Visibility Analysis on the Web as an Indicator for Public Relations and Marketing Evaluation

Peter Kiefer and Klaus Stein

Laboratory for Semantic Information Technologies

Otto-Friedrich-University Bamberg

96045 Bamberg, Germany

peter.kiefer@wiai.uni-bamberg.de

Abstract

One componant of an effective marketing strategy is the evaluation of public relations and publicity campain efforts. We provide instruments for web-based marketing evaluation by introducing the concept of topic visibility based on search engine hit counts. This allows to monitor changes in public awareness of certain topics like products or companies in time. Using covisibilities and domain restrictions provides fine-granular evaluation regarding selected markets. A number of case studies illustrate our approach.

1. Introduction

Today's information society causes an information glut that is both, hard to manage for the individual but also an aggravation for organizations when trying to attract people's attention (see also work on attention economies, e. g. [5]). This aggravation arises not only for companies competing for customers, but also for nonprofit organizations that strive to arouse public awareness for their concerns. With lots of money and effort spent on public relation and marketing campaigns, the evaluation of these attempts becomes of vital importance. Certainly, methods for such evaluation tasks have been developed (see e. g. [7, 19] for work on public relations evaluation), but they are usually based on surveys and therefore relatively expensive and time-consuming.

On the other hand, the internet has been discovered as an easily accessible and rich source of information for knowledge discovery. Web Mining has already proven to be useful for marketing issues, e. g. for the evaluation of internet shops with web usage mining (see [16]).

Our paper focuses on web content mining as one indicator for the evaluation of public relations and marketing campaigns. It is inspired by current findings from [8] showing that events from the real world leave traces on the web and can be monitored by visibilities on the internet. We propose an accordant approach for the areas of public relations and marketing and present some examples indicating its usability. As computer scientists, we do not claim our approach to be exhaustive or standalone. We rather try to give experts from these fields indicators they can combine with their standard methods. We are aware that the results gathered with our approach of visibility analysis may easily be misinterpreted without substantiated background knowledge and encourage experts from these areas to combine visibility analysis with their standard approaches.

The rest of the paper is structured as follows: Sec. 2 introduces the concept of topic visibility on the web, shows that real-world events influence visibility and presents a small case study from market research. Ideas how to exploit domain dependent visibilities for public relations and marketing evaluation are subject of Sec. 3. Section 4 extends the concept of visibility to covisibilities, which allows even more sophisticated usage for evaluation purposes. Again, a small case study is presented. Finally, Sec. 5 gives an overview on related work and points out issues for future research.

2. Visibility on the web

2.1. Public attention and online visibility

We aim at measuring public attention on certain topics. We call a topic anything that can draw public attention on itself, ranging from typical discussion group topics like 'climate policy' to persons like 'George Bush' or even something basic like 'christmas'. We restrict our view on topics that are expressible by some kind of simple term, i. e. complex topics like 'US foreign policy during the cold war and its impacts on the German economy' are ignored (at least for the scope of this paper). For measuring public attention

we introduce the idea of public visibility, which denotes a topic's presence in media. For example, a topic being on the front page of important newspapers and discussed on many channels on TV would have a high public visibility (at a certain date). In other words: it would be a topic that people are currently talking about, concerned about and willing to spend their attention on. Manually analyzing daily news while trying to monitor public visibility of topics across different media has already been done by social scientists, e. g. [6].

What we strive to do is to monitor public visibility of topics on the web, i. e. somehow automate the process of measuring visibility. In [8] a quite simple measurement for web visibility has been introduced: the hit count of a search engine[1] for the topic's search term gives it's visibility on the web. Monitoring a topic's visibility over time we get time series, for an example refer to Fig. 1 showing the visibility for the topics 'Weihnachten' (christmas) and 'Fasching' (carnival) from Dec. 28, 2004 to Jan. 19, 2005 (all analysis for this paper was done in German using the Google Web API: http://www.google.com/apis/). This example clearly demonstrates how events from the real world leave their traces on the internet: Especially the rapid decrease in visibility of 25% for the topic 'Weihnachten' was not expected, for often web pages are created for a certain event but not necessarily removed afterwards. The continuous growth of the web (see e. g. [9]) suggested that most of the webpages are kept, which makes this finding not as trivial as may seem.[2]

In addition to christmas and carnival, in [8] another example on the Kyoto protocol enforces the conclusion that real world events concerning a topic change that topic's visibility on the web.

