Designing for cultural perceptual factors in information retrieval

Research proposal · SI 531: Human Interaction in Information Retrieval
Bryan Smith · besmit@umich.edu

Introduction

In many ways, the world is becoming smaller. Not only are individuals more mobile, but the impacts of globalization means that distributed work environments and international markets are bringing culturally diverse people together.

At the same time, online activities are becoming more intertwined with daily life, and enabling technologies such as laptops, smart phones and tablets offer more ubiquitous access to the internet as well as blurring the distinction between our online and offline lives. Additionally, the amount of information available online is ever increasing, and as a result it is likely that search will play and increasingly important role in satisfying our personal and professional information needs.

There is a clear value in understanding how software is used; companies invest heavily in usability research, and HCI is a vibrant area of academia. Much of the research originates from Europe and North America, which begs the question whether we are seeing the whole picture on how software is used. It is important for us to determine whether HCI concerns such as usability must be contextualized to account for the cultural differences of the global web community.

There is a rich body of research within psychology and anthropology that supports this investigation. The influential (if controversial) work by Hofstede includes the cultural dimensions theory, which proposes five dimensions that help explain cultural values; this work is particularly helpful in understanding the challenges of communication and cultural sensitivity. Later work by Nisbett offers insights into how culture impacts our perception; particularly, it implies that our cognition and perception are impacted by culture rather than being a product of our universal nature.

While this and related research offers a useful foundation, we must establish clear implications for the field of information retrieval. Research the helps us design search for wider audiences must first lead us to an understanding of how individual and cultural differences impact how we search. This study aims to investigate cultural factors in search; in particular, this pilot study is concerned with determining whether research into these differences can be used to inform the design of search.

Background

Barber and Badre (1998) argue that usability must be defined within a cultural context. Investigating culture and the genre of web content (e.g., government sites, news and media), they identified many high frequency cultural markers by country, such as a high frequency of the color green in Israel and Lebanon, cityscape photos in Japan and motion in France.

Simon (2001) further notes that since the World Wide Web matured in the US and Europe, it tends to exemplify the values and norms of these countries. Using 160 participants recently arriving within the US from Asia, Europe, Latin & South American and North America, he found significant cultural and gender differences. For example, Europeans and North Americans were more likely to offer feedback that suggested changes to the navigation of the sites to makes them simpler and easier to use, and whereas Asians tended to prefer less bright colors, Europeans and North Americans preferred lighter and brighter colors. Simon also found that Asians registered the highest level of trust, followed by Latin Americans, with Europeans and North Americans exhibiting the lowest degree of trust.

Chau et al (2002) found that consumers in different countries with different ethnic origins used the web for different purposes, and that furthermore this resulted in users of the same site often having different

impressions. The researchers found that US users were more likely to use sites for information search and e-commerce, while Hong Kong users were more likely to use sites for social communication and hobbies. In two separate exploratory studies, Faiola and MacDorman (2008) found that Americans and Chinese participants tend to prefer and perform faster on sites designed by people of their own national culture.

There are two traditions of research that offer various models for investigating cultural differences, including the cultural dimension models as proposed by Hall, Hofstede and others, and investigations into cultural impact on cognition by researchers such as Nisbett.

Recent research by Lee, Tran and Lee (2007) found that cultural differences between the Netherlands (which has a high individuality score in Hofstede's model) and Korea (which has a high collectivism score) impacted performance in usability research. Marcus and Alexander (2007) conducted an usability study on an undisclosed financial website, and showed that individuals from cultures that score high on individualism have particularly strong desire to see imagery representative of their own country, and that countries with a short-term time orientation score need a quick and concise insight into the purpose of a site.

Nisbett et al (2001) report that they have found that East Asians tend to be *holistic*, which they characterize as being more attentive to the entire field while making relatively less use of categories and formal logic, while Westerns tend to be more *analytic*, paying more attention to objects and their categories, and using more rules. They conclude that cognitive processes once considered to be universal are instead impacted by the cultural in which the individual participates. Later, Nisbett and Miyamoto (2005) report that similar differences in perception have been identified between eastern and western Europeans, southern and northern Italians, and the working and middle classes, respectively.

Research in support of Nisbett's perceptual dichotomy often focuses on performance metrics regarding visual perception. Ji et al (2000) use a *rod and frame machine*, in which a rod and a square frame rotate independently; the task is to judge when the rod appears to be vertical independent of the square frame. The researchers found that Americans tend to spend less time on the task as well as make fewer mistakes, and while Asian males demonstrated a decreased performance when given control of the machine, Americans performance increased. The researchers conclude that the American participants exhibited more field independence. Similarly, Kitayama et al (2003) used the *frame-line test*, in which participants were shown a box with a vertical line, and given another box and told to draw a vertical line of the same absolute size or similar size relative to the box. The researchers found that Japanese participants were more accurate in the relative task, and that American participants were more accurate in the absolute task.

