

Cultural Determinants of Search Behaviour on Websites

Anett Kralisch & Bettina Berendt
Humboldt University Berlin, Germany
Institute of Information Systems
{[kralisch.berendt](mailto:kralisch.berendt@wiwi.hu-berlin.de)}@wiwi.hu-berlin.de

Abstract

Our research aims to detect the impact of various cultural variables on search behaviour. We tested our hypotheses by examining users' navigation steps recorded in a large logfile of a widely frequented multilingual website. Results indicate that the users' culturally determined thinking patterns are reflected in their search behaviour, mainly by the selection of search options that are used for information retrieval.

1 Introduction

Public information offerings, such as in the field of public health care, are increasingly realised over the Internet. However, whether or not patients/consumers/citizens will favour the World Wide Web as an information source depends to a large extent on how easily they can find the appropriate information (e.g., Savolainen 1999). Research has shown how various personal and situational variables determine the ease of a search task, yet little research has been done regarding cultural determinants of information seeking behaviour.

Therefore, we examine to which extent the users' cultural backgrounds have an impact on their information seeking behaviour, in particular on the use of search engines and hyperlinks. Culture, defined as mental concepts influencing the relationships with other people, the environment and the concept of time (see Hofstede 1991, Hall 1989, Trompenaar 1995), is important for

understanding search behaviour because it is one of the most relevant aspects of a user's personal context – a factor that has been shown to be significant but under-researched in recent studies of information accessing behaviour (Rice, McCreadie, & Chang, 2001, pp. 132ff.).

Empirical data is provided from logfiles of a website in the field of public health care. Results from our analyses provide helpful instructions for the conceptualization of search services and interface design on websites in the fields of public information aiming to offer worldwide information accessibility. Users preferences for information retrieval found in our analysis can be met on a website by focusing attention in interface design on clear structures, on providing appropriate background information or highlighting search engine options – depending on the target groups and their cultural backgrounds.

Various personal and situational variables appear to affect search behaviour. One group of variables, investigated by Shneiderman (1997) and Braham (1997), is the impact of the type of search task. Other researchers, such as Scaife and Rogers (1996), focused their attention on information presentation. The most important variable appears to be the user's cognitive capacity, which has been so far examined mostly under the aspects of topic knowledge and web experience (e.g., Navarro-Prieto 1999, Vakkari 2000).

Early research on cultural determinants in the field of computer mediated communication focused on “visible” manifestations of culture such as symbols and layout (Baber and Badre 1995, Sheppard & Scholtz 1999). More recent research started to investigate less visible concepts of culture. Findings from Chau et al. (2002) illustrate how users from different countries differ in their perception of the purpose of the Internet, and consequently exhibit differences in their user behaviour and general attitudes toward the Internet. Marcus et al. (2000) examined a number of Hofstede's cultural dimensions and their possible impact on user-interface design.

Our main aim in extending previous research is to explore how cultural influences lead to variations in users' navigation behaviour, mainly information seeking behaviour, on a given website. We will focus on preferences concerning the use of search engines and alphabetically and content-organized hyperlinks. Furthermore, in contrast to previous studies that rely on laboratory experiments with a very limited number of participants, the logfiles generated by our website represent an unusually large empirical data set. The analysis of Web server logfiles ensures a high ecological validity and allows the investigation of a wide range of – frequent and infrequent –

authentic behaviours that occur in the use of a Web site and that may not be “producible” in a laboratory setting (Reips, 1999).

The paper is structured as follows: First, we will give a short introduction to Hofstede’s and his colleagues’ cultural dimensions and their possible effect on information retrieval patterns. From this, we infer hypotheses about the preferences of use of these search options depending on the users’ cultural backgrounds. The following sections describe the methodology applied in our research and present the results obtained from the data analysis. We will then discuss the outcome of our empirical analysis, illustrate the limits of a logfile analysis in the field of cultural internet research, and draw conclusions for application of this study’s results. An outlook on future research completes our article.

