

# DESIGNING AN ON-LINE HELP AND FAQ SYSTEM FOR FINAL YEAR PROJECTS

David Inman  
London South Bank University  
Faculty of Business, Computing and Information Management  
Borough Road  
London SE1 0AA  
dave@lsbu.ac.uk

## ABSTRACT

*Providing help on a potentially very broad range of questions is problematic for any domain. On-line help attempts to address this problem, but navigation and search often do not yield the desired answers. Students tend to use email in such situations, and it can be very time consuming to provide answers, even though many questions are similar. A system is presented here which attempts to build an on-line FAQ or help system on top of an existing web based help and information space. The FAQ library is easily indexed and graded so that effective re-use can be made of individual replies to questions. By using the existing web space and database as the portal for question asking it is hoped to reduce the number of unnecessary questions asked.*

## Keywords

*Help system, On-line Help, Projects*

## 1. INTRODUCTION

The growing trend to provide on-line help is based upon the imagined efficiency of such an approach. Yet providing help on a potentially very broad range of questions can be problematic since navigation and search often do not yield the desired answers. Phone support or e-mail is then used in such situations, which can negate the imagined productivity gains in an on-line help system. In addition, these 'special' questions once answered are not re-used, which is unproductive in a domain

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

© 2006 Higher Education Academy

Subject Centre for Information and Computer Sciences

where questions may be very similar. A system is presented here which attempts to build an online FAQ or help system on top of an existing web based help and information space. By using soft-match search, answers to existing questions can be provided relatively easily. The administrator is thus freed to answer the genuinely unique questions, and to maintain the database of questions and answers.

This paper reports on about 200 undergraduate projects annually within 3 departments at London South Bank University, and a similar number of postgraduate projects. The web space which supports these students with information and help comprises about 70 static web pages, and a similar number of dynamic pages generated from a PHP/MySQL database. These web pages are accessed on average about 250 times a day:

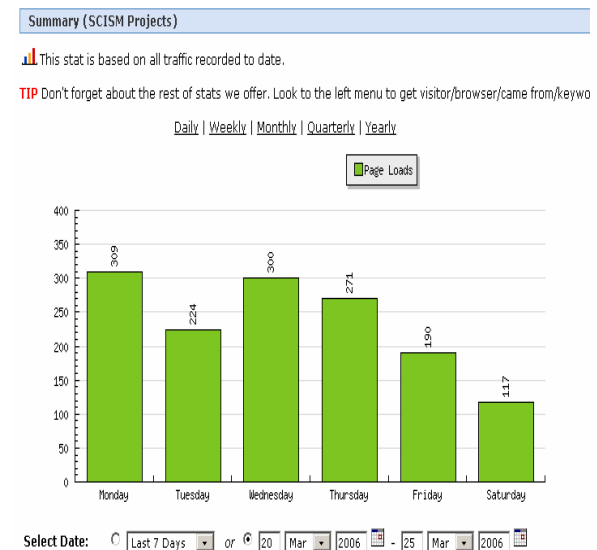


Figure 1: Hits on project web space in March 2006

In the last year 1211 email queries based on the undergraduate projects alone were sent by students or around 23 a week. This obviously takes a few

hours each week to answer adequately and highlights the need for some FAQ system to reduce these questions. Many of these questions are on similar topics, and often the web space could have provided answers.

The motivation for this work was to reduce the administrative workload in answering individual questions whilst still providing timely and effective help.

## 2. STUDENT INPUT

Students ask questions by one of two methods; via a web form or from a link via the web help space directly. The second is preferred, as the student is likely to find help directly in the web space in the first instance. As it is slightly quicker to use the second method, it is hoped that fewer questions will be asked.

Both methods take the user to the same form shown below, but if the student arrives at this web form from the web help files, the Main Topic and Sub Topic are completed automatically depending on where they came from. Of course the student can change these settings manually if they want.

### Help

Use this page to ask for more information about your project.

