

Learning Information Retrieval through an Educational Game

Is Gaming Sufficient for Learning?

Kai Halttunen and Eero Sormunen

The Department of Information Studies, University of Tampere, Finland

E-mail: kai.halttunen@uta.fi

E-mail: eero.sormunen@uta.fi

Address:

Kai Halttunen, Department of Information Studies, FIN-33014 University of Tampere, Finland

E-mail: kai.halttunen@uta.fi ; Fax: +358 3 215 6560 ; Tel.: +358 3 215 6957

Published in *Education for Information* 18(2000):4, 289-311

Abstract

Computer-supported learning environment (CSLE) Information Retrieval Game (IR Game) is described and evaluated. The IR Game is based on idea that test collections used in laboratory-based IR experiments could be used in instruction as a rapid query analysis tool. The goal is to provide a realistic environment for demonstrating the performance of queries in different types of search situations. The outline of the IR Game is introduced both at the technical and instructional levels. An evaluation experiment is reported revealing that students found different characteristics of the IR Game both enhancing and inhibiting learning. The perceived usefulness of features like query performancefeedback, visualization of search results, comparison of queries, document browsing and help facilities are discussed. Several ideas for the elaboration of information retrieval learning environments are put forward. These include design of search tasks, implementation of feedback, social negotiation and articulation of student's working theories.

1. Introduction

Information retrieval skills are a key area of expertise for information professionals. Information retrieval is taught routinely at universities and other educational institutions, and a wide variety of textbooks about the basics and principles of searching has been published [e.g. 1-4]. The educational material covers four main areas focusing on representing (1) the context of information retrieval as a part of information seeking activities, (2) basic principles of information retrieval systems, (3) general search strategies applicable in all ordinary retrieval settings, and (4) specific search strategies for particular retrieval settings and information sources. The main goal of instruction is to develop learners' practical capability to perform successfully any search task appearing in the professional work situation.

"Hands-on-keyboard" classroom exercises are a common method for teaching and practising practical searching skills. Operational search systems and databases or their training versions are

typically used. The advantage of using operational systems is that the learner gets a realistic feeling of the particular system used. (S)he also learns typical operations of a retrieval system, and appropriate ways to formulate queries. There are very few studies on learning and instruction of information retrieval in the traditional context but it is easy for professional IR educators to identify some shortcomings of this approach.

One of the shortcomings is that operation IR systems do not give feedback about the performance or success of user queries. For instance, if the expansion of a query by synonymous search keys is the theme of an exercise, the trainees have to browse through a long list of documents to find out the effect of different query expansion moves on retrieval performance. If one wished to compare the performance of different search systems, e.g. Boolean and best-match algorithms, one should even switch between systems to compare their effectiveness.

Browsing text documents and assessing their relevance is time consuming and causes high cognitive stress. In addition, creating an overall image of query success in different queries from the large set of observations is sometimes difficult. It is not possible to make a general observation of how many relevant documents a retrieved document set contains or where the relevant items are located. In theory, an overall image of success can be created while practising searching in an operational IR environment but the assessment process simply takes too much time and effort to work effectively in practices.

Information retrieval instruction is an obvious application area for computer-supported learning (CSL) systems where the problem of performance feedback can be solved. CSL has been developed for, and studied in the instructional settings of particular online services. The focus has been mainly on the characteristics of individual services and databases rather than on generally applicable principles of IR systems and searching strategies. The ideas of simulation and intelligent tutoring systems have been applied but not the idea of query performance feedback. [5-11].

One exception is Markey & Atherton [12] who used evaluation of query performance in Dialog's ONTAP Eric database. In this setting it was possible to compute recall and precision scores for any user query in the given ONTAP exercises by comparing the searcher's results with the "perfect answer set". Although the system by Markey and Atherton was based on the limited technologies of the late 1970's, and recall and precision scores are only a modest way to represent performance feedback is an obvious innovation that can be elaborated.

In this paper, a prototype of a computer-supported learning environment, the *Information Retrieval Game (IR Game)* is introduced together with evaluation results from an evaluation study. The IR Game was based on the innovation to apply a test collection created for laboratory-based IR experiments as an instructional tool. Standard test collections offer a mechanism to generate search exercises with performance feedback since they provide a large set of documented search tasks (topics) and reliable estimates of the relevant documents available for each search task. The IR Game is a Web-based CSL application that can be used to demonstrate and test the effectiveness of any individually formulated query for a search task in textual and image databases. It was developed at the University of Tampere and has been in routine use in all searching oriented IR courses for more than two years.

The paper is organised in three main sections. First, we shall address the basics of computer-supported learning environments to create an appropriate background for the paper. Next, the outline of the IR Game is presented both at the technical and instructional levels. Finally, the results of an evaluative study are reported, and some conclusions made.

2. Conceptions of the computer-supported learning environments

Computer-supported learning environments can be divided into two main categories, namely *open learning environments* and applications of information technology based on the behavioristic tradition of *programmed learning*. However, these two approaches are just extremes, and there are many parallel models of how information technology can be utilized in constructing learning environments.

Jonassen [13] makes explanatory notes about the usage of computers in instruction. He divides usage to learning *from*, learning *about* and learning *with* computing. Learning from computing implies the delivery of computer-assisted instruction, including drill and practice, tutorials and, on a very limited scale, intelligent tutors. Learning about computing represents computer literacy. Learning with computing highlights the use of software packages as mindtools, cognitive tools for learning. Computing is seen as an intellectual partner for the learner in order to engage in and facilitate critical thinking and higher-order learning.

Theories and conceptions of learning have affected the instructional design of computer-supported learning environments. Behavioristic conceptions of learning created computer-based instruction programs consisting of presentation of learning material, drill and practice, tutorials, simulations, games and tests. Such of learning activities are based on an expository model of instruction. According to the model, for instruction to be effective the following four phases should be present: presenting information, guiding the student, having the student practise and assessing student learning. [14.]

Information processing theory or cognitivism introduced tutoring features into the instructional software. The design of CBI programs was greatly affected by advances in artificial intelligence and expert systems. Software was seen as an intelligent tutor whose activities were based on the modeling of a teacher, a student and learning material. [15.]

The design of instructional software described above saw learning as an individual process. Recent approaches based on situational cognitivism and constructivism have emphasized social co-operation in learning. Learning is seen as a social construct. The main functions of instructional software are associated with information gathering and arrangement, discussion and presentation of learning outcomes. This can be seen as the transition from learning *from* computing to learning *with* computing. [13.]

The emphasis on learning being constructive [16], situated [17-19] and collaborative [20,21] has set new challenges for instructional design of learning environments [22-26] According to these views computer-supported learning environments should support constructive and situated learning processes and promote more active learning strategies. According to Kintsch [27] tutoring systems can be seen in this context as becoming gradually "unintelligent", which gives only temporary support to students and allows them to perform at a level just above their present level of competence. When the help is no longer needed, it will be gradually withdrawn and finally removed [28].

The term "learning environment" has several meanings, but it appears often in the literature without definition. It is largely taken for granted. The need for a definition is emphasized by the development of information technology and its introduction to education. On one hand, a learning

environment can be seen as a natural living environment as a whole. In this case it is important to accept the division into informal, non-formal and formal learning and education. On the other hand, a learning environment can be seen as a construct of space or place and different kinds of services which foster learning. Perkins [29] suggests that all learning environments, including traditional classrooms, include the following key components or functions: 1) *information sources*, 2) *symbol pads* i.e. tools to manipulate symbols and language, 3) *phenomenaria* as area for presenting, observing and manipulating phenomena, 4) *construction kits* such as math-manipulation or multimedia authoring software, and finally 5) *task managers* such as teachers, tutors and co-learners.

