Breadth vs. Depth

Version of 06 Dec 2011

Roman Mauhart
Christian Hofer
Christopher Gallè
Tomas Karas

Abstract

This report deals with the Breadth vs. Depth issue on both desktop and mobile devices. In Chapter 1 Roman and Christian give a short introduction into this topic and explain relevant terminology. Then they cover some significant studies in chapter 2, summarize their findings and give a recommendation. After that Christopher investigates in Chapter 3 differences on mobile devices regarding the breadth-vs-depth tradeoff. Finally Tomas covers age differences and their effect on search performance and error rates in hierarchical layouts. He also takes a look at people with poor eyesight and which special needs they have.

Contents

List of Figures

Chapter 1

Introduction

This report covers the most important studies and summarizes their findings. Starting with some definitions of related terminology, the report will also cover some studies which have dealt with the breadth vs. depth tradeoff. Additionally, insight into this topic regarding the mobile sector is given. Last, but not least the affect of age differences will be discussed.

1.1 What is Breadth vs. Depth?

1.1.1 Difference Between a Wide and Narrow Layout

Menus can either be broad or deep. A broad menu has many options per category whereas a narrow menu has fewer. In a narrow structured menu the user has only few choices at each level therefore he has to make more choices in order to reach his expected information. The aforementioned example would be considered as a deep menu structure.

If a structure has few levels it is usually called a shallow hierarchy. The main question now is how a menu - whether it is in a web or application context - should be presented to a user. This problem has been recognized early as an important factor for a well performing menu structure and was essential in the design of such a user interface.

1.1.2 Information Scent

The process when a human being is on the search for information is called information foraging. It refers to a theory which tries to understand how users hunt for information. This does not only apply to common tasks like their daily life, but is also a very important aspect of information architecture and web usability. When information foraging is taken into account, information scent plays a decisive role.

In this context it is considered how well links and navigation options match a visitor's information need and can be seen as a quality indicator. This is crucial for information systems, because when users are seeking for information they would like to know whether or not they are getting closer to the desired information. If they are not getting closer they are likely to be on the wrong path and tend to move to a different information source, because they are not confident anymore. In other words they doubt their choices and revert previous taken decisions [Kalbach, 2008].

1.2 Measure Layout Performance

Determining which layout performs best is usually achieved by a formal experiment, a study or survey.

First of all, test subjects are needed in order to conduct a survey. There is no fixed number how many participants are required. However, usually at least 20 subjects, sometimes up to one hundred, are recruited. Some studies prefer the presence of domain experts, because a familiarity with domain terminology may be required.

When enough test subjects have been recruited, they are asked to perform several tasks. These tasks usually consist of extracting a piece of information in a given hierarchical structured menu. During the tasks data about error rate and response time is recorded. Afterwards, the test subjects are asked which layout they preferred most. This satisfaction is not measured in a statistical way, but rather their personal opinion is gathered.

In conclusion the obtained data is analyzed and statistically evaluated. The criteria for a well performing menu structure is a low error rate and a fast response time. There are many different methods how to analyze and statistically compare such designs.

Examples of methods to analyze such data are as follows:

- Analysis of variance (ANOVA)
  - This method is useful to compare several means.
- Multivariate analysis of variance (MANOVA)
  - This method is useful to determine search time and search efficiency.

Chapter 2

General

2.1 Seven plus/minus Two Rule

In the early days, user interface designers applied the seven plus/minus two rule. It originated from a study of psychologist George Miller published in 1956. The study showed that human short-term memory is limited to retain only seven plus/minus two items [Miller, 1956].

Therefore, this rule was applied over a long time in the creation of menu layouts and web navigation. But it turned out that Miller's study contained no significance for this field of research. Subsequent studies have revealed that this rule cannot be directly applied to web navigation.

2.2 Broad and Narrow Layouts

One of the earliest studies which dealt with the impact of breadth versus depth in hierarchical menu structures was D.P. Miller. In 1981 he conducted an experiment where he tested four different menu structures, namely $64^1, 2^6, 4^3$ and 8^2. This factors refer to the number of items per level, i.e. 2^6 means that there are six levels and two items per level. No significant results were obtained, but he concluded that if the depth of the menu structure increased, so did the response time. D. P. Miller referred this finding to the limitation of human short-term memory [Miller, 1981].

