

The Influence of Adaptation on Hypertext Structures and Navigation

Vinicius F. C. Ramos
Federal Univ. of Rio de Janeiro
Rio de Janeiro, Brazil
vfcr@cos.ufrj.br

Paul M. E. de Bra
Eindhoven Univ. of Technology
Eindhoven, The Netherlands
debra@win.tue.nl

ABSTRACT

In adaptive hypertexts the user is guided in two ways: through the existence of links and through link annotation or hiding. Link structures have been investigated, starting with Botafogo et al, and the effect of link annotation has been studied, for instance by Brusilovsky et al. This paper studies the *combined* effect of link structure and annotation/hiding on the navigation patterns of users. It defines *empirical hubs* and studies their correlation with *hubs* as defined by Kleinberg without considering adaptation. The data for the analysis have been extracted from the logs of the course “Hypermedia Structures and Systems,” an online adaptive course offered at the Eindhoven University of Technology.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation (I.7)]:
Hypertext/Hypermedia—*Navigation*

General Terms

Experimentation, Human Factors, Measurement

Keywords

Adaptive Hypermedia, Navigation, Evaluation, Structural Analysis

1. INTRODUCTION AND RELATED WORKS

Adaptive Hypermedia (AH) systems create an individual model for each user based on her goals, preferences, and knowledge, and use the model to adapt the system to the needs of the user [2]. AH systems have been widely used in online distance learning, where adaptation is employed as a means to improve the learning experience of each student. In the adaptive educational hypermedia (AEH) system AHA!, developed by De Bra et al. [7], the on-line learning material is adapted in two ways: the adaptation engine can present links in different colors depending on whether it considers

the student ready to study the link destination (page), and it can conditionally include fragments and objects in the page, for instance to show a prerequisite explanation. Link colors or colored icons suggest which links are relevant to each student; this approach is for instance discussed in Romero et al. in [13].

Recent years have witnessed much effort on the development of frameworks for adaptive Web-based educational systems [3] and on the evaluation of AEH systems [9]. The work on AHA!, used for the course “Hypermedia Structures and Systems” at the Eindhoven University of Technology, started in 1996 [6] and its considered a pioneer work in the area. This AEH system is under continuous development [7]. For this paper AHA! version 3.0 was used [8]. We briefly describe the AHA! system in Section 2.

Empirical studies carried out by Kaplan et al. [11] and Brusilovsky and Pesin [5] suggest that adaptive navigation support can speed up the navigation and learning of the students. A good survey about the state of the art of AEH systems can be found in [4], while a more specific treatment on evaluations of AEH systems is presented in [9].

In this work we analyze the structure of an adaptive course created with AHA!. We use data obtained from the access log of the students. To understand the way users navigate through an adaptive course, our goal is to identify which pages and links influence the choices of the students and contrast this with the rules employed by the adaptation engine. Another goal is to verify whether the adaptation rules, created by the author of the course, influence the navigation of the students.

In non-adaptive hypertext the first significant research on the navigation structure was done in [1]. An analysis of link structures makes sense as the structure is independent of how the hypertext is used, unlike with adaptive hypertext where the link structure (of *available* links) may change while the user is navigating the hyperspace. While [1] concentrated on *compactness* and *stratum* more recent link structure analyses have focused on *hubs* and *authorities* in the context of hyperdocuments as proposed by Kleinberg [12]. Typically, hubs have a lot of outgoing links, and authorities a lot of incoming links. In this paper we also introduce the notion of *informative pages*, which are pages that not only have a large number of links, thereby being good candidates for *hubs* but also carry information that *guides* the student through the links in the page, an *alphabetical index* of keywords is an example of a strong *hub* that is not an *informative page* as it has no textual information to help the user in selecting a link to follow.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'10, June 13–16, 2010, Toronto, Ontario, Canada..

Copyright 2010 ACM 978-1-4503-0041-4/10/06 ...\$10.00.

The goals of this paper are twofold: we want to come up with a definition of structure that takes the existence of adaptation into account, but that can be verified on the definition of an adaptive hyperdocument (not considering actual use), and we want to verify the structure that emerges from the actual use of the adaptive hyperdocument by the students and analyze the influence of link existence versus the adaptive link annotation of hiding.

