Since 2001 the chair of Media Informatics at the University of Bamberg provides a range of eLearning courses within the field of Information Retrieval with more than 50 students each year. These courses are integrated within the curricular of the University of Bamberg as well as within two virtual study programs, a virtual master program (www.vawi.de) and the Bavarian Virtual University (www.vhb.org).

The courses are designed to be conducted as distance learning. Using a self developed course authoring system based on DocBook in conjunction with an open source eLearning System (moodle), we provide the textual course content in HTML and PDF. Besides the content and forums for the communication between students and teachers we make strong use of automated quizzes and manual assignments, giving students feedback about their learning advance. For a deeper understanding of several concepts and algorithms we have created several Java-applications which are integrated within the course and the quizzes. These applications cover f.e. pattern matching algorithms, string distance computation, cluster algorithms, stemming algorithms, and so forth.

Based on more than five years of providing Information Retrieval within different eLearning courses the paper presents a mature concept for distance learning in Information Retrieval together with usage statistics and evaluation results.

Keywords: eLearning, Information Retrieval

1. INTRODUCTION

The growing dominance of the Internet in the last decade in conjunction with the every day usage of common and specialized search services like Google\(^1\) or Technorati\(^2\) leads to an evolving interest in the research area of Information Retrieval. According to this evolution and an upcoming demand for distance learning the chair of Media Informatics has developed a range of eLearning courses in German based on a former presence course at the University of Bamberg.

Since 2001 we provide Information Retrieval courses on different channels. Besides the classical presence learning which we support by eLearning at the University of Bamberg we offer courses for distance learning within the curricular of the virtual master program in Business Information Systems (VAWI\(^3\)) as well as the Bavarian Virtual University (vhb\(^4\)).

2. COURSE STRUCTURE AND CONTENT

Due to the number of different research areas in the field of Information Retrieval as well as a given course size we decided to set up a two part Information Retrieval course. The first part (IR1) contains an overview over the research area, basic definitions, methods as well as fundamental algorithms and data structures. An overview of current search services for the internet and an outlook concludes this course. In the second part (IR2) we provide a detailed insight into different current research areas, such as Peer-to-Peer information retrieval. Table 1 gives an overview of the syllabuses of the two parts.

3. DIDACTIC CONCEPT: FROM PRESENCE TO DISTANCE LEARNING

Different curriculums in which the individual courses are embedded as well as different general conditions require an adequate didactic concept. Considerable types are [2]:

---

1. www.google.com
2. www.technorati.com
3. www.vawi.de
4. www.vhb.org
1. Presence learning supported by eLearning materials: A traditional lecture conducted by presence teaching and exercises will be supported by eLearning materials (scripts, self-assigned tests, mid-terms).
2. Distance learning with physical on-site presence for discussion, often called low-residence program: Students are using the provided eLearning materials for self-organized learning. A weekly or biweekly meeting is reserved for discussing open questions and assigning the learning matter for the following week.
3. Distance learning with supervision: This option considers a spatial separation of the students which makes periodical meetings difficult or impossible. This fact will be compensated by the usage of communication through forums, e-mail or chat.
4. Distance learning without supervision: Temporal studies abroad or internships are often a good reason for students to take the exam in a semester where the course is not offered. In this case, it has been shown useful to provide the eLearning materials still after the date of an exam and the following semester.

<table>
<thead>
<tr>
<th>Information Retrieval 1 (IR1)</th>
<th>Information Retrieval 2 (IR2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Motivation and Introduction</td>
<td>1. Further IR-Models</td>
</tr>
<tr>
<td>3. Considering the Vagueness of Natural Languages</td>
<td>3. Context-based Information Retrieval</td>
</tr>
<tr>
<td>5. Vector Space Model and its Implementation</td>
<td>5. Peer-to-Peer Information Retrieval</td>
</tr>
<tr>
<td>6. Alternatives to Global Searching: Classification, Clustering, Browsing</td>
<td>6. Geographic Information Retrieval</td>
</tr>
</tbody>
</table>