Based upon Miller's research Snowberry, Parkinson & Sisson (1983) did a similar experiment. First they held an initial screening session where the memory capabilities of the test subjects were measured. Afterward Miller's experiment was replicated and they collected response times and error rates to evaluate the performance of the given menu layout. Additionally to the categories provided by Miller, they also introduced a fifth menu structure. Here the items were categorized in groups by relevance. The collected data of the experiment revealed that the memory span was less important than the visual scanning process of the subjects. Furthermore the fifth category where similar menu items grouped together, performed significantly better in regard of search time and error rate [Sisson et al., 1986].

In 1984 Kiger altered Miller's study. He kept all menu structures except the 64^1 was removed. Additionally he provided two new different structures, namely 16x4 and 4x16. Like previous research efforts, he came to the same conclusion that broader menu structures performed best in terms of error rate and response time. Participants had a significant faster response time and fewer errors with the 4x16 structure, but favored the 8^2 layout. In addition he identified slower search times when the menu structure had more than eight items [Kiger, 1984].

E. Eugene Schultz Jr. et. al. (1986) investigated the effects of alphabetically ordered menu structures. They tried to show that such ordering would improve search performance. However they came to the conclusion that this approach did not show a significant improvement over a random order [Schultz and Curran, 1986].

Jacko and Salvendy (1996) tried to understand how complexity in a hierarchical menu structures is perceived by the user. They came to the conclusion that with increasing depth, the perceived complexity also increases significantly. This is due the fact that users tend to get more uncertain when they navigate deeper and deeper in a menu structure [Jacko and Salvendy, 1996].

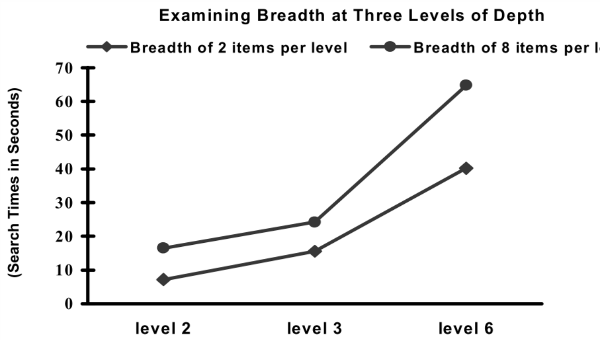

Figure 2.1: A figure that shows increasing perceived complexity with increasing depth.

figure source: [Bernard, 2002]

Conclusion

This was only a small extract of existing surveys that dealt with the breadth vs. depth issue in menu structures. Although there is no general recommendation how to solve this problem, there is a common understanding that has been proven to be successful. Namely that broader menus outperformed narrower menus significantly in error rates and response time.

However there is a constraint to this statement. That is that there exists a certain point where even broader menus introduce higher error rates and response times. This is in contrast to the subjective perception of the users who usually preferred layouts with the least amount of levels.

2.2.1 Where to Draw the Line?

Mathews J. Miles and Scott J. Bergstrom [Miles and Bergstrom, 2010] tried to answer this question in their 2009 study where they conducted a formal experiment with 120 students. The subject of this study was how the number of labels, used to categorize books of libraries, influence the response time.

The participants were asked to navigate to a library website and find a certain piece of information in the database. They collected lists from thirty university library websites in the UK, US and Canada. The number of labels to describe the subjects varied significantly. Some universities used smaller lists (e.g. five labels) to categorize their books, whereas others used up to 72 labels to describe a subject.

All participants were divided in twelve experimental groups and were assigned to a representative subject list from a library website. They were asked several questions, for example:

- Which category would most likely have information about modern graphical design?
- Which category would most likely have information on skateboarding?

The statistical analysis of the collected data resulted into two observations:

- The authors found no relevant pattern when all questions were considered individually. No significant result, whether a narrower or a broader list performed better, was gathered.
- However, if all questions were combined in an overall average graph, a clearer pattern emerged. That is that after approximately fifty items, the response time increases significantly. This can be seen in figure 2.2.