2.2. Visibility for public relations and marketing evaluation

Evaluation is always concerned with an analysis of actions undertaken and consequences these actions had, i. e. if (or to which degree) some goals were reached. The study [19] establishes a 2step model which reflects the hierachical goals of public relations: The influence of public relations expense on the company's reputation as a first step, and the effect of the company's reputation on its revenues as a second step. On the other hand, marketing (see e. g. [10] for an overview) usually is more focused on products and markets, paying less attention on reputation. We are neither experts in marketing nor in public relations, so in this paper we do not make a strict separation between the two areas. In
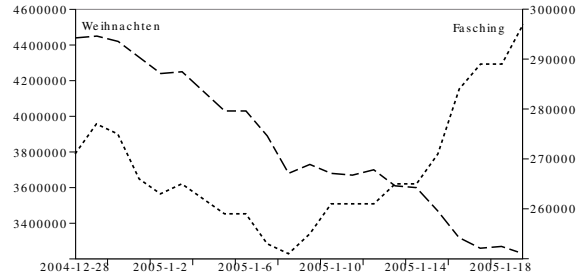

Figure 1. Estimated hit counts (Google) for 'Weihnachten' (christmas) and 'Fasching' (carnival) in time.

fact, we make the assumption that gaining high public visibility for certain topics may be a goal for both, marketing and public relations. The topics that may be of interest for a monitoring of visibilities include but are not limited to: the name of our organization/company, the name of a competitor, a product we have recently launched, a product category we concentrate on, new technology we have developed, our new funny marketing slogan, as a oneperson company (e. g. a rock star): the name of that person, as a nonprofit organization: the topic we are concerned about, e. g. an environmental topic, any other topic solely or mainly related to our organization.

Furthermore, we can divide visibility monitoring into shorttime monitoring in the context of certain events on the one hand and longtime monitoring without reference to a specific event on the other hand. A typical example for shorttime monitoring is a product launch where hopefully the product name and even the company name will gain visibility. Longtime monitoring might monitor the visibility of several competitors over some years to compare longterm shifts in public attention, in other words: the developing of 'attention shares' (somehow analogical to market shares).

We must keep in mind that high visibility does not necessarily mean high reputation, but rather sometimes even the opposite. For example, if an accident on an oil rig causes pollution, the reputation of the oil company will decrease while its visibility will surely increase. In this case, caution must be taken to interpret the results correctly. Anyway, some 'organizations' like stars from the music industry seem to be content with gathering any kind of public attention, not caring if it is good or bad reputation, or in the words of Oscar Wilde [18]: 'The only thing worse than being talked about is not being talked about.' Additionally, we are aware that high visibility or reputation does not necessarily lead to high revenues, but we leave this question to public relations experts.
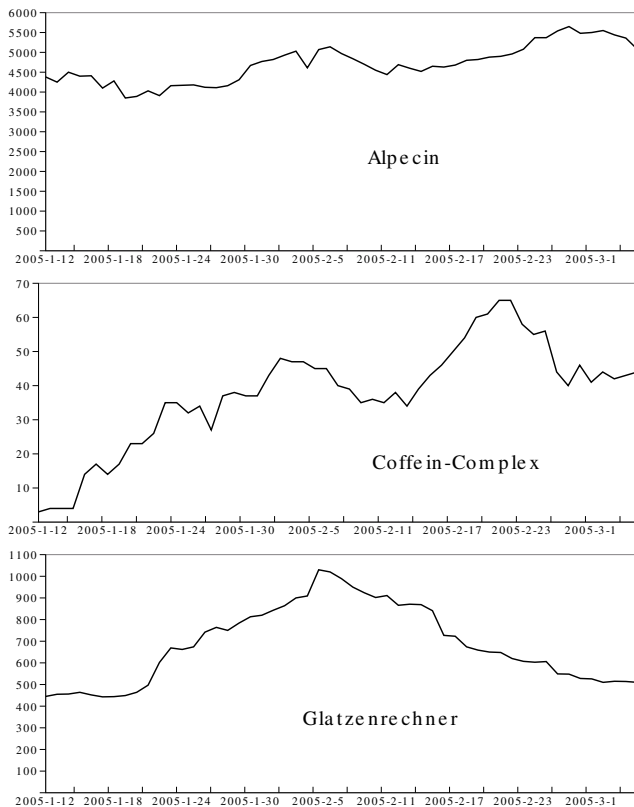
---

[1]i. e. the number of pages reported by the search engine to be found on a given search query. Note that this normally is an estimated value.

