

Evaluating the effectiveness of electronic Special Interest Groups

by

Carmel Ryan B.B.S., D.C.S

Supervisor: Micheál Ó hÉigearthaigh

A dissertation submitted for the degree of Masters of
Science

School of Computer Applications, Dublin City
University

OCTOBER 1997

Declaration

I hereby certify that this material, which I now submit for the assessment on the program of study leading to the award of Masters of Science in Computer Applications, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my own work.

Signed: Carmel Ryan
CARMEL RYAN

Date: 30/9/22

Acknowledgements

I would like to thank my supervisor, Micheál Ó hÉigearthaigh, for his assistance with this thesis and indeed for all his help and words of encouragement during my (long) student life in DCU.

A special word of thanks goes to Dr. Fergus Kelledy, the IR guru of the CA lab. Thanks for all your advice and assistance for the duration of this thesis, it is greatly appreciated.

To my parents, many thanks for the support and patience while waiting all those years for me to finish college. All I can say is - it's payback time!

To Gavin Gollogley, the CA Web Wonder - thanks for all your help with the imagemaps and all those soccer conversations!

To all the postgrads who were present during my time in the lab, thanks for putting up with the slagging, particularly my two foreign friends. Allen, my favourite African - thanks for trying to teach me Swahili and Oscar, the happy Catalanian, thanks for being and having a great laugh! To both of you, thanks for letting me correct your English, for laughing at my jokes and for cooking some absolutely delicious meals!

To Fergus thanks for chauffeuring me around everywhere and also for dealing admirably with all the slagging, laughing at my jokes and eating my deliciously cooked meals. Finally, thank you very much for your help with this thesis, your encouragement and endless patience. Long may you retain all those super qualities.

**Evaluating the effectiveness of
electronic Special Interest Groups**

by

Carmel Ryan

Abstract

The Centre for Teaching Computing aims to facilitate co-operation between third level colleges in Ireland and Europe. Such co-operation will result in the development, use and publication of best practices and material for teaching computing. A number of electronic discussion groups were set up by the CTC to assist in providing a forum for exchanging ideas and materials and promoting co-operation between those involved in the field of computing at third level. The eSIGs need to have members communicating on a regular basis to be successful. This dissertation explains the technology of the eSIGs and details the support measures that have been put in place to ensure that they deliver on their potential.

TABLE OF CONTENTS

LIST OF FIGURES.....	7
INTRODUCTION.....	8
CHAPTER ONE - TECHNOLOGY OF ESIGS	11
1. Background Information	11
1.1 Newsgroups	13
1.2 How the eSIGs operate.....	13
1.2.1 MAILBASE Server	14
1.3 Objectives of the CTC eSIGs	14
1.4 Management of the eSIGs	14
CHAPTER TWO - INFORMATION FILTERING.....	16
2. Introduction	16
2.1 What is Information Filtering?.....	17
2.2 Information Retrieval	19
2.2.1 Automatic Indexing.....	20
2.2.1.1 Word Frequency.....	21
2.2.1.2 Suffix removal	21
2.2.1.3 Term Weighting.....	22
2.2.1.4 Similarity Measures.....	24
2.2.1.5 Phrase formation and Thesaurus Classes.....	26
2.2.2 File Structures.....	26
2.2.2.1 Indexed files.....	26
2.2.2.2 Signature Files	27
2.2.2.3 Inverted Files	27

2.2.2.4 Clustered Files	28
2.2.3 Information Retrieval Models.....	30
2.2.3.1 Boolean Retrieval.....	31
2.2.3.2 Vector Space Model.....	32
2.2.3.3 Probabilistic Retrieval.....	32
2.2.3.4 Latent Semantic Indexing.....	33
2.3 Information Filtering and Information Retrieval.....	34
2.4 Current Information Filtering Systems.....	36
2.4.1 Personal Filters	37
2.4.1.1 Elm - Email Filter	37
2.4.1.2 Borges	37
2.4.1.3 SIFT	38
2.4.1.4 Connectionist Filtering System.....	40
2.4.2 Group Filters.....	41
2.4.2.1 Information Lens System.....	41
2.4.2.2 Tapestry.....	43
2.4.2.3 GroupLens Filtering System.....	44
2.4.2.4 Active Collaborative Filtering System.....	45
2.5 Summary.....	46
CHAPTER THREE - MONITORING THE eSIGs.....	47
3. Overview	47
3.1 Monitoring the eSIGs using IF techniques.....	47
3.1.1 Content-based versus Social Filtering.....	48
3.2 Preliminary Appraisal of eSIGs.....	48
3.2.1 Members and Traffic on the eSIGs.....	49
3.3 Data Collection.....	51