TABLE 1: Syllabuses of the two course parts
One requirement for our course offers was the coexistence of these different types of didactic concepts. Further we had to support the variation of the course content, for example the delivered eLearning materials and included self assigned tests. Table 2 gives an overview of our courses and the respective didactic concept.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Course</th>
<th>Implemented Didactic Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Bamberg</td>
<td>IR1</td>
<td>1, 4</td>
</tr>
<tr>
<td>University of Bamberg</td>
<td>IR2</td>
<td>1, 4</td>
</tr>
<tr>
<td>Virtual Master Program (VAWi)</td>
<td>IR1</td>
<td>3</td>
</tr>
<tr>
<td>Bavarian Virtual University (VHB)</td>
<td>IR1</td>
<td>2, 3</td>
</tr>
<tr>
<td>Bavarian Virtual University (VHB)</td>
<td>IR2</td>
<td>2, 3</td>
</tr>
</tbody>
</table>

**TABLE 2:** Courses and their didactic concept

### 4. CREATING MULTI-CHANNEL E-LEARNING

Providing eLearning courses for different audiences on different eLearning platforms or course management systems lead us to the development of an approach which allows the creation of eLearning content without a binding to a specific system. Further requirements were the ability to define parts within the course content only for a specific audience as well as the support for specific features, e.g. tests, for all supported eLearning platforms.

Available standards like LOM\(^5\), IMS\(^6\) or AICC\(^7\) are primarily targeting the structure and description with metadata of a whole course, individual learning objects or their resources. Several, non standard approaches present the usage of XML not only for structuring but also for describing the eLearning content [1]. Standards like SCORM\(^8\) define a reference model for content which should be interchangeable between different course management systems. As a trade-off SCORM only supports a limited set of text formats or display styles with respect to mathematical formulas or source code.

---

5 Standard for Learning Object Metadata, ltsc.ieee.org/wg12/
7 AICC Aviation Industry Computer-Based Training Committee, www.aicc.org
8 Advanced Distributed Learning Initiative (ADL), www.adlnet.org
4.1. A Single Source, Multiple Publishing Approach

Our approach is based on the well known and accepted standard DocBook\(^9\) [3] which contains already some of our required features like an audience-specific publishing. Furthermore, DocBook is by design expandable with new features (e.g. multimedia objects, applets, self tests). Besides a broad range of available tools for authoring and generating online and printable documentation from a DocBook XML source, we created a special tool chain for the creation of online and printable eLearning content as well as the deployment into eLearning platforms. The main reasons for this tool chain were as follows:

- Support for our own DocBook extensions like margins, self tests as well as the integration of interactive elements like Java Web Start applications, applets or Flash.
- Lack of proper formula rendering in freely available tools for rendering DocBook into HTML or PDF\(^{10}\).
- Creation of deployable packages for eLearning platforms.

Our tool chain – especially the creation of HTML content – is partly based on available tools using XSLT\(^{11}\). We only added different tasks via the Java-based Ant build-tool\(^{12}\) for reformatting and packaging the resulting HTML, image and further multimedia files into the destination format and a package for our main course management system Moodle\(^{13}\). Splitting the reformatting and packing tasks from the rest of the authoring work chain, it is possible to publish eLearning content for different eLearning platforms as well as to migrate from one to another course management system.

For the creation of printable PDF from DocBook we preferred a LaTeX-based approach against the widely existing approaches based on formatting objects (i.e. Apache FOP\(^{14}\)). Besides the lack of a proper rendering for mathematical formulas in freely available FO-processors, those processors often show also an improper formatting quality, for example within page breaks and hyphenation.