You can always come to the workshop to ask for more detailed help. Or look in the [project web guide!](#)

Your student number as shown on your student card

#### Your question:

Please input a brief summary of your question

#### More information on your question:

Please give some more detail on your question here - faster response likely!

Please describe what your question is about (faster response likely!):

Main Topic	Sub Topic
Introduction	Academic Integrity
Overview	Comparison with level 2 Project
Flowchart	Start Project
Activities	Choose Project
Deadlines	Choose Supervisor
Find a Project	Supervisor Interests
Supervisors	Summary of roles/tasks of Supervisor
2nd Assessors	Register Project
Register/Edit	Registration Examples
Project List	Allocation of Supervisor

Dave Inman Monday 03rd of April 2006 03:22:56 PM

Figure 2: Web form for asking questions

## 3. QUESTION ANSWERING

When a question is submitted, a copy and link to an answer form are emailed to the project coordinator, so that notification is carried out as soon as the question has been asked. An example is shown here:

From: **projects@lsbu.ac.uk**  
 Date: **29-Mar-2006 07:47**  
 Subject: **Project question from XXX**

Question: Final activities for last week of semester 1

More Info: As you already know I will be in London for only one week..

Main topic: Activities

Sub topic: Deadlines

Av. Quality of this student's Qus = 4.2000

Sup Notes (Good): 2006-01-13(Good):  
 The student seems to be on course with her project

Sec Notes (OK) :2005-12-14(OK): XXX  
 has made a good start to a promising project.

Use the link below to answer this question directly: [link](#)

There are 1 questions unanswered in total now: [link](#)

Figure 3: Email notification of a new question

From this email the coordinator can see the question, more information about the question and the topic area of the question. To help prioritise answering these questions, additional information is provided on the quality of the student's previous questions. This is a manual assessment by the project co-coordinator of how universal and important a question might be to other students. Progress reports by staff are also available so that conscientious students might receive priority over those who perhaps have not used the web space very effectively. In this way, high quality, well-indexed questions receive a prioritized response, helping to build a reliable FAQ database as quickly as possible.

The email contains links go directly to a web form to answer the specific question, as well as a link to all questions currently unanswered, if the coordinator prefers that mode of operation.

The interface to this system is show below:



Figure 4: Web form for viewing / answering questions

## 4. FAQ DATABASE

Questions are entered into a simple MySQL database using PHP scripts. This database is then used to display potential answers to any new question asked. The difficult problem here is to match two syntactically different questions that have the same semantics, and is a central problem in natural language processing research. The rest of this paper looks at potential ways this can be performed.

### 4.1 Suggesting Answers

If the student arrives at the question input form from the web help page links, we know the topic and sub-topic of their question automatically. This allows us to suggest answered questions to them, in the same sub-groups, and by presenting these before the question input form we encourage them to look at the answers to these questions before inputting their own question. The topic and sub-topic are automatically set as shown below:

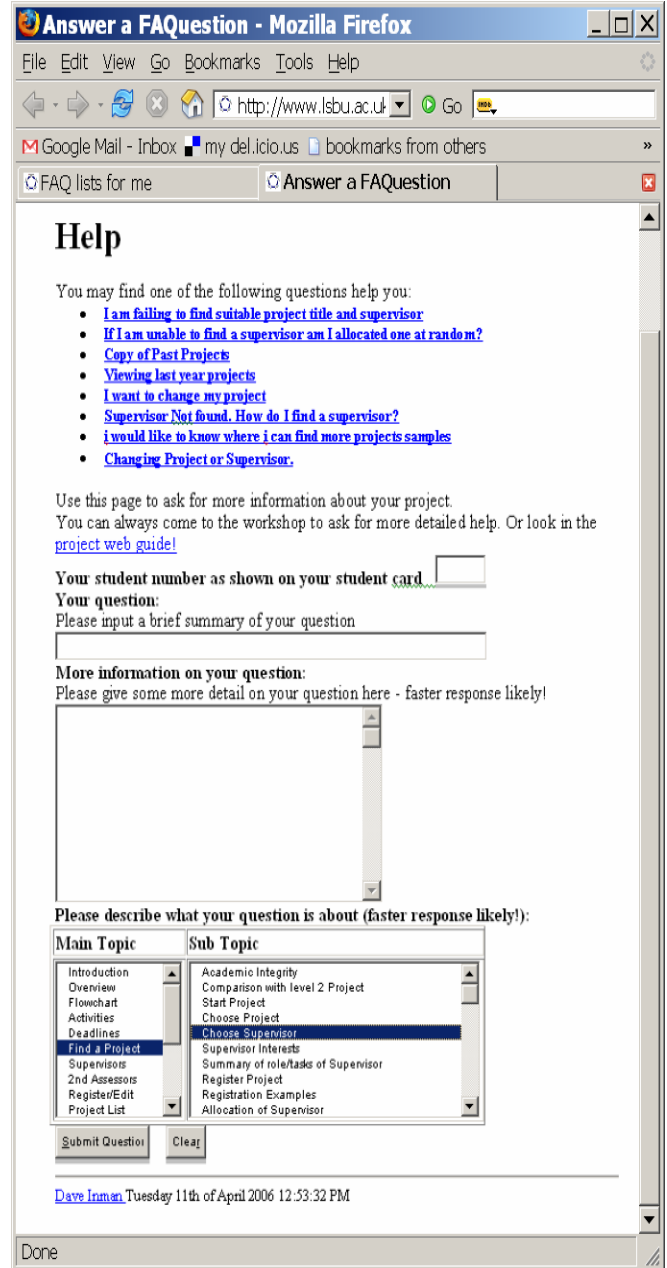


Figure 5: FAQ suggestions before Question Input

It is of course much simpler to click on a FAQ than to input a question. If the student is encouraged to click on similar questions, they may find an answer without any further action.

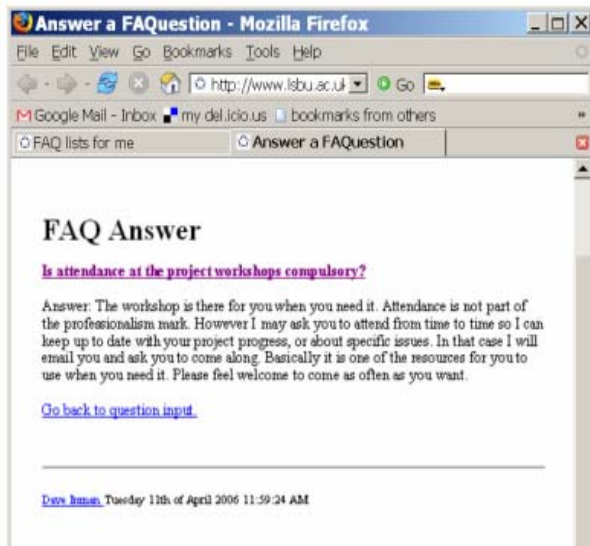


Figure 6: Answer to a FAQ

The key to the success of this is of course the granularity of the main topics and sub-topics. If we have too many topics duplication and problems of navigation become problematic. Too few topics do not allow a student to find similar questions easily enough. The main constraint here is the size of the chunks of content presented on the web help pages. This constraint really limits the benefit that can be gained from simple topic matching to find suitable FAQs to suggest to the student. The next section looks at information retrieval and natural language processing approaches we are investigating to attempt to achieve better performance.

## 5. BETTER ANSWER MATCHING

We are investigating 4 approaches to improve the match of a new query against the FAQ database.

### 5.1 Addition of Key Words

If the project co-ordinator adds key words to each question, at the time it is answered, these key words give an extra dimension to allow matching. The problem is the choice of a controlled vocabulary when selecting key words. Too many in the set generates too few matches, too few provides no benefit. A further problem is the effort required to add key words to each question. Two approaches are being investigated to assist here.