Computer and network-based learning environments are clearly differentiated or integrated tools and applications, providing resources, learning tasks, feedback as well as tools for individual and group reflection. In this perspective, a learning environment consists of learning material and tools for problem solving and communication. The basic principle in designing computer or network based open learning environments has been the emphasis on learner activity and the "learning by doing" approach. The student's self-directness is equally important in this design paradigm. Computer-supported learning offers many opportunities to implement different kinds of learning activities as well as tools for constructing and reflecting learning and thinking processes. [30-32.]

On the whole, computer-supported learning environments should not be considered only as technological applications that affect learning but rather as a complex system of interconnected instructional goals, technologies, materials, activities, settings, personal roles etc. exerting combined effects. Powerful learning environments integrate tools, resources, and pedagogical methods that enhance comprehension and understanding.

3. The IR Game

The Information Retrieval Game was developed at the University of Tampere as a rapid query analysis tool [33]. The goal of the IR Game is to provide a realistic environment for demonstrating the performance of queries in different types of search situations.

3.1. Experimental test collections and IR instruction

The basic idea of the IR Game arose from the insight that the test collections used in the laboratory-based IR experiments could be used in instruction. A traditional test collection consists of a database, a collection of test topics (search tasks), and relevance assessments indicating which documents are relevant in respect to a given search topic [34,35]. The main focus of experiments has been on average performance, e.g., measuring the precision of queries at some standard point of operation and averaging the results across the whole set of search tasks. In contrast, Hull [36] has encouraged researchers to analyze individual queries but this approach has not been widespread. Evaluation results averaged across a large set of search topics reflect general trends in IR but searching skills are developed by experiencing success and failure in individual queries. Thus the instructional use of test collections is closely related to the idea introduced by Hull to analyse individual queries.

As far as the authors know, experimental test collections have not been used in instruction, nor as a component of any IR learning environment. This is quite surprising, since relevance data provided by the test collections are an obvious source of feedback on a person training one's searching skills.

The topics of a test collection typically represent a specified search task [see e.g., 37,38], and could serve as a description of an exercise. Further, the relevance data of the topic could be used to measure the level of success that a person has achieved in queries to pass the exercise, and to indicate his/her progress.

Traditional test collections have been designed for experimental research, and they possess some features that limit their usefulness as instructional material. One of the limitations is that search tasks and relevance assessments are heavily tied to the notion of topicality. For instance, the utility aspects of documents or the degree of relevance seldom have any role in experienced relevance assessments. Another limitation is that typical experiments have focused on evaluating matching techniques for unstructured natural language texts (textual queries and textual documents) excluding other types of metadata like author names, source information, dates, manually assigned index terms, etc. However, these types of metadata have an important role in everyday searching, and should be included in the learning process. Thus, a traditional test collection may offer some opportunities to extend instructional materials, but the main point is that they offer a general model for organising instructional materials in the IR Game.

3.2. An outline of the IR Game

The IR Game consist of four major components: (1) a set of well-specified search tasks for retrieving documents from a database, (2) judgements explicating which documents of the database match the relevance requirements of a search task, (3) a front-end system supporting and monitoring searching in all appropriate retrieval systems and databases, and (4) a feedback system for measuring and visualizing the performance of any query executed. The fundamental requirement for the search task is that it be well-specified. If a search task is not specifiable in detail, neither is it possible to formulate indisputable relevance requirements nor identify the set of documents that should be found. The chances of the Game giving feedback to the user are totally based on “knowing” the set of “right” documents. This may appear restrictive since real search tasks are often ill-defined or muddled [see e.g., 39]. However, this is not totally the case in the game because issues of user’s uncertainty can be simulated by selecting search tasks appropriately, for example, so that the user is not familiar with the domain of the task. The unambiguity associated with the description of the search task and the certainty felt by the user in interpreting it are different issues.

The search tasks and relevance judgements are both biased towards topicality aspects and restrict the possibility of using traditional test collections in instruction. If new search tasks are designed for the IR Game, the problems of collecting relevance data become a major issue in large databases. Assessing all documents in the database is too laborious to be applied. The extensive pooling method introduced by Sparck Jones and van Rijsbergen [34], and applied in TREC [37] is also too resource intensive for individual educational units. Sormunen [40] has evinced an approach based on *inclusive query planning* and *extensive queries*. Boolean and probabilistic queries tend to retrieve only partially overlapping document sets, and both types of matching should be applied in extensive queries to create the set of documents assessed. However, the hunt for relevant documents still requires a lot of effort, especially in subject-oriented search tasks containing broad concepts. On the other hand, when non-topical relevance requirements are applied in a search task (date, source, and other metadata), a professional searcher familiar with the database may substantially reduce the number of documents to be assessed. In addition, the number of search tasks may be smaller in instruction than in experimental test collections such as TREC. It is only necessary that different types of search tasks be available to provide a versatile basis for designing exercises.

Several types of front end-systems have been proposed and experimented with for IR providing more or less sophisticated intermediary functions [41]. Their goal is typically to automate some steps in the formulation and execution of queries. For instance, the ExpansionTool developed at the University of Tampere automatically expands the query formulated by the user, and generates different query versions for selected IR systems and databases. The behavior of the Expansion Tool can be changed by adjustable parameters [42]. In a learning environment, the goal of the front-end system is slightly different. Some routines may be automated (e.g., login/logout) and a unified user interface may also be provided to hide unnecessary peculiarities in the target IR systems. The major difference is that the user should face the genuine search functions as they are in an operational retrieval system to learn query formulation strategies. The task of the front-end system is also to represent the exercise (the assignment), search results, and performance feedback to the user in the form defined by the tutor.

The capability of the IR Game to generate performance feedback for the user is based on the analysis of search results. In its simplest form, this means that the output of the target system is parsed to identify the documents retrieved by the user-generated query. Performance measures can be calculated by comparing the list of retrieved documents to the list of relevant documents. The next step is to generate feedback to the user. Several options are available, for instance:

1. relevant documents can be highlighted in the output list
2. the user can be informed about the overall performance (e.g., $\frac{m}{n}$ % of the n relevant documents were found)
3. the performance can be visualized as a traditional Precision/Recall curve
4. other ways of visualization can be used (e.g., based on Document Cut-off Values or positions of relevant documents on graphical output list)

How much and in what form feedback is given is a didactical issue. The point is that the IR Game should support different forms of feedback, and it is up to the tutor to make the decision how the learner is informed in a particular exercise.

A tool like the IR Game can be used for different purposes: *For a tutor in a classroom*, the IR Game is a tool to show the overall effectiveness of any query. It is easy to demonstrate how any reformulation of a query, or any change in the retrieval system reflects on query performance. A tutor may also apply the IR Game to design learning environments where students may perform exercises and learn IR techniques at their own pace. *For an advanced student*, the IR Game is an environment for learning by doing, for example the query formulation tactics in Boolean or best-match retrieval systems. *For a researcher*, the IR Game is a tool for developing experimental research ideas and analyzing how the averaged results hold in individual search tasks.

3.3. The prototype of the IR Game (v. 3.0)

Version 3.0 of the IR Game was completed at the end of 1998. The main goal was to create a prototype in order to test the general idea of the IR Game in a realistic operational context, and to develop the very basic functions of the system, such as how to visualize the effectiveness of queries. Auxiliary features that are important in all-inclusive learning environments, such as the sophisticated course tools of the tutor, or the answering tool of the student, were regarded as of secondary importance.

The implementation was based on using a set of HTML pages as the WWW interface for executing CGI programs and connecting the functions of the IR Game. A general view on the flow of

operations between the WWW interface (the WWW browser side), CGI programs, and UNIX-based programs/resources (the WWW server side) is presented in Fig. 1. The user operates the system by switching between six HTML main pages designed for different operations: (1) setting system parameters, (2) selecting the search task, (3) query formulation and input, (4) precision/recall feedback, (5) display of query result titles, and (6) display of a selected document. In addition, separate pages are used for login, and for showing the list of best queries executed so far by any of the users in a particular search task.

