Measurement of Online Visibility

Nadine Schmidt-Mänz and Wolfgang Gaul

Institute of Decision Theory and Management Science, University of Karlsruhe, Germany

Abstract. To attract web visitors via the internet it is fundamental for all kinds of online activities to be "visible" in the net. Visibility measurement is important for web sites: it helps to define benchmarks with respect to competition and allows to calculate visibility indices as predictors for site traffic.
This paper discusses a new approach to measure online visibility (OV) and compares it with one known from the literature.
We describe physical and psychological drivers of OV and suggest a measurement of OV that works automatically as a robot in the internet. Managerial implications to make web sites "smelly" or "visible" are also discussed.

1 Introduction

Nearly 80% of all internet users find new web sites with the aid of search engines. With this information in mind, search engines appear to be very important instruments to get in contact with new customers for whatever online business or visitors of web sites in this medium. For example, an insurance broker listed at the first position of the result page of a search engine may expect to be frequented by up to 10,000 visitors per month when web users look for "insurance" [6]. A traditional marketing campaign would cost several thousand dollars to have the same effect. Hence, it is very important to "maximize" what could be called online visibility (OV) of web sites, or -at least- to improve OV relative to competition.
In section 2 we describe main drivers of OV and focus on facts about human online searching behavior. This leads to a new measure of OV and possibilities for its determination as discussed in section 3. In section 4 we give some managerial implications. Section 5 presents conclusions.

2 Drivers of Online Visibility

OV is composed of different parts. Here, we will concentrate on the following two: There are psychological drivers of OV derived by human online searching behavior. Here, questions like how humans use the internet and how they interact with search engines have to be taken into consideration. Psychological drivers can decrease OV. Additionally, physical drivers of OV such as links to a web site, banner ads, listings in search engines or directories, etc., are of importance. They can be influenced by administrators of web sites themselves and increase OV.

2.1 Psychological Drivers of Online Visibility

To measure OV it is essential to understand psychological drivers derived by human online searching behavior:
As already mentioned about 80 % of all internet users find new web sites with the aid of search engines. Thus, *Google*, *Teoma*, etc., are effective instruments to reach new visitors or customers [6].
In course of time, people become more efficient in using the web by navigating directly to a web site they already know, but they use search engines to find new ones (see table 1 according to an independent report by *WebSideStory*, Inc. [14]).

Table 1. Global Internet Usage

Referral Type	**2002**	**2003**	trend
Direct Navigation	50.21%	65.48%	↗
Web Links	42.60%	21.04%	↘
Search Engines	07.18%	13.46%	↗

As global internet usage with respect to search engines has nearly doubled from 2002 to 2003 the question is how do internet users interact with search engines? There are three main studies covering web searching strategies with the help of search engines by analyzing query logs: The *Fireball* [8], the *Excite* [9], and the *AltaVista* study [12].
In the following we only concentrate on the *AltaVista* study, since conclusions of all three studies are nearly the same. [12] is based on the largest data set (one billion queries submitted to the main *AltaVista* search engine over a 42-day period) and provides a broad spectrum of information: given number of queries in the data set, length of the collection period and analysis, it is the most complete web searching study to date.

Among the facts worth mentioning about human online searching behavior the following are interesting:
Nearly 77.6% of all query sessions consisted of only one request. 85.2% of searchers examined only one result screen per query (7.5% two, and 3.0% three screens). The average number of terms in a query added up to 2.35 ($\sigma = 1.74$) and that of operators (AND, OR, NOT ...) to 0.41 ($\sigma = 1.11$). According to the total number of queries, 63.7% occurred only once, i. e. the formulated queries were "almost" unique. The most popular term in a query was "sex" with an appearance of 1,551,477 times. This is equal to 2.7% of the total number of non-empty queries.

2.2 Physical Drivers of Online Visibility

Drèze and Zufryden [5] defined visibility as the extent of presence of a brand or a product in the consumer's environment. Thus, one can view OV as an indicator with respect to potential web site traffic in the same way as awareness is a precursor to purchase. A web site can draw attention to its content and attract potential internet surfers by both offline (television, radio, newspapers, etc.) and online (banner ads, links from search engine result pages, online directories or other web pages, etc.) means.
The authors identified three main physical drivers which increase OV and can be measured by taking a snapshot of the internet.