In a paper especially relevant to this study, Dong and Lee (2008) conducted an eye tracking study with Chinese, Korean and American participants, in which they showed participants the same static web page, but translated to their native language, for thirty seconds. They found that Koreans moved their eyes more than American or Chinese participants, and were more likely to read in a circular pattern. Both Chinese and Korean participants were more likely to scan back and forth, while American participants were more likely to read the page's title information. They concluded with design implications, which are listed below in the *Systems Descriptions*.

Fitzgerald (2004), Nisbett and Miymoto (2005) and Faiola and MacDorman (2008) offer particularly helpful summaries of this and related research.

Purpose of study

The purpose of this study is to determine whether research related to Nisbett's analytic and holistic perception can be used to inform the design of search systems. Given the marked difference observed in Dong and Lee (2008) between Koreans and Americans, combined with the knowledge gained from the eye-tracking study (particularly, the circular tracking observation for Korean participants), this study is limited to investigating how Koreans and Americans engage in web search.

This study is designed to address two research questions:

1. To what extent are cultural differences demonstrated between how Americans and Koreans perceive and interact with search system interfaces?
2. To what extent can research inform the design of search system interfaces appropriate for cultural differences?

Systems Descriptions

To investigate the above research questions, I designed two novel search systems, intended to support two broad classes of cultural perception (analytic and holistic), as well as make use of eye-tracking observations in Dong and Lee (2008), and incorporating the design implications offered by Faiola and MacDorman (2008) as well as Dong and Lee (2008). The design implications are listed in Table 1.

Analytic	***Holistic***
• Distinction of parts[1]/Layout facilitates focus on each group[2] • More reliance on text[1] • Simplicity[2]	• Favors a global view[1]/show whole context of website[2] • Discourage isolation of part[1] • Content placed more freely[2] • Koreans more likely to read in circular pattern[2]

Table 1. Design implications for experimental systems. From Faiola and MacDorman[1] (2008) and Dong and Lee[2] (2008)

Both systems use the *Bing search API* to retrieve the top hundred results for each query. (They also support the Google Custom Search API; while Google provides up to 100 free searches daily using the API, the Bing search API is free unless exceeding six searches a second.)

These results are ordered by relevance ranking, and submitted to a local *Carrot2* Clustering Document Server for dynamic clustering. (Throughout the discussion of the systems, I will refer to the clusters as categories; it is worth mentioning that the clusters contain overlapping results, and should not be considered to be exclusive classifications.) Search (including clustering) takes approximately three seconds or less, and search results are cached for up to a day, so that viewing different clusters or using pagination is fast.

When clicking on a result, the document opens in a new tab. This is done because all actions (performing search, viewing clusters and clicking on results) are recorded in a database; the prevents duplicate records resulting from returning to search result page. Users were trained for both systems, and no users exhibited or expressed any difficulty with the tabs; one participant mentioned preferring this to standard search engine behavior of opening results in the same window.

I code named the search engines so as to not bias participants: the *Banana search system* was designed to support analytic perception, whereas the *Mango search system* was designed for holistic perception. Both of these search systems are open source, and are available for download at:
http://bryanesmith.com/projects/cultural-search

Banana search system

The Banana search system was designed to support analytic perception. The priorities are visually distinct search results, as well as densely-displayed category labels. In the following screenshots, a total of 15 categories can be viewed without scrolling, but only five search results.

The results are paginated so that there are ten results per page. Vertical tabs represent the categories; clicking on a category tab will open the results for just that category in the main content area. The default category is “All”, and if the user prefers, they can ignore the categories entirely and just browse the results for this default category.

