Article

Design of a digital dissertation information management system

William Bradley Glisson and Gobinda G. Chowdhury

The authors

William Bradley Glisson is a former MSc student, University of Strathclyde, Glasgow, UK.
E-mail: Glisson@aol.com
Gobinda G. Chowdhury is a Senior Lecturer in the Department of Computer and Information Sciences, University of Strathclyde, Glasgow, UK.
E-mail: gobinda@dis.strath.ac.uk

Keywords

Digital libraries, Information systems, Systems design, Information retrieval, Databases

Abstract

This paper describes a prototype dissertation information system. The various components of the system include the database, the information retrieval engine, dynamic Web pages via markup languages, and client side JavaScripts. Managerial and information retrieval issues associated with a digital dissertation information system are discussed and the database configurations and codes written for performing various operations online are described. The paper ends with a discussion of the project results and future developmental opportunities.

Electronic access

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/0033-0337.htm

Program
Volume 36 · Number 3 · 2002 · pp. 152–165
 · ISSN 0033-0337
DOI 10.1108/00330330210440449

1. Introduction

Dissertations contain extremely valuable information since they report the results of first-hand research information conducted by students within any specific department. It is often necessary for a system to be developed and implemented in order to allow other researchers, students and academic staff, easy and prompt access to this information. Along with the development of an access mechanism, there is also a need to develop mechanisms for managing the system that will aid the students, supervisors, dissertation co-ordinator(s) and support staff in electronic communication. Such a system would facilitate improved correspondence between the supervisor(s) and the students and would have the ability to produce statistics and generate reports on a real-time basis. The research reported in this paper is concerned with the design and implementation of a simple digital dissertation information system applicable to any academic institution of higher education. Various digital dissertation management systems have been developed in the recent past, the most prominent one being that developed at Virginia Tech, in the USA, which has given rise to the Networked Digital Library of Theses and Dissertations (NDLTD) (Fox *et al.*, 1997; Suleman *et al.*, 2001a, b).

2. Potential benefits

Traditionally the department maintains a file with minimum information on dissertations. Typically the file contains the dissertation titles along with the students' and supervisors' name. This information is entered as soon as a student submits a title, and a supervisor is allocated. After that, the student and the supervisor work in their own way: in most cases the student submits draft chapters to the respective supervisor, and gets back the versions with corrections/comments, etc. The documents are sent back and forth either as e-mail attachments, or in printed form. Printed copies of communications, such as e-mail messages, etc., between a student and his/her supervisor, and notes etc. from the supervisor, are usually kept in hard copy files maintained for each student.

The proposed digital dissertation system aims to automate the entire process such that

everything takes place in a digital environment. The new system would allow the students to submit a topic electronically, and would allow the department office to store the dissertation registration information in a database that would not only show the title, date of registration, supervisor's name, etc., for a given student, but would also have room for storing specific notes and comments that may be useful for various administrative purposes. The next most important aspect of the system is that it would allow the student and the supervisor to work together in a secured digital environment. The system would allow the student to submit a draft chapter, or any other form of communication, and would allow the concerned supervisor to work on the file electronically and submit his/her corrections, suggestions, specific comments, etc. This would therefore allow both the parties to keep track of the progress of the work, etc. Finally, when the dissertation is submitted and examined, it would be included into the digital library of theses and dissertations that can be accessed by anyone. Different levels of security will allow only authorised users to get access to the respective data, files and documents. For example, a given student can only see his works with all the comments, etc.; a given supervisor can see files for all his/her students; the dissertation co-ordinator, head of the department, and office staff, may be able to see certain parts of the system to facilitate their activities, and so on.

Such a digital dissertation information system should have three major functions:

(1) to help managers (dissertation co-ordinators, support staff, etc.) in the management of information on past dissertations;
(2) to help supervisors and students perform current dissertation-related activities online;
(3) to provide online access to (search and browse) the dissertations.

To illustrate the issues associated with project development, design, and implementation, the construction of a Dissertation Information System (DIS) at the University of Strathclyde in Glasgow, UK was chosen as a case study. The potential benefits of such a system to the university and the department for administrative functions are many. The system will help to:

- locate and identify individual dissertations by year, author, title, course, subject/topic and keywords;
- prevent duplication of research;
- identify, via reports, when dissertations are sent for examination, when students were granted extensions and the reasons for the extension, and so on;
- save time by producing quick reports on the assignment of dissertation topics and dissertation advisors;
- run plagiarism detection software in the future on the system to prevent plagiarism in the department;
- help influence the university policies on the presentation of digital dissertations.

The potential benefits to the faculty staff and dissertation supervisors are that the system:

- can be used as a source of information for suggesting topics to potential dissertation candidates;
- can be used as a source for tracing the development of research tracks within a department, on area of research, and so on;
- can provide an easily accessible source of information regarding a standard dissertation format and style;
- can be used as a resource for identifying the areas of expertise of lecturers by quickly generating reports on the dissertation topics supervised by an individual over a period of time;
- will be available for access any time of the day and year;
- will facilitate communication between the advisor and student;
- will provide records of every communication between the advisor and student that can be used for future reference at any point of time;
- will provide ideas for future research topics which can be generated and dispersed conveniently.

For students undertaking dissertations such a dissertation information system will be useful for providing:

- convenient access to research dissertations from anywhere – on or off campus;
- convenience to do research based on the student's time schedule i.e. a student would not need to be physically present to hand in a chapter (or the whole report), get feedback from supervisors, and so on;

- a convenient form of communication between student and supervisor(s);
- an appropriate mechanism for keeping records of communication between the supervisor(s) and the student which can be referred to for corrections and/or modifications, etc.;
- better control on the research schedule.

The other benefits to all parties involved in implementing such a system include savings in time and effort. The hidden cost behind the storage of hard-copy material, the cost of dispersing paper to students and the time lost for travelling (even within the campus), waiting, for student-supervisor meetings etc. can be reduced. It can also be argued that the reduction in time, effort and cost will allow the students to exert a greater focus on other avenues of research and other aspects of their dissertations.

3. Methodology

The following outline was followed to accomplish this project:

Problem definition. This involved defining the specific scope of the problem, establishing stakeholders and conducting interviews using structured and open-ended questionnaires:

- Defining user requirements. This involved developing questionnaires, determining the target audience, setting an interview schedule and conducting interviews.
- Outlining options to solve the problem. This stage involved exploring languages, exploring Database Administration (DBA) Systems and exploring operating systems.
- Defining system specifications based on the user requirements.
- Developing a prototype.
- Testing the prototype.

A survey of the needs of 16 dissertation students and seven advisors/supervisors within the Department of Information Science at Strathclyde are shown in Table I.

A survey of the technical resources at hand revealed an inventory consisting of the following:

(1) The Department had a Web server with available space.
(2) The Department also had a server configured to run NT 4.0 with service pack 6.
(3) Several different database software packages (Access, FileMaker Pro, older versions of Microsoft SQL Server) were available in the Department with FileMaker Pro being the preferred database due to ease of use by the system's administrator.
(4) Some dissertation information was stored on a FileMaker Pro database. For each dissertation student the following information was recorded on completion of the dissertation:
    - author – student's name;
    - title – title of the dissertation;
    - course;
    - date – the date that the dissertation was submitted;
    - keywords;
    - supervisor.
(5) Several different Web-authoring tools were used in the department: Front Page, Word HTML, Text Editors, and Dream Weaver with the latter being the most common.
(6) Templates were being used in the Department for the design of the Web pages.
(7) A search engine (HT://Dig) was being used in the Department that had the ability to index HTML pages, text, and PDF documents.