4.2. Integrating Interactive Elements

Within our eLearning courses we integrated a range of interactive elements primarily of Java applets and self tests. Within the different Java applications (see fig. 3) which will start via Web Start students can experiment with and evaluate different methods and algorithms. Self tests in the flavour of automated quizzes and manual assignments give students feedback about their learning advance.

\[\text{FIGURE 3: Java Applications for Simulating Pattern Matching and Cluster Algorithms}\]

\(^{9}\) www.docbook.org  
\(^{10}\) for example using the XSL stylesheets from The DocBook Project (docbook.sourceforge.net) [4] with Apache FOP (xmlgraphics.apache.org/fop/)  
\(^{11}\) docbook.sourceforge.net  
\(^{12}\) ant.apache.org  
\(^{13}\) www.moodle.org  
\(^{14}\) xmlgraphics.apache.org/fop/
With respect to the applications we intensely discuss the potential didactic goals. For example the pattern matching application on the left side of figure 3 shows the attempt to combine the understanding of the principle algorithm with its implementation in Java. To this end, the progress of the algorithm is visualized in the upper area of the application window and the current state of the Java execution together with the actual variable values is shown in the lower area. This way the students can correlate both levels of abstraction in one application. Other applications like the clustering application shown on the right side of figure 3 give students the opportunity to visually explore the impacts of different algorithms.

5. USER FEEDBACK AND EVALUATION

At the end of each course a user evaluation is conducted. These evaluations show that on a scale from 1 (unimportant) to 7 (very important) most students think that applets or Java-applications as well as self tests are rather important. Nevertheless the enthusiasm for the applications is not as high as we expected. A score of 5 is the most common judgment for the usefulness of applets which shows that the applets can still be improved.

More importantly our web server statistics show that students constantly prefer printable course materials (PDF documents) against online content. Due to the improvement of the online reading support facilities in tools for PDF reading, like Acrobat Reader, this preference seems to become even stronger over the years.

With respect to communication channels forums are clearly preferred compared to chats or e-mail. In our experience chats are rather useless for supervision because technical discussions are much more intense in forums.

6. LESSONS LEARNED AND CONCLUSION

Besides the user feedback one important influencing factor for the further development directions of the courses is the effort necessary for course creation and maintenance. Obviously, this effort has to be related to the number of students participating in the courses and the change rate of the course content. Especially our single source approach and the high change rate of the content in current research fields led to a situation where the creation of sophisticated courses with written textual descriptions in conjunction with the design of images, animations or applications is counterproductive in a way that makes it impossible to consider current research trends. This led us to the decision of stepping back within the IR2 course to a common lecture style which is technically based on PowerPoint slides and a special screen recording system (Lecturnity15). Positive evaluations from students of a different course addressing multi media techniques confirm that decision.

15 www.lecturnity.de

FIGURE 4: Lecturnity in action
In opposition to the fast changing content in IR2 our IR1 course contains basics which are mostly stable. Considering the rare usage of online eLearning content and the strong demand for printable documents based on PDF we will create our further course versions based on LaTeX. Using LaTeX is the most common way for lecturers for authoring documents in this research area. Mature and freely available tools like editors with integrated proof reading, preview and roundtrip facilities confirm our decision.

Using a common LaTeX to HTML converter seems to allow for the easy creation of online versions in HTML. However, there are still problems with complex formats and mathematical formulas.

In addition we will still provide animations, applications and self tests within our course management system Moodle which will be referred by embedded links at the appropriate positions within the PDF documents. On the other side online elements like self tests can refer so called named labels within PDF documents. Using those mechanisms allows for the same interaction provided by HTML in combination with the print quality known from LaTeX with the expectation of a lower effort in the technical creation of eLearning content.

To sum up, based on our experiences and the user feedback we refactored our eLearning strategy for IR considering the maturity of PDF documents and the linking opportunities from PDF to course management systems and vice versa. For the rapidly changing IR2 content we use PowerPoint presentations inducing rather low creation efforts (many slides are already available from recent research talks) together with rapid eLearning techniques based on screen recordings. On the other hand the more stable IR1 content is further presented in an online and in a printable version produced using an environment comprised of LaTeX, PDF and LaTeX2HTML ensuring comparably low creation efforts.

REFERENCES