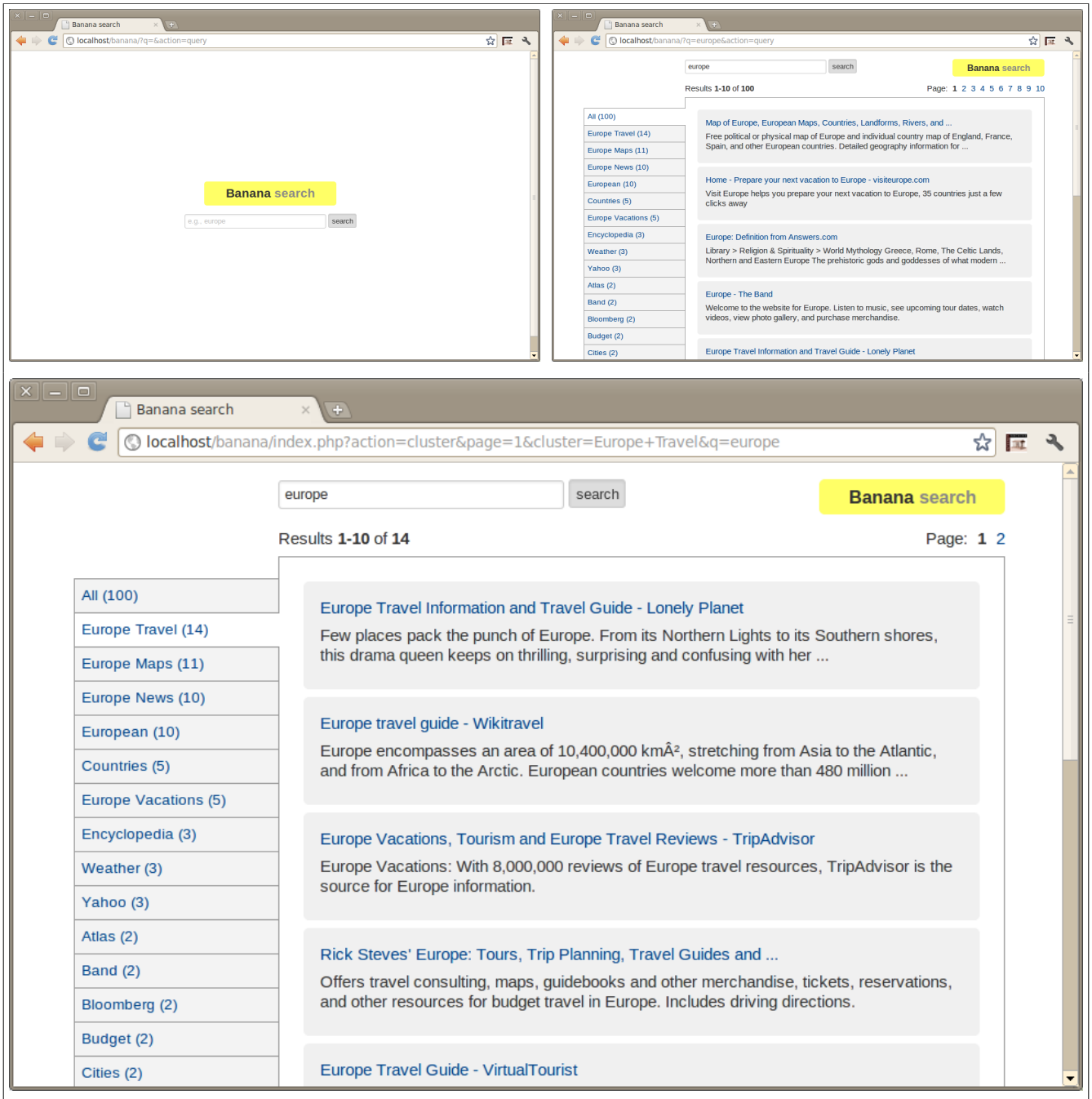

Table 2. The Banana search system. Clockwise from the top: the home page; the default “All” category for the search “Europe”; and the “European Travel” category for the same search.

Mango search system

The Mango search system was designed to support holistic perception. The priorities are a high density of search results as well as supporting the aforementioned circular scanning pattern mentioned in Dong and Lee (2008). This, however, comes at the expense of the number of categories “above the fold”: in the following screenshots, a total of 18 search results can be seen without scrolling, but only six categories.

The home screen is simple, much like the Banana search system. However, the search results page is quite different; category clusters are rendered in boxes, three per line, rather than as tabs, and there is no pagination. Each cluster box shows the top three results for that cluster. If the user click “View all...” for a

cluster, a modal dialog box opens with the results, which are formatted similarly to Banana, but without pagination. This search is a little more complex, resulting in three different views rather than the two of Banana, but there is only one search page which offers a meaningful overview of each category.