The technical skills being utilised, within the Department, that have an impact on this application included:

- programming skills for C++, PHP and HTML;
- administrative knowledge of the Unix operating system;
- software knowledge of FileMaker Pro and Dream Weaver.

4. Proposed solutions

The solutions were considered from a long-term and a short-term point of view. The short-term solution was developed into a fully functional prototype along with the specific aspects of the long-term solution. The proposed short-term solution used the following tools and techniques:

Table I User requirements summary

	Students	Percentage	Advisors	Percentage	Total response	Total percentage
Would use the system	16	100.00	6	85.71	22	95.65
Would not use the system		0.00	1	14.29	1	4.35
Would use a system to help manage current dissertation work		0.00	6	85.71	6	26.09
Would not use a system to help manage current dissertation work		0.00	1	14.29	1	4.35
Ability search abstracts only	3	18.75	2	28.57	5	21.74
Ability search dissertations only		0.00		0.00	0	0.00
Ability to search both abstracts and dissertations	13	81.25	5	71.43	18	78.26
Would use a directory for stored communication	13	81.25	5	71.43	18	78.26
Would not use a directory for stored communication	2	12.50	1	14.29	3	13.04
Not sure if they would use a directory for stored communication	1	6.25	1	14.29	2	8.70
Would use a calendar booking system	3	18.75	3	42.86	6	26.09
Would not use a calendar booking system		0.00	2	28.57	2	8.70
Would like a general itinerary	5	31.25		0.00	5	21.74
Would like to have both a general itinerary and a calendar system	6	37.50		0.00	6	26.09
Search by year	14	87.50	7	100.00	21	91.30
Search by author	7	43.75	7	100.00	14	60.87
Search by title	16	100.00	7	100.00	23	100.00
Search by subject	16	100.00	7	100.00	23	100.00
Search by course	15	93.75	5	71.43	20	86.96
Search by words within text	16	100.00	7	100.00	23	100.00
PDF file format	9	56.25	5	71.43	14	60.87
HTML file format	9	56.25	4	57.14	13	56.52
XML file format	1	6.25		0.00	1	4.35
RTF file format	2	12.50		0.00	2	8.70
Word file format	7	43.75	1	14.29	8	34.78
Latex file format		0.00	1	14.29	1	4.35
Would like to see links to search engines	12	75.00	5	71.43	17	73.91
Would like to see links to Internet dissertation and thesis systems	14	87.50	6	85.71	20	86.96
Would like to see links to an online dissertation guide	15	93.75	5	71.43	20	86.96
Would use an online dissertation topic application	16	100.00	7	100.00	23	100.00
Online feedback form	7	43.75	4	57.14	11	47.83
E-mail address for feedback	8	50.00	2	28.57	10	43.48
Would like to see both a feedback form and e-mail	1	6.25		0.00	1	4.35
Use digital calendar		0.00	4	57.14	4	17.39
Critical compatibility with new system for use		0.00	1	14.29	1	4.35
Type of calendar used		Two palm pilots, HTML calendar, Auspice				

Notes: Total number of students interviewed 16, total number of advisors interviewed 7, total number of interviews 23

- HTML (Hypertext Markup Language) and Dream Weaver were used to develop the seven front-end Web pages;
- CDML (Claris Dynamic Markup Language) was used to create 14 dynamic Web pages that interact with the FileMaker Pro database;
- appropriate database security was set up;
- current security available on the Unix departmental server was used;
- client-side validation was written in JavaScript;
- abstracts of dissertations were stored in HTML format and in a separate directory;
- dissertations were stored in PDF format and in a separate directory;
- HTML was used as the file format for bibliographies;
- a new FileMaker Pro database was created to support the Dissertation Information System with the appropriate database files;
- a file structure was created to hold the advisor/student correspondence with the

appropriate security. The transfer of data to and from the server for correspondence between the advisor and the students takes place via a File Transfer Protocol (FTP) account. A directory had to be created within the advisor's directory that would allow the students to upload and download correspondence as desired. The students would not have the right to delete from this directory. These directories would serve as a repository for holding information and proof of correspondence to protect both parties. The file server had also been configured to allow both parties the appropriate access. This would allow for direct access from the University without going through the Web page.

- the HT://Dig search engine in conjunction with the Adobe Acrobat Acoread was used to automatically index Portable Document Format (PDF) files.

The long-term solution would involve the investment in an information retrieval (IR) system and implementation of a relational database. The prototype also examined the information retrieval system dtSearch. A trial copy of dtSearch was downloaded (http://www.dtsearch.com), installed and configured to search dissertations. The modifications to the design of the short-term solution should be minimal when it comes to the implementation of the long-term solution. The HTML abstracts and the PDF dissertations used in the short-term solution can be moved to separate directories on the Windows server. The data in the database will need to be converted from FileMaker Pro to a relational database. This could prove to be a very tedious process depending on the export capabilities in FileMaker Pro.

It was proposed that in the long run, the Department should consider converting the departmental database to a relational database. The departmental database would be the master database, while the dissertation database would be the child database. Fields would need to be defined in the master database so that they could be mapped on to the child database as needed. This would reduce potential coding problems and increase the long-term reporting potential of the system by allowing for data to be shared between the two systems. This sharing of data would necessitate the implementation of appropriate security measures.

The major change would be to purchase an information retrieval system like dtSearch. One can, at this stage, use the new software to construct a Web front-end to the IR system, and use the existing design that would allow the end-user to choose an abstract search or a full-text dissertation search. The links to the Department's front-end (Web page) could be modified to point to the new software. Separate indexes could be built for the abstract search and the full dissertation search thereby improving the retrieval features. New code would need to be developed for the application pages to be able to communicate with the relational database. The JavaScript that is used for the current client validation could be used to validate the new application forms. The short-term and long-term system design overviews are shown in Figure 1.

5. Implementation and testing

The overall design of the system can be viewed through the use of Data Flow Diagrams (DFD). A data flow diagram is defined as "a graphical technique that depicts information flow and the transforms that are applied as data move from input to output" (Pressman, 1992, p. 292). The various levels of information flow are represented in Figures 2, 3 and 4. The first DFD, Figure 2, illustrates a high level overview of the system. The next level DFD, Figure 3, begins to break this down into greater detail. The subsequent levels from this point target individual aspects of the system as depicted in Figure 4. Some screen shots, Figures 5-13, have been provided to demonstrate how the system functioned upon completion of the prototype.

It should be noted that the link for MSc Statistics only appears if the individual is supposed to see this link as indicated in Figure 6.