2. AHA! AND THE ADAPTIVE HYPERTEXT COURSE

The *AHA!* system is a Web-based adaptation engine and authoring tool developed at the Eindhoven University of Technology, starting in 1996. In this paper we consider *AHA!* version 3.0.

An adaptive course in *AHA!* consists of concepts that are connected to pages. The pages contain information and links to concepts. (Links must refer to concepts instead of pages, and the link destination page may be adaptively selected by the adaptation engine.) A page may also contain links to “external” pages on the Web, but we shall not consider these here. One approach used in *AHA!* to adapt a course is to use different colors for the links of the pages. The author of the course defines rules to determine the conditions under which a presentation class is associated with a link. The default presentation classes are *bad*, *good*, and *neutral*, and have the following meaning and presentation style:

- The *bad* links point to *non-recommended* concepts, which means that according to the rules defined by the author, the student is expected to study something—do some reading or perform some tests—before accessing these concepts. *Bad* links are colored in black and are not underlined, which implies that they are indistinguishable from the textual information of the page. So, bad links are hidden within the text, though they are fully operational and can be clicked on at any time.
- The *good* links point to a recommended concept that the student has not yet visited after it became recommended. Good links are colored in blue.
- The *neutral* links point to a recommended concept that the student has already visited after it became recommended. Neutral links are colored in purple.

Because of the choice of adaptation and colors the hypertext course looks like a standard (non-adaptive) website with pages and blue and purple links.

The data considered for this paper consists of the access log of 127 students of the hypertext course offered during 2006 and 2007. The course was offered only through distance learning and had students from Dutch and Flemish universities. The access log contains 62993 entries, but we used only 34850 entries in our study. We found that 19886 entries referred to 51 students for which information regarding the test they performed was missing. We also disregarded 9257 entries for which the “referrer” information necessary to identify the link used to reach a concept or page was missing.¹

¹There are several possible causes for this problem such as the Web browser (e.g., Opera) may reuse pages from its local cache or the students may open a page in a new tab or window.

3. EMPIRICAL ANALYSIS OF LINK USAGE

We start our analysis by noticing that even though links associated with the presentation class *bad* were hidden within the text, the number of accesses via bad links was substantial: roughly 6.58% of all the links followed were *bad* links, while 42.29% were *good* links and 51.13% were *neutral* links. The issue of users following non-recommended links has been observed in other systems but was never studied closely. Our goal in the sequel is to analyze this issue further in order to discover why students go against the recommendation of the author of the course and access *bad* links.

Consider Figure 1, which shows the presentation classes used to access the 12 main concepts of the course. We consider main concepts to be the concepts that introduce a topic in the course; the other concepts are subconcepts of some main concept. The main concepts are listed in (and linked from) the course’s welcome page. Since this page provides a bullet list of topics the *link hiding* technique does not really succeed in hiding links here, but it does provide annotation by not showing blue links for the non-recommended links.

Note in Figure 1 that the concepts *definition*, *history*, and *intro* are never accessed through *bad* links because these links are always recommended. These concepts represent the first three main concepts or chapters of the course. Each chapter contains a multiple choice test. The “advanced” topics only become recommended after the student has completed the tests of the introductory chapters. (Only taking the tests is considered, not the score that is obtained.)

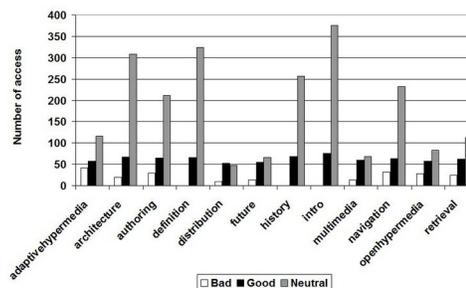


Figure 1: Total number of accesses performed by the students to each main concept of the course. The colors represent the presentation class of the link used to access the concept.

The eight “advanced” concepts shown in Figure 1 have sometimes been accessed through a *bad* link. The *Adaptive Hypermedia* concept has 42 accesses via *bad* links and 174 accesses via *good* or *neutral* links, giving roughly 20% of its accesses via *bad* links. We now focus on the *bad* accesses to this *Adaptive Hypermedia* concept. One might think that most *bad* accesses to any concept would come from the “welcome” page where the non-recommended links are still visible in a bullet list. However, the collected data shows that less than 12% of the *bad* accesses to “Adaptive Hypermedia” came from the welcome page; hence, most bad accesses came from pages in which the link to “Adaptive Hypermedia” is effectively invisible (colored black and appearing in the main text). This suggests a behavior that we refer to as curious browsing: through the welcome page the students are hinted about the existence of the *Adaptive Hypermedia* concept,

but they are not immediately drawn to it. But throughout the text, on different pages, the concept is mentioned again and again, and at some point the student’s curiosity prevails and she finds out that the term is actually a hidden link and clicks on it. In addition to “being mentioned” the term *Adaptive Hypermedia* is named as a technology used in the hypertext course itself and students may wish to find out what that actually does.