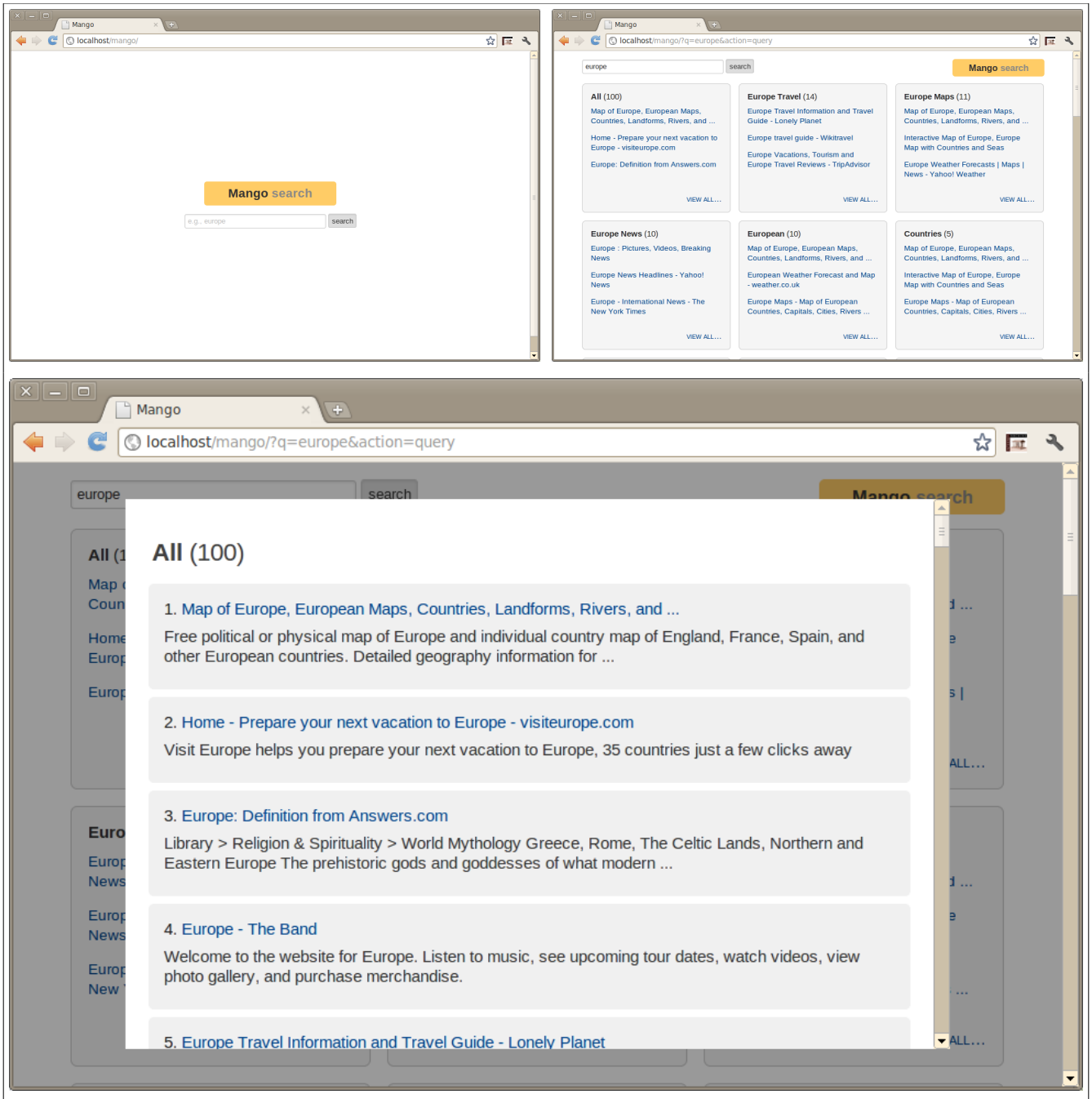

Table 3. The Mango search system. Clockwise from the top: the home page; the results page for the search "Europe"; the "All" category modal dialog after clicking on "View all..." for "All" category.

Methodology

Because culture is not likely to be the primary factor, observing the impact of culture over other factors (e.g., search history, familiarity with existing search technology, willingness to adopt new technology, familiarity with the topics for the search tasks, gender, etc.) will require many participants. For this study, between sixty to eighty participants will be recruited to perform four search tasks, two for each system.

Korean and American students would be recruited from within US and Korean universities, respectively, to avoid complicating factors such as cross-cultural priming. Since participation takes around one hour, $15 / ₩18000 in compensation would be provided.

Each participant will fill out a consent form and demographic questionnaire (Appendix 1), and will receive a brief description of how the experiment will proceed. Then the students will be assigned a system, and receive standard training for the system. After receiving training, the student will perform two search tasks (one simple, one complex), with a ten minute limit on each so that activities can be compared and to prevent participant burnout on later tasks. After completing the first tasks, the participant will then be introduced to the second system, receiving training, and perform a simple and complex search task with the same ten-minute restriction. Following the tasks, there is a brief six-question interview (Appendix 2), and then the user fills out a sixteen question exit questionnaire (Appendix 3).