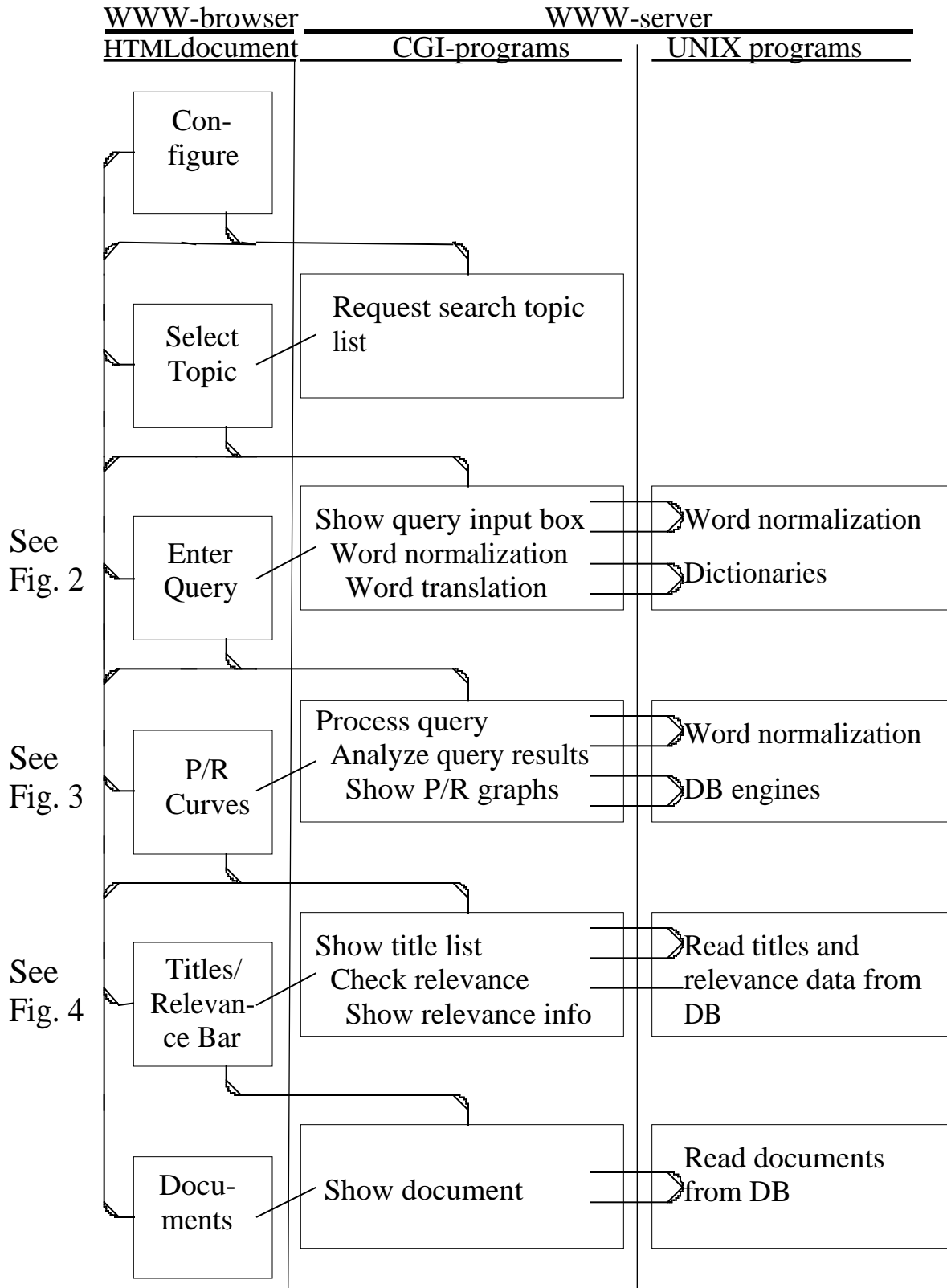


Figure 1. General view of the flow of operations in the IR Game

The IR Game can be used over the network with a standard WWW browser in two modes: the *direct mode*, and the *exercise mode*. In the direct mode, the user is allowed to access all functions of

the Game for instance to set the parameters of the system as s/he prefers, to select which search topic, IR software and database is used, and to review all feedback produced by the system. In the exercise mode, the tutor creates an external HTML exercise page representing the search task, and decides what operations of the IR Game will be available to the student by pre-selecting the system parameters. The student enters directly to the query-input page by activating a link on the exercise page, and is served by those operations and feedback features that the tutor has selected for that exercise.

The point of departure in designing the IR Game prototype was that the experimental test collections should be used as the primary instructional material. As stated, this restricted the type of exercises to topically oriented search tasks but this was expected to be adequate at the preliminary testing stage. One aim was to interface the system both to a Boolean and to a best-match retrieval system. Both of these are important from the educational viewpoint, and set different requirements for the system, for example, how to visualize the search results. The IR Game (v. 3.0) integrated five main resources:

1. IR test collections
 - a) A Finnish test collection of about 54,000 newspaper articles with 35 test topics and relevance judgements for about 17,000 articles
 - b) A Finnish database of about 51,500 newspaper articles, sample search tasks with relevance judgements for instructional purposes
 - c) An English test collection (a subset of TREC) of about 514,000 newstext documents, and corresponding TREC test topics and relevance judgements
 - d) A database of newspaper photographs with textual metadata, sample search tasks with relevance judgements for instructional purposes
2. Retrieval systems
 - a) A Boolean IR system TRIP¹
 - b) A probabilistic retrieval system InQuery²
3. Applications for computing recall-precision information
4. MOT translation dictionaries in Finnish, Swedish, English, and German³
5. Morphological analyzers for word normalization³

The prototype of the IR Game followed the general guidelines described in the preceding section. The query input page contains only the description of the search task, an empty field for query input, and links to the query language manuals. It is up to the user to formulate the query, and to check that the syntax of the query language is followed (Fig. 2). When the user judges that the query is completed, s/he enters the query by clicking the “Submit query” button, and the IR Game sends the query to the target IR system (here TRIP), parses the result set, compares the list of retrieved documents to the list of relevant documents, and calculates precision and recall values for the query.