We dwell on the idea of curious browsing a little further, and show in Figure 2 that 24 students accessed the *Adaptive Hypermedia* concept via *bad* links out of a total number of 42 accesses, which shows that some students accessed this concept *more than once* via *bad* links. We also point out that 76 students appear in the access log, so roughly 30% of the students accessed *Adaptive Hypermedia* via *bad* links.

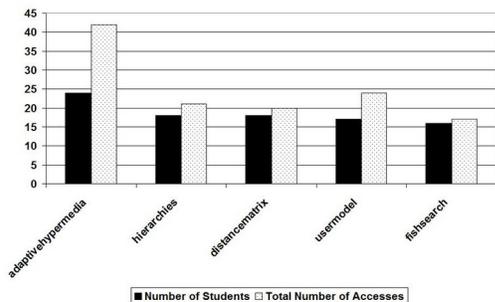


Figure 2: Number of students that accessed a concept via *bad* links. We show the five concepts with the largest number of student accesses via *bad* links.

The concepts *hierarchies*, *distancematrix*, and *usermodel* can be reached from the concept *Adaptive Hypermedia*; so Figure 2 shows that students who follow a *bad* link to read about *Adaptive Hypermedia* do not stop at the first page but look for more hidden links to detailed information from that concept. Aside from *Adaptive Hypermedia*, the other concepts had roughly an average number of 1.5 accesses via *bad* links per student, which suggests that after revisiting the page, the students refrained from following the *bad* link again and waited until the link became *good*. This gives evidence that the students generally abided by the adaptation rules and that the adaptive engine of *AHA!* was to a large extent successful in guiding the students’ navigation. In order to reinforce our suggestion that curious browsing for “Adaptive Hypermedia” was caused by the topic being mentioned in several places, we note that the in-degree (number of links pointing to the concept regardless of the link’s presentation class) of the “Adaptive Hypermedia” concept is 14, which is the fifth largest in-degree in the course.

The observations in this section show a connection between the link adaptation and link structure: the adaptive guidance offered by the system is generally followed by the students, except for concepts that turn out to be an *authority*, i. e. that have many incoming links. In the next section we study link structure further to find out how adaptation and link structure influence each other.

4. HUBS AND INFORMATIVE PAGES

Kleinberg [12] created an important algorithm called *HITS* that uses ideas from random walks and measures each page

according to two classes, called *hubs* and *authorities*. Intuitively, good *authorities* are considered the interesting Web pages (i. e., Web pages that people want to visit), while good *hubs* are the Web pages that have links to good *authorities*.

In the previous section we noted that *authorities* may counteract the effect of link adaptation: *authorities* are visited even when the system recommends against them. Here we study the role of *hubs* in adaptive courses. In general, *hubs* are pages containing a large number of links (i. e., a large out-degree), but since in an adaptive course the number of links can change, we need a new definition for *hubs*.

Figure 3 shows a rank of the concepts whose links were most frequently used to access *other* concepts. We remark that twelve of the concepts presented in Figure 3 have the largest out-degree in the course. This shows how a concept is used as a *hub*, since some concepts not only have large out-degree, but also contain links that were indeed used by the students to access another concept. In Section 4.1, we investigate the idea of *hubs* for an adaptive course with the notion of *empirical hubs*. In Section 4.2 we introduce the idea of *informative pages* to illustrate why the existence of many pages with a large out-degree was not a problem in the hypertext course.

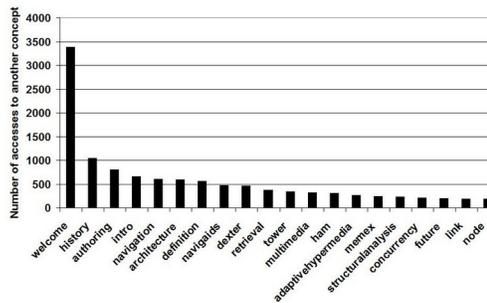


Figure 3: Concepts whose links were most frequently used to access another concept.