[2]Note that when doing long term studies the continuous growth of the web has to be taken into account and the hit count has to be normalized.

Figure 2. Estimated hit counts in time.

2.3. Case study: Alpecin

The following example of Alpecin indicates how visibility analysis can be used in the context of a product launch: Alpecin is the brand name of the Dr. Kurt Wolff GmbH & Co. KG, a German company from the pharmaceutical branch researching on products for hair and scalp care like shampoos or hair liquids. They concentrate on avoiding early loss of hair, e.g. with a hair liquid called 'After Shampoo Liquid' which has a special chemical compound as new ingredient, the 'CoffeinComplex'. In January 2005 there were strong marketing attempts in German media to promote this 'After Shampoo Liquid'. Commercials were emphasizing the 'CoffeinComplex' and encouraging consumers to visit the company's website and try the 'Glatzenrechner' ('balding calculator', http://www.alpecin.de/en/balding-calculator/), a web page taking input from the user (e.g.: How old are you? Is anybody in your family bald headed? Are you regularly under stress?) and returning the personal expected age of bald head.

A successful marketing campaign should draw public attention on the product and the company and therefore raise public visibility. We monitored the three topics 'CoffeinComplex', 'Glatzenrechner' and 'Alpecin' (Fig. 2)

from Jan. 12, 2005 to Mar. 5, 2005 and detected significant changes in visibility: 'CoffeinComplex' started with a visibility of 3 and increased up to 65 on Feb. 22 before going down to the level of around 43. This developing shows how a product-related term or technology that did almost not exist on the internet can gain visibility through marketing actions.

'Glatzenrechner' was already present with a visibility of 445, but more than doubled its visibility to reach a maximum of 1030 on Feb. 5 before it approached a visibility aroung 500

'Alpecin' certainly also was present on the internet and showed a slight increase from around 4000 to around 5200. Note that the peeks in the curve on Feb. 7 and Feb. 27 are similar to those of 'CoffeinComplex' with a delay of 3 to 5 days.

It is not possible to judge if the marketing compaign was successful without knowing the goals the company had and the effort that was spent on the campaign. There remain many open questions, some of them are: Can the slight rise in visibility for 'Alpecin' be seen as a success? Why does 'CoffeinComplex' seem to leave permanent traces on the internet while 'Glatzenrechner' returns to the initial level of visibility? We leave this interpretation to marketing experts and believe that for a deeper analysis a cooperation with the company would have been required.

In the following we introduce concepts for visibility analysis that are a little more complex and allow more sophisticated analysis.

3. Domain-dependent visibility

The first enhancement is domain-dependent visibility: The visibility of a topic in a domain is defined as the hit count of a search engine when searching for the string of that topic in only that domain. Google, for example, allows this with the search parameter 'site:' and then restricts its search to the domain specified, e.g. 'Klimapolitik site:greenpeace.org'. Fig. 3 illustrates the result of a search that was restricted to two domains, www.greenpeace.org and www.wwf.de. It shows the simultaneous change of visibility of one topic in different places.

In the context of public relations and marketing evaluation, domain-dependent visibility can be used in several ways: First of all it allows a geographically focused search on interesting markets by using top-level domains. Second, a focusing on special target groups is facilitated by restricting search on specialized domains, e.g. large blogs or discussion groups mainly used by kids. Third, a comparison of visibilities in several domains is potentiated, especially the comparison of technology adoption in different markets ('How does our public relations campaign work in Germany [.de], and Austria [.at]?'). An-
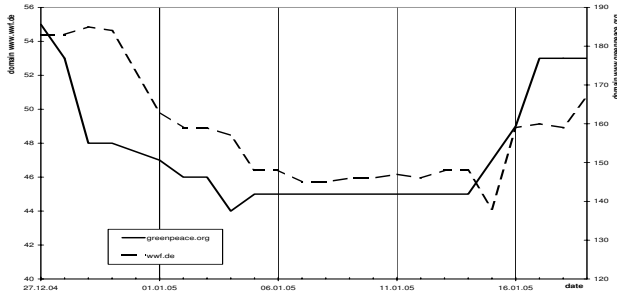
Figure 3. Estimated hit counts (Google) for 'Klimapolitik' (climate policy) on www.greenpeace.org and www.wwf.de in time.