2 Cultural Effects on Information-Seeking Behaviour

National cultural differences were first investigated by Hofstede (1980), who, based on a large research project, developed the concept of cultural dimensions. Later on, Hall (e.g., 1989) and Kluckhohn, Strodbeck and Trompenaar (1991) and other authors extended Hofstede’s concept adding further dimensions.

For the purposes of our study we categorize the cultural dimensions and their impact into three groups: (1) differences in amount of information needed, (2) differences in time perception, and (3) differences in space perception.

All users of our website could select between the following search options: search engine, alphabetically organized links and content-organized links.

2.1. Impact of differences in the amount of information needed

We propose that divergences in the amount of information needed are due to differences concerning the two cultural dimensions context specificity, and uncertainty avoidance (UA). Context, first described by Hall (1989), refers to the amount and specificity of information in a given situation. High-context communication is where most of the meaning is in the context while very little is transmitted in the message. With low-context communication, information has to be stated explicitly otherwise the meaning is distorted (Hall 1989).

UA deals with a society's tolerance for uncertainty and ambiguity. It indicates to which extent "members of a certain culture feel threatened by unknown or uncertain situations" (Hofstede 1991).

We expect high UA and context specificity to be similar in their effects, leading to a more extensive collection of information. Since navigation through hypertext provides more context information than search engine results do, we suggest that members of low-context cultures and low-UA cultures tend to opt more for the use of search engines than members of high-context and high-UA cultures do. Browsing through an information collection, as opposed to searching, is an activity that is characterized by the inclusion of many resources into consideration, and that allows an evaluation of the various items found, thus reducing the ignorance or uncertainty one has about a topic area (Rice et al., 2001, p. 302).

A similar suggestion was made by Luna et al. (2002), who believe that hierarchical structure is preferred by high-context cultures and search-based structures by low-context cultures.

Similarly, since content-organized links provide the most context information of all search options, we suggest that this type of hyperlinks is preferred by members of high-context cultures and by members of high-UA cultures. These propositions lead us to the following hypotheses:

H1 (H2): Members of low-context cultures (low-UA cultures) have a stronger preference for search engines than members of high-context cultures (high-UA cultures).

H3 (H4): Members of high-context cultures (high-UA cultures) have a stronger preference for content-organized links than members of low-context cultures (low-UA cultures).

2.2. Impact of differences in time perception

There are two major cultural time related dimensions. Here we will only consider the aspect of long term orientation (LTO), analyzed by Hofstede. "Long Term Orientation stands for the fostering of virtues oriented towards future rewards, [...]. Its opposite pole, Short Term Orientation, stands for the fostering of virtues related to the past and present [...]" (Hofstede 2001).

As Marcus (2000) points out, in contrast to cultures with a high level of LTO, cultures with a low level of LTO will consider the immediacy of

results and achievement of goals an important issue. When searching for information, these short-term oriented cultures should therefore opt more often for the use of the search engine than long term oriented cultures do. Long-term oriented cultures, on the other hand, should exhibit stronger preferences for hyperlink navigation, which requires more patience to achieve navigational and functional goals, but also enables the browsing user to evaluate whether items will be of interest (Rice et al., 2001, p. 256) – a more long-term oriented view of the utility of information resources than a search for a specific and known item.

In a similar manner, we believe alphabetically organized links to be more preferred by short term oriented cultures than by long term oriented cultures. This is due to the fact that alphabetically organized links are perceived as offering faster access to the desired information than content-organized hyperlinks do.

According to these statements we hypothesize that:

H5: Members of cultures with a short-term orientation have a stronger preference for search engines than members of cultures with long-term orientation.

H6: Members of cultures with long-term orientation have a stronger preference for search content-organized links than members of cultures with short-term orientation.

2.3. Impact of differences in space perception

Within the context of cultural dimensions, the issue of power distance is the aspect that appears to have the most important impact on space perception. Power Distance (PD) is one of Hofstede's Dimensions and focuses on the nature of human relationships in terms of hierarchy. High-PD countries exhibit strong hierarchical systems. Low-PD countries have flatter hierarchies and tend to view subordinates and supervisors as closer together (Marcus 2000).

