, 2006.