Students can submit a topic for their dissertation through a form as shown in Figure 9. Once it is done successfully, a confirmation message can be seen by the student (see Figure 10). The student can revisit the page and edit the topic or any other

Figure 1 System design

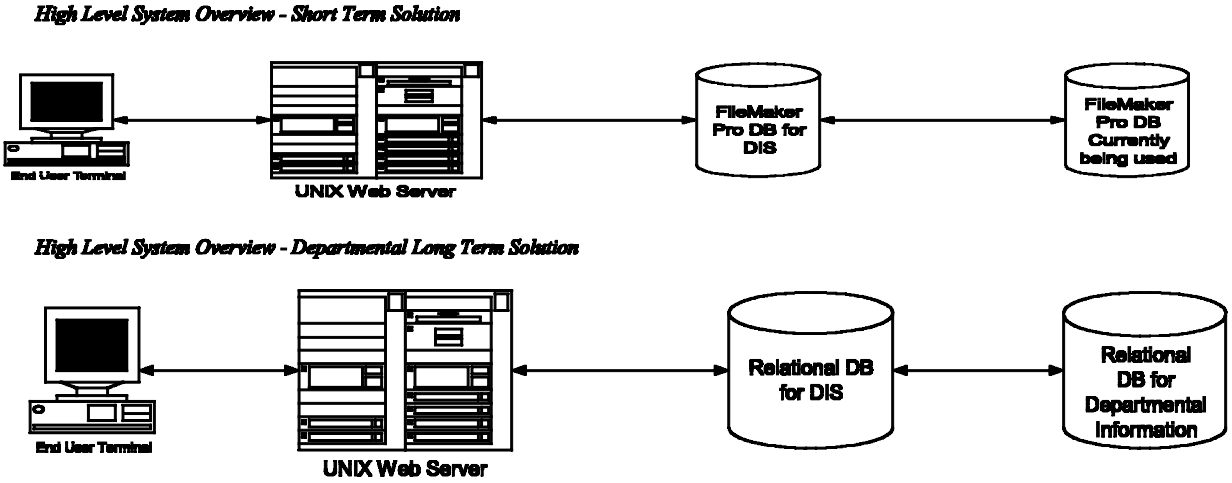
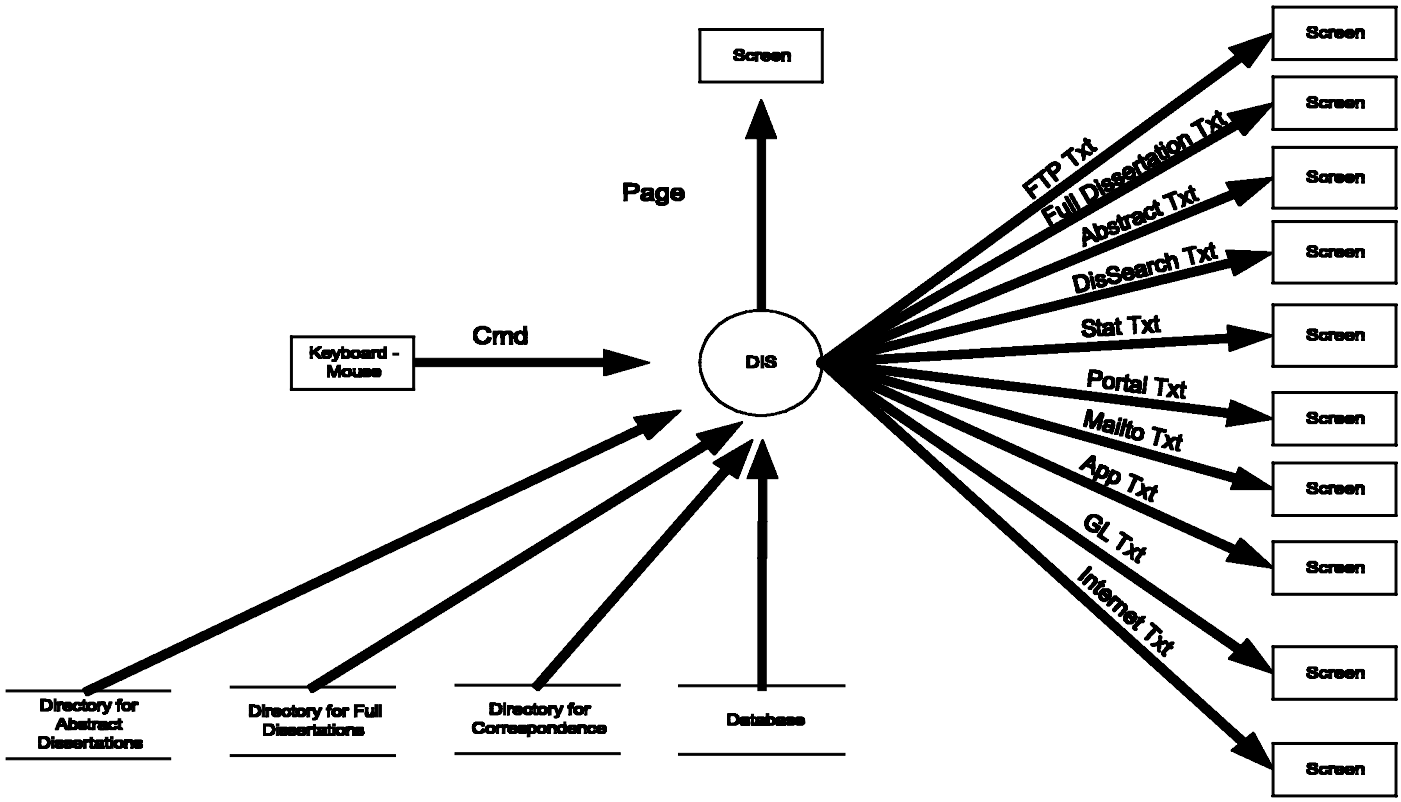


Figure 2 Dissertation information system level 0



field until the topic and the supervisor is approved by the Department.

Staff members can check the number of students being supervised by them at any given point of time, and can also see the current status of each of their students. Figure 11 shows the screen for staff login and Figures 12 and 13 show the search results screens for all the students, and one particular student, respectively.

5.1 Database design issues

Due to a need to increase database security on particular fields, the main file was divided into two separate files and Web security was added to the site. The Web security applied to the field reg_id made this field an exact search field. An exact search field in the FileMaker Pro database means that the field's contents have to be exact to bring back a positive response from the database. Web security was