In addition it was evaluated whether or not label names have an impact on users' performance. A short test with three lists, each list consisting of items with different label names, was performed. The authors found no significant deviation for different label names.

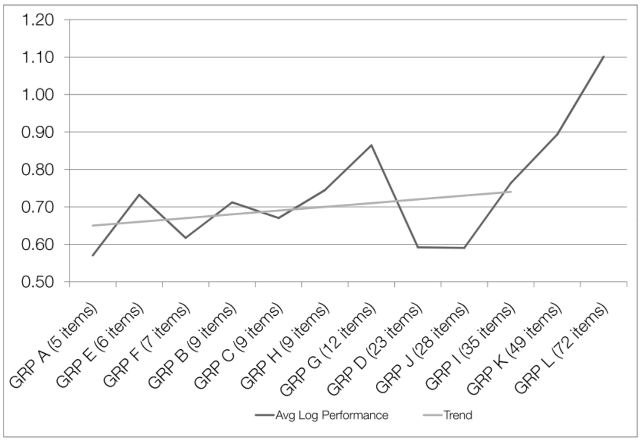

Figure 2.2: A figure that illustrates the error rates of all questions combined.

figure source: [Miles and Bergstrom, 2010]

2.3 Additional Shapes

2.3.1 Norman and Chin (1988)

The previous discussed studies have only dealt with the number of levels in a layout and the shape was not taking into account. Norman and Chin were interested in how they affect the search performance.

They examined the following five different menu structures as shown in figure 2.3:

- (4 x 4 x 4 x 4): menu structure with constant breadth
- (8 x 8 x 2 x 2): Broader selection at the top level with decreasing broadness at the bottom.
- (2 x 2 x 8 x 8): Narrower selection at the top level with increasing broadness at the bottom.
- (8 x 2 x 2 x 8): concave menu shape
- (2 x 8 x 8 x 2): convex menu shape

The total number of nodes varied heavily. They ranged from 294 up to 456 nodes, whereas the number of terminal nodes was fixed to 256 [Bernard, 2002].

Additionally, this study distinguished between explicit and implicit search tasks. For example in an explicit task the test subject was instructed to search concrete information like named items. Whereas an implicit task

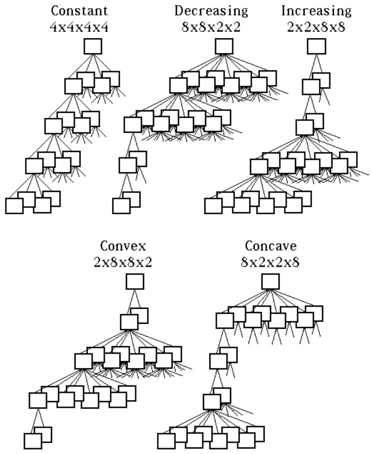

Figure 2.3: An overview of different layouts.

figure source: Norman, Kent L., http://lap.umd.edu/poms/chapter9/chapter9.html

participants were asked to find subjective information, like *"Search for the best web browser out there."*.

The collected data showed that the convex shaped layout performed worst. In addition, participants used the back button more often, which is attributed to their lostness. On the other hand, structures with the increasing and concave shape were performing best with the fastest search times.

There were two differences between those two structures:

- Participants using the concave shape needed the back button less often.
- Participants were less disoriented when using the increasing structure (2 x 2 x 8 x 8) for explicit tasks.

All in all a concave menu layout was considered better, although the increasing layout performed best in explicit tasks as seen in figure 2.4. This is due the fact that in a real world implementation implicit and explicit search terms always occur together and thus they are equally important.

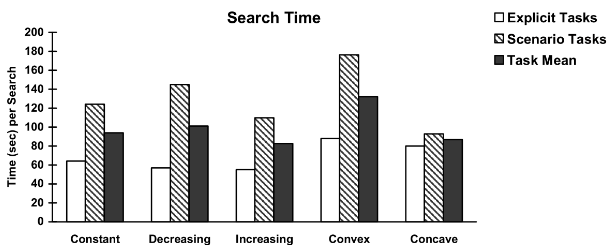

Figure 2.4: Mean search time results [Norman, 1991]

figure source: [Bernard, 2002]

Norman and Chin came to the conclusion that shape has a significant impact on search performance in hierarchical structures.