Simple search tasks:

$Q_{s}1$: Find evidence that supports or refutes claims that radio waves from radio towers or car phones affect brain cancer occurrence. (*Source*: TREC 2003 Robust Track Topics, 310)

$Q_{s}2$: What hard evidence proves that the Vikings visited or lived in Scotland? (*Source*: TREC 2001 web ad hoc topics, 503)

Complex search tasks:

Qe1: What are the costs and quality of defense in death penalty cases? What are the arguments for and against the death penalty? (*Source*: modified version of TREC 2003 Web Topic Distillation Topics, TD25)

Qe2: Is there any evidence that US workers are retiring earlier or later? What are some explanations of retirement trends, and how does it impact the job market? (*Source*: self-generated for this study)

Table 4. The four search tasks. Every participant will be assigned one simple task and one complex task per search system, and is allowed up to ten minutes to perform the task.

The order of the tasks as well as the order of the systems will be balanced, so that half the participants start with the Banana system and half with the Mango system.

Metrics

To determine whether or not the search systems support the culture perception styles for which they are designed, the following will be measured: *Ease of use*, *Satisfaction*, *Effectiveness* and *Efficiency*.

Qualitative and quantitative data will support the investigation of each. (See below for more information.) Note that *efficiency*, however, is a particularly challenging aspect to measure; while the number of actions to complete a task is directly measurable, this cannot be assumed to be an indicator of efficiency since different participants might cease their search for different reasons.

Also, *familiarity* will be measured principally because it cannot be controlled; it is hypothesized that a familiar interface might have more influence on preferences and performance than any cultural cognitive impact, though this would be challenging to determine experimentally.

Qualitative measurements

The exit questionnaire will provide 16 responses, all using a four-point Likert scale. Additionally, actions performed by the users, including searches, browsing categories and clicking on search results are automatically recorded. The investigator will ask the participants whether each result is relevant, which will later be collated with the database data.

Quantitative measurement

Qualitative data will include observations from the study (recorded in a legal pad by investigator) as well as

responses during the post-task interview.

Pilot study

For this pilot study, four participants were recruited – two Korean, two American – but due to unforeseen circumstances, a participant was unavailable. As a result, one Korean female, one Korean male and one American male (ages between 31 and 36) participated. Both Koreans had lived in the US for a while (four years and eight years), which would likely mitigate any potential cultural effect.

All were highly experienced with web search, and none showed any difficulty learning how to use either system. All participants responded that they have used the web from 13-16 years, and spend between 2-4 hours on the web each day.

All three primarily used Google for search, though one of the Korean participants mentioned that they also used Naver and Daum. **Satisfaction with their preferred search systems was high** (Table 5), though they also believed that their search systems could offer additional features or functionality.

In Table 6 and Table 7, I show aggregate information about actions performed, categorized by task and system. This small sample does not permit a statistical analysis, but it is worthwhile to explore these preliminary results to further clarify potential areas of interest for future investigation.

	Disagree	Somewhat disagree	Somewhat agree	Agree
I am satisfied with the web search system(s) I use.	-	-	-	3
My preferred web search system(s) is easy to use.	-	-	-	3
I find it is not difficult to find what I want using web search system(s).	-	-	1	2
I believe that my preferred search system(s) is missing features and functionality that would help me.	-	-	3	-
I believe that I generally use web search systems effectively.	-	-	2	1
I generally enjoy searching the web.	-	-	1	2

Table 5. Satisfaction with preferred search systems for the participants was high.

The **number of searches and viewed results is roughly similar** for **all tasks** as well as for **both search systems**, perhaps with the exception of the task on retirement ages, which resulted in more query reformulations. (Mango has more searches because two of the three participants performed this search task using the Mango search system.)

While I expected this might be true for the search systems, I expect that the “difficult” tasks to require more searches (here, I observe +3) and documents viewed (here, I observe -1). In the study, it will be worthwhile to correlate these numbers with task difficulty judgments to determine whether these numbers can be used as a measurement of efficiency.

	Searches	Viewed results
Q_{s1}	5	16
Q_{s2}	4	13
Q_{e1}	5	15
Q_{e2}	7	13

	Searches	Viewed results
Banana	9	27
Mango	12	30

Table 6. Actions performed, aggregated by task and by system.

For the relevance judgments, the *percentage scores* are calculated as the percentage relevant documents out of all documents; a lower score indicates lower performance. The *difficulty score* is calculated by assigning a numeric value (1 for easiest, 4 for hardest) based on the rank order of difficulty provided by participant. (See *Limitations* for a discussion on the problem with the difficulty score.)