Figure 3 Refinement of DIS DFD – level 1

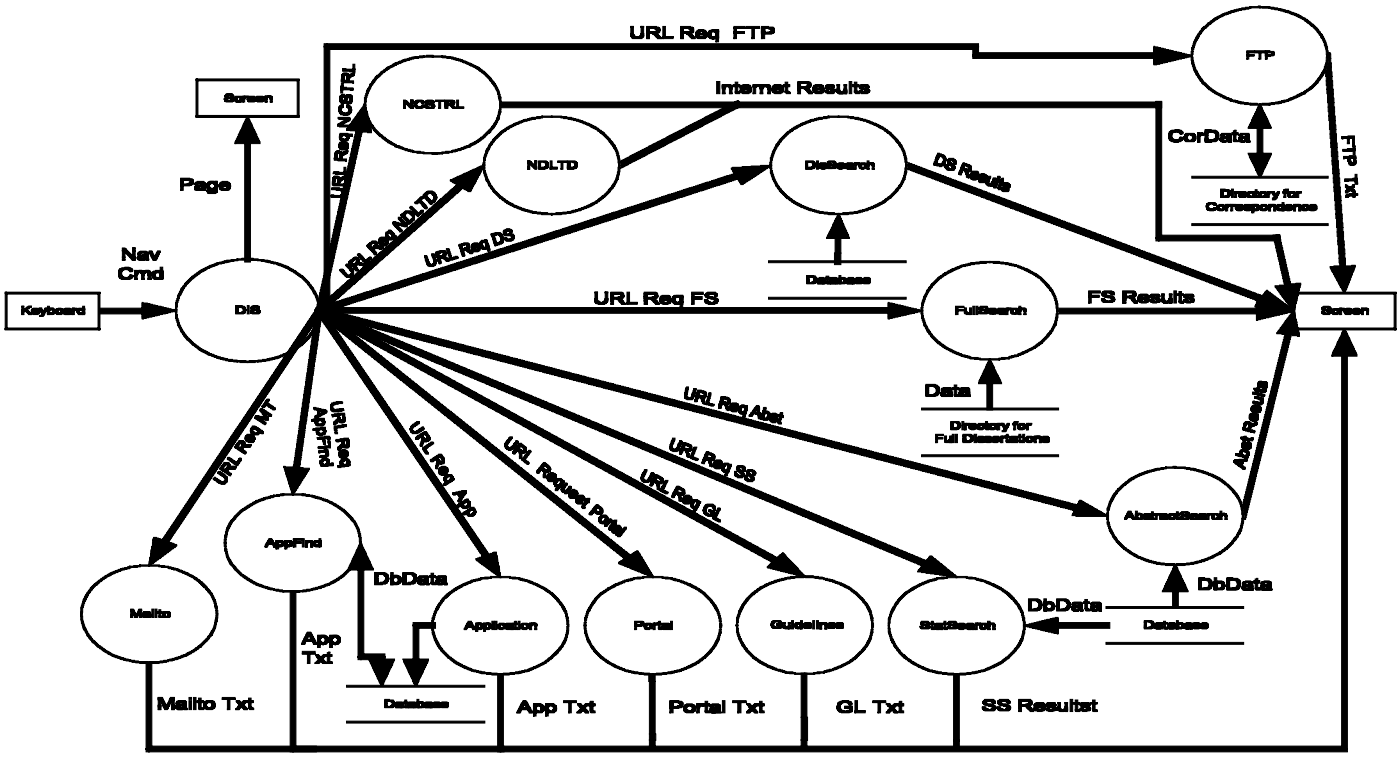
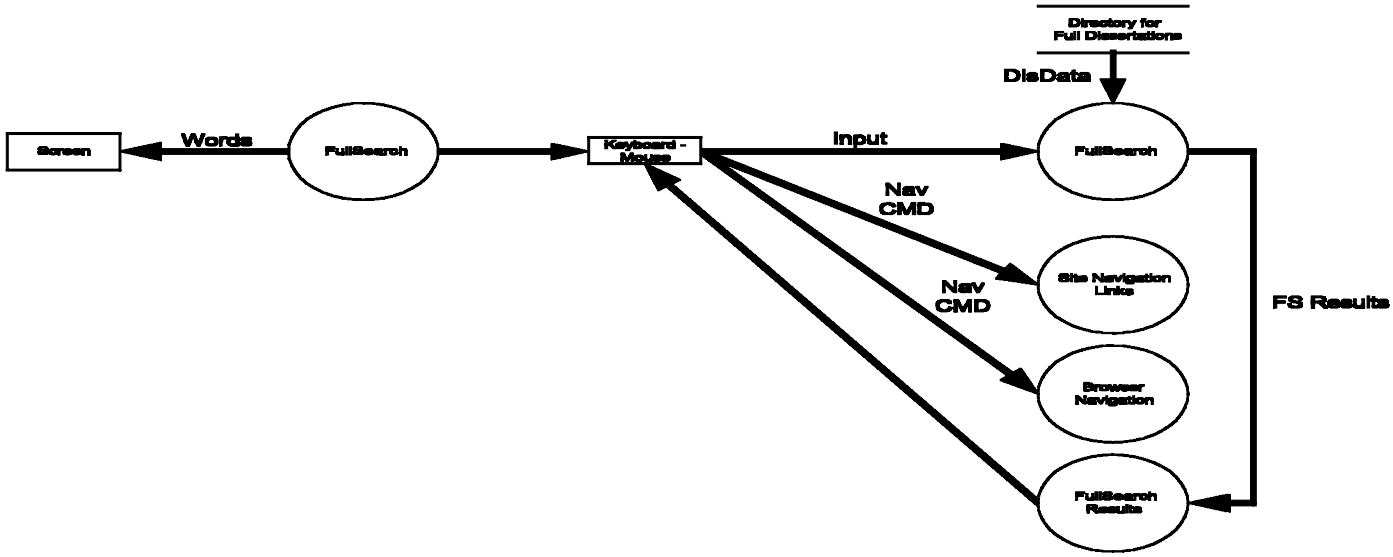


Figure 4 Refinement of FullSearch



also applied to the field password making this field an exact search field, a field that is not displayed, and an exact update field. This resulted in the file, Dissertations, being altered to create two files, which are Dissertations and DisData. The security applied to DisData initiates a login prompt box for anyone trying to view the data in this file.

The increased field security also made the Dissertations file non-searchable by the general public. To solve the search issue, a third file was created, which was DisSearch. The effect of creating DisSearch is similar to what a relational database would call a "view". The creation of the third file means that the creation of a new record in the database now requires information updates in three separate files. The final solution is to create a new record in the DisSearch file. The DisSearch file needs a field that will automatically create a serial number to act as