Links from other web sites: One can increase OV by increasing the number of incoming links from important or frequently visited web pages. To find the amount of incoming links from web pages to the underlying site one can use the brute force approach, i. e. crawl the net and keep track of all links located on web sites. This is the common approach of most search engines which provide users with a list of these links (fan-ins, inbounds, incoming links). Users can retrieve this kind of information by a special command and the corresponding URL in the search interface (e.g., "link:www.xyz.com" with respect to *Google*) and should repeat this procedure for different engines to ensure maximum coverage [2].
There is a serial position effect in the HTML code that relates to effectiveness of links (the higher the rank order of a link the higher the click through rate on that link [1]) The hypothesis that the effectiveness of links is also affected by the depth of the page on which the link was located (with the home page of a web site having a depth of 1) was statistically not significant.

Online Directories: To become visible in the net it is important to be listed in the adequate categories of online directories. Directories or subject catalogs that represent the "yellow pages" of the internet are characterized by their hierarchical structure. The most important property of directories is that they are validated by human beings and not by robots. Experts look at every page submitted for registration in the index and review it. This procedure is generally referred to as *manual indexing*. One advantage is that a human indexer can detect relations between seemingly different contents. But it is very time consuming and expensive. Well known examples for directories are *Yahoo!* [15] and *ODP* [13].

Search Engines: Web search engines (also called web indices, index servers, or simply search engines) work exclusively with automated methods [11]. A collection of high rankings in such engines is one of the best prerequisites to yield enhanced OV for new visitors. Drèze and Zufryden [5] took the fact into consideration that search engines always return result pages which are formatted in a special way (e.g. 10 links per result page) and that the frequency

of looking at the first, second, third ... result page is decreasing.
In addition, they characterized web sites as trees that surfers explore by going up and down branches where less weight was assigned to deep links and more to links that are close to the root of the respective site.
These results gave some insights in how people process online information. First, the position of a link on a web page is important for OV. Second, as it has not been proved that people view web sites structured in a arborescent manner, the depth of a page in the site seems to have little impact on its importance. This may be due to the ability of search engines to route people directly to that page where the desired information is located. Additionally, it should be noted that, in the meantime, it is legalized that search engines use these "deep links" [4]. Third, the results show that visibility of web sites in search engines is dependent on both position of the link on and depth of the result page on which the underlying link is situated.

3 A New Approach to Measure OV

Based on the knowledge of human online searching behavior and the functioning of ranking algorithms of search engines such as PageRank [3,7,10] we can define several (new) impacts on OV.

3.1 Impacts of Human (Online) Searching Behavior on OV

One aspect to define a new measure of OV is the importance of fan-ins and directory listings on ranking algorithms of search engines. Another aspect is related to the searching behavior of internet users. If OV is measured by the three physical drivers mentioned before (search engine ranking, incoming links, and directory listings) in the same way, some problems may arise. To calculate a precise measurement one will have to subtract all overlappings and as such a measure doesn't exist we have estimated OV based on what people will definitively see.

1. The first search engine results visible on the answer screen as a cutout of the first result page (in the majority of cases ten entries) will have the best chance to be clicked on by a searching person
2. Human searchers don't browse every result page of a search engine. They only browse a few with exponentially decreasing intensity (*AltaVista* study [12]).
3. We do not incorporate the number of results returned by search engines, as most search engines return ten entries per result page.
4. In addition, we consider a measurement of Adwords' appearance (a special feature of *Google*) because this is a good instrument for enterprises to become visible via up-to-date activities (special offers, sales) in a short time with the aid of bundles of keywords.

5. We do not use general keywords from keyword databases but propose an individualized set of keywords to differentiate web sites by content.
6. We do not only measure OV with one set of keywords but use a set of up to three keywords and all subsets of variations.
7. As it is not possible to measure OV in directories in an objective way (the best-known directories list URLs or names in alphabetical order like yellow pages) we omit the determination of visibility via listings in online directories.
8. We count fan-ins to a certain extent: we include incoming links with a small factor in order not to overestimate this effect.