4.1 The Empirical Hub Coefficient

We define the *Empirical hub coefficient* (EHC) of a page X as the ratio between the number of times that students clicked on a link of X to go to a different concept and the number of times that students accessed X . Intuitively, pages with large EHC are the ones used as hubs. Note that the EHC is a number between 0 and 1, where 1 means that each time a student visited page X she clicked on a link to another concept and 0 meaning that nobody ever clicked on a link in X that leads to a different concept (most likely because X has no links).

Table 1 contrasts the out-degree with the EHC of the pages of the course. The rows represent the EHC split into intervals of 0.05, and the columns represent the out-degree. Entry (i, j) of the table then corresponds to the number of pages having EHC falling into the interval represented by row i and out-degree represented by column j .

In the hypertext course, links to a given concept could be found in many other concepts. The course is quite densely linked; links to a single concept can even be found in the subconcepts of that concept, which are called *chapters*. This tends to decrease the EHC of the concepts since students are

| EHC \ Out-degree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|------------------|----|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
| 0 - 0.05 | 19 | 3 | | | | | | | | | | | | | | | | |
| 0.05 - 0.1 | 11 | 4 | 2 | | 1 | | | | | | | | | | | | | |
| 0.1 - 0.15 | 1 | 8 | 1 | 2 | 2 | | | | | | | | | | | | | |
| 0.15 - 0.2 | | 1 | 1 | | 1 | 1 | | | | | | | | | | | | |
| 0.2 - 0.25 | | 5 | 5 | | 2 | | | | | 1 | | | | | | | | |
| 0.25 - 0.3 | | 1 | 3 | | 1 | 1 | | 1 | | | | | | | | | | |
| 0.3 - 0.35 | | 1 | 2 | | 1 | 1 | 1 | | 1 | 1 | | | | | | | | |
| 0.35 - 0.4 | | 2 | 1 | 3 | | 1 | | 1 | | | | | | | | | | |
| 0.4 - 0.45 | | 3 | 3 | 1 | 1 | | | 1 | | | | | | | | | | |
| 0.45 - 0.5 | | 2 | 1 | 1 | 2 | 1 | 1 | 2 | | | | 1 | | | | | | |
| 0.5 - 0.55 | 1 | | 1 | | | 2 | 1 | | 1 | 1 | | | | | | | | |
| 0.55 - 0.6 | | | 1 | 1 | 1 | 2 | | 1 | 1 | | 2 | | | | | | | |
| 0.6 - 0.65 | | | | 1 | 3 | | 1 | 1 | | | | | | 1 | | | | |
| 0.65 - 0.7 | | | | 2 | 1 | 1 | | 2 | 1 | | | | | | 2 | | | |
| 0.7 - 0.75 | | | | | | 1 | 1 | | | | | | | | | | 1 | |
| 0.75 - 0.8 | | | | | | | | | | 1 | | | | | | | | 1 |
| 0.8 - 0.85 | | | | | | | | | | | | | | | | | | |
| 0.85 - 0.9 | | | | | | | | | | | | | | | | | | |
| 0.9 - 0.95 | | | | | | | | | | | | | | | | | | |
| 0.95 - 1 | | | | | | | | | | | | | | | | | | |

Table 1: Number of pages with EHC according to the row and out-degree represented by the column.