Figure 5 Login prompt for system access

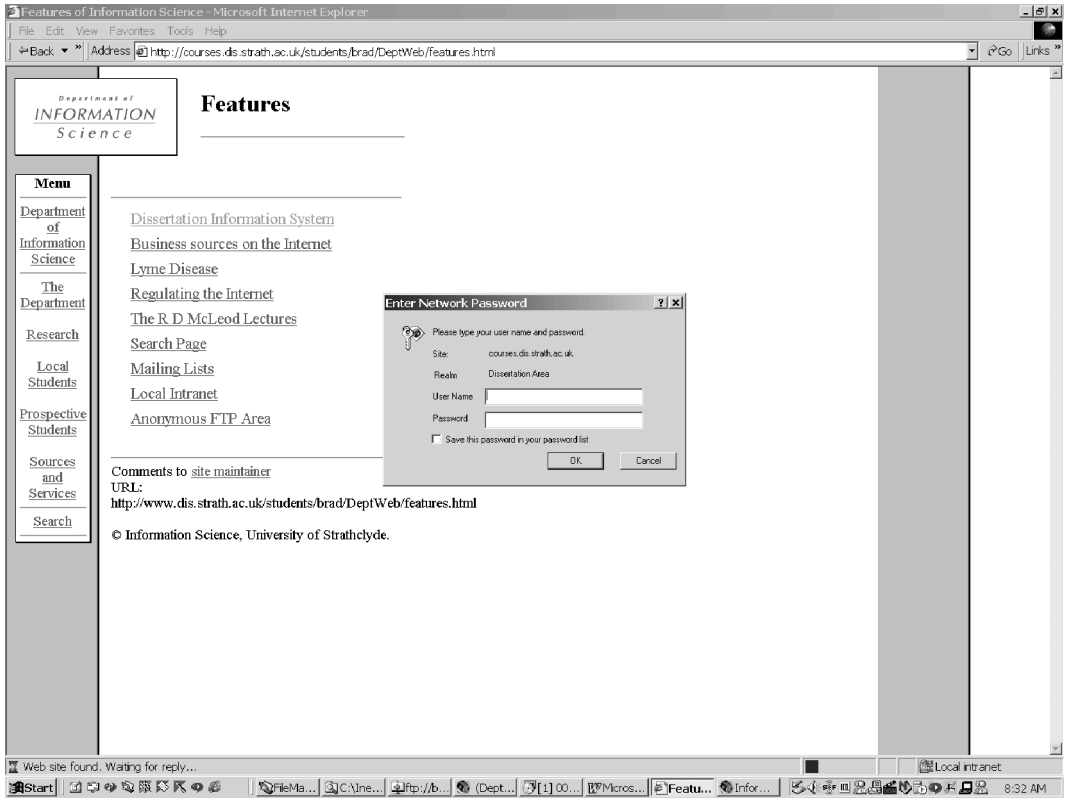


Figure 6 Dissertation information system home page

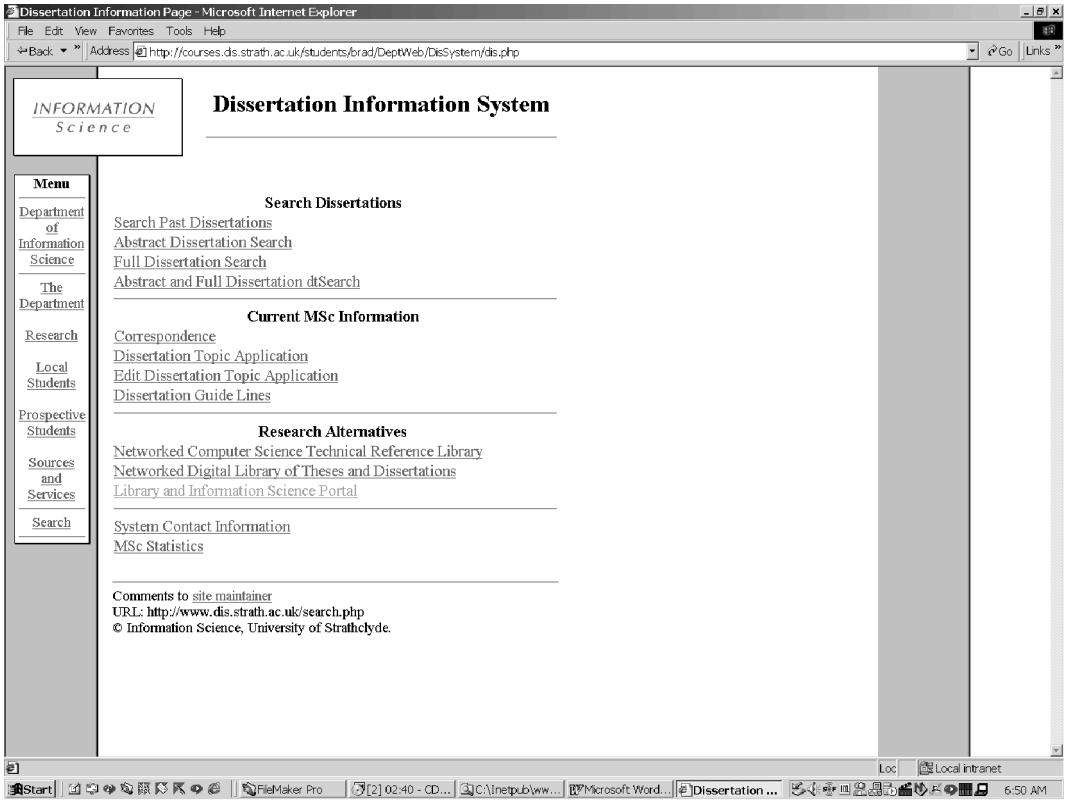


Figure 7 Dissertation search form

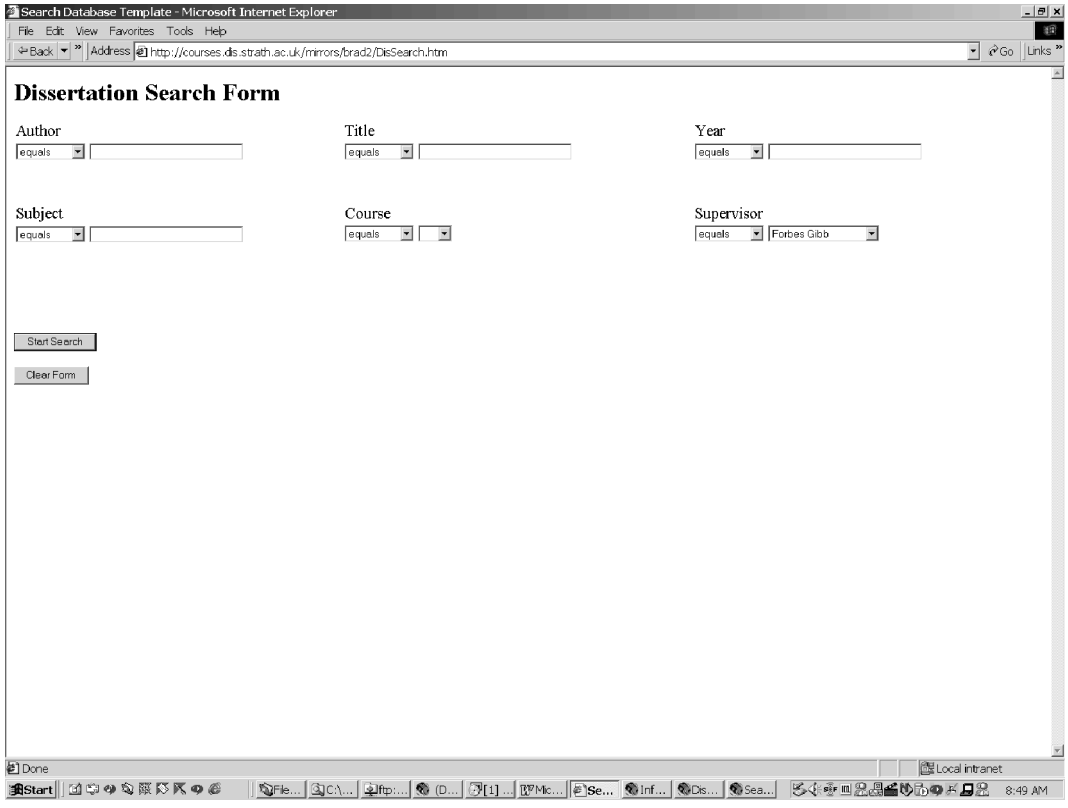


Figure 8 Search results