	Relevant	Not relevant	Percentage	Difficulty
Qs1	12	4	75%	3.0
Qs2	10	3	77%	2.0
Qe1	13	2	87%	2.2
Qe2	9	4	70%	2.8

	Relevant	Not relevant	Percentage
Banana	21	7	75%
Mango	23	6	79%

Table 7. Relevance judgments, aggregated by task and by system.

Qe2 (task about retirement) had the lowest performance, and Qs1 (task about radio waves and cancer) the second lowest performance; these were also ranked as the most difficult questions. Higher performance was noted with less difficult questions. This preliminarily supports the notion that **performance might be negatively impacted by task difficulty**. This will be controlled by ensuring that both systems will receive a balanced number of each task.

Table 8 shows the average number of actions performed and relevance judgments per participant by nationality. The sample is too small to find a statistically-significant trends, but it would be interesting to find out whether the study finds any significant differences in the amount of activity based on system as well as the percentage of relevance judgments based on national identity or system. In particular, the study should determine **whether there is any interaction between national identity and system,** which will support the primary purpose of this investigation.

Banana

	Searches	Viewed results
Korean	2.5	7.5
American	4	12

Mango

	Searches	Viewed results
Korean	3.5	7
American	5	16

Total

	Searches	Viewed results
Korean	6	14.5
American	9	28

Banana

	Relevant	Not Relevant	Percentage
Korean	6.5	1	87%
American	8	4	67%

Mango

	Relevant	Not Relevant	
Korean	5.5	1.5	79%
American	12	4	75%

Total

	Relevant	Not Relevant	
Korean	12	2.5	83%
American	20	8	71%

Table 8. Average tasks performed and relevance judgments for Korean and American participants.

Table 9 shows responses to the post-task questionnaire by each participant, labeled by nationality. Interestingly, **responses to the Mango search system show more variability than the Banana search**. Overall, differences were by a very small margin, but Banana scored higher on *ease of use* and *familiarity*,

and the Mango search system scored higher for *satisfaction*, *efficiency* and *effectiveness*. Willingness to use Banana for future searches was higher than Mango for one of the Korean and American participants; however, the difference in willingness to use these two systems was less for the two Korean articipants.

Mango search

	Disagree	Somewhat disagree	Somewhat agree	Agree
I found that the Mango search system was easy to use.	K		A	K
The Mango search system had a similar interface to web search tools I currently use.	K	K	A	
The Mango search system offers some benefits that my preferred search system does not.		AK		K
My preferred search system offers some benefits over the Mango search system.			AK	K
I believe that I used the Mango search system effectively.		K		AK
I enjoyed using the Mango search system.	K		A	K
The Mango search system was an effective tool for performing the search tasks.		K	A	K
The Mango search system was an efficient tool for performing the search tasks.		K	A	K
I would voluntarily use the Mango search system for future web searches.	AK		K	

Banana search

	Disagree	Somewhat disagree	Somewhat agree	Agree
I found the Banana search system was easy to use.			A	KK
The Banana search system had a similar interface to web search tools I currently use.	A			KK
The Banana search system offers some benefits that my preferred search system does not.		K	K	A
My preferred search system offers some benefits over the Banana search system.			K	AK
I believe that I used the Banana search system effectively.			K	AK
I enjoyed using the Banana search system.			AKK	
The Banana search system was an effective tool for performing the search tasks.		K	AK	
The Banana search system was an efficient tool for performing the search tasks.		K	K	A
I would voluntarily use the Banana search system for future web searches.		K	AK	

Table 9. Responses to the post task questionnaire, including nationality. “A” is an response by an American, and “K” is a response by a Korean.

Given the small sample size, the qualitative feedback from the interview is of more value in this preliminary analysis. One person (Korean) mentioned a preference for the Mango system, while the other two people

preferred the Banana system. All participants said that given the choice, they would use the preferred of the two systems exclusively for future tasks.

The participant that preferred Mango said that it is more effective than Banana, and that the participant prefers the (modal) dialog because it is "more seamless". This participant also mentioned that with Mango, you can "see the headline and contents" and that "**you can instantly determine whether the category is relevant... it's more efficient.**" This participant did not like the positioning of the vertical tabs in Banana, and mentioned that they required clicking to determine the benefits of the categories.

Remarks by participants preferring Banana include a strong preference for the vertical tabs on the sidebar by the American participant, noting that they offer a **compact way to view the categories**. This participant also stated that they preferred how the Banana search system categorized content, though it is not clear whether they meant visually or by the actual categorization. The Korean participant stated that the categories were not that useful, and **preferred to rely on the relevance rankings**, which they "trusted". This participant also stated that **because they didn't understand how the documents were clustered, Mango was harder to use**. This participant made this observation that **Mango required more scrolling**.