George W. Bush   *Brasil*

2.916.000   184.000   38.416.000

George W. Bush   *Iraq*

1.870.000   1.230.000   20.970.000

Coca Cola   *make it real*

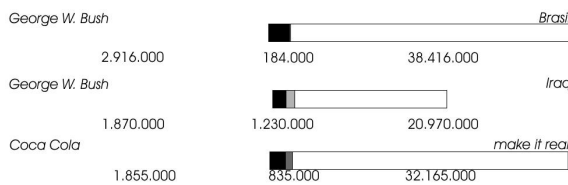1.855.000   835.000   32.165.000

Figure 4. Bar visualization of visibilities and covisibilities (August 25, 2005).

other issue solved with domain-dependent visibility is a danger that arises when a campaign like that of Alpecin is evaluated, namely that of self-created online visibility, i. e. web pages the company itself has created for its campaign. For example, if 50 of the 65 pages containing 'Coffein-Complex' were pages Alpecin had created for their marketing campaign, this would tamper the results. The self-created visibility in the company's domain has to be subtracted from the overall visibility. In our use case of Alpecin the visibilities in the domain www.alpecin.de remained constant and small ('CoffeinComplex'=2, 'Glatzenrechner'=4, 'Alpecin'=24), so in this case the self-created visibility did not play an important role. Nevertheless, a deliberate proceeding is still needed to handle the danger of self-created visibility, because some text advertisement might have been made in other domains. In this case, these pages have to be subtracted as well, leading to the question 'Where did we make online advertisement containing our search strings?'. To handle this, the online advertisement should contain a recognizable unique string, so we can, for example, take the visibility of 'CoffeinComplex' and subtract the visibility of 'Bothered of hair loss? Use Alpecin's CoffeinComplex!'

4. Covisibility

In this section, we try to measure the dependencies between different topics and introduce the measure of covis-

ibility of two topics[3] based on cooccurrence: Two topics cooccurring in a large number of documents should have *something* in common, whatever this 'something' is. In the context of this paper, we define the covisibility of two topics as the hit count of a search engine when searching for documents that contain the strings of both topics, e. g. searching for 'Alpecin AND CoffeinComplex'. We expect, for example, the covisibility of 'George W. Bush' and 'Iraq' to be higher than that of 'George W. Bush' and 'Brasil', because the former are more often discussed together than the latter ones. Figure 4 shows this example in a bar visualization: The numbers in the center are the covisibilities, the numbers on the left and on the right are *not* the visibilities, but the number of pages that contain *only* one topic. The visibility can be calculated by adding the center number to that on the left (or on the right respectively). Obviously, Brasil has the highest visibility while Bush has a quite low visibility, but the covisibility of Bush and Iraq is higher than that of Bush and Brasil.

4.1. Covisibility for public relations and marketing evaluation

Covisibilities, if applied deliberately, allow several kinds of detailed (at least more detailed than using only visibilities) analysis for public relations and marketing evaluation. The following listing of covisibilities gives an impression of its usefulness:

The covisibility of the name of our company with a product we concentrate on. This number is much more meaningful than the simple visibility of that product, because we are especially interested in our share of attention according to that product.

Accordingly the covisibility of our competitors with a product to compare the public covisibility. Which company is more associated with the new product/new technology in public opinion?

The covisibility of a company's name with a marketing slogan. It is often the goal of public relations to connect the slogan with the company in the mind of people. A high web covisibility is a first step to this connection. Figure 4 shows the covisibility of 'Coca Cola' and 'make it real'.

The uncoupling of a unique slogan, e. g. a neologism, from our company in the course of time. When we first launch our slogan, almost all pages containing the slogan will also refer our company. Now we can monitor, how the slogan slowly diffuses into the net and becomes common sense, for other people will like our slogan and use it for their purpose.

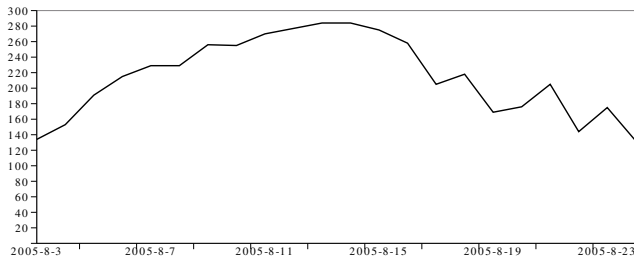[3]we restrict ourselves to two topics, generalization for three or more topics is obvious.