As indicated by Marcus, members of high-PD cultures are characterized by taller hierarchies in their mental models and by a higher ability and desire to structure their access to information (Marcus 2000). In the context of information seeking behaviour, users from low-PD cultures should tend to avoid following links through a deep hyperstructure and therefore favour the use of search engines (and vice versa for members of high-PD cultures).

Thus, we formulate the following hypothesis:

H7: Members of high-PD cultures have a stronger preference for hyperlink search than members of low-PD cultures.

Current research has not yet revealed an apparent impact of other cultural dimensions on information-seeking behaviour. They are therefore not considered in our analysis.

3 Method

3.1. Data

Logfile data of a large international website in the field of public health, recorded between November 2001 and November 2002, provides the necessary empirical data for the verification of our hypotheses, stated above. The website offers information in the field of public health care, and was available in four languages (English, German, Spanish, Portuguese) at the time when the data were collected. Navigation behaviour of users from more than 185 countries is documented in our data.

Sessions that describe each user's navigation path through the website are inferred from the logfile by using information about IP addresses, time data and data about the referrer page (e.g., Cooley et al. 1999). All the information necessary for our analysis is provided by the logfile (e.g., access of search engine, hyperlinks etc.). We are furthermore able to exclude robots from the data set by detecting them by their IP addresses or their navigation patterns (e.g., regularity of access, number of page requests – see Tan, Kumar 2002). The majority of this work was done automatically using the sessionizing tool WUMPREP developed at Humboldt University Berlin and Leipzig Graduate School of Management (www.hypknowsys.de). For easier data processing we transferred the logfile data to a database.

Geographic information is obtained from the IP address using specialized software (Geoselect – www.geobytes.com) and contains the following items: country, province code, city, data about latitude, longitude, time zone and certainty of information. Due to the detailed geographic data, we can assign the cultural indices (Hofstede 1991, Hall 1989) to each session, if the concerned country information is provided. Concerning Hofstede's indices, the boundary between the low and high groups is the highest cultural index divided by two. Concerning all cultural indices, we were able to include more

than 200,000 sessions with more than 3 million page requests in our analysis.

The website whose logfiles are examined offers its users search engines, alphabetically and content-organized hyperlinks. All language versions are presented with the same interface design.

3.2. Measures

We determine the preference for the various search options considering two main aspects: (1) selection of search options and (2) usage patterns. Measures from the first group analyse all users, and will answer the following questions: (a) Which search options were used?, (b) In which combination with other search options were they used?, (c) In which order were the various search options used? (search option access ranking).

Measures from the second group take into consideration only those users who actually opted for a particular search option. We will measure (a) how many browsing steps were accomplished prior to accessing that search option and (b) how many times it was used (total and (c) relative to number of page requests).

4 Results

Table 1 shows the selection of search options by cultural dimensions. Concerning three cultural dimensions out of four, evidence is provided from our data that the users' selection of the search options occurs in the predicted way. UA seems to influence navigation behaviour in the opposite direction than predicted. Groups with the highest percentage of users who opted for the use of the search engine are the following: high UA, low context, low long term orientation and low PD. We obtained analogous results when analysing the exclusive use of search options within a session (see table 1). These outcomes confirm our hypotheses.

Table 1: selection of search options

	UA		Context		LTO		PD	
	Low	High	Low	High	Low	High	Low	High
SE	29.1%	30.5%	29.8%	21.0%	29.7%	23.8%	29.8%	28.0%
CL	48.3%	45.0%	47.0%	49.7%	47.1%	53.6%	46.9%	48.9%
AL	55.1%	58.8%	55.5%	55.2%	55.8%	52.7%	56.1%	57.3%
SEO	10.1%	10.4%	10.8%	7.6%	10.5%	6.3%	10.6%	8.0%
CLO	32.3%	28.8%	31.3%	35.5%	31.2%	39.3%	30.8%	32.6%
HLO	70.9%	69.5%	70.2%	79.0%	70.3%	76.2%	70.2%	72.0%

(only users who used at least one search option were considered)

SE = users within one user group who used the search engine

CL= users within one group who used content organized links

AL= users within one group who used alphabetically organized links

SEO= users within one group who used only the search engine

CLO= users within one group who used only content-organized links

HLO = users within one group who used hyperlinks only

Grey : no hypothesis about the distribution

However, an analysis of the usage patterns (in table 2) reveals partially inconsistent preferences for the diverse search options within the various groups. Concerning the impact of UA, all significant measures confirm our results obtained from the first step and contradict therefore again our hypotheses. Significant results concerning the cultural variables context specificity, LTO, and PD confirm previous results and hence our hypotheses when they refer to content-organized/ hyperlinks. Outcomes concerning the use of the search engine are ambiguous.