3.2 The New Measure *GOVis*

Our new approach to estimate OV has been called GOVis (Gage of Online Visibility) to differentiate it from other measures. It is given by

$$GOVis = \sum_{k=1}^{\sum_{n=1}^{N}\binom{N}{n}\cdot n!} \left[\alpha \cdot \sum_{p=1}^{2} \frac{1}{e^{p-1}} \cdot \sum_{r=1}^{R} X_{kpr} + \beta \cdot \sum_{a=1}^{A} Y_{k1a}\right] + \gamma \cdot \ln(Z_L) \tag{1}$$

with

* $\mathcal{K}$ is a set of interesting keywords, with $\mid \mathcal{K} \mid = N$ (normally $N \leq 3$), $\sum_{n=1}^{N}\binom{N}{n} \cdot n!$ is the quantity of all ordered subsets of $\wp(\mathcal{K}) \setminus \{\emptyset\}$ and k is the kth subset of keywords with which a query in *Google* is performed,
* p is the depth of the result pages of the search engine used (we have restricted depth to $p \leq 2$),
* R is the quantity of results per result page and r is the rth ranking position on the result pages (we, normally, use *Google* with $R = 10$),
* A is the maximum quantity of Adwords per result page (*Google* standard is $A = 8$) and a is the ath Adword ranking position,
* L is the corresponding URL, Z_L is the number of corresponding fan-ins,
* h_{kpr} is the hyperlink at the rth ranking position on the page with depth p generated by a query with the kth subset of keywords,
* w_{k1a} is the Adword link at the ath Adword ranking position on the page with depth 1 generated by a query with the kth subset of keywords,
* $X_{kpr} = \begin{cases} 1, & h_{kpr} \text{ links to } L \\ 0, & \text{otherwise} \end{cases}$ $\quad Y_{k1a} = \begin{cases} 1, & w_{k1a} \text{ links to } L \\ 0, & \text{otherwise} \end{cases}$
* $\alpha + \beta + \gamma = 1$.

In (1) we excluded measures of OV in directories, on portals, in chat rooms or banner ads, etc. We didn't consider overlappings, but we used different numerical values for α, β, and γ to consider how GOVis depends on differences with respect to the meaning of fan-ins, Adwords or rankings in search engine result pages.

3.3 Implementation and Results of *GOVis*

We implemented (1) in PHP as a parser. We used *Google* and Adwords for queries (because *Google* has the most pages in the index). For the quantity of fan-ins we considered *AltaVista* (as in *Altavista* it is possible to exclude links from the home domain (link:http://www.xyz.com -host:http://www.xyz.com) while *Google* only saves fan-ins of pages with a high PageRank). In this first version of *GOVis* it is possible to enter up to three keywords, a special URL and to customize α, β, and γ.
For results of OV measured by *GOVis* we used scenarios with different α, β, and γ for $\mathcal{K}$ ={Bücher, Dvds, HiFi}, see table 2.

Table 2. GOVis results

scenario$_x$ (α, β, γ)	enterprise$_x$	$\sum X_{kpr}$	$\sum Y_{k1a}$	$\ln(Z_L)$	*GOVis*
0.15, 0.00, 0.85	***Amazon*$_1$**	3+1	4	14.1774	**12.5560**
	Ebay$_1$	0+6	1	11.7464	10.3155
0.40, 0.00, 0.60	***Amazon*$_2$**	3+1	4	14.1774	**9.8536**
	Ebay$_2$	0+6	1	11.7464	8.3722
0.85, 0.00, 0.15	***Amazon*$_3$**	3+1	4	14.1774	**4.9893**
	Ebay$_3$	0+6	1	11.7464	3.6381
0.80, 0.15, 0.05	***Amazon*$_4$**	3+1	4	14.1774	**4.0032**
	Ebay$_4$	0+6	1	11.7464	2.5030

In scenario$_1$ and scenario$_2$ we chose α and γ on the basis of adjusted figures from table 1 of the years 2002 and 2003 without "Direct Navigation". The number of users that find new web sites with the aid of search engines increases while on the other hand fewer users reach new pages over web links.
In scenario$_3$ we defined α and γ according to PageRank. Here, γ is the probability that a random surfer continues to click on links while with $1-\gamma$ (s)he will jump to any page in the web. In this scenario search engines are the best possibility to jump to any visible page in the web (where we have the information in mind that 80% of all internet users find new web pages with the aid of search engines).
In scenario$_4$ we also included the visibility of Adwords to a small extent and lowered the impact of incoming links on ranking algorithms.
As it is important to know more about the navigational behavior of target segments to which the efforts to increase OV are directed the calculation of α, β and γ is essential for managerial implications.

3.4 Improvements by *GOVis*

First of all, *GOVis* can be individualized, i. e. that for every company and their competitors OV is measured by content specific keywords. As *GOVis* does not use a general list of most common used keywords one can select keywords depending on special up-to-date activities (sales, new products). This and the fact that we base visibility measurement on human searching behavior (only a few keywords for the formulation of a query, only a few result pages of the search engine are actually considered) lead to short running times to compute OV via GOVis.
GOVis is also an effective instrument to control visibility of online activities in relation to competitors. It represents a relative measure of OV based on special keywords concerning the content of and activities related to the corresponding web site.
Since α, β, and γ are customizable one can calibrate these parameters on the basis of new values from new online surveys concerning human searching behavior and the way how target segments interact with the internet to find web sites.