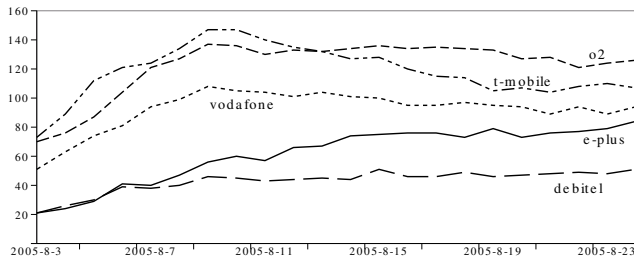**Figure 5. Visibility for 'handyflatrate' on the domain 'de' in time.**

Figure 6. Covisibilities for different carriers and 'handyflatrate' on the domain 'de'.

4.2. Case study: Flatrate for mobiles

In August 2005 all German carriers in the mobile market started offering contracts with flatrate for the first time, i. e. 'pay a constant amount of money per month and phone as long as you want'. This concept is called 'handyflatrate' (note that the German term for a mobile phone is 'handy'). Figure 5 shows the visibility of 'handyflatrate' from Aug. 3 to Aug. 24: It doubled in the beginning and then returned to around 150, so our guess is that this new product handyflatrate did not settle in public attention for a long time. All these monitorings were restricted on the domain 'de' to focus on the German market.

In a second step we try to distinguish between the main German carriers (T-Mobile, Vodafone, O2, E-Plus and Debitel) and use covisibilities of these carriers with 'handyflatrate', refer to Fig. 6 for the results: First of all, we see that all carriers gained covisibility with 'handyflatrate'. Secondly, it is obvious that the three biggest carriers generally had higher values. This is not very surprising, but if we take a closer look, we find the curve of E-Plus to grow steadily and almost reach that of the very big carrier Vodafone, while all other carriers settle on their level or even decrease like the biggest carrier T-Mobile.

Figure 7 fosters this finding: It compares the relative covisibilities ('attention shares') for carrier plus 'handy' and carrier plus 'handyflatrate' on Aug. 25. Note that we could
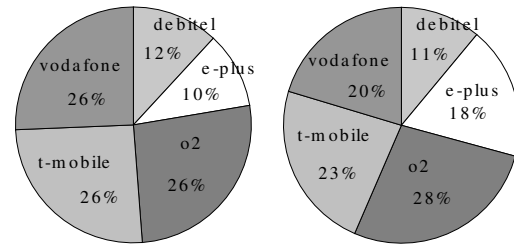

Figure 7. Percentual covisibilities of carriers and 'handy' (left), carriers and 'handyflatrate' (right) on the domain 'de' on Aug. 25.

not use pure carrier visibilities because of the special string 'O2' (we do not want to count pages related to oxygen). This diagram shows that E-Plus could increase their visibility in the field of handyflatrates, compared to the overall visibility of the company itself.

There are two possible interpretations: Either we claim that E-Plus somehow managed to impress people more with their product, leading to more public attention and visibility, or we might argue that E-Plus continued with their marketing campaign for a longer time, while the other carriers turned towards other subjects. At this point we again cannot decide about the correct interpretation without knowing about the goals the companys had. However, this example illustrates how covisibilities can be used to link products with companies to analyze a market with different competitors under the aspect of public attention.

5. Related Work and Outlook

The aim of our paper is supported by [13], another article from the wide field of public relations evaluation. Here, the authors suggest public relations practitioners not to overestimate the behavioural effects of campaigns and to set more moderate and alternative goals: 'instead of measuring the direct influence on attitudes, evaluation should also incorporate indirect aspects such as the stimulation of interpersonal discussion'. That is exactly what our method allows: To measure the amount of interpersonal discussions on the internet, especially when measuring visibility in discussion groups. They furthermore claim that 'evaluation of PR should include an element of environmental monitoring of social, political and economic factors', which also can be facilitated by using visibility analysis with adequate topics.

These environmental factors are usually not standalone topics, but rather strongly semantically related. The company-related topics about products and technology often are semantically related as well. In [8] it was shown that often it is not enough to monitor one topic, but rather a whole field of semantically related topics. For these cases a

method to adjust visibilities using semantic networks and Spreading Activation to mostly comply with an intuitive view of visibilities has been proposed.

In this paper we treated the web as an unstructured collection of documents by using simple hit counts as visibility. More sophisticated visibility measures that take link structures between documents into account should be developed. The most wellknown measure in this context, that may possibly be integrated, is Google's PageRank ([1]). Additionally, special measures for discussion groups are suggested. This is already done by the approach of communication oriented modeling (COM) [12], which investigates largescale communication processes with message/reference-networks.