It should be noted, that not even half of the between subjects effects are significant and in three cases also Pillai's trace indicates non-significant results.

Table 2: MANOVA (Between Subjects Effects and Pillai's Trace¹)

		UA	Context	LTO	PD
Search engines	Number of page requests prior to access	.586	.494	.046 (- / H) [†]	.209
	Number of uses	<.001 (+/H)	.002 (- / L)	.407	.447
	Number of uses per page requests	.705	.041 (+ / H)	.148	.100
	Ranking(Mann-Whitney U-Test)	<.001 (- / H)	.179	.605	.523
Cont.org links	Number of page requests prior to access	<.001 (+ / L)	.909	.991	.053 (- / H)
	Number of uses	.405	.188	<.001(+/H)	<.001(+/ H)
	Number of uses per page request	.611	<.001 (+ / H)	.004 (+ / H)	.049 (+ / H)
	Ranking (Mann-Whitney-U-Test)	<.001 (+ / L)	.653	.090 (- / H)	<.001 (+ / L)
Alph. Links	Number of page requests prior to access	.821	.504	.203	.720
	Number of uses	<.001(+/H)	<.001(- / L)	.051(+ / H)	.001 (+ / H)
	Number of uses per page request	.232	.433	.816	.042 (- / L)
	Ranking (Mann-Whitney-U-Test)	<.001 (- / H)	.001 (- / H)	.568	.847
Pillai's Trace	Number of page requests prior to access	.003	.655	.017	.109
	Number of uses	.001	<.001	<.001	<.001
	Number of uses per page request	.538	<.001	.017	.031

[†] **Table Interpretation:** all numbers represent significance levels

.046 = p-value from between subjects effects of MANOVA

- = direction of increasing values (e.g., here: higher number of page requests prior to access for the low-LTO group)

H = indicates the group with the stronger preference (our interpretation; only for significant results)

Grey: no hypothesis about the distribution

5 Discussion

Concerning the selection of search options, we can state that with the exception of the impact of UA, our hypotheses have been confirmed. Preferences

¹ Pillai's Trace shows whether the effect of each factor (independent variable) on the dependent variable group is significant (Tabachnik & Fidell 2001).

for both search engines and hyperlinks search occurred as predicted and group variances were significant. Hence, there is evidence that culturally determined information need (as determined by the context index), time perception (as determined by long term orientation) and space perception (as determined by PD) influence users' search behaviour on websites.

If only those who actually use a search engine or content-organized links are considered, results about usage patterns confirm only to some extent these outcomes.

The clear stronger preferences for search engines found among the high-UA group and for content-organized links among the low-UA group are at a first glance surprising, since they seem to contradict the information need of both groups. Yet, in the second group of measures we only consider those users who regard the search engine as an appropriate search device. Therefore, a higher use among the high-UA group is consistent with the higher amount of information needed by these users. A further indicator confirming high amount of information needed is the highest number of total page requests among the high-UA group. Concerning the measures from the first group, it appears that preference for a limited number of choices among the high-UA group (Marcus 2000) has a stronger effect than the need for information and is reflected by the lower use of content-organized links.

To sum up, there is evidence that cultural dimensions – in particular amount of information needed and the perception of time and space – have an impact on the users' search behaviour. These differences in search behaviour are likely to be caused by inherent thinking patterns which are determined by the users' cultural backgrounds.

Website providers offering information to an international audience should take these results into consideration when designing search options and information access on their website. In deciding whether to emphasize internationalization (a common design for all as the "smallest common denominator") or localization, more and more sites opt for the latter: an accommodation of regional differences in domain conceptualizations and the concrete Web resources that embody these conceptualizations. Localization constitutes a significant investment and should therefore be based on an informed basis of interface preferences in the respective region (usually defined in terms of a country).