4 Managerial Implications

The scenarios in table 2 show that it becomes more and more important to increase search engine visibility, but that it is also important to take into consideration the searching person itself. Thus, it is not the ultimate to optimize a web site for better search engine ranking. In the end, this can be a tit-for-tat strategy because search engines will adapt their ranking algorithms to block attempts that only optimize web sites with respect to a high ranking position in their result pages. An honest way (with customers/visitors and search engines in mind) to improve OV or to make web sites "smelly" is to observe the log files of the corresponding web site to detect search engine referrals that include important keywords of searching persons. Another possibility is to sift online keyword databases to compare already used keywords with descriptions or text content of the corresponding web site to meet customer needs with respect to, e.g., content, special topics or product descriptions. In case of up-to-date activities it is advisable to buy bundles of Adwords as search engines crawl web sites only every few months, so temporal activities won't be visible.

5 Conclusion

In total the measurement of OV should represent a benchmark for specialized content or activities and not lead to a strategy to become listed at the first position of a search engine result page based on a query corresponding to

keywords describing a general topic. There is an impact of the order of keywords in a query on visibility, thus, long term optimization of content of the corresponding web site has to be considered. Visibility has to be measured in constant short term periods as the internet changes "quickly".
The bundle of content visibility, Adwords visibility, search engine visibility, and visibility based on incoming links describes the most important instruments to account for OV. GOVis is a first step in order to measure online visibility affects.

References

1. Ansari, Asim and Mela, Carl F. (2002) e-Customization. Forthcoming in Journal of Marketing Research,
   Preprint: http.//www.cebiz.org/downloads/ecustomization.pdf.
2. Bradlow, Eric T. and Schmittlein, David C. (2000) The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines. Marketing Science, Vol. 19, No. 1, 43-62.
3. Brin, Sergey and Page, Lawrence (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. Proceedings of the 7th International World Wide Web Conference (WWW7), 107-117.
4. Bundesgerichtshof, Mitteilung der Pressestelle (Nr.96/2003) Internet-Suchdienst für Presseartikel nicht rechtswidrig. Urteil vom 17. Juli 2003 - I ZR 259/00.
5. Drèze, Xavier and Zufryden, Fred (2003) The Measurement of Online Visibility and its Impact on Internet Traffic. To appear in: Journal of Interactive Marketing, Preprint: http://www.xdreze.org/Publications/Visibility.html.
6. Fischerländer, Stefan (2003) Websites Google-gerecht gestalten - Ganz nach oben. iX 08/2003, 84-87.
7. Henzinger, Monika, Heydon, Allan, Mitzenmacher, Michael, and Najork, Marc (1999) Measuring Index Quality Using Random Walks on the Web. Proceedings of the 8th International World Wide Web Conference (WWW8), 213-225.
8. Hölscher, Christoph (1998) How Internet Experts Search for Information on the Web. In: H. Maurer & R. G. Olson (Eds.), Proceedings of WebNet98 - World Conference of the WWW, Internet & Intranet. Charlottesville, VA: AACE.
9. Jansen, Bernard, Spink, Amanda, and Saracevic, Tefko (2000) Real Life, Real Users, and Real Needs: A Study Analysis of User Queries on the Web. Information Processing and Management, 36: 207-227.
10. Kleinberg, Jon (1998) Authoritative Sources in a Hyperlinked Environment. In: Journal of the ACM, Vol. 46(5), 1999, 604-632.
11. Michael, W. Berry and Browne, Murray (1999) Understanding Search Engines - Mathematical Modeling and Text Retrieval. Software, Environments, and Tools, SIAM, Society for Industrial and Applied Mathematics, Philadelphia.
12. Silverstein, Craig, Henzinger, Monika, Marais, Hannes, and Moricz, Michael (1999) Analysis of a Very Large AltaVista Query Log. *SIGIR Forum*, 33(1): 599-621.
13. www.dmoz.com, last visited August 10, 2002
14. www.websidestory.com. Search Engine Referrals Nearly Double Worldwide, According to WebSideStory, March 2003, last visited August 10, 2003.
15. www.yahoo.com, last visited August 10, 2003.