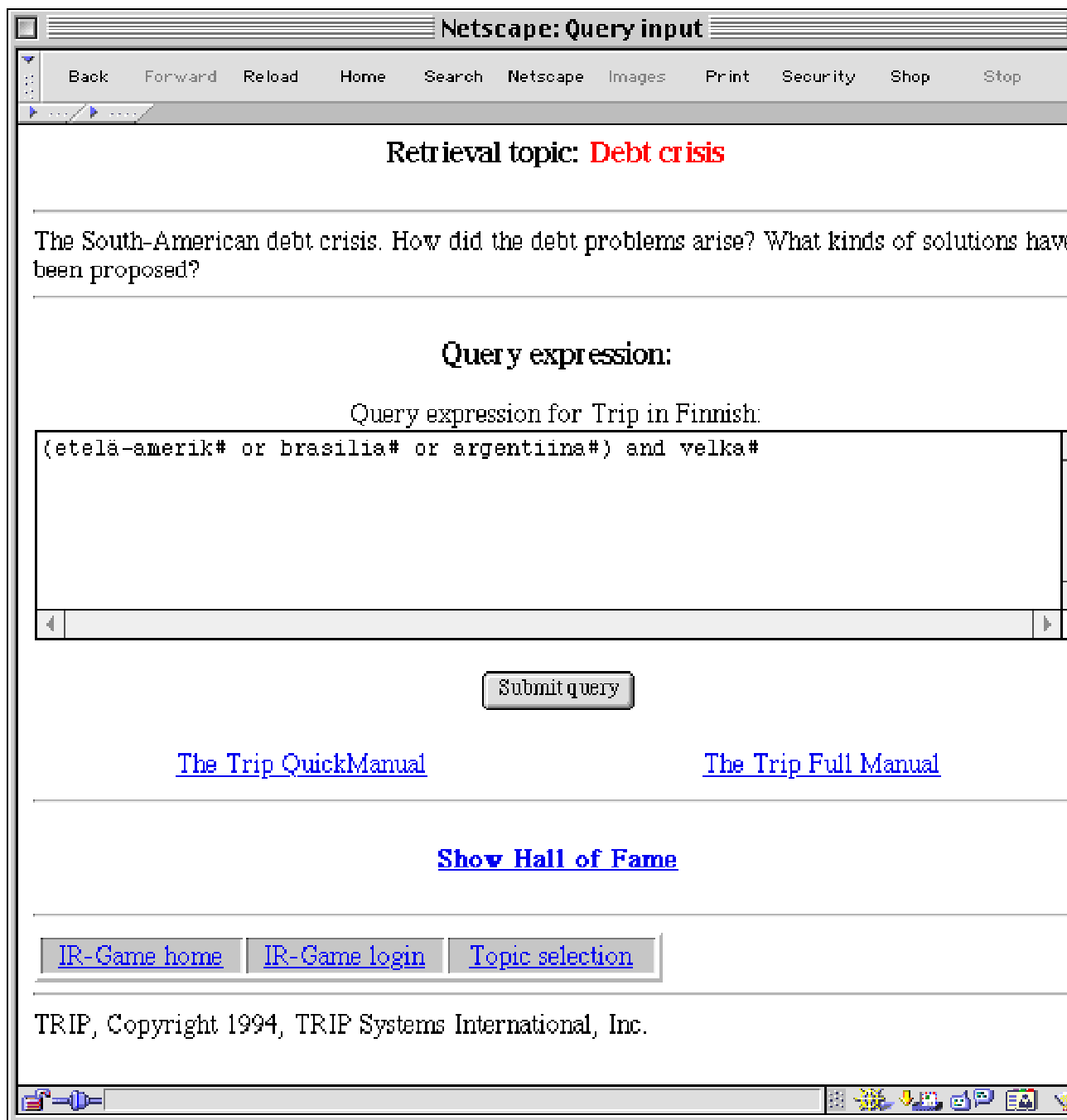


Figure 2. Query input page in the IR Game

Representing precision as a function of recall is a traditional way to illustrate the effectiveness of a query in best-match IR systems. This is one of the visualization techniques applied in the IR Game (Fig. 3). One problem with the P/R curves is that novice users may not be familiar with the concepts of precision and recall, and feel unsure or irritated in interpreting them. Another problem is that only one value for precision and recall can be calculated for a single Boolean query while in best-match systems precision can be estimated for any query across the whole recall range. Fortunately, it turned out that different P/R measures can be merged fluently into a single graphical representation.

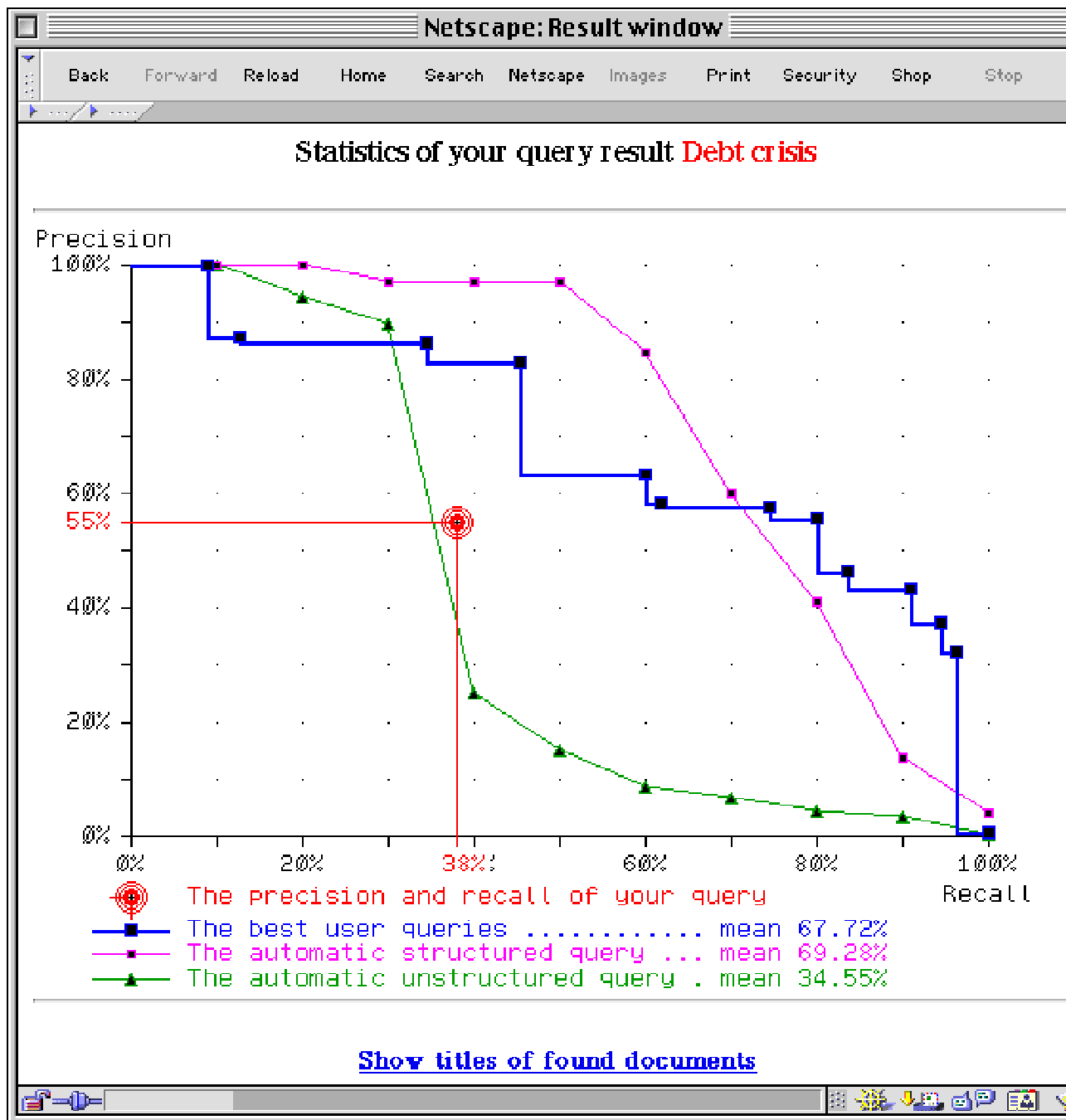


Figure 3. Precision-Recall page showing the performance curves

The precision of the last Boolean query is shown as a dot (see “The precision and recall of your query” in Fig. 3), and for the best-match query as a chain of dots (joined with a hair-line) across recall levels $R_{0,1}$ - $R_{1,0}$. The system automatically generates from the list of best queries (called the “Hall of Fame”) a graphical reference in relation to which the user may compare his/her personal P/R result. In Boolean queries, the reference is a stepped curve representing precision/recall achieved in the best queries so far (overall or within the present course) and stored in the Hall of Fame (“The best user queries” in Fig. 3). In best-match queries, the reference represents the P/R curve of the very best query in the Hall of Fame. In addition, any performance curve can be attached to the background of the chart. For instance, two curves (“The automatic ...”) in Fig. 3 are derived from a particular InQuery study.