It should be pointed out that Hofstede's cultural dimensions, which underly our variables, have been criticized by some authors as too simple, stereo-

typic and ethnocentric: dimensions and indices refer not to societies but to nations which themselves consist of several different cultures, societies and subgroups (e.g., Fleming & Sjøborg 1999, Garsten 1994, Holden 2002). Future research should take these factors into account. Nonetheless, our results offer a descriptive account of behavioural differences between audiences from different countries, and this constitutes a first *operational* basis for Web site design decisions that aim at a localization by country.

6 Limitations of our study

As mentioned above, an analysis of the Web server logs of a heavily frequented, internationally used Web site offers unprecedented opportunities in terms of the amount and variability of data and users studied. The investigation of their search behaviour taps authentic search histories, producing a high ecological validity. However, a logfile analysis also presents a number of methodological challenges.

First, we encounter problems of session and navigation path reconstruction due to difficulties such as cached and not recorded webpages, use of proxy servers or firewalls, etc. Some IP addresses do not allow the reconstruction of the country from which the user accessed the site. These cases were indicated by the Geoselect software, and the sessions were removed from future analysis.

Logfile analyses like the one presented in this paper are field studies and therefore share the typical advantages and disadvantages of this kind of study. In particular, personal variables (such as an individual user's culture) and situational variables (such as potential problems caused by the Internet connection, distraction by other tasks, etc. as well as the satisfaction with search outcomes) cannot be directly observed and are thus subject to a margin of uncertainty.² In an investigation of search behaviour, uncertainty about a user's goals and strategies pose a particular challenge – as Rice et al. (2001, p. 254) note, these are difficult to determine even when observing users walking through a library's book shelves. We limit our attention to a user's choice of navigational tools, more specifically, search options.

² If they are of interest, they may – with caution – be inferred from the logfile data.

It is well-known that the choice of navigational tools, and the strategies in which these tools are employed, are related to a user's search goal (e.g., Allinson & Hammond, 1989; Shneiderman, 1997). For example, it is generally acknowledged that known-item searches tend to be performed using search engines, while more open-ended searches tend to encourage hyperlink browsing. However, there is no reason to assume that the distribution of search goals differs systematically across countries and cultures. Thus, the uncertainty about goals is likely to introduce unsystematic variation, which would strengthen rather than weaken the interpretability of a statistically validated impact of culture on search behaviour.

With respect to the generalization of results, our study shares the weakness of many studies of hypertext navigation – the limitation to one hypertext resource or Web site. Various studies have shown how the website structure affects navigation and search behaviour (e.g., Larson & Czerwinski 1998, Thüring et al. 1995). A website's topic field may have a similar effect, as may interface elements other than those investigated here. Navigation and search behaviour analyses might therefore be regarded to a certain degree as site and domain specific. Also, cultural variables might be confounded with socio-economic aspects, Internet distribution and experience. In this case differences in search behaviour could have other causes. Further research is clearly needed here.

7 Outlook

We see five major fields of investigation for future research: First, we suggest to complement our research with experiments where personal and situational variables can be controlled and included in the investigation.

Second, our examination of the impact of cultural dimensions on search behaviour could be extended to an empirical analysis of how culture affects navigation behaviour in general. This would include aspects such as depth of navigation or time-related variables.

Third, since we have seen how differences in amount of information needed affect search behaviour on a website, research should investigate how culturally determined information needs affect the website selection process.

Fourth, future research should investigate a possible impact of those cultural variables that – for the reasons mentioned above – were not integrated into our analysis.

Finally, our study only took into consideration the effect of cultural variables on behavioural measures. Future research should also investigate their effect on attitudinal variables.