| | Out-Degree | Concept | Total Number Of Users | First Access | | First Click | | First Access with Click | |
|----|------------|----------------|-----------------------|--------------|------|-------------|------|-------------------------|------|
| | | | | Students | Time | Students | Time | Students | Time |
| 1 | 18 | history | 76 | 59 | 86 | 68 | 73 | 57 | 79 |
| 2 | 17 | authoring | 76 | 60 | 89 | 68 | 88 | 57 | 98 |
| 3 | 15 | intro | 76 | 50 | 182 | 73 | 116 | 49 | 174 |
| 4 | 14 | dexter | 75 | 57 | 135 | 64 | 148 | 56 | 145 |
| 5 | 14 | welcome | 76 | 61 | 71 | 75 | 62 | 61 | 71 |
| 6 | 12 | openhypermedia | 74 | 52 | 99 | 57 | 74 | 46 | 87 |
| 7 | 11 | definition | 76 | 49 | 159 | 70 | 145 | 49 | 171 |
| 8 | 11 | www | 76 | 62 | 159 | 59 | 178 | 51 | 184 |
| 9 | 10 | link | 76 | 60 | 106 | 60 | 128 | 51 | 119 |
| 10 | 10 | microcosm | 66 | 48 | 78 | 19 | 53 | 17 | 48 |
| 11 | 10 | navigaids | 72 | 58 | 24 | 64 | 26 | 54 | 25 |
| 12 | 10 | scripting | 68 | 50 | 89 | 14 | 81 | 14 | 81 |
| 13 | 9 | dynamicview | 67 | 45 | 109 | 33 | 119 | 29 | 123 |
| 14 | 9 | hyperties | 75 | 59 | 130 | 55 | 115 | 49 | 129 |
| 15 | 9 | intermedia | 74 | 61 | 93 | 57 | 92 | 51 | 94 |
| 16 | 9 | navigation | 76 | 54 | 86 | 69 | 80 | 52 | 93 |
| 17 | 9 | sculptural | 66 | 46 | 113 | 40 | 103 | 37 | 114 |
| 18 | 9 | wwwserver | 68 | 54 | 91 | 48 | 85 | 43 | 87 |

Table 2: First access information for some concepts. The table shows the out-degree, the number of students who accessed each concept, and three types of first access: *first access*, which consists of the very first access of a student to the concept provided she followed a link or pressed the back button of the browser after visiting the concept, *first click*, which is the first time a student followed a link from the concept, and *first access with click*, which is the intersection of the previous two. For each type of access, the table shows the number of students and the average time in seconds they spent on the page.

unlikely to revisit concepts that they have already visited while studying another chapter. On the other hand, Table 1 shows that there are many pages with out-degree between 3 and 9 and with EHC larger than 0.5. This may be another consequence of the absence of a centralized menu. In this course, when finding a link, students seemed to prefer to follow it immediately instead of waiting for the link to appear in subsequent pages. Such behavior (“let’s visit the page now because we risk not finding it again later”) contributes to having pages with large EHC.

A standard statistical measure that can be used to see the correlation between the out-degree of a page and its EHC is the correlation coefficient [10]. For the data shown in Table 1, the correlation coefficient is 0.71, which means a large out-degree implies large EHC.

4.2 Informative Pages

As Table 1 shows, this course has many pages with large out-degree, which can be a problem since a myriad of links in a page may generate confusion about which link the student should follow. On the other hand, pages with large out-degree are beneficial for decreasing the depth of the link structure of the course, i. e., the minimum number of links

that must be followed between any two pages that are furthest apart in terms of number of links.²

The hypertext course has 18 concepts with out-degree larger than 8, which is a quite large number. We examine this issue further with Table 2, which shows concepts ordered by out-degree. For each concept, it shows the total number of students that accessed that concept and three types of first access: first access, first click, and first access with click. For each of the three types of access, Table 2 shows the number of students that had an access of that type to the concept, and the average time in seconds that the students spent in the Web page during such access.

Table 2 reveals an important characteristic that motivates us to define *informative pages*. We classify a page as an *informative page* if it contains information that can guide the student on the decision about which link she should follow. Therefore, the presence of *informative pages* mitigates the problem inherent to pages with large out-degree. An example of an *informative page* is the *history* concept. This concept contains textual information regarding the order according to which the student should follow its links. Note that for the *first access with click*, students spent 79 seconds on the page on average, which gives evidence that the student indeed read the textual information in the page before

²In the terminology of [1] this would imply a high *compactness* value.

following a link. On the other hand, consider the concepts *welcome* and *navigaids*. The concept *welcome* is special because on the first visit (when just starting the course) it actually shows a “readme” page which has a lot of information and just one link, to itself. After that it becomes a page with almost no information and a large number of links. The “readme” is studied carefully, requiring 71 seconds on average. The concept *navigaids* is a typical example of a non-informative page: it has very little text, and a large number of links. There is not enough text to really guide the student into choosing a particular link. The reading time of the first access is only 25 seconds.

Now, consider the concepts *microcosm* and *scripting*. For *first access*, note that 48 students accessed *microcosm* and 50 students accessed *scripting*. However, for *first click*, only 19 and 14 students accessed *microcosm* and *scripting*, respectively. This means that most students that accessed these concepts never followed their links; they most likely clicked in the back button of the browser. We see this kind of concepts as (de-facto) dead ends.

