Guidance Performance Indicator - Web Metrics for Information Driven Web Sites

Carsten Stolz
Kath. Universität
Eichstätt-Ingolstadt, Germany
carsten.stolz@ku-eichstaett.de

Maximilian Viermetz
Heinrich-Heine-Universität
Düsseldorf Germany
maximilian@viermetz.net

Michal Skubacz
Siemens AG, Corp. Technology
Munich, Germany
michal.skubacz@siemens.com

Ralph Neuneier
Siemens Corporate Research
Princeton, USA
ralph.neuneier@siemens.com

Abstract

For the evaluation of web sites a multitude of metrics are available. Apart from general statistical measures, success metrics reflect the degree to which a web site achieves its defined objectives. Particularly metrics for e-commerce sites based on transaction analysis are commonly available and well understood. In contrast to transaction based sites, the success of web sites geared toward information delivery is harder to quantify since there is no direct feedback of user intent. User feedback is only directly available on transactional web sites[1].
We introduce a metric to measure the success of an information driven web site in meeting its objective to deliver the desired information in a timely and usable fashion. We propose to assign a value to each click based on the type of transition, duration and semantic distance. These values are then combined into a scoring model describing the success of a web site in meeting its objectives. The resulting metric is introduced as the GPI and its applicability shown on a large corporate web site.

1 Introduction

The fast development of the internet has established it as a new medium for private and business communications and transactions. Private as well as corporate entities use the internet as a platform for presentation, communication, sales transaction, entertainment, service and information. Almost every company maintains a web presence and invests considerable amounts of time and money into increasing the attractiveness of their web site. The profitability of this investment is most easily measurable by the monetary value of the transactions handled by the web site. Many metrics have been developed for such web sites [7] [3]. But these metrics are not directly applicable for web sites without sales or purchase activity. This suggests a development of new metrics making the success of an information driven web site measurable.

The challange in developing a new metrics rises from the main difference between transactional and information driven web sites the user feedback. By purchasing a product on a web site, the user gives feedback about the web site, the and the whole purchasing process. He does not stay anonymous and provides personal billing data, making him identifiable when returning to the web site. In contrast to that, a user on an information driven web site stays anonymous and it remains uncertain whether he was interested in the content he visited. In the following we will present an approach to overcome this gap, starting with the objective of information driven web sites.

The objective of information driven web sites is to provide the user with desired information. First we have to find out when we can judge a user to have found what he or she was looking for. The objective of these sites is to guide the user to the desired content pages. Navigation pages support the user in his search.

We analyze the transition between navigation and content pages as well as within both page categories. We de-

 IEEE/WIC/ACM International Conference on Web Intelligence 2005, Proceedings, pages 186–192

fine the objective of information driven web sites to lead the user to content pages as fast as possible and hold her there as long as possible. Therefore we use the time that the user has stayed on a web page as a factor in the metric. We also consider the web page topic in order to incorporate user interest into the metric.

The **contribution** of this work is to combine transitions between page types, duration and topic mixture in order to evaluate the positive or negative contribution to the objective of a web site. On her way through the web site the user collects positive and negative scores. Beyond the transition described by a single click we will extend our analysis to a transition composed of several consecutive clicks. This increase in scope will give the scoring model additional accuracy. This approach to generating the metric can be considered generic and is applicable to all kinds of web sites focused on providing information.

We will first give an **overview** of existing web metrics and point out their inability to accurately measure the success of information driven web sites, since existing web metrics focus on transaction based web sites. In our approach we will address this limitation. Combining the available structure, content and usage data of a web site, we develop a scoring model and describe its applications. By extending the model from clicksequences of two to more clicks, we create an extended model that allows better interpretation and application. In a case study we use real world data gained from a corporate web site for evaluation.

2 Metrics for Web Sites

The following sections will provide an overview of existing metrics and measurements for web site success. Before web measures can be developed, the objectives of a web site have to be specified.

2.1 Web Site Objectives

Regarding corporate web sites, the business models range from:

- sell or purchase,
- inform,
- entertain or
- communicate

via the internet. The objectives of a web site can be derived from its business model, i.e. increase sales or improve usability or attractiveness. Consecutively, measures have to be defined to quantify or qualify the achievement of the respective objectives. A general overview of web metrics is provided by Dhyani [7], NetGenesis [11] and Schwickert[15]. The most recent study about existing web metrics as been performed by Calero in 2004 [3]. Measuring and analyzing web sites one can distinguish between *structure*, *content* and *usage* data and their specific measures, as Cooley proposed in [6].