Another visualization technique called the *relevance bar* was designed especially for novice users, and integrated into the page that displays the titles of documents retrieved (the horizontal bar in Fig. 4). The relevance bar also serves as a control device in browsing the list of titles. The titles are displayed in sets of ten, and this set can be changed by clicking an appropriate segment of the relevance bar. Colour coding (green for relevant, grey/white for non-relevant) is used to show the relevant documents both in the relevance bar and in the list of titles.

Netscape: Title list

Back Forward Reload Home Search Netscape Images Print Security Shop Stop

The list of document titles from topic **Debt crisis**

There are **55** documents relevant for your topic. Among the **38** documents found there are **21** relevant documents.
Click on the the diagram to browse document titles.

| Rank | Relevant | Document # | Picking <u>1/1/2</u> | Document title |
|------|----------|------------|----------------------|---|
| 1 | 1 | 17905 | | Talouteen USA lainaa Meksikolle |
| 2 | 2 | 43228 | | Otsikko puuttuu |
| 3 | 3 | 68890 | | Velat köyhien maiden ongelma |
| 4 | 4 | 68948 | | aloittivat kiistellen |
| 5 | No | 69166 | | Gorbatshev: Yhteistyötä lisättävä Köyhät maat varjokokoukseen Presidentti George Bushilla ja pääministeri Margaret Thatcherilla oli hauskaa aamiaispöydässä ennen päivän neuvotteluja. |
| 6 | No | 69238 | | Argentiinaan uusi talous- ministeri |
| 7 | 5 | 79276 | | Rahastoa ehdotettu Yhdysvallat näkisi mielellään nämä kymmenen Kolumbian huumekaupan johtohahmoa oikeuden edessä. Kolme kaivatuinta näiden kymmenen joukossa ovat Pablo Escobar Gaviria (alarivissä, vasemmalla), Gilberto Jose Rodriguez (alarivissä, oikealla) sekä Jorge Luis Ochoa (ylärivissä, toinen vasemmalta). |
| 8 | No | 80634 | | Yhteistyöhön EY:n kanssa Kokouspaikasta köydenvetoa |
| 9 | 6 | 115337 | | Menetetty vuosikymmen Itä-Eurooppa lisäuhkana Naisten asema edelleen heikko Vaihtoehtoja huumeviljelylle |
| 10 | 7 | 147391 | | 'Demokratiasta tullut uskonto' 'Vapauksista tingittävä' |

[Give a hint!](#)

Figure 4. Document titles page showing the relevance information

In Boolean queries, the relevance bar with color coding is an efficient tool to demonstrate the size and the content of the result set. For instance, in a database of news articles it is easy to illustrate how relevant articles occur in clusters in the chronologically ordered result set, sometimes quite far from the top. The text above the relevance bar reminds the user that the database may contain many relevant documents that were not retrieved by the present query. The observation that so many relevant documents are not retrieved although a query was carefully designed can be a disturbing experience for a user who has worked only with traditional searching environments.

In best-match queries, the relevance bar is very useful when demonstrating the changes in the relevance ranking of documents from one query to another. The forepart of the bar full of green indicates success in ranking while a sparse chain of green poles indicates bad ranking. It is much more difficult for the user to gain control over the search results in best-match systems than in the (exact-match) Boolean systems. Nor is there an established corpus of expertise in everyday best-match searching comparable to that of Boolean searching. Thus, the relevance bar is also an excellent instrument for teachers to learn and develop searching strategies for best match systems in different search tasks and databases. In this respect both teachers, students and researchers face an obvious need for a learning environment.

Precision/recall curves and the relevance bar emphasize different aspects of query effectiveness. The former illustrates the overall and especially high recall performance of queries. On the basis of the P/R curve or a single P/R value it is easy to say exactly which query performed best at a given recall level (a system-oriented view). The relevance bar is useful in illustrating what the content of the result set is studying it top-down. The user easily discovers what has been the success in terms of user effort in browsing the title list (a user-oriented view). By pressing the "Give a hint" button at the bottom of the title display page the user is displayed a relevant document not retrieved by the present query (see Fig. 4). The idea of this function is to encourage the user to study the content of non-retrieved relevant documents and discover new query terms. Only one randomly selected document is shown per query so as not to make query expansion too easy. The user may also study any of the retrieved documents by clicking the title. Documents are displayed one by one, and the user may mark them relevant or non-relevant. Picked documents are collected into a list, and the selected documents have also a special color code displayed on the relevance bar and in the column "Picking" (see Fig. 4). The tutor can use this feature to collect the set of relevant documents when building a new search task. The student can apply it in collecting the candidate documents in composing an answer in open exercises where relevance data are not available.

Version 3.0 was the first prototype of the IR Game launched for operational use. It was not intended to be a finished product or a self-contained learning environment for IR instruction. Typical of prototypes, the IR Game (3.0) is not easy to expand or modify for new user groups, and not much attention was paid to the user interface. However, the prototype was designed with the perceived educational needs in mind, and received quite an enthusiastic response from teachers at the University of Tampere. The prototype was adopted for all appropriate courses either as a demonstration tool or as a tool for creating WWW-based exercises.

IR Game has been used as a routine tool on four IR courses: 1) Methods in IR (4 ECTS credits, 1998-1999) 2) Text retrieval methods (4 ECTS credits, 1998-1999) 3) Course in journalistic information acquisition (6-10 ECTS credits, 1999), 4) Basics of information retrieval (6 ECTS credits, 1998-1999). All empirical results in this paper are based on Version 3.0. The empirical study reported in this paper is the first systematic attempt to collect a response from the students during one of the courses.

4. An evaluation experiment

Because the educational use of the IR Game has not previously been studied or evaluated we decided to use open and low structured methods in order to provide an exhaustive and rich view of the phenomenon at hand. We tried to construct an overall picture of the use of the IR Game in an instructional setting. Research questions for user evaluation were: 1) What kind of learning experiences did subjects have while using the IR Game? 2) What functions of the game enhance and inhibit learning? 3) How do students use the opportunities in the game and proceed in their search process? 4) Are there recognizable failures in system functions leading to inappropriate user behavior? The analysis of findings on these questions offers insight for the discussion of how the IR Game could be developed to enhance the learning of information retrieval. It should also be possible to discuss essential functional elements that should be incorporated into the game in order to talk about a real information retrieval learning environment (IRLE).

4.1. Data and methods

The use of the IR Game was systematically evaluated in the course "Basics of Information Retrieval" (6 ECTS credits) 1999 at the Department of Information Studies, University of Tampere. The basic course consists of a series of lectures (30 contact hours), tutored exercises (6 contact hours) and weekly distant exercises, which were based on different kinds of activities on the web. Feedback on both of these exercises was given in the following lecture. The IR Game was used in the fifth tutored training session. The basic features of the game were introduced to students in the preceding lecture. The lecturer used the IR Game to demonstrate concepts of precision and recall. The game was also used earlier to demonstrate image retrieval to the same audience.

Three kinds of data were gathered during the course. First, we observed the instructional design of the lectures in order to have background information on the instruction of the system. Through observation it was possible to construct a view of advanced organizers which learners adapted. Observation was continued in tutored exercises with four groups using the IR Game. These groups consisted of ten learners each who carried out information retrieval tasks in small groups of two to three people or in some cases alone. Second, the four search sessions in tutored exercises were recorded on video as a log session. Video also included discussion in the small group preparing, performing and evaluating their query. A microphone was situated beside the computer screen.

The third body of data consists of stories written by students attending this course. In these stories they describe good and bad learning experiences while using the IR Game. The method of empathy-based stories was applied [43,44]. The methodology is also called non-active role-playing [45,46]. This method involves the writing of short essays according to instructions given by the researcher. The respondent is given some orientation, which is called the script. This script should be used in conjunction with the respondents' imagination in the writing of the story. The writer of the story either continues the situation detailed in the script or describes what must or may have taken place prior to that situation. Variation is crucial to the use of this method, and there are at least two different versions of the same script which vary with regard to a certain key issue. Variation distinguishes the method of empathy-based stories from many other methods of acquiring data.