Table 2 corroborates the idea that *informative pages* play an important role guiding the students through the maze of links that Web pages with large out-degree can create. We observe that students spent a large amount of time in their first access to the concepts, suggesting that they stopped to read the information provided in the page. *Informative pages* combine the functionality of a *hub* with a normal course page, whereas other pages that are just *hubs* but have a short reading time offer a lot of choice but no guidance. Such pages are to be avoided. The author should add information to the pages that can guide the student through the many links they offer.

5. CONCLUSION

In this work we have presented a case study of an adaptive course implemented with *AHA! version 3.0*. Our focus has been to empirically check whether the adaptation mechanism influences the students’ behavior. We have approached this issue from different angles. We have performed an analysis of the presentation classes of the links in the course, focusing on detecting if students followed the so-called *bad* links and if their navigation behavior was affected by the presence of *bad* links. We found that pages that can be classified as *authorities*, having many incoming links, make the students curious and thus cause students to investigate them even when they are not yet recommended. When an author wishes for students to follow the adaptive guidance, *authorities* that are not recommended from the start should be avoided. We have also carried out a structural analysis of the course, comparing the definition of *hubs* to new empirical measures, called *empirical hub coefficient* and *informative pages*. We have observed that *empirical hubs* arose in the course, most of which turned out to also be *informative pages* that guided the navigation of the students through the large number of links that *empirical hubs* contain. We also found non-informative *hubs* that give the student many choices but no guidance. An author should ensure that all *hubs* are also *informative pages*.

In the near future, we plan to carry out another case study of the hypertext course that is now re-implemented with the newer version of *AHA!*, called *GALE* and part of the EU FP7 project GRAPPLE. The newer version has a menu to navigate through the course; therefore, a comparison be-

tween the two versions may shed more light on the issue of verifying how the structure of the course and the navigation aids that are offered impact students’ navigation.

6. REFERENCES

- [1] R. A. Botafogo, E. Rivlin, and B. Shneiderman. Structural analysis of hypertexts: identifying hierarchies and useful metrics. *ACM Transactions on Information Systems*, 10(2):142–180, 1992.
- [2] P. Brusilovsky. Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3):87–129, 1996.
- [3] P. Brusilovsky. Adaptive educational hypermedia. In *Tenth International PEG Conference on Intelligent Computer and Communications Technology - Learning in On-Line Communities*, pages 8–12, Tampere, Finland, 2001.
- [4] P. Brusilovsky, A. Kobsa, and W. Nejdl, editors. *The Adaptive Web: Methods and Strategies of Web Personalization*, volume 4321 of *Lecture Notes in Computer Science*. Springer, 2007.
- [5] P. Brusilovsky and L. Pesin. Adaptive navigation support in educational hypermedia: An evaluation of the ISIS-Tutor. *Journal of Computing and Information Technology*, 6:27–38, 1998.
- [6] P. De Bra. Teaching hypertext and hypermedia through the web. *Journal of Universal Computer Science*, 2(12):797–804, 1996.
- [7] P. De Bra, A. Aerts, B. Berden, B. de Lange, B. Rousseau, T. Santic, D. Smits, and N. Stash. AHA! The adaptive hypermedia architecture. In *HYPertext: Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 81–84, New York, NY, USA, 2003. ACM.
- [8] P. De Bra, D. Smits, and N. Stash. The design of AHA! In *HYPertext: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 133–134, New York, NY, USA, 2006. ACM.
- [9] C. Gená. Methods and techniques for the evaluation of user-adaptive systems. *Knowledge Engineering Review*, 20(1):1–37, 2005.
- [10] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, August 2001.
- [11] C. Kaplan, J. Fenwick, and J. Chen. Adaptive hypertext navigation based on user goals and context. *User Modeling and User-Adapted Interaction*, 3(3):193–220, 1993.
- [12] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.
- [13] C. Romero, S. Ventura, J. A. Delgado, and P. De Bra. Personalized links recommendation based on data mining un adaptive educational hypermedia systems. In *Creating New Learning Experiences on a Global Scale. Second European Conference on Technology Enhanced Learning, EC-TEL 2007*, pages 293–305, Crete, Greece, 2007. Springer.