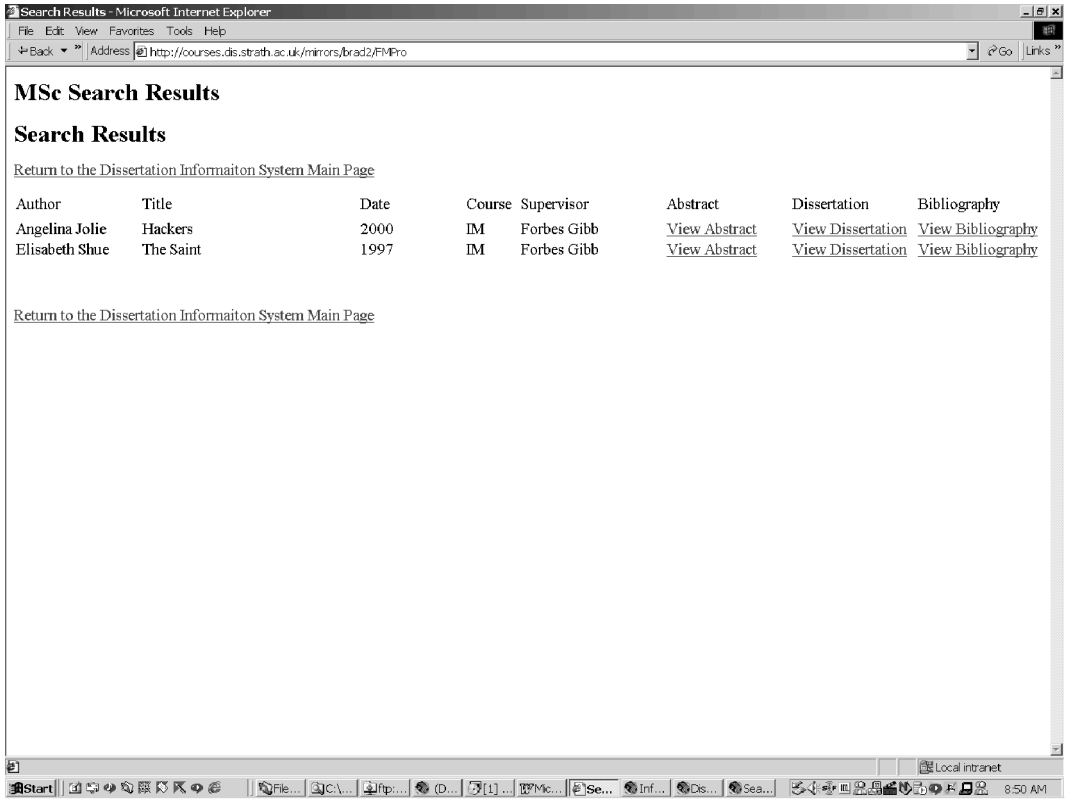


Figure 9 Topic application

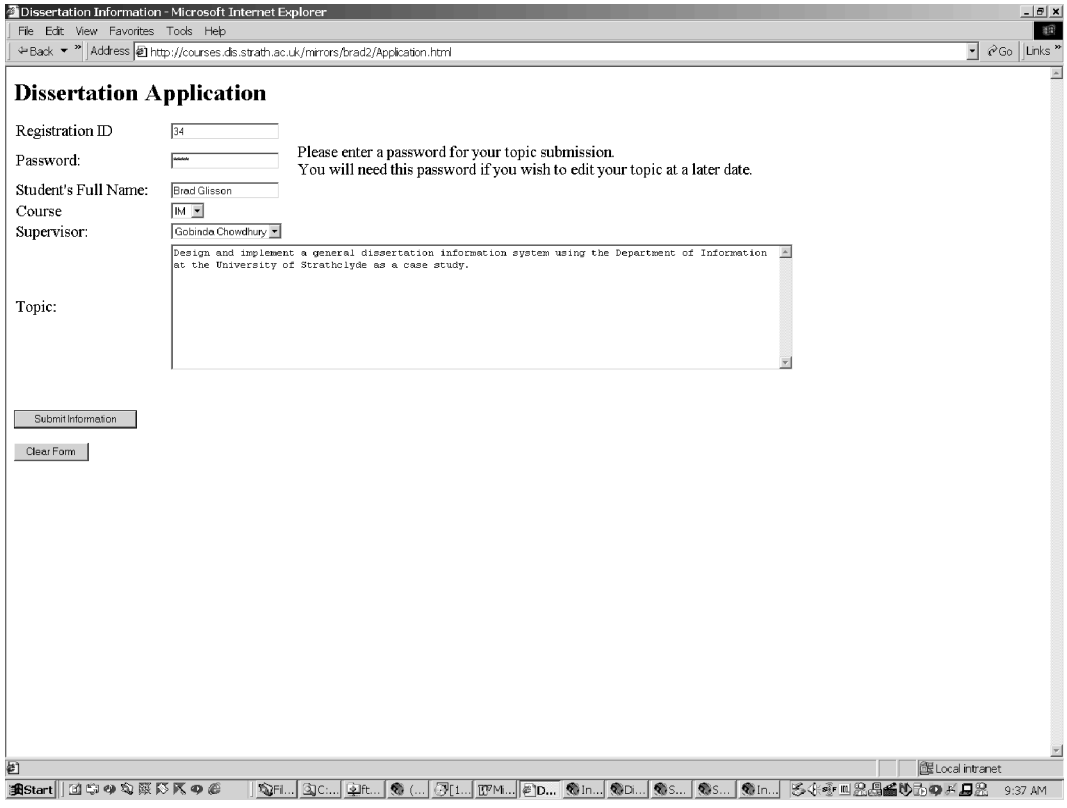


Figure 10 Application confirmation page

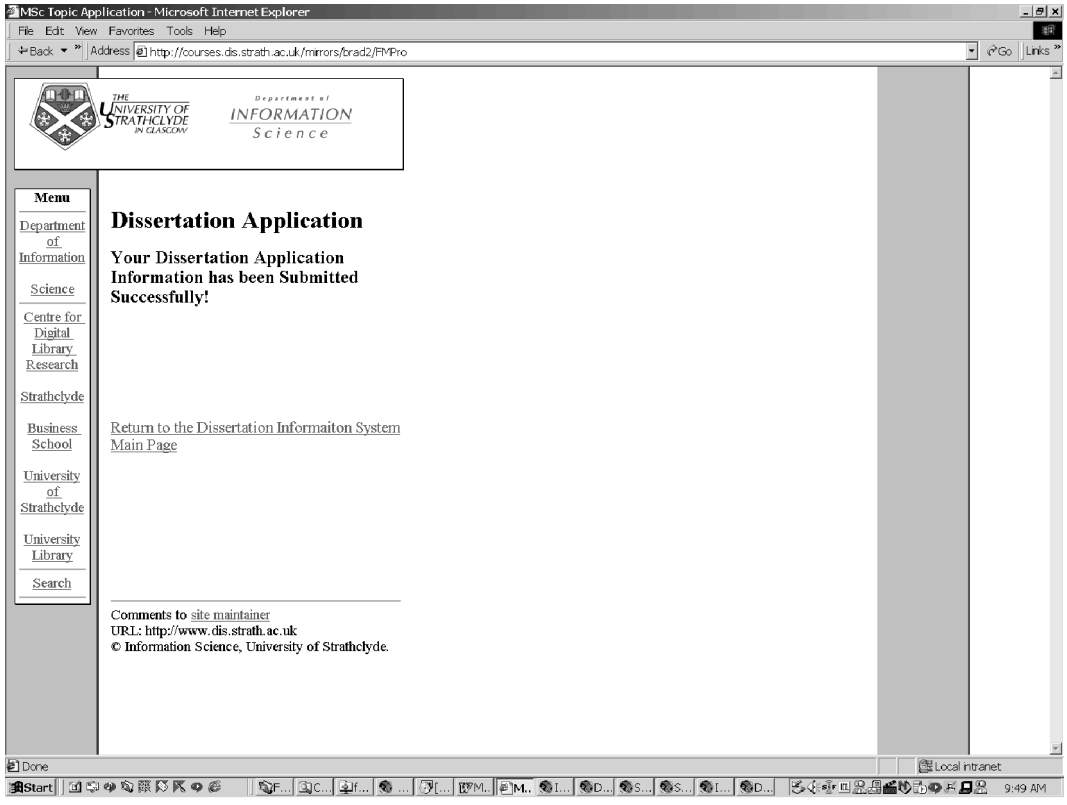


Figure 11 Staff login prompt box

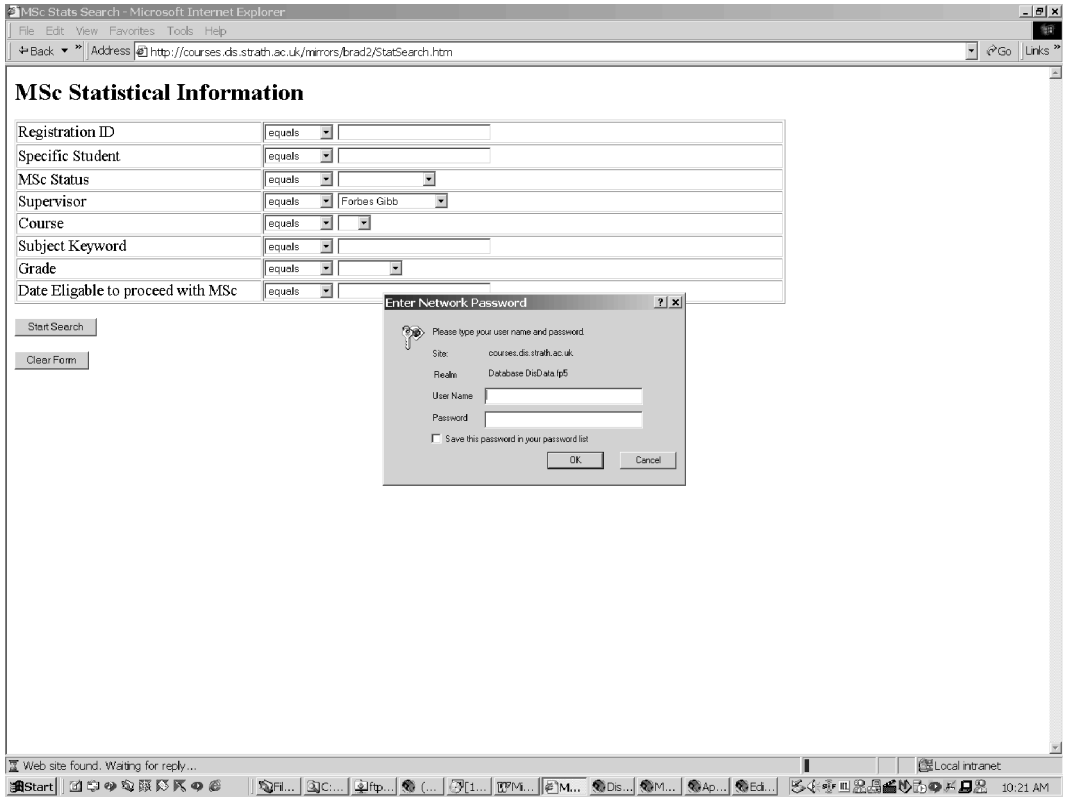


Figure 12 MSc search results for students