2.2 Web Structure Measurement

Dhyani describes Graph Property Measures like **centrality** or **compactness** in [7], which describe the position and relation of one web page in comparison to the whole web site or **link similarity** to compare it with other web pages directly. Calero mentions in [3] other measures, for example **depth**, **breadth**, **in-out-link count** or **prestige**. These can be applied to improve the web site structure and navigability.

2.3 Web Content Measurement

Evaluating the composition and formatting of web pages Ivory et al. evaluated in [9] expert-rated design metrics. The information and semantics provide valuable insight into a web site. In [17] we analyzed web content to identify topics on web sites. This is a prerequisite to understand what a user is interested in, not yet a metric by itself.

2.4 Web User and Usage Measurement

The measures for content and structure by themselves cannot confidently measure the success of a web site. The users determine the success of a web site. Apart from the basic usage measures **page impression, page view or event, sessions, unique users and duration**[13], Dhyani et al. [7] describe **frequency** and **recency** as usage metrics. All above mentioned measures are used for statistical analysis but can not exhaustively measure the success of a web site, since the users intention is not considered. The problem to identify user intention is described by Pather [12] as a conceptual study to evaluate user satisfaction by gap measurement, but no metrics are developed. Like Buys and Brown in [2] user enquiry is used instead of discovering user perception by analyzing user behavior.

The user intention and by extention the web site's success can be easily measured if the user completes a transaction on an e-commerce site. On transaction based web sites a variety of success measures have been developed.

2.4.1 Transaction based Success Measures

The users declare their intentions as soon as they are willing to pay for a product and purchases it, revealing also the monetary value or utility. If a purchase is conducted on

a web site, this makes the success measurement straightforward and allows deeper analysis of the whole purchase process.

NetGenesis [11] and Schwickert [15] show how different measures can describe the customer lifecycle. From **reach**, **acquisition**, **conversion** and **retention** a customer can reach **loyalty** status. Other metrics like **abandonment**, **attrition** and **churn** describe migration of users. Schonberg et al. [14] describe **clickthrough** and **look-to-buy** metrics.

2.4.2 Information Provision based Success Measures

The absence of transactions makes it challenging to discover the user's intention and satisfaction. Most approaches combine structural, content and usage data to uncover the user intention. Heer and Chi propose in [8] a way to discover user interest by clustering usage and content data. Heer and Chi continued their research on web site usability in [8] and [5]. Like we did in [17][18] Jin et al. [10] analyze and compare usage data and content in order to discover hidden semantic information. Barnard and Wesson outline three measurements to evaluate the usability of a web site [1], **effectiveness** for the user determining her task, **efficiency** measures the time it took the user to complete her task and **satisfaction**, inquired from the user by questionnaire. Chi combines in [4] usage and content data to predict user action and evaluate web site usability. Their objective is to simulate hypothetical user behavior. In contrast to our approach they create user agents with given information goals, whereas we evaluate user action in combination with web site structure and content in order to estimate whether users accomplished their target.

In [18] we compare the user's intention and web site perception with the intentions of the web author identifying inconsistencies between both user and web author. This metric regards the whole web site and all users as a whole, resulting in an indicator for improvements in web site design. This allows a qualitative but not success oriented judgment about a web site.

Summarizing the existing success based metrics for information driven web sites, the utility of a new quantification of the web site's success is considerable.

3 A Metric for non-e-commerce Web Sites

Motivated to discover user perception of a web site, we present an approach that combines user behavior, site content and structure while also harnessing user feedback. Developing a measurement for information driven web presences, we must first determine the objective of the web site. Afterwards we identify available information that can contribute to a good estimation of a successfull user session.

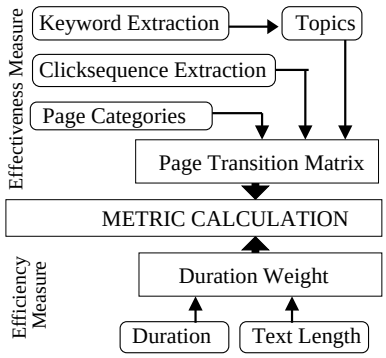

Figure 1. Metric Building Process

The **Objective** of a non-e-commerce web site is not the facilitation of client transactions but rather providing information to users. This objective is reached when a user has found the piece of information he or she was looking for.