### 5.2 Parsing to find Head Noun Phrase

Natural language processing has produced effective parsers for most major languages [1] and for

English we can find the head-noun phrase for a sentence in a reasonable time. The hypothesis being tested here is that a head-noun phrase somehow captures the essence of the sentence, and by stripping out all but the noun from this phrase we may generate a keyword for each sentence. These can then be expanded by using a thesaurus for example, such as that used for testing the ambiguity and semantic category of words used in web search [2].

### 5.3 Vocabulary generation for topics

Previous projects within the Natural Language Processing group here at London South Bank University looked at generation of vocabulary sets for nodes within a manually classified web hierarchy, such as Yahoo! It works by taking manually classified web pages in an existing hierarchy, and then automatically generates a vocabulary to describe the important terms that should appear in a web page within each classification group. This vocabulary, as a weighted vector, can be used in comparison with an unclassified document to find where in the hierarchy the unclassified document belongs. This allows many web pages, currently unclassified, to be given a place in the hierarchy [3].

There are clear comparisons with the task here. By looking for a vocabulary set for each FAQ group (e.g. by main topic or sub-topic) we hope to automatically classify new questions by matching against these weighted vectors. It also introduces the possibility of using virtual topic categories, not linked to the web help pages.

### 5.4 Latent Semantic Indexing

In order to achieve soft-matches, we are attempting using a thesaurus [2] but better results are indicated by using an approach called Latent Semantic Indexing [4]. The approach works by looking for co-occurrences between key words in documents. These links are found by a matrix decomposition and simplification process originally developed for use in physics, called Singular Value Decomposition (SVD). Google has started recently to use this approach to generate hits on web pages that do not contain the key words entered by the user [5]. This enables Google to rank the Nokia web site top when entering the search term "~phone", but 46<sup>th</sup> if the LSI ~ is omitted. Nokia uses a vocabulary rich in words like 'mobile' but not so rich in the use of the word 'phone'.

In our application we could use this approach to make soft matches between new query and the existing database. Soft matches range from allowing minor errors in spelling to be

accommodated, up to a full semantic similarity assessment even if the words used are quite different. Often the words will not match, and whilst thesaurus expansion can generate more matches, we hope that LSI can work even better in the domain of a selected and limited vocabulary, such as within our help pages.

## 6. EVALUATION AND FURTHER WORK

There will be two approaches to evaluation; user surveys (both staff and students) and more quantitatively, by monitoring traffic on project queries, both by email and the FAQ system described here. Initial trials of the system this year have been incomplete, but indications are that students need to be made very clear that the FAQ system will give them a better turnaround time than using email. Current attempts to persuade them to abandon email queries in favour of the FAQ system have only been of very limited success.

Apart from further investigation of the approaches described here, we are also looking to expand the FAQ database using existing email queries. This could of course be carried out manually, but would be time consuming. We are hoping to use some of

the automated approaches described in section 5 to assist here.

In addition we are considering more extensive NLP techniques to attempt semantic classification of queries and their answers.

## 7. REFERENCES

- [1] Allen, James, *Natural Language Understanding*, Benjamin Cummins, 1988
- [2] Inman D., *Ambiguity in Web searches*, in Proc. Internet and Multimedia Systems and Applications pp. 52-57 Pub: IASTED, 2000,
- [3] Jenkins C. & Inman D., *Adaptive automatic classification on the Web*, in Database and Expert Systems Applications pp 504 - 512 Pub: IEEE Computer Society ,2000,
- [4] Berry M.W., Dumais S.T., and Letsche T.A. *Computational Methods for Intelligent Information Access* , Proceedings of Supercomputing'95, San Diego, CA, December 1995.
- [5] Wilson,Dr. Ralph F., *How Does Google Use Latent Semantic Indexing (LSI)?* accessed at <http://www.wilsonweb.com/art/promote/google-lsi.htm>,2005