Their recommendation for a good performing concave menu structure is:

- a broader selection at the root level.
- a broader selection at the leaf level.

- a narrower selection between root and leaf level.
- reduce as many menu frames as possible.

2.3.2 Bernard 2002

In a more recent study Michael L. Bernard (2002) took a closer look on symmetrical and asymmetrical menu structures. He tried to address whether or not a correlation between shape and depth does exist. Like Norman and Chin's study from 1988, he examined many different shapes with the same depth. However, he had a slightly different approach. In the aforementioned study the total number of nodes varied heavily. Thus Bernard argued that:

> ”...this study sought to have approximately equal number of nodes across hypertext conditions since smaller structures are generally easier to search than large structures.” [Bernard, 2002]

He was interested whether or not this slightly changed condition would have an impact on Norman and Chin's findings. He managed to even out the total amount of terminal level nodes to 330 ± 10. In his study 120 participants took part. Like in the study by Norman and Chin a task consisted either in finding an implicit or explicit target.

He found that broader structures performed significantly better. Bernard also confirmed Norman and Chin's findings that participants had a higher search performance when looking for explicit rather than implicit tasks, although this finding was not significant.

His study also showed that symmetrical shaped layouts were not performing as well as asymmetric shapes. For example, the (4 x 4 x 4 x 4) layout did not only perform worse than shapes of the same depth, but also worse than more deeper structures like (3 x 2 x 2 x 2 x 12).

Thus he agreed with Norman and Chin that shapes have a significant effect on search performance and concave shapes are more navigationally efficient and lead to less errors.

A good starting path for a concave shape consists of a broad root level where all labels should be well distinguishable. This helps the user to gain a good overview without being too overwhelmed. Furthermore like Bernard said, it *”helps the user form a more exact match between the concept related to the target item and the actual target item itself”* [Bernard, 2002]. The levels at the middle should be more restricted in order to reduce uncertainty and limit complexity. Broader middle levels would decrease the users confidence and increase error rates, because it is more likely that a user would take a wrong path. This is in contrary to the terminal level. At this stage a broader menu structure leads to more information certainty and information scent. He argues that the association between a descriptor at a the entry point and the descriptor at the terminal level should be as explicit as possible. This concept refers to information scent and was seen very important by Bernard.

Concluding he tried to empathize that relying solely on the tradeoff between breadth vs. depth may not be enough, even for web design. Bernard also pointed out that shapes play an equally important role and should not be ignored. His findings came to the same conclusion as Norman and Chin, namely that concave shaped layouts reduce uncertainty and shall be preferred.

2.4 What Approaches are there to Organize Layouts?

2.4.1 Card Sorting

In this approach all items which should appear in the final menu must be identified first. After that is done all collected items are written to cards and are distributed incoherent on a table. Then the test subjects were instructed to arrange related cards into groups. This task can be achieved with two different methods, either a bottom-up or top-down approach.

Chin (1986, 1987) conducted a study to find out which of them performs better. Therefore he recruited test subjects and divided them into two groups. The first group should sort the cards in a bottom-up and the second in a top-down manner.

From his findings could be concluded that the top-down approach resulted in greater menu depth and the bottom-up approach resulted in greater menu breadth.

These results have not led to prefer one of the approaches, but that both extract important bits of information. The final conclusion was to combine the results of both methods and it turned out that this can be done very effectively.

A disadvantage of card-sorting is that it only provides good results if the test is performed with test subjects out of the intended audience. If there are multiple target audiences no ideal result can be achieved with this procedure [Norman, 1991].

2.4.2 Semantic Space

Another approach was to group items which were most frequently used or were most similar. How these arrangements affect the menu layouts performance was investigated by McDonald, Stone, Liebelt & Karat (1982). In their study items were grouped together based on similarity and frequent occurrences. In addition they allowed their test subjects to arrange the menu items according to their taste.