In 3.1 and 3.2 we will explore whether a user has found the sought after piece of information. We will then estimate if the user was interested in the piece of information offered by the web site.

3.1 Effectiveness Measure

The effectiveness of an information driven web site depends on successfully leading the user to pages providing the sought after content. We distinguish between pages helping the user to find information and navigating the web site, particularly navigation pages, as opposed to content pages, considered to be sources of information.

3.1.1 Web Page Categories

Information can be found on pages filled with content. Other pages are used to structure the web site and allow easy and intuitive navigation. We distinguish between **navigation and content pages** and evaluate the transitions between both categories creating a matrix of every possible category transition.

In the following we combine these aspects into a new metric and explain how this approach helps to estimate the degree of fullfillment of a web site's purpose. Since the intended target page is reachable via navigational pages, the transition from navigation pages to content pages can generally be regarded as successfull for a web site's objective. An overview of web page category transitions is depicted in table 1, rows showing the source page of a clicksequence and columns the target page.

The pages dedicated to guiding the user through the web site are *Home and Search Page, Sitemap*. The content pages represent the major part of a web site and need further attention by identifying topics of the content pages and the topic

Table 1. Page Category Transition Rating

		Destination (2nd click)				
		Home	Sitemap	Search	Content	Session End
Start (1st click)	Home	−	−	−	+	−
	Sitemap	−	−	−	+	−
	Search	−	−	−	+	−
	Content	−	−	−	+	0

transition per click in order to quantify successful transitions. Due to the fact that the last click can not be evaluated (see section 3.2.2), we assign a value of 0 to all clicksequences *page* → *end*.

3.1.2 Web Page Topics

Since only few pages are designed for the Semantic Web, we assume to have to analyze the text body of each web page and extract key words, apply stopword lists and stemming. We performed different clustering techniques to identify content groups building topics. We have evaluated hierarchichal and Kmeans clustering as well as maximization of likelihood by EM-Algorithm. The quality of the results have been evaluated in detail in [17, 18]. We calculate a probability for each page of belonging to each topic, resulting in a topic vector for each web page. Finally we calculate distance between all web pages based on content similarity.

The user behavior depends on his interest. Some users are interested in one topic, others want to have an overview over the whole web site. We consider the user type by comparing each topic transition with the topic mixture of the user session. The resulting matrix measures the degree of change in content allowing an estimate of the content the user is interested in.

3.2 Efficiency Measure

Having a measure for the effectiveness, measures for efficiency will focus on evaluating the time it took the user to complete her task. A web site is efficient if the user is guided to the content quickly, demanding a short duration on navigation pages.

3.2.1 Duration

This indicator is the time the user stays on each web page. Table 2 assigns bonus values to long durations on content pages and negative penalty values to navigational pages and vice versa for short durations.

Table 2. Web Page Duration Rating

	short duration	long duration
Navigation	+	−
Content	−	+

The duration depends not only on the user, but also on the text length. In order to make all durations comparable we normalize all values as described in the following section 3.4.

3.2.2 Limitations

The duration of the last click in a user session is not measurable since we do not receive any further user actions after the last click.

We can measure the time between the user's actions, which means clicks, but it is technically not possible to measure either the amount of attention the user payed to the web page content or whether he read the content at all. It is generally only possible to achieve this with any accuracy under controlled conditions.

In case of tabbed browsing, allowed by a newer generation of browsers, the duration and sequence does not necessarily resemble the sequence and duration the user is reading the content.

3.3 Extending the scope of transitions

Since we consider the transition from navigation to content pages to be crucial for a web page's success, we will now focus on these transitions. So far we have evaluated each click or page sequence on its own. The success of a sequence Navigation → Content can be more accurately judged by taking the next click into account. Extending the transition from two to three pages allows an evaluation of user feedback to the interestingness and relevance of the offered content of the first evaluated click from the navigation page to the content page.

For the transition type Nav → Cont we extend the second degree scoring model described in sections 3.1 to 3.2 to a third degree model. The second degree model assigns positive values to the nav-cont-transition. But it is not considered whether the user was guided to the desired content. Hence the user feedback has to be considered. It can be estimated by analyzing the duration how long the user looked at the offered content in combination with the subsequent page type.