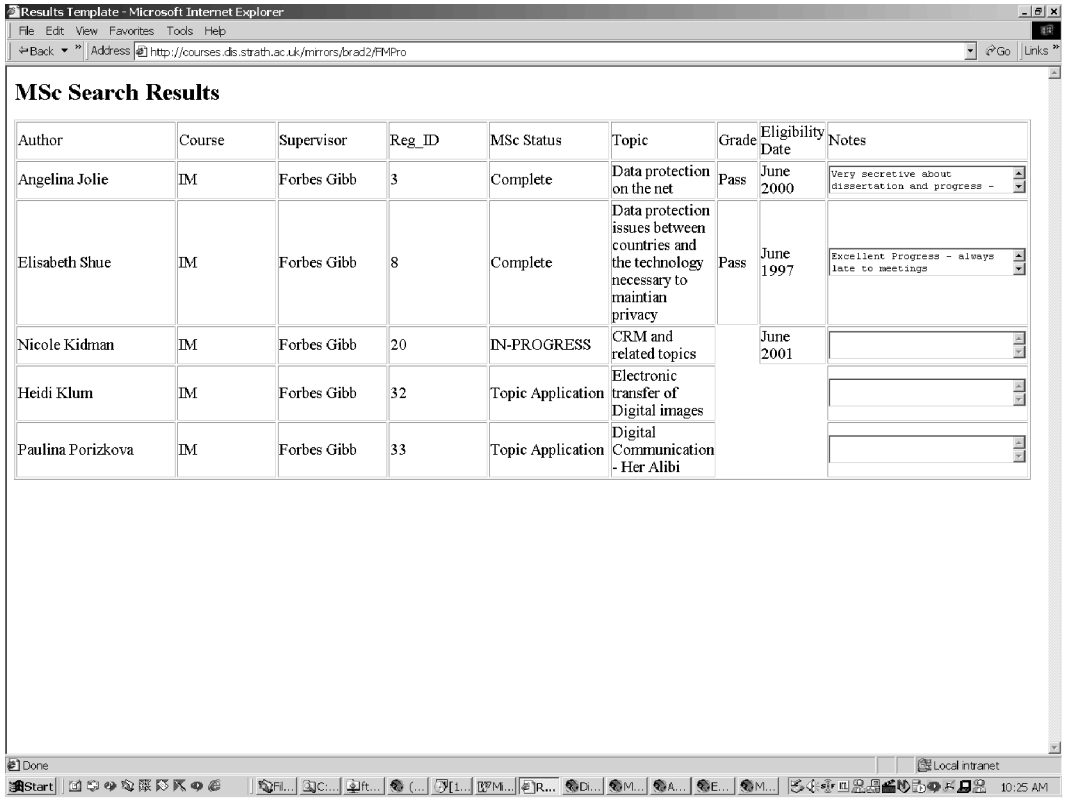
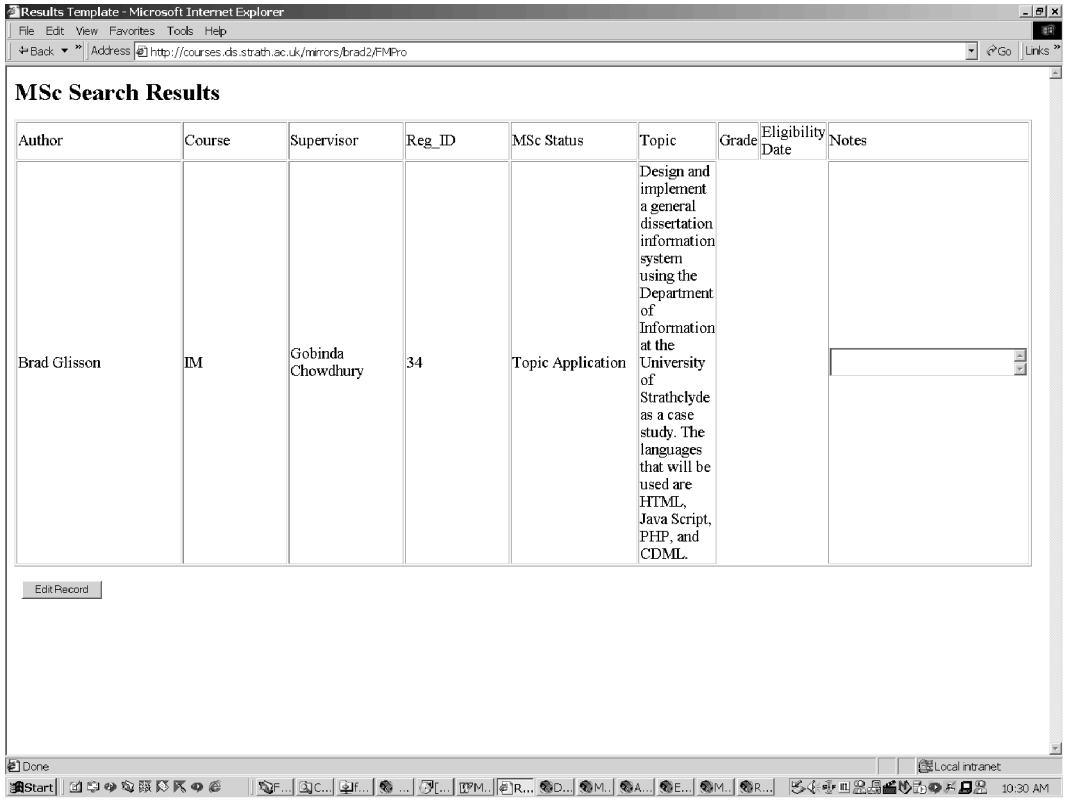


Figure 13 MSc search results for individual student



the primary key for the DisSearch file and foreign keys for the other files. This information will need to be propagated to the files Dissertations and DisData.