References

- Allinson, L., & Hammond, N. (1989). A learning support environment: The hitchhiker's guide. In R. McAleese (Ed.), *Hypertext: Theory into Practice* (pp. 62-74). Norwood, NJ: Ablex.
- Barber, W. & Badre, A. (1998). Culturability: The Merging of Culture and Usability. *Proceedings of the 4th Conference on Human Factors and the Web*, June 1998.
- Chau , Patrick Y.K., Melissa Cole, Anne P. Massey, Mitzi Montoya-Weiss & Robert M. O'Keefe (2002). Cultural Differences in the Online Behavior of Consumers. *Communications of the ACM*, 45(10), 138-143.
- Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Journal of Knowledge and Information Systems*, 1(1), 5-32.
- Fleming, Daniel & Henrik Sjøborg (1999). Human Resource Strategies as a Technology of Governance in Danish Subsidiaries in Southeast Asia. *Nordisk Arbejdslivskonference. Norden i verden. Verden i Norden*. Stavanger.
- Garten, C. (1994). Apple Computer. Stockholm.
- Hall, Edward T. (1989). *Beyond Culture*, New York.
- Hofstede, G. (1980). *Culture's consequences: international differences in work related values*, Beverly Hills.
- Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind*, London .
- Hofstede, G. (2001). *Culture's consequences. Comparing Values, Behaviors, Institutions and Organizations Across Nations*, Thousand Oaks .
- Holden, Nigel J. (2002): „Cross-Cultural Management. A Knowledge Management Perspective. London.
- Larson, K., & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent from information retrieval. *Proceedings of the Association for Computing Machinery's CHI '98*, 18-23.

- Luna, David, Laura A. Peracchio & Maria D. de Juan (2002). Cross-Cultural and Cognitive Aspects of Web Site Navigation. *Journal of the Academy of Marketing Science*, 30 (4), 397-410.
- Marcus, Aaron & Emily West Gould (2000). Cultural Dimensions and Global Web User-Interface Design, *Proceedings of the 6th Conference on Human Factors and the Web*, June 2000.
- Navarro-Prieto, Raquel, Mike Scaife & Yvonne Rogers (1999). Cognitive strategies in web searching. *Proceedings of the 5th Conference on Human Factors and the Web*, July, 1999.
- Pedersen, Per E. & Herbjørn Nysveen (2003): Search mode and purchase intention in online shopping behavior. To appear in: *International Journal of Internet Marketing and Advertising*.
- Reips, U.-D. (1999). Theorie und Techniken des Web-Experimentierens. In B. Batinić, A. Werner, L. Gräf, & W. Bandilla (Eds.), *Online research: Methoden, Anwendungen und Ergebnisse* (pp. 277-295). Göttingen, Germany, etc.: Hogrefe.
- Rice, R.E., McCreddie, M., & Chang, S.-J. L. (2001). *Accessing and browsing information and communication*. Cambridge, MA: MIT Press.
- Savolainen, Reijo (1999). The role of the Internet in information seeking. Putting the networked services in context. *Information Processing and Management*, 35(6), 765-782.
- Scaife, M. & Rogers, Y. (1996). External cognition: how do graphical representations work ? *International Journal of Human-Computer Studies*, 45, 185-213.
- Sheppard, Charles & Jean Scholtz (1999). The Effects of Cultural Markers on Web Site Use, *Proceedings of the 5th Conference on Human Factors and the Web*, July 1999.
- Shneiderman, B. (1997). Designing information-abundant Web sites: issues and recommendations. In: S. Buckingham Shum and C. McKnight (eds.). *Web Usability* (special issue). *International Journal of Human-Computer Studies*, 46.
- Tabachnik, B.G. & L.S. Fidell (2001). *Using Multivariate Statistics*, Boston.
- Tan, Pang-Ning & Vipin Kumar (2002). Discovery of Web Robot Sessions based on their Navigational Patterns. *Data Mining and Knowledge Discovery* 6, 9-35.
- Thüring, M., Hanneman, J. & Haake, J.M. (1995). Hypermedia and Cognition. Designing for Comprehension. *Communications of the ACM* 38 (8), 57-66.
- Trompenaar, Fons (1995). *Riding the waves of culture. Understanding Cultural Diversity in Business*, London.

- Vakkari, P. (2000). ECognition and changes of search terms and tactics during task performance: A longitudinal case study. Proceedings of the RIAO 2000, Paris, 894-907.