The test results revealed that there was no significant difference between the frequency and the similarity layout. Also it was shown that the personally arranged layout performed worst. When considering real-world application menu layouts, they came to the conclusion that frequency based layouts are in general better performing than similarity based layouts [Norman, 1991].

Chapter 3

Mobile

When we are moving away from the desktop we suddenly enter a completely different world concerning usability and navigation. In this chapter we will discuss some of the problems we encounter and we will discuss a study to this topic.

3.0.3 Problem

In general we can narrow down the problem to two factors:

- the reduced screen size and
- the restricted input capabilities

Due to the restricted screen size the reading speed decreases up to 25% and also the information retrieval capabilities reduce. The lack of keyboard and mouse also make it harder for the user to navigate through structures.

3.1 Depth and Breadth away from the Desktop

In this study by Geven et. al. in 2006 Geven et al. [2006] (done before the release of the iPhone and smartphones in general) they made an experiment with 15 novice users from whom 14 returned a second time as intermediate user and a third time as expert user. The age varied between 17 and 39.

3.1.1 Structures tested

The study considered 4 different hierachy structures: 4 x 6, 8 x 4, 16 x 3 and 64 x 2 and the tests were executed on following devices 3.1.

3.1.2 Results

For each test the maximum of time to find the target item was 150 seconds. As a Result they found that not only the time differed significantly between the three devices but that the narrow structures 4 x 6 and 8 x 4 performed better that the others.
As seen in the figure 3.2 there also seems to be a correlation between the size of the used device and the preference towards a broader structure.

Figure 3.1: Devices used for the study.

figure source: [Geven et al., 2006]

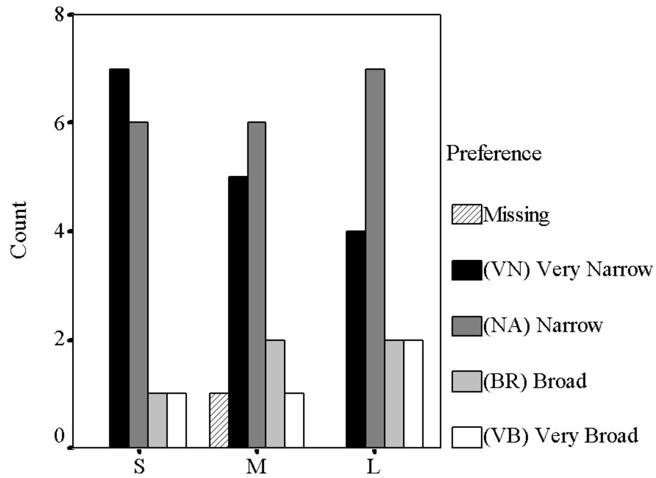

Figure 3.2: User preference for hierarchy, for each of the devices.

figure source: [Geven et al., 2006]

3.2 Conclusion

Concerning mobile devices we can conclude that narrow structures are preferred because of the limited screen size and input capabilities. However, in times of smartphones and tablets the handling of such devices differ from those used in the study and there is room for plenty of new studies.

Chapter 4

People

Regarding the Breadth vs. Depth topic in Information Architecture it is also very important to examine if the previously mentioned findings hold true for different user types in terms of different abilities and different models of computer usage. As not everybody is using or can use computers in the same way there might be significant differences between optimal depth/breadth parameters for particular subpopulations.

Differences in abilities and computer usage models may be found in various properties of the users. One might distinguish people by their age, where disrepancies in terms of experience, working memory performance or eyesight might occur, which could lead to other information retrieval strategies and thus to optimum depth/breadth parameters differing from those for typical users. Also, users with poor or no eyesight form a special group of interest because they experience the computer interface in another way than sighted users.

In this chapter we are going to present work regarding these subpopulations, this means age-related and eyesight-related differences in particular.

4.1 Age-related differences

Snowberry et. al. (1983) [K. Snowberry, 1983] isolated in their study three 'kinds of forgetting', namely

- forgetting the target word,
- forgetting the pathway in deep menus,
- and wrong association of options,

which due to age-related impairment of short time/working memory may be relevant especially for older computer users. However, in their study they found that broad and shallow menus are generally preferable, what replicates previously described results, but does not answer any questions concerning age-related differences amongst users.

