

PROJECT – CIS616 S06

Assigned: March 27, 2006

Preliminaries

This project is to be carried out in teams of 1-3 persons. You need to select a web application domain that requires a back-end image database and build the application from start to finish. There is a main core that includes designing a database that has ‘query by image content’ capabilities. However, you are free to select a specific topic for your project. For example, the database can be a medical image database or a database with images of products (e.g., an eBay like e-commerce website). You are required to design the underlying database and define the application functionalities you will provide with the database. You also need to implement the application: web interface, underlying database, and code for database operation. The implemented system should be publicly accessible to everyone. The project involves three parts that are graded: **a proposal (15%)**, **a final report (80%)** and **a short presentation (5%)**.

Example data items and operations

An example of major data items and functions follows: The medical system should include information about patients, exams, diagnosis, medical images, etc. Specifically, for each patient or doctor we may store their name, address, electronic address, and a unique login and password if they need to access the online database. For each medical exam or clinical test they performed we need to store information such as the date and time, the type of exam, the doctor that ordered the exam, the diagnosis, etc. The system should allow patients to view their record and search other records without providing identifying information. The doctors should be able to view the full records of their patients and records of other patients but should not provide identifying information for patients that are not theirs. We should also be able to store some optional fields such as a JPEG file with a medical image or a URL with more information about the disease. For each image, we derive a number of numerical ‘features’, which we shall use for similarity searching in images. I will provide more details in class about the feature vectors that can be derived using existing feature extraction techniques. The images should also be annotated to allow more standard types of queries (not just content-based). We also want to keep track of the history of each patient.

The following type of functions should be handled by the target system: Registration: each new patient or doctor has to provide his/her name, and address; our system should return a (unique) login name and a password. For an e-commerce application the same information should be stored for each customer. For simplicity, assume that users are not allowed to change their logins or passwords; also assume that we do not check for duplicate or near-duplicate names and addresses. Image browsing: patients are allowed to look at their information including their medical images. In the case of e-commerce application, customers can ask for products to be sorted on alphabetically by product name, etc. Images may be searched by text once patients/doctors or customers have specified keywords (the system has to return all items that contain all these keywords in their description). Searching by image: doctors or patients may ask queries on the images, then the system should return the N closest matches, according to the Euclidean distance of their feature vectors. For the e-commerce application customers may also place a bid for a product they like or sell that product.

Project Proposal

Once you have selected an application area for your project, you should do some background reading so that you are capable of describing, in some detail, exactly what you expect to accomplish. You should definitely take a look at the Scientific Image DataBase (SIDB) project [3]. Once you have read up on your topic, you will be ready to write your proposal.

In the proposal you should describe what you plan to do and how you plan to do it. You should describe the problem that you will be considering, how you plan to address it, what tools you will use, what you expect to produce as a result of this work, and anything else that you think the instructor should know to evaluate your plans. The proposal should be approximately 3-4 pages long, self-contained, describing clearly the alternatives that you will be considering and the milestones. Please provide a plan of activities and time estimates, per group member. The proposals will be evaluated by the instructor. After approval of the choice of topic you should start working on your project.

Final Report

Your final report is expected to be a detailed report, treating in depth the agreed topic. It should be in the form of a paper. You should describe what you did, what results you obtained, and what you have learned and/or can conclude from your work. It should include an introduction, a survey, your contributions (major design decisions, assumptions, rationale, etc), problems for future research, a list of references, a brief description of the developed modules, a user's manual and a programmer's manual. You should explain the database design of your program and state reasons why you took this approach. Careful packaging for distribution and a *demonstration* is expected.

Milestones

(a) Database design (E-R diagram and descriptions about tables); (b) Description of functionalities; (c) Create & populate tables; (d) Web interface for uploading, displaying and searching image data; (e) Functionalities as in real life applications, such as online shopping;

Software

Although you can use an RDBMS of your choice it is suggested that you implement your project using Postgresql [1] and Linux (lucas system on which everyone should have access). It is also suggested to use PHP [2], an HTML embedded server-side scripting language, to implement the application. A nice example of such an application is the Scientific Image DataBase (SIDB) project [3]. You will need to take a close look at the SIDB project before starting to implement your own project.

Grading

The project involves three parts that are graded: **a proposal (15%)**, **a final report (80%)** and **a short presentation (5%)**. You should submit electronically all these parts to vasilis@temple.edu.

The final report will be graded according to the following guidelines:

- a description of the purpose of the project, justification for the solutions, a report on the database design [20 pts.]
- code and documentation produced: a source program listing [20 pts], a user's manual for the system [15] pts, your testing efforts to evaluate correctness and completeness [35 pts] (e.g., erroneous cases that your system can detect and handle reasonably), a description of the system's limitations and the possibilities for improvements [10 pts].

Due dates

Proposal: April 4, 2006

Final report: May 5, 2006

Short presentation: May 9, 2006

References

[1] Postgresql official website: <http://www.postgresql.org/>

[2] Php official website: <http://www.php.net/>

[3] Scientific Image DataBase project: <http://sidb.sourceforge.net/>