Limitations

- **Categories were confusing.** This partly due to the fact that clustering was limited to relatively small snippets, but ultimately clustering is more problematic for users than manually-assigned categories. Since the Mango search system relied on categories more, it is possible that it was more negatively impacted by the clustering quality.
- **One participant experienced a really unstable and slow connection for the first question.** As a result, I extended the task time to a total of twenty minutes. However, this participant viewed 67% more documents for this task than other tasks with a ten minute limit. Hence, for the study, it is important to have a (1) stable internet connection and to (2) enforce the time limit. Using an on-screen timer would help ensure that time limits are uniformly enforced.
- **I did not explain what makes a result relevant or not relevant**. In the study, I would start the study with a brief explanation.
- **Difficulty scores for questions were calculated from ranks of tasks**. The easiest task was assigned a 1, and the most difficult a 4. If a person stated that two tasks had similar difficulty, then they were given an average of two positions, e.g., 2.5. This sort of averaging is not admissible for ordinal data; it would have been better to ask the user to select a difficulty score from a continuous "thermometer" to give ratio data, which has more analytic affordances.
- **Familiarity cannot be controlled**. Given the global nature of much technology, it is possible that familiarity with technology might conflict and possibly override cultural cognitive impacts.

Conclusions

Studying cultural impacts on search is important considering the growing international community of web users as well as the growing importance of search; even though cultural cognitive impacts on search is probably not the primary factor, understanding its role will afford us a better overall picture of how search varies for different users. This research is not without its challenges, the primary one being familiarity with existing technology, which might conflict with perceptual tendencies as discussed by Nisbett et al.

From observing these participants as well as conducting an exit interview, it is clear that **individual differences were more important than nationality**. Unsurprisingly, these preliminary results also demonstrate that individual differences will make finding potential cultural impacts difficult. The only solution will be a large sample size.

It is worth mentioning that:

- None of the participants had difficulty using the systems
- Though the participants all shared their preferences without being prompted, they did not show any frustration with either systems
- All the participants used the categories for both systems, even though one user remarked that the categories were only a distraction

Regardless, the quality of the categories impacted the usefulness of both systems, though it undoubtedly did more for Mango. This is a problem without a solution if these search systems are going to use categories; the best hope would be to improve the clusters by somehow providing the clustering server more representative content for each result.

Lastly, since the goal is to test for differences based on culture (using nationality as a proxy for culture), a cheaper form of preliminary testing would help select a design for a search system promoting holistic perception. Simple preference tests for a variety of design variants could be conducted with a large number of participants online far more easily than in individual one hour sessions, and would potentially allow for design iteration before conducting the actual study.

References

Barber, W., and Badre, A. (1998). Culturability: The merging of culture and usability. Presented at the Conference on Human Factors and the Web, Basking Ridge, New Jersey: AT&T Labs. Retrieved from http://zing.ncsl.nist.gov/hfweb/att4/proceedings/barber/

Chau, P.Y.K., Cole, M., Massey, A.P.M, Montoya-Weiss, M., and O'Keefe, R.M. (2002) Cultural differences in the online behavior of consumers. communications of the ACM, 45(10), 138-143.

Dong, Y., Lee, K. P. (2008). A cross-cultural comparative study of users' perceptions of a webpage: With a focus on the cognitive styles of Chinese, Koreans and Americans. International Journal of Design, 2(2), 19-30.

Faiola, A., and MacDoman, K.F. (2008) The influence of holistic and analytic cognitive styles on online information design: Toward a communication theory of cultural cognitive design. Information, Communication & Society, 11(3), 348-374.

Fitzgerald, W. (2004) Models for cross-cultural communications for cross-cultural website design. ERB-1108. NRC 46563 National Research Council Canada, 1-11.

Ji, L., Peng, K., and Nisbett, R.E. (2000) Culture, control, and perception of relationships in the environment. Journal of Personality and Social Psychology, 78(5), 943-955.

Kitayama, S., Duffy, S., Kawamura, T., and Larsen, J.T. (2003) A cultural look at new look: perceiving an object and its context in two cultures. Psychological Science, 14, 201-206.

Lee, J., Tran, T., and Lee, K. (2007) Cultural Difference and Its Effects on User Research Methodologies. HCI and Culture, Lecture Notes in Computer Science, 4559/2007, 122-129. DOI: 10.1007/978-3-540-73287-7_16

Marcus, A., and Alexander, C. (2007) User validation of cultural dimensions of a website design. Proceedings of the 2nd international conference on Usability and internationalization, July 22-27, 2007, Beijing, China

Nisbett, R.E., Peng, K., Chai, I., and Norenzayan, A. (2001) Culture and systems of thought: holistic versus analytic cognition. Psychological Review, 180(2), 291-310.