The following example shows the difference between the second degree and the third degree model. The user has reached a content page via the search page. Now he stays

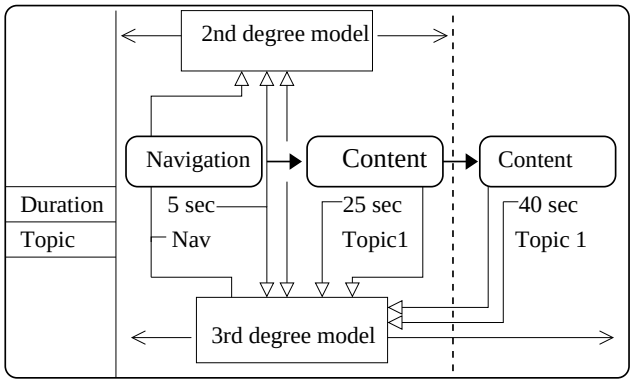

Figure 2. Metric Construction

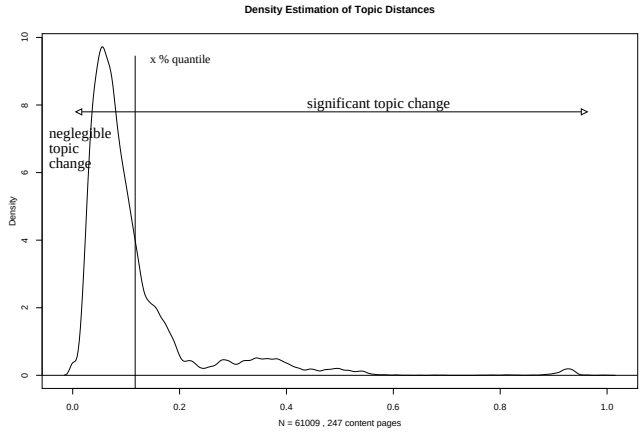

Figure 3. Weight Determination

long and proceeds to another content page. This is considered positive, since the user has been led to a content page and kept there for a longer time, allowing the estimation that he was interested in the content.

A change in this scenario leads to a different judgment about the web sites' success. If the user returned from the content page after a short time to the search page, the offered content page did not show any desired information.

As depicted in figure 2, the third degree model includes information of the user action (click) following the evaluated click. First the duration on the content page where the user has been led to and second the following page type. The revised scoring model regards the following scenarios:

positive: Nav → Cont (normal duration)→ subsequent pages . . .

negative: Nav → Cont (short duration) → subsequent pages . . .

If the user was actually interested in the content, he will respectively stay longer. Like Spool [16] points out, users skim web pages first and recognize undesired content very quickly. The user will leave the page and the duration will be short.

So far the third degree model allows to integrate user feedback on one specific scenario. A general consideration of transitions over more pages should make the model more accurate.

3.4 Building the Metric

Effectiveness Measures We beginn with the calculation of the effectiveness measures described in 3.1. First we extract the textual information from the web pages with the help of a crawler. After cleaning the data we determine the number of topic clusters and apply standard clustering methods to identify groups of related web pages.

Now we have all the information needed to construct the transition matrix assigning values to each possible click. The basic value for each click is derived from the categories of the source and destination pages, as described in table 1. The multiplication factors are directly derived from the transition table.

Definition 1 (Transition Type) *Let τ be a transition within the web site and T the associated transition table. The transition type χ_τ is defined as*

$$\chi_\tau = 1 * sign(T_{i,j}) \tag{1}$$

where i is the source type of τ, and j the destination page type.

The next aspect we want to capture in the metric is the type of transitions occurring between web pages bearing content. We want to emphasize transitions which stay within a topic, and deemphasize transitions between topic areas. Having assigned values to each page category transition, the transitions between content pages are evaluated in greater detail. We calculate the distance between all possible content page transitions based on the comparison of both topics. In other words we calculate a similarity measure by comparing the topic affiliation of all pages.

Normalization Instead of using continuous values as gained from the distance between individual pages, we determine a transition to be significant or not. A significant transition distance would signify a change in topic, while an insignificant shift would leave the user perusing the same general topic area. Since we want a web site to guide a user quickly to the content he desires, we introduce a characterization of this topic change.

Definition 2 (Transition Weight) *Let τ be a transition between content pages. The effectiveness of a transition is captured by an assigned weight factor μ_τ.*

$$\mu_\tau = \begin{cases} negligible\, topic\, shift & \Rightarrow 1 < \mu_\tau \\ significant\, topic\, shift & \Rightarrow 0 < \mu_\tau < 1 \end{cases} \quad (2)$$

The evaluation of the degree of topic change is performed by regarding all occurring topic transitions and comparing them with the topic transition in focus, as can be seen in figure 3. Any change larger than x is considered significant, while any change smaller would leave the user within the to same general topic.