The students wrote their stories describing the learning experience one week after the tutored exercises. The researcher introduced the method and reviewed the situation to be described. Two

kinds of varied scripts were used; one describing a good learning experience and the other reporting an inadequate one. Fifteen minutes was allowed to write these descriptions. The data consisted of 67 stories from which four had to be excluded due to incomplete content. There are 32 stories describing positive learning experiences and 31 stories describing negative ones.

4.2. Evaluation results

How did students experience the learning situation and used the possibilities of the IR Game? The analysis of empathy-based stories revealed three major themes that were constructed by the students: 1) The IR Game as an educational software, 2) databases used and 3) tutored exercises as a learning environment. The analysis was based on an emic analysis, in other words, it was based on the conceptual framework created by the subjects studied. Themes found in stories construct an analytical frame, which is used in the detailed interpretation of the data. The analysis frame is presented in Table 1.

Table 1

Analytical frame for empathy-based stories and the number of subject mentioning a theme.

| Theme | Enhance learning | Inhibit learning |
|--|-------------------------|-------------------------|
| IR Game as an educational software | | |
| The idea of feedback | 20 | 5 |
| Visualization of search results | 10 | 2 |
| Comparison to other queries | 7 | – |
| Document browsing | 11 | 2 |
| "Give a Hint" feature | 5 | 6 |
| Database content and type | 4 | 3 |
| Tutored exercises as a learning environment | 19 | 31 |

4.2.1. The IR Game as an educational software package

The first theme of the empathy-based stories dealt with the description of the IR Game as educational software. The basic function of the system – the query effectiveness feedback – was naturally seen to promote learning significantly. Feedback concerning the performance of one's own query, the chance to freely reformulate the query and to further evaluate the effect of changes on performance was seen as a highly motivating and learning advancement. Furthermore the feedback mechanism allows students to pay attention to the analysis and evaluation of query formulation and search keys. This was contrasted with the heavy browsing and evaluation of search results, which is typical when operational databases are used for educational purposes. Respondents outlined the usefulness of feedback in building a basic understanding of central phenomena of information retrieval. A few of the respondents constructed mental models of the impact of query formulation on search results. Several students reported that the graphical presentation of search results more informative than numerical values.

Preconceptions of the IR Game affected the learning situation and students' learning outcomes. Some of the students seemed to have been waiting for a much more entertaining game than the IR Game in fact is. The most game-like features such as the hall of fame were also kept in the background in this course.

Surprisingly the feedback mechanism and presentation of query performance was also seen as a factor inhibiting learning. Some students stressed that their attention was fixed on precision -recall estimates and they tried to improve on their previous results mechanically, without analysis and reflection of their preceding queries and results. In this respect, the feedback mechanism tempts searchers to pay attention to the performance measures achieved, not to the analysis of the search task situation. This approach underlines surface level coping and passion for gaming. Obviously the risk of surface level learning was greatest in the probabilistic IR system where queries were unstructured sets of search keys. The reason for gaming and surface level coping can be found in prototype design. Presenting p/r-curves just after query fixes attention on them. In educational applications search results and documents could be shown first and p/r-curves as an additional view and visualization of results. Several stories pointed out the difficulty of transferring the acquired knowledge to real life search tasks and operational systems.

The IR Game offers the opportunity to see the best query formulations and the achieved results of other searchers. This feature of the system fosters learning. Presenting other searchers' success creates a subtle competition and desire to improve one's own search results. Surprisingly this feature was not mentioned as a factor inhibiting learning despite the overall criticism of gaming features and attraction to pay attention solely to search results. Subjects were able to see the "Hall of fame", but the tutor deliberately did not highlight it.

Equally important, the chances to display and browse the documents found, newspaper articles and photos, was mentioned as a feature promoting learning. Both relevant and non-relevant documents can be displayed and analyzed to find new search keys or new ideas. The video log and observation revealed that in some cases only a couple of relevant documents were examined briefly and new search keys were picked up quite sporadically without deeper reflection and articulation. This may indicate that searchers seldom read long documents on the screen. The fact that the primary motivation for browsing documents was not to absorb new information could also be a barrier to gathering new search terms from documents. Since the analysis of documents was a highly stressful cognitive activity it was not used as often as expected. Evidently the gaming function was emphasized in such situations - students were desperately trying to invent new search keys to improve their search results.

Lastly the IR Game offers a "Give a Hint" function which displays one relevant unfound document. This function was seen as a learning promotion feature for the same reason as document browsing. However, in the majority of stories the hint function was seen as an obstacle to learning because the users interpreted this feature differently than the designers of the system. With good reason students thought that this feature of the IR Game would offer them interactive, situational help for query formulation and ideas for improving a search. It was hoped to integrate system guidance into the situation encountered and the tasks performed. This finding clearly supports the need and importance of scaffolding arrangements in computer supported learning. How can the instructional software support the learner to reach the next level of performance? How can the situational and interactive learner support and guidance be implemented?

4.2.2. Database content and its impact on learning

The second and the minor theme in empathy-based stories dealt with the content of databases. The databases used in the tutored exercises support learning because respondents regard them as real, operational systems used in real-life settings. The authenticity of databases was one factor supporting the chance to transfer learning outcomes to forthcoming work situations. In addition, the journalistic approach and content of the databases was also motivating for several students because they will probably find themselves working within the media industry. The use of the image database was stimulating to some subjects because of its novelty. In fact, for most of the students this opportunity to perform image retrieval was their first.

The database content and the databases used in exercises were mentioned only once in stories describing obstacles to learning. In this situation, the lack of descriptive meta-information about databases and their content was seen as a factor inhibiting learning.

4.2.3. Tutored exercises as a learning environment

The third theme of stories described the exercises as a learning environment. As has been noted the learning environment at hand consisted of lectures and tutored exercises, learning materials, search exercises, IR systems and databases used, system help facilities and instruction offered by tutors, information retrieval tasks and instruction in them, and finally overall feedback compiled by the instructor. Students were novices in database searching and used the IR Game for the first time. The whole learning environment as such was emphasized in students' empathy-based stories in a very interesting way, revealing important issues that should be addressed in future research and development.

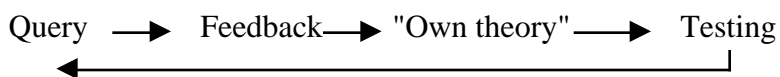
First, introduction and briefing for the learning situation and also the first exercise in the IR Game are of vital importance as a motivating and orientating factor. The first search task was designed as an easy and quickly approachable image retrieval task in order to find a picture of an Icelandic woman writer. Most of the searchers succeeded well, reaching 100 % recall/precision. This ultimate result was very astonishing and motivating for the searchers. An opposite observation revealed that some of the searchers could not orient towards image retrieval. They used similar strategies as in large full-text databases and their queries were far too focused and specific. The first search results were in this case obscure and the motivation to proceed declined. For instance, the video-log revealed that it was very difficult to change one's search orientation in such a situation.

Second, tutoring in the exercises was seen to be very important and supportive of learning. Advice, specific questions and the opportunity to ask questions were naturally seen as an important factor. The activity of the tutor replaced and complemented the inadequate help and hint facilities of the IR Game. Furthermore, the limited selection of predefined search tasks was evidently an obstacle to a motivating learning experience. Search requests were unfamiliar and uninteresting to the students. The linguistic expressions in the requests seemed to be too predefined and artificial.

Third, social situation of the learners was an important element in the learning situation. Exercises were done in groups of two or three students. In a few situations a student completed exercises all by her/himself. Many stories describing successful learning situations dealt with the importance of collaborative learning, discussion and articulation of problems and solutions at hand. Collaboration challenged the "playing the game" approach where the game-like features attracted the subjects to a groundless pursuit of high query performance without articulating the goals and strategies.