Another field of research called diffusion of technology (see [14] for a comprehensive reading) is concerned with explaining how innovations find their way into social systems. The diffusion of an innovation is often measured by counting the number of purchased products and usually follows an Sshaped pattern. Comparing the Scurve of an innovation with the developing of its visibility would be a quite interesting task for future research.

The usage of search engine hit counts is not new in the field of Web Mining. Search engine queries were also used by [17] for an automatical detection of synonyms and by [11] for the validation of questionanswering systems, which both are further areas of application for covisibilities. Another possible application of covisibilities is the automatic extraction of facts. This has recently been done by Etzioni et al. who used hit counts from a search engine for their system called KnowItAll to automatically extract facts from the WWW [2, 3, 4]. Furthermore, they used hit counts to improve the precision of an information extraction system [15]. What none of these Web Mining approaches addressed are visibilities in relation to realworld events.

We are currently monitoring visibilities of topics related to the September 2005 election in Germany. With that study we try to expand the method of visibility analysis to other fields to get a deeper understanding of reasonable applications.

Our main concern for future work is the missing cooperation with experts from other fields, e. g. the field of public relations. We would highly appreciate the cooperative conduction of a case study that is running on a largescale, for long time, covering all aspects of the analyzed field and hope to be able to carry out a deeper evaluation of the future results.

References

[1] S. Brin and L. Page. The anatomy of a largescale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[2] O. Etzioni, D. Cafarella, D. Downey, S. Kok, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. Webscale information extraction in knowitall (preliminary results). In *Proceedings of the 13th Intl. World Wide Web Conference*, 2004.

[3] O. Etzioni, D. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. Methods for domain-independent information extraction from the web: an experimental comparison. In *Proceedings of the 19th National Conference on Artificial Intelligence*, 2004.

[4] O. Etzioni, D. Cafarella, D. Downey, A.M. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. Unsupervised namedentity extraction from the web: An experimental study. *Artificial Intelligence*, 2005.

[5] J. Falkinger. Attention economies. CESIFO WORKING PAPER NO. 1079, ifo Institut für Wirtschaftsforschung, München, 2003.

[6] H. J. Gans. *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, 'Newsweek' and 'Time'*. Northwestern University Press, Evanston, IL, 25th anniversary edition, 2005.

[7] L. C. Hon. Demonstrating effectiveness in public relations: Goals, objectives, and evaluation. *Journal of Public Relations Research*, 10(2):103–135, 1998.

[8] P. Kiefer, K. Stein, and C. Schlieder. Visibility analysis on the web using covisibilities and semantic networks. In *Proceedings of the European Web Mining Forum (EWMF)*, Porto, Portugal, October 2005. accepted.

[9] J. Kleinberg and S. Lawrence. The structure of the web. *Science*, 294:1849–1850, 2001.

[10] P. Kotler and G. Armstrong. *Principles of Marketing*. Prentice Hall, Englewood Cliffs, New Jersey, USA, seventh edition, 1996.

[11] B. Magnini, M. Negri, and H. Tanev. Is it the right answer? exploiting web redundancy for answer validation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 425Ű–432, 2002.

[12] T. Malsch and C. Schlieder. Communication without agents? from agentoriented to communicationoriented modeling. In *Regulated Agent-Based Social Systems: First International Workshop, RASTA 2002, Bologna, Italy, July 16*, pages 113–133. SpringerVerlag, Berlin, Heidelberg, New York, 2004.

[13] M. McCoy and O. Hargie. Implications of mass communication theory for assymetric public relations evaluation. *Journal of Communication Management*, 7(4):304–316, 2003.

[14] E. Rogers. *Diffusion of Innovations*. The Free Press, New York, USA, fourth edition, 1995.

[15] S. Soderland, O. Etzioni, T. Shaked, and D. Weld. The use of webbased statistics to validate information extraction. In *Papers from the AAAI-2004 Workshop on Adaptive Text Extraction and Mining(ATEM-2004)*, San Jose, CA, 2004.

[16] M. Spiliopoulou. Web usage mining for web site evaluation. In *Communications of the ACM*, volume 43(8), pages 127–134, 2000.

[17] P. Turney. Mining the web for synonyms: Pmiir versus lsa on toefl. In *Proceedings of ECML2001*, pages 491Ű–502, Freiburg, Germany, 2001.

[18] O. Wilde. *The Picture of Dorian Gray*. Modern Library, Modern Library Paperback edition, 1998.

[19] K. Yungwook. Measuring the economic value of public relations. *Journal of Public Relations Research*, 13(1):3–26, 2001.