Mead et. al. (1997) [S. Mead, 1997] also did a study regarding age (and training) of web users. They expected seniors to be slower, have limitations in spatial abilities and get more easily lost in depth and breadth of hierarchical information structures.

Indeed, older users have been observed to undergo much longer wanderings through structures than younger and middle-aged users, and they also often used exhaustive search strategies like clicking through the whole test web link by link, visiting all possible pages until the target has been found.

In this study they found seniors to

- be less likely than younger users to complete all given tasks
- have overall higher task completion times,
- but to be as likely to take the optimal path to the target as young and middle-aged users.

They state that there are (significant) differences in task completion performance between age-resolved user groups, but they found that there are only minor differences of optimal breadth and width parameters between these groups.

Zaphiris et. al. conducted a study in 2003 [Zaphiris P., 2003] where they also found significant Depth*Age group effects, namely older users being slower in all depth/breadth configurations without effect on optimal parameters respective to those of common users. They also found that seniors prefer non-expandable menus while younger users prefer expandable hierarchies.

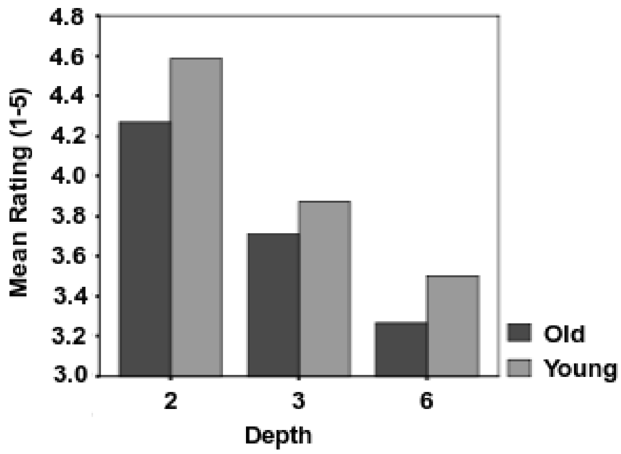

Figure 4.1: Mean rating for ease of navigation by depth for young vs. older users (higher is better) [Norman, 1991]

figure source: [Zaphiris P., 2003]

4.2 Eyesight-related differences

In their study from 1983 Snowberry et. al. [K. Snowberry, 1983] also found out that instead of memory span, visual scanning abilities are predictive of performance, especially in the deepest hierarchies. This is quite interesting regarding users with little or no eyesight, as they cannot rely on visual scanning methods while navigating through information hierarchies, but have to use other techniques instead.

As Hochheiser & Lazar [Hochheiser H., 2010] stated in their publication from 2010 considering breadth vs. depth in menu structures for blind users, menu search is generally not a strictly linear process. The common menu search process involves partially serial scanning as well as random hopping through the variety of possible choices.

However, blind users have to rely on text-to-speech software, which reads the menu items serially. Here, a model of auditory menu selection is applied, which is rather different from the common strategies. Using this method, retention of just two menu items in the working memory is needed. The user only has to keep a 'best-of item' (strongest candidate) in memory, to which he compares the other serially presented choices item-by-item. Thus, the short time memory load is constant and independent of menu breadth, substantiating the recommendation to use broad and shallow structures, where depth is minimised while the increase in breadth is not affecting memory load.

4.3 Conclusion

Considering the work mentioned in this chapter one can conclude that there are differences in overall performance between particular user groups, which result from group specific properties or inabilities of the different users. More interestingly, the optimum depth and breadth parameters have not been found to differ significantly for those groups respective to common users. Given that there are only minor differences in the optimal parameters, we can recommend the usage of parameters found to be optimal for the general population also for older users or those with impaired eyesight.

Chapter 5

Concluding Remarks

Now that some of the most important studies have been discussed, it is time to ask whether or not there is a common practice on how to create hierarchical structures. Unfortunately, this question can not be answered in a general way.

However there are some recommendations given. Like Miller and successional studies have shown, broader menus are usually performing better than narrower ones in regard of search performance. Moreover, there are clearly some preferences towards asymmetrical shapes, especially the concave shape, as discussed in chapter 2.3.