Although collaboration seemed to enhance learning it could also be an obstacle to the development of understanding. Some members of a group might be too well-informed or dominating.

Some stories of good learning experiences presented the learning process in the form of a simple model. For instance, one subject modeled learning activities in the following manner:



In these stories the performance feedback of the IR Game as well as the documents was actively analyzed. In these situations students did also put the "give a hint" feature to a good use. Through the analysis of the preceding query and the relevant documents found, student formed a theory of their own in order to improve their queries. This theory was tested by reformulating the query. The option to actively reformulate a query and test especially it was seen as a learning advancement.

5. Discussion

At the present stage of development, the IR Game is an important novel prototype tool in IR learning environment construction. Compared to categories in Jonassen [13], the IR Game clearly has functions that enhance learning *with* and *about* computing. In respect to components of learning environments according to Perkins [29], the IR Game represents phenomenaria, i.e., an area for presenting, observing and manipulating phenomena of IR. The game can also be seen as a construction kit for query modeling and analysis. Its offers chances to simulate different kinds of settings of search tasks, databases and retrieval systems. The descriptive evaluation of the IR Game revealed several functions and components that should be implemented into a more comprehensive IR learning environment.

The user evaluation revealed that the same features of the IR Game could be seen as factors enhancing or inhibiting learning. This may be a result of a individual learner's prior experience of IR or, for example, the learner's personal learning style. This is also a matter of the instructional design of exercises, and the learning situation as a whole. The IR Game leaves many options to the tutor to design a suitable learning environment for information retrieval. The tutor can modify the feedback mechanism, game-like features and help functions and also decisions concerning presentation and articulation of learning outcomes. The formulation and design of individual search tasks is dependent on the test databases used. In the present setting the database consisted of news stories from the beginning of the 90's. The first year students seldom had any interest in the themes presented in these.

Although the feedback mechanism of search success is the key function of the system it can also lead to surface level coping and gaming in order to reach high precision - recall scores. The solution to this problem could be found in the articulation and social negotiation of search statements and the learner's own theories on them. The rotation of excercies where performance feedback is given, not given or given delayed to the user could also constrain the adoption of surface level learning behavior. Browsing of search results and the use of the "Give a Hint" function could be developed to be more situational and interactive.

Database content, and the meta-information describing it, seem also to be important in an instructional setting. A real-world setting and up-to-date content of databases create motivation and improve the opportunity to transfer learning outcomes to real-life situations. It is possible to interface the IR Game to new databases and design new search task sets for them. Search tasks could also be built from the perspective of different kinds of learners and their areas of interest. Inclusion of harvested web-databases could certainly increase the use of the IR Game in open learning environments and distance education.

The search tasks in the test collections are well specified. This exactness and predefined linguistic expressions were seen as an obstacle for learning. One solution to this problem could be the integration of search requests within the framework of simulated real-life activities or simulated work-tasks, which are proposed in evaluation of interactive information retrieval systems [47]. This kind of approach is similar to anchored instruction, an instructional approach developed by the Cognition and Technology Group at Vanderbilt [48-50]. Anchored instruction is strongly associated with situated learning and constructive learning environments. The major goal of anchored instruction is to overcome the problem of inert knowledge by teaching problem solving skills and independent thinking. Anchored instruction in a learning environment is intended to permit sustained exploration by students and teachers and it enables them to understand the kinds of problems and opportunities that experts encounter and the knowledge that experts use as tools.

As was mentioned earlier, the IR Game is intended to support different types of learners from the beginner in a classroom to a researcher in an IR laboratory. A realistic search task situation can be created for each learner group by applying the basic tool appropriately. The social construction of understanding of the phenomena in IR was present in several stories describing successful learning experiences. In its current phase, the IR Game does not support collaborative learning and study, but these functions can be included in the setting with the help of instructional design in classroom setting or integrating communication tools and shared workspaces in to the IR Game.

5.1. Future research and development

Our aims for further research will be designing, implementing and evaluating a computer-supported, web-based, constructive learning environment for information retrieval. This includes the design and evaluation of alternative instructional methods, relevant to constructive learning. Some methods can easily be implemented in the computer-based environment, but some of them need to – and should – be maintained by human tutors. In the evaluation phase, learning outcomes in the different kinds of environments are evaluated.

The technical and instructional development of IR Game continues, based on the findings of user evaluation and observations from learning situations complemented with conceptions of learning environments. The four main areas of development are: 1) development of tools to support learner and tutor, such as communication, exercise building and answering, search and learning history analysis (e.g. electronic search portfolios), 2) tools for the seamless integration of search systems (Boolean and best-match) and analysis tools for their comparison both at system and individual query levels, 3) identification of the instructional options of different search settings and tasks, 4) integration and evaluation of the usage of textual, image, multimedia and web-document databases into the IR Game.

Scripts of Method of Empathy Based Stories

Student uses IR Game in the basic course in information retrieval. Student conducts searching on both image database and text database. After the exercises student thinks that he/she has learned a great deal about information retrieval. Use your imagination and write a realistic story or description of elements which enhanced learning while using the IR Game.

Student uses IR Game in the basic course in information retrieval. Student conducts searching on both image database and text database. After the exercises student thinks that he/she had not learned a lot about information retrieval. Use your imagination and write a realistic story or description of elements which inhibited learning while using the IR Game.

References

- [1] S. Harter, *Online information retrieval : concepts, principles, and techniques*. Orlando: Academic Press, 1986.
- [2] W. R. Hersh, *Information retrieval : a health care perspective*. New York: Springer, 1996.
- [3] F. W. Lancaster and A. J. Warner, *Information retrieval today*. Arlington: Information resource press, 1993.
- [4] A. Large, L. A. Tedd and R. J. Hartley, *Information seeking in the online age : principles and practice*. London: Bowker, 1999.
- [5] E. Caruso, Online training for searching online. Paper presented at the 2nd International Online Information Meeting. London, 5-7 December, 1978.
- [6] E. Caruso, *TRAINER : Computer assisted learning and practice for users of DIALOG/ORBIT*. Final report. Pittsburg University, PA, 1979.
- [7] E. Caruso, Computer aids to learning online retrieval. In M. E. Williams, ed. *Annual Review of Information Science and Technology (ARIST)*. White Plains: ASIS & Knowledge Industry Publications, 1981. pp. 317-336.
- [8] J.L. Chapman, A state transition analysis of online information-seeking behavior. *Journal of the American Society for Information Science* 32 (1981) 325-333.
- [9] J. C. Griffith and N. P. Norton, A computer assisted instruction program for end users on an automated information retrieval system. In National Online Meeting. Proceedings of the Second National Online Meeting, New York, USA, 24-26 March 1981. Medford (NJ): Learned Information, 1981. pp. 239-248.
- [10] C. T. Meadow, E. S. Aversa and T. T. Hewett, A computer intermediary for interactive database searching. I. Design. *Journal of the American Society for Information Science* 33(1982) 325-332.
- [11] K.H. Stirling, A micro-based emulator for on-line search services. *Drexel Library Quarterly* 20(1984) 87-97.