Efficiency Measures The necessary information for the efficiency measures, specifically duration and text length, have been collected along with the other data. The time necessary to perceive, read and understand textual information depends apart from design issues mainly on the length of the text. We characterize the interest of a user in a given content page by calculating the time spent per word. By contrasting this value to the average time spent on all words for all users, we can get a feel for where a user's focus is acknowledged, and where not.

Definition 3 (Efficiency Factor) *Let τ be a transition within the web site. We have the duration d_τ and the text length l_τ of the source page. We also maintain the global average of all users reading a word δ_{word}:*

$$\delta_{word} = \frac{\sum\limits_{\forall\tau} d_\tau}{\sum\limits_{\forall\tau} l_\tau} \quad (3)$$

The efficiency ϕ of a transition τ can now be characterized with respect to δ_{word}

for $content \xrightarrow{\tau} content$:

$$\phi_\tau = \begin{cases} 0 < x < 1 & \delta_{word} > \delta^\tau_{word} \\ 1 < x & \delta_{word} < \delta^\tau_{word} \end{cases} \quad (4)$$

for $navigation \xrightarrow{\tau} content$:

$$\phi_\tau = \begin{cases} 0 < x << 1 & , if\ \delta_{word} < \delta^\tau_{word} \\ 1 << x & , if\ \delta_{word} > \delta^\tau_{word} \end{cases} \quad (5)$$

With the help of the defined characteristics, we can now construct our metric. Since the metric reflects how well a web site guides a user to the content he seeks, we refer to the metric as the Guidance Performance Indicator.

Definition 4 (Guidance Performance Indicator) *We determine the effectiveness of a session σ with $\sigma = \{\tau_1, \tau_2, \ldots \tau_m\}$ by combining the transition type χ_τ, transition weight μ_τ and transition efficiency ϕ_τ for each transition $\tau \in \sigma$.*

$$GPI_\sigma = \sum_{i=1,\tau_i\in\sigma}^{|\sigma|} \chi_{\tau_i} * \mu_{\tau_i} * \phi_{\tau_i} \quad (6)$$

Table 3. GPI applied to example pages

	GPI	clicks	GPI / clicks
simple Search	-350,5	443	-0.79
Advanced Search	14,9	75	0.20
Products: Leasing	775,5	1193	0,65

4 Evaluation

We analyse usage data from a corporate web site dealing with financial products and services over a time period of one month, consisting of 247 content pages and 8 navigation pages with 10290 page impressions or clicks and 4301 sessions. We have identified 10 topics.

Web Page Assessment One application of the GPI is to evaluate and compare web pages. We regard two navigation pages and one content page as comparison, since one GPI value of its own reveals no user feedback. Both navigation pages are search pages, namely the simple search and the advanced search offering more selection possibilities.

Table 3 shows the absolute GPI values. In order to compare the GPI values of pages with different traffic density, we contrast their average GPIs per click. The negative average GPI per click for the simple search page can have from several reasons. Users need more time to find a relevant search result within all search results or the search leads to information not of interest to the user. Whereas the Advanced Search reached a positive GPI by leading users faster to desired information. This is a reasonable result, since more selection criteria allow better search results. But 443 times the simple search was used compared to 75 times usage of advanced search.

By making the advanced search more prominent on the web page, the development of the GPI over time would give valuable hints to the effectiveness of this design decision.

Session Assessment Table 4 shows two sample sessions. Both users have each accessed a search page and have been directed to a content page. According to the GPI the web site provided better guidance in the first instance, session It took the second user (who was accessing the simple search page) two clicks to reach a content page, which he left after a short duration and retreated to the home page. The negative GPI reflects the assumption that this user did not find what he was looking for.

Extended Scoring Model Comparing the basic model with the extended model in table 4, we can see that the first session is not measured differently. The extended model assigned a much lower GPI to the second click in the session (search $\rightarrow$ content), because the user did not stay long on

Table 4. Example Sessions evaluated by GPI

	GPI_1	GPI_2	GPI_3	total
Basic Model GPI				
Home→Adv. Search→ Content → Content	-1.1	1.7	1.5	2.1
Search → Search → Content → Home	-1.2	1.44	-1.4	-1.16
Extended Model GPI				
Home→Adv. Search→ Content → Content	-1.1	1.7	1.5	2.1
Search → Search → Content → Home	-1.2	-1.44	-1.4	-4.04

the content page. The extended model penalizes the misleading search results.