Unfortunately this may not be the best case for every application. If, for example the number navigation mistakes are high, the search performance might be better in a narrower layout. Likewise, as mentioned in chapter 3.

References

Bernard, Michael Lewis [2002]. *Examining a metric for predicting the accessibility of information within hypertext structures*. PhD Thesis, Wichita, KS, USA. AAI3063270.

Geven, Arjan, Reinhard Sefelin, and Manfred Tscheligi [2006]. *Depth and breadth away from the desktop: the optimal information hierarchy for mobile use*. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pages 157–164. MobileHCI '06, ACM, New York, NY, USA. ISBN 1-59593-390-5. doi:http://doi.acm.org/10.1145/1152215.1152248. `http://doi.acm.org/10.1145/1152215.1152248`.

Hochheiser H., Lazar J. [2010]. *Revisiting breadth vs. depth in menu structures for blind users of screen readers. Interact. Comput.*, 22, pages 389–398.

Jacko, J. A. and G. Salvendy [1996]. *Hierarchical menu design: breadth, depth, and task complexity. Perceptual and Motor Skills*, 82. ISSN 1187-1201.

K. Snowberry, N. Sisson, S.R. Parkinson [1983]. *Computer Display Menus. Ergonomics*, 26, pages 699–712.

Kalbach, James [2008]. *Uncertainty in Information Architecture*. `http://experiencinginformation.wordpress.com/tag/breadth-v-depth/`.

Kiger, John I. [1984]. *The depth/breadth trade-off in the design of menu-driven user interfaces*. *Int. J. Man-Mach. Stud.*, 20, pages 201–213. ISSN 0020-7373. doi:10.1016/S0020-7373(84)80018-8. `http://dl.acm.org/citation.cfm?id=217.222`.

Miles, Mathews J. and Scott J. Bergstrom [2010]. *Classification of Library Resources by Subject on theLibrary Website: Is There an Optimal Number of Subject Labels? InformationTechnology & Libraries*, 28.1. `http://www.ala.org/lita/ital/sites/ala.org.lita.ital/files/content/28/1/miles.pdf`.

Miller, D. P. [1981]. *The Depth/Breadth Tradeoff in Hierarchical Computer Menus*. pages 296–300. `http://www.mendeley.com/research/the-depthbreadth-tradeoff-in-hierarchical-computer-menus`. Annual Human Factors Society Meeting (1981).

Miller, George A. [1956]. *The magical number seven, plus or minus two: some limits on our capacity for processing information*. *Psychological Review 63*, 101(2), pages 81–97. doi:10.1037/h0043158. `http://www.psych.utoronto.ca/users/peterson/psy430s2001/Miller%20GA%20Magical%20Seven%20Psych%20Review%201955.pdf`.

Norman, Kent L. [1991]. *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*. Greenwood Publishing Group Inc., Westport, CT, USA. ISBN 089391553X.

S. Mead, A. Sit, A. Spaulding [1997]. *Effects of age and training on world wide web navigation strategies. Proc. Human Factors Society*, pages 152–156.

Schultz, E. Eugene, Jr. and Patrick S. Curran [1986]. *MENU STRUCTURE AND ORDERING OF MENU SELECTION: INDEPENDENT OR INTERACTIVE EFFECTS? SIGCHI Bull.*, 18, pages 69–71. ISSN 0736-6906. doi:http://doi.acm.org/10.1145/15683.1044100. `http://doi.acm.org/10.1145/15683.1044100`.

Sisson, N., S. R. Parkinson, and K. Snowberry [1986]. *Considerations of menu structure and communication rate for the design of computer menu displays*. *Int. J. Man-Mach. Stud.*, 25, pages 479–489. ISSN 0020-7373. doi:10.1016/S0020-7373(86)80018-9. `http://dl.acm.org/citation.cfm?id=30043.30045`.

Zaphiris P., Ellis R.D., Kurniawan S.H. [2003]. *Age Related Differences and the Depth vs. Breadth Tradeoff in Hierarchical Online Information Systems. Lecture Notes in Computer Science Universal Access Theoretical Perspectives Practice and Experience Proceedings of 7th ERCIM Workshop User Interfaces for All.*