Nisbett, R.E., and Miyamoto, Y. (2005) The influence of culture: holistic versus analytic perception. TRENDS in Cognitive Sciences, 9(10), 467-473.

Simon, S. (2001) The impact of culture and gender on websites: an empirical study. Database for Advances in Information Systems, 32(1), 18-37.

Appendix 1

Consent to participate

In this pilot study, you will be performing two web search tasks in two search systems, totaling four search tasks. Participation will take approximately one hour. I am not testing your ability to perform these searches; rather, the purpose of the study is to observe how participants use the search systems.

For both search systems, there will be a brief training session; feel free to ask questions so that you are comfortable using the systems. After completing the four tasks, there will be an interview and a questionnaire. Following the interview, you will be offered an opportunity to find out more about the study. You may also request a copy of the final report.

Participation is voluntary, and you are free to leave the study at any time. If you are have any questions, you may ask for clarification as well as decline to answer any particular questions or perform any tasks. Your responses will be anonymized, but might be included in the final report.

Name: _____ Date: _____

Signature: _____

Please answer the following questions as best as you can. You may decline to respond to any questions.

What is your age? _____

What is your gender? _____

What is your nationality? _____

Please describe the countries in which you have lived, including when and how long:

Please list any web search systems (e.g., Google, Bing) you use at least weekly:

Approximately how many hours do you spend on the web each day? _____

Approximately how many years have you been using the web? _____

Please circle or underline the response that most closely represents how you would answer each of the following questions.

I am satisfied with the web search system(s) I use.

[Disagree] [Somewhat disagree] [Somewhat agree] [Agree]

My preferred web search system(s) is easy to use.

[Disagree] [Somewhat disagree] [Somewhat agree] [Agree]

I find it is not difficult to find what I want using web search system(s).

[Disagree] [Somewhat disagree] [Somewhat agree] [Agree]

I believe that my preferred search system(s) is/are missing features and functionality that would help me.

[Disagree] [Somewhat disagree] [Somewhat agree] [Agree]

I believe that I generally use web search systems effectively.

[Disagree] [Somewhat disagree] [Somewhat agree] [Agree]

I generally enjoy searching the web.

[Disagree] [Somewhat disagree] [Somewhat agree] [Agree]

Appendix 2

Exit interview

Participant: _____

Which tasks were easiest for you? Most difficult?

Which tasks did you enjoy the most? Which the least?

How did you feel about using the Banana search system?

How did you feel about using the Mango search system?

If you were limited to using either or both of these two systems for all of your web searches, would you choose one or use both?

What changes would you make to the systems?

Appendix 3

Post task questionnaire

Participant: _____

For the following questions, please circle or underline the response that most closely represents how you would answer each of the following questions.

Mango search system

1. **I found the Mango search system was easy to use.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

2. **The Mango search system had a similar interface to web search tools I currently use.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

3. **The Mango search system offers some benefits that my preferred search system does not.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

4. **My preferred search system offers some benefits over the Mango search system.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

5. **I believe that I used the Mango search system effectively.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

6. **I enjoyed using the Mango search system.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

7. **The Mango search system was an effective tool for performing the search tasks.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

8. **The Mango search system was an efficient tool for performing the search tasks.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

9. **I would voluntarily use the Mango search system for future web searches.**
   [disagree] [somewhat disagree] [somewhat agree] [agree]

Banana search system

10. **I found the Banana search system was easy to use.**
    [disagree] [somewhat disagree] [somewhat agree] [agree]

11. **The Banana search system had a similar interface to web search tools I currently use.**
    [disagree] [somewhat disagree] [somewhat agree] [agree]

12. **The Banana search system offers some benefits that my preferred search system does not.**
    [disagree] [somewhat disagree] [somewhat agree] [agree]

13. **My preferred search system offers some benefits over the Banana search system.**
    [disagree] [somewhat disagree] [somewhat agree] [agree]

14. **I believe that I used the Banana search system effectively.**

[disagree] [somewhat disagree] [somewhat agree] [agree]

15. **I enjoyed using the Banana search system.**
[disagree] [somewhat disagree] [somewhat agree] [agree]

16. **The Banana search system was an effective tool for performing the search tasks.**
[disagree] [somewhat disagree] [somewhat agree] [agree]

17. **The Banana search system was an efficient tool for performing the search tasks.**
[disagree] [somewhat disagree] [somewhat agree] [agree]

18. **I would voluntarily use the Banana search system for future web searches.**
[disagree] [somewhat disagree] [somewhat agree] [agree]