From 882 possible transitions falling into the category "Nav → Cont → . . . " 164 transition weights or 18% have been recognized as misleading by the extended scoring model. The average click value of 0.66 was slightly reduced to 0.65.

In order to evaluate this extended model, a longer evaluation period is necessary as well as an empirical user enquiry. We are working on a more accurate model, using a Markov chain which regards longer click sequences and user scenarios.

5 Conclusion

Though there are numerous general metrics to measure web site usability and success available today, those including user feedback are limited to analyzing transaction based e-commerce web sites.

Therefore we have introduced a new metric geared towards assessing the success of information driven web sites. By modeling desired user behavior patterns, the metric assigns positive as well as negative values according to the perceived success of a user session. We approach the challenge of making user feedback available without enquiring the user directly, by analyzing her behavior and visited content.

With help of the GPI metric a web site editor can discover important elements in the website structure and content, which influence user behavior. For example attractive content presentation, positive web site design or misleading navigation. Monitoring the development of the metric over time can reveal user acceptance of the web site and reactions to changes in content and design.

The GPI can be applied to all kind of web sites without and can be adjusted to specific web site characteristics by fine tuning of the transition matrix.

References

[1] L. Barnard and J. L. Wesson. Usability issues for e-commerce in south africa: an empirical investigation. *SAICSIT '03*, pages 258–267, 2003.

[2] M. Buys and I. Brown. Customer satisfaction with internet banking web sites: an empirical test and validation of a measuring instrument. *SAICSIT '04: Proc. Conf. South Africa IT research in developing countries*, pages 44–52, 2004.

[3] C. Calero, J. Ruiz, and M. Piattini. A web metrics survey using wqm. In *Web Engineering, ICWE 2004, Munich, Proc.*, pages 147–160. Springer, 2004.

[4] E. H. Chi, P. Pirolli, and J. Pitkow. The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a web site. *CHI '00: Proc. SIGCHI*, pages 161–168, 2000.

[5] E. H. Chi, A. Rosien, and G. S. et al. The bloodhound project: automating discovery of web usability issues using the infoscentπ simulator. *CHI '03: Proc.*, pages 505–512, 2003.

[6] R. Cooley. The use of web structure and content to identify subjectively interesting web usage patterns. *ACM Transaction on Internet Technology*, 3(2):93–116, May 2003.

[7] D. Dhyani, W. Keong NG, and S. Bhowmick. A survey of web metrics. *ACM Computing Surveys*, 34(4):469–503, December 2002.

[8] J. Heer and E. Chi. Separating the swarm: Categorization methods for user sessions on the web. *ACM*, 2002.

[9] M. Y. Ivory, R. R. Sinha, and M. A. Hearst. Empirically validated web page design metrics. *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 53–60, 2001.

[10] X. Jin, Y. Zhou, and B. Mobasher. Web usage mining based on probabilistic latent semantic analysis. *KDD '04: Proc. of ACM SIGKDD*, pages 197–205, 2004.

[11] NetGenesis. E-metrics business metrics for the new economy. www, 2000.

[12] S. Pather, G. Erwin, and D. Remenyi. Measuring e-commerce effectiveness: a conceptual model. *SAICSIT '03*, pages 143–152, 2003.

[13] J. Pitkow. Web characterization activity characterization metrics. Technical report, W3C, Zugriff: 04.05.2004 1998.

[14] E. Schonberg, T. Cofi no, and R. H. et al. Measuring success. *Commun. ACM*, 43(8):53–57, 2000.

[15] A. C. Schwickert and P. Wendt. Controlling kennzahlen fuer web sites. Arbeitspapiere WI 2, Justus-Liebig-Universitt Gieen, 8 2000.

[16] J. Spool, W. Schroeder, T. Scanlon, and C. Snyder. Web sites that work: Designing with your eyes open. *Chi 98*, pages 147–148, April 1998.

[17] C. Stolz, V. Gedov, K. Yu, R. Neuneier, and M. Skubacz. Measuring semantic relations of web sites by clustering of local context. In *Springer LNCS: Int. Conf. on Web Engineering, ICWE 2004, Munich*, pages 182–186. Springer, 2004.

[18] C. Stolz, M. Viermetz, M. Skubacz, and R. Neuneier. Improving semantic consistency of web sites by quantifying user intent. *Springer LNCS: Int. Conf on Web Engineering, ICWE 2005, Sydney*, 2005.