Another area that prompted necessary attention from a database perspective is the issue of errors and the need to trap and decipher errors as they occur. The FileMaker Pro database returns an error number when an error occurs. This indicates that code is needed to trap for this data and then convert this data into meaningful information. To achieve this goal, a database was created that holds all of the error codes and the equivalent text translation currently used in FileMaker Pro 5.5.

5.2 Proxy implementation

It should be noted that, for the purpose of this prototype, the Web server is acting as a modified proxy server. A proxy server generally acts as an intermediary between the requesting computer and an internal computer on the destination network. The basic process for an incoming message going through a proxy server involves the receipt of an incoming message, permission validation to enter the network, destination address alteration to that of the internal computer and forwarding the message (Fitzgerald and Denis, 1999, p. 408). This process is reversed for messages leaving a network so that the address of the original computer is substituted with that of the proxy server. This intermediary can have the ability to act as a type of gateway, handling security issues and cache frequently requested pages.

The Web server for the Department is an Apache Unix Web server that is also configured as the proxy server. This server has been configured so that certain URLs have a "ProxyPass directive" (Apache.org, 2001). This directive allows the server to mirror to the remote server. When a URL is called, that has a "Proxypass directive", the calling IP (Internet Protocol) address is altered to the proxy server's IP. This means that the remote server only recognises one calling IP address, although, in reality, several individuals may be viewing the information. There is also a "ProxyPassReverse directive" that changes virtual URLs passed from a local server (http://httpd.apache.org/docs/mod/mod_proxy.html). In the case of the FileMaker Pro implementation, the URLs that interact with the FileMaker Pro database have been designated as having ProxyPass and ProxyPassReverse directives.

5.3 Client-side scripting issues

Splitting the files and establishing the appropriate relationships had an adverse effect on the JavaScript used in the client validation for the application. The creation of the relationships altered the names of the fields to include the symbol "::". This symbol is used by CDML to connect the name of the relationship with the appropriate field, e.g. SearchDissertations::reg_id. The use of the double colon causes JavaScript to send an error message indicating that it is looking for a semicolon. The solution to this dilemma is to pass all of the elements to a function that creates an array of temporary objects and then to use the substring method to parse the desired portion of the name. Once a temporary object has been defined, via the desired portion of the name, then data validation can be performed on the object.

6. System testing

This was one of the weakest areas of this work. Due to lack of time and resources, the testing for the system was restricted to alpha testing and limited beta testing. The alpha testing consisted of the following elements:

- testing of all links;
- entering data via the Web interface to the database to test displays without client-side validation enabled;
- data were entered via the Web interface to test program operation without JavaScript enabled;
- testing fields for JavaScript where relevant.

The alpha testing revealed the following changes in the system development:

- the need to implement FileMaker Pro's Web database security option;
- the need to implement the additional files to preserve system functionality;
- the need to trap for errors generated by the database;
- some computers have trouble accessing the initial FTP page while others do not. The computers that do have a problem with initial access to the FTP page can select "file", then select "new" on their Internet browser and the page operates correctly.

The beta testing was conducted in the following manner. Four users, MSc students at the University of Strathclyde, were allowed access to the system but received no instructions on how to operate it and were encouraged to see if they could cause the system to fail. A questionnaire was developed for feedback, via end-user interviews, on the system design and functionality.

The initial results from the end-user beta testing and evaluation indicated that the overall system meets the needs of the students with two students rating the system as good and two students rating the system as excellent. Concerns and suggestions were raised regarding search terminology displayed and the displaying of the database errors. End-user suggestions for the use of new terminology in all search pages included the use of "All", "Any", and "Boolean". It was suggested that the display of database errors be combined with the initial page indicating that the error existed. All of the participants in the student user group feedback indicated that they preferred the dtSearch software. Their reasons ranged from the presentation of the search results and the documents on one page, to hit count indicators, to highlighting the found word, to the variety of searches available via dtSearch. A future suggestion raised by one student is for the system to indicate the total number of dissertations available for searching by each class.

7. Summary

The final element of the project, which is the administrative side, is addressed through dynamic reports and the ability to add notes on individual students. The question that develops from this project is whether the supervisor and the student will correspond electronically via the system. This brings up the need to address the cultural aspects of a digital environment and the cultural changes necessary to support that environment. The cultural aspects are beyond the scope of this research.

Although the initial evaluation produced encouraging results, further evaluation may help to improve the design and performance of the system. This system can be scaled to multiple intranets by simply building a common interface with links to the individual dissertation information system on the various intranets. The division by intranets is due to the fact that one department, theoretically,

could have multiple intranets. That is, a common interface could be constructed with links to departments ranging from information science, to computer science, to engineering, to philosophy. The current design of the system lends itself to a decentralised or federated search vs a centralised search. A centralised search occurs when all of the data are stored in a single location and searched from that location; and a decentralised search is when data are stored in separate locations and a search has to be initiated from that location (Miller, 2000). A federated search occurs when a search is initiated from one location, searches are conducted in other locations and the results are compiled, integrated and presented. A multiple intranet roll out of the application would prompt a decision on the design and the type of search, i.e. a federated search or a decentralised (site-by-site) search.

References

Apache.org (2001), "Apache module mod_proxy", *Apache HTTP Server Version 1.3 28*, available at: httpd.apache.org/docs/mod/mod_proxy.html (accessed 14 May 2002).

Fitzgerald, J. and Denis, A. (1999), "Network security", *Business Data Communications and Networking*, 6th ed., John Wiley & Sons, New York, NY.

Fox, E.A., Eaton, J.L., McMillan, G., Kipp, N.A., Mather, P., McGonigle, T., Schweiker, W. and DeVane, B. (1997), "Networked digital library of theses and dissertations: an international effort unlocking university resources", *D-Lib Magazine*, available at: www.dlib.org/dlib/september97/theses/09fox.html (accessed 14 May 2002).

Miller, T. (2000), "Why data storage is like inventory management", *NetworkWorldFusion Newsletters*, available at: www.nwfusion.com/newsletters/stor/2000/0619stor1.html (accessed 14 May 2002).

Pressman, R. (1992), *Software Engineering, A Practitioner's Approach*, McGraw-Hill, New York, NY.

Suleman, H., Atkins, A., Gonçalves, M.A., France, R.K., Fox, E.A., Cachra, V., Crowder, M. and Young, J. (2001a), "Networked digital library of theses and dissertations: bridging the gaps for global access – part 1: mission and progress", *D-Lib Magazine*, available at: www.dlib.org/dlib/september01/suleman/09suleman-pt1.html

Suleman, H., Atkins, A., Gonçalves M.A., France, R.K., Fox, E.A., Cachra, V., Crowder, M. and Young, J. (2001b), "Networked digital library of theses and dissertations: bridging the gaps for global access – part 2: services and research", *D-Lib Magazine*, available at: www.dlib.org/dlib/september01/suleman/09suleman-pt2.html (accessed 14 May 2002)