- [12] K. Markey and P. Atherton, ONTAP online training an practice manual for ERIC data base searchers. Syracuse, New York: ERIC Clearinghouse on Information Sources, Syracuse University, 1978.
- [13] D. H. Jonassen, Computers in classroom : mindtools for critical thinking. Englewood Cliffs (NJ): Prentice Hall, 1996.
- [14] S.M. Alessi and S.R. Trollip, Computer based instruction : methods and development. 2nd edition. Englewood Cliffs (NJ): Prentice Hall, 1991.
- [15] D. Kopec and R. B. Thompson (Eds.), Artificial intelligence and intelligent tutoring systems : knowledge-based systems for learning and instruction. New York: Horwood, 1992.
- [16] R. Glaser, The maturing the relationship between the science of learning and cognition and educational practice. Learning and Instruction 1(1991) 129-144.
- [17] J.S. Brown, A. Collins and P. Duguid, Situated cognition and culture of learning. Educational Researcher 18 (1989) 32-42.
- [18] J. Lave and E. Wenger, Situated learning : legitimate peripheral participation. Cambridge: Cambridge University Press, 1991.
- [19] B.G. Wilson and K.M. Myers, Situated cognition in theoretical and practical context. In D. H. Jonassen & S. M. Land, eds. Theoretical foundations of learning environments. Hillsdale (NJ): Lawrence Erlbaum Associates, 2000. pp. 57-88.
- [20] S. A. Barab and T.M. Duffy, From practice fields to communities of practice. In D. H. Jonassen & S. M. Land, eds. Theoretical foundations of learning environments. Hillsdale (NJ): Lawrence Erlbaum Associates, 2000. pp. 25-56.
- [21] S.D. Teasley and J. Rochelle, Constructing a joint problem space : the computer as a tool for sharing knowledge. In by S.P. Lajoie and S.J. Derry, eds. Computers as cognitive tools. Hillsdale (NJ): Lawrence Erlbaum Associates, 1993, pp. 229-257.
- [22] A. Collins, J. S. Brown and S.E. Newman, Cognitive apprenticeship : teaching the crafts of reading, writing and mathematics. In L. B. Resnick, ed. Knowing, learning and instruction : essays in honor of Robert Glaser. Hillsdale (NJ): Lawrence Erlbaum Associates, 1989. pp. 453-494.
- [23] C. Dede, The evolution of constructivist learning environments : immersion in distributed virtual worlds. In B. G. Wilson, ed. Constructivist learning environments : case studies in instructional design. Englewood Cliffs: Educational technology publications, 1996. pp. 165-178.
- [24] D.C. Edelson, R.D. Pea, and L. Gomez, Constructivism in collaboratory. In B. G. Wilson, ed. Constructivist learning environments : case studies in instructional design. Englewood Cliffs: Educational technology publications, 1996. pp. 151-164.

- [25] J. Lowyck and J. Ellen, Transitions in the theoretical foundations of instructional design. In T.M.Duffy, J. Lowyck and D.H. Jonassen, ed. *Designing environments for constructive learning*. Berlin: Springer, 1993. pp. 213-226.
- [26] W. Winn, A constructivist critique of the assumptions of instructional design. In T.M.Duffy, J. Lowyck and D.H. Jonassen, ed. *Designing environments for constructive learning*. Berlin: Springer, 1993. pp. 189-212.
- [27] E. Kintsch, Principles of instruction from research on human cognition. In J.M. Spector, M.C. Polson & D.J.Muraida , eds. *Automating instructional design : concepts and issues*. Englewood Cliffs (NJ): Educational technology publications, 1993. pp. 23-44.
- [28] L. Vygotsky, *Mind in society : the development of higher psychological processes*. Cambridge (Mass.): Harvard University Press, 1978.
- [29] D.N. Perkins, Technology meets constructivism : do they make a marriage? *Educational Technology* 31(1991) 18-23.
- [30] M. Hannafin, *Learning in open-ended environments : tools and technologies for the next millennium* [Online]. Available: <http://dragon.ep.usm.edu/~dsurry/ole.htm>. (1999) [2000 January 27]
- [31] M. J. Hannafin and S. Land, The foundations and assumptions of technology enhances, student-centered learning environments. *Instructional Science* 25(1997) 167-202.
- [32] M. J. Hannafin, S. Land and K. Oliver, Open learning environments : foundations, methods, and models. In C. Reigeluth, ed. *Instructional theories and models* (Volume II). Mahwah (NJ): Lawrence Erlbaum Associates, 1999.
- [33] E. Sormunen, J. Laaksonen, H. Keskustalo, J. Kekäläinen, H. Kemppainen, H. Laitinen, A. Pirkola, and K. Järvelin, IR Game : a tool for rapid query analysis in cross-language IR experiments. In Proceedings of the Joint Workshop on Cross Language Issues in Artificial Intelligence and Issues of Cross Cultural Communication, 24th November 1998. Singapore: Kent Ridge Digital Labs, 1998. pp. 22-32.
- [34] K. Sparck Jones, and C. J. van Rijsbergen, Information retrieval test collections. *Journal of Documentation* 32(1976) 59-75.
- [35] J. Tague-Sutcliffe, The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management* 28(1992) 467-490.
- [36] D. Hull, Stemming Algorithms: a case study for detailed evaluation. *Journal of the American Society for Information Science* 47(1996) 70-84.
- [37] E. Voorhees, & D. Harman, Overview of the sixth Text REtrieval Conference (TREC-6). *Information Processing & Management* 36(2000) 3-35.
- [38] K. Sparck Jones, Further reflections on TREC. *Information Processing & Management* 36(2000) 37-85.

- [39] P. Ingwersen, Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation* 32(1996), 3-50.
- [40] E. Sormunen, *A method for measuring wide range performance of boolean queries in full-text databases*. Doctoral Thesis. Acta Electronica Universitatis Tampereensis. Tampere: University of Tampere, 2000. Available: <http://acta.uta.fi/pdf/951-44-4732-8.pdf>. [2000 May 30]
- [41] H. Drenth, A. Morris & T. Gwyneth, Expert systems as information intermediaries. In M. E. Williams, ed. *Annual Review of Information Science and Technology (ARIST)*. Medford: ASIS & Learned Information, 1991. pp. 113-154.
- [42] K. Järvelin, J. Kristensen, T. Niemi, E. Sormunen and H. Keskustalo, A deductive data model for query expansion. In Proceedings of the 19th International Conference on Research and Development in Information Retrieval (ACM SIGIR '96), Zürich, Switzerland, August 18-22, 1996.
- [43] A. Eskola, Non-active role-playing : some experiences. In A. Eskola in collaboration with A. Khilström, D. Kivinen, K. Weckroth and O. Ylijoki. *Blind alleys in social psychology : a search for ways out*. Amsterdam: North Holland, 1988.
- [44] J. Eskola, *Eläytymismenetelmä sosiaalitutkimuksen tiedonhankintamenetelmänä* [The method of empathy-based stories as a method of acquiring data in social research]. Doctoral Thesis. University of Kuopio. Tampere: Taju, 1998.
- [45] G.P. Ginsburg, Role-playing and role performance in social psychological research. In M. Brenner, P. Marsh & M. Brenner, eds. *The social context of method*. London: Groomhelm, 1978.
- [46] G.P. Ginsburg, The effective use of role-playing in social psychological research. In G. P. Ginsburg, ed. *Emerging strategies in social psychological research*. New York: John Wiley & Sons, 1979.
- [47] P. Borlund, Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation* 56(2000) 71-90.
- [48] Cognition and Technology Group at Vanderbilt, Anchored instruction and its relationship to situated cognition. *Educational Researcher* 19 (1990) 2-10.
- [49] Cognition and Technology Group at Vanderbilt, Technology and the design of generative learning environments. *Educational Technology* 31 (1991) 34-40.
- [50] Cognition and Technology Group at Vanderbilt, The Jasper experiment : an exploration of issues in learning and instructional design. *Educational Technology Research and Development* 40 (1992) 65-80.

¹ Marketed by TietoEnator, <http://www.tietotech.se/trip/>.

² InQuery is an information retrieval system developed at the Center for Intelligent Information Retrieval, Computer Science Department, University of Massachusetts, <http://ciir.cs.umass.edu/>.

³ An extension designed for Cross-Lingual IR research, omitted here.