3.4 Questionnaire Design.....	51
3.5 Results of Questionnaire.....	53
CHAPTER FOUR - USER INTERFACE DESIGN	56
4. Overview	56
4.1 User Interface Design.....	56
4.2 Design Issues.....	57
4.2.1 Primary Design Goals.....	58
4.2.2 Secondary Design Goals.....	59
4.2.3 Interactive Styles.....	61
4.2.4 Data Display	61
4.3 Developing the CTC Tool.....	63
4.3.1 World Wide Web.....	63
4.3.2 Geographical Web Tool.....	65
4.3.2.1 Reasons for using the World Wide Web.....	66
CHAPTER FIVE - CTC WORLD WIDE WEB TOOL.....	68
5. HTML.....	68
5.1 Information Sources	69
5.2 Designing the Web site.....	69
5.3 Development of the Tool.....	71
5.3.1 Imagemaps.....	72
5.3.1.1 The Picture File.....	73
5.3.1.2 The Map File.....	73
5.3.1.3 Writing the HTML Code	75
5.3.1.4 Making the Imagemap Work.....	76
5.3.2 Usage of Frames	76

5.3.3 Search Facility	79
5.3.4 Using the WWW Tool.....	80
 CHAPTER SIX - CONCLUSIONS.....	83
6. Introduction	83
6.1 Conclusions.....	83
6.2 Future Work.....	84
6.3 Summary.....	85
 BIBLIOGRAPHY	86
 APPENDIX A.....	90
 APPENDIX B.....	95
 APPENDIX C.....	109

List of Figures

FIGURE 2-1 - SIMPLE FILTER ARCHITECTURE	16
FIGURE 2-2 - OVERVIEW OF INFORMATION RETRIEVAL.....	20
FIGURE 2-3 -MATRIX OF DOCUMENT VECTORS	24
FIGURE 2-4 - GENERAL STRUCTURE OF AN INVERTED FILE	28
FIGURE 2-5 - GENERAL MODEL OF INFORMATION RETRIEVAL	30
FIGURE 2-6 - INFORMATION FILTERING MODEL	34
FIGURE 2-7 - INFORMATION FILTERING/RETRIEVAL COMPARISON	35
FIGURE 2-8 - OVERVIEW OF SIFT	39
FIGURE 3-1 - CURRENT STATUS OF eSIGs.....	50
FIGURE 5-1 - HOME PAGE OF GEOGRAPHICAL INTERFACE TOOL	78
FIGURE 5-2 - MAP OF EUROPE WEB PAGE.....	80
FIGURE 5-3 - MAP OF FRANCE PAGE.....	81
FIGURE 5-4 - UNIVERSITY OF COMPIEGNE (COMPUTING DEPT)	82

Introduction

The Centre for Teaching Computing (CTC) was established in 1992 by the University of Ulster and Dublin City University. It was set up in response to the rapid growth in demand for personnel with computing knowledge and skills [CTC-94]. As the computer industry is a rapidly changing one, there is a continuing need for new computing skills and hence frequent changes in existing courses are required, as is the introduction of new programmes. Supporting computing lecturers in institutions of higher and further education throughout Ireland in the shared development, evaluation and dissemination of teaching materials and methodologies should ensure that the problems of coping with these frequent changes will be minimised.

By facilitating this co-operation between third level colleges, the CTC helps to avoid the problems associated with :

- the duplication of effort in the design of courses and course material
- variation in standards
- an absence of effective means of recognising and sharing best thinking and practices
- failure to recognise and exploit the emerging international market for computing education materials at third level.

By acting as a forum for exchanging ideas and materials the CTC hopes to facilitate the development, use and publication of best practices and materials for teaching computing, resulting in :

- improved availability of high quality computer personnel to service the needs of the computer industry in Ireland
- improved cost efficiency due to less duplication of effort in course design and preparation
- high quality courses that keep fully abreast of developments in the rapidly changing area of computing
- the promotion of computing education as an industry in itself

In order to fulfill its mission, the following activities are being undertaken by the CTC:

- Development and maintaining of databases of syllabi, examination questions and the research and teaching interests of computing lecturers
- Organisation of an annual conference on Computing Curriculum Development & Delivery
- Hosting a series of workshops, colloquia and seminars
- Publishing of a monthly newsletter
- Promotion of postgraduate research in the teaching of computing
- Provision of assistance to computing lecturers in courseware development

Through these activities, the CTC has built up many contacts not only in Ireland but all over Europe. As a result of making these contacts, the CTC's Annual Conference has grown each year and delegates now come from all over Europe.

As the main thrust of each conference is the development and delivery of the curriculum, two of the most important activities are :

- the establishment of a database of European recommended syllabi
and
- the development of courseware written to the specifications of these syllabi.

During the 2nd Annual Conference, held in September 1994, it was decided that in order to achieve these goals, discussion between academics in the EU, on all aspects of the syllabi would need to be stimulated and fostered. As a result, a number of electronic discussion groups were established to encourage such discussion. These groups would enable computing academics across Europe to collaborate and discuss topics of mutual interest. Eight groups - known as electronic Special Interest Groups (eSIGs) were set up, each to represent a different area of computing.

The provision of a forum, in addition to the annual conference, for academics to discuss topics of interest will provide members with:

<ul style="list-style-type: none"> • pointers to new developments/resources on the Internet
<ul style="list-style-type: none"> • access to evaluations of public domain and commercial products
<ul style="list-style-type: none"> • access to lists of conferences, software reviews, reports and good practice as drawn up by the group
<ul style="list-style-type: none"> • a facility to make informal comments and exchange views on their specialist areas
<ul style="list-style-type: none"> • access to centres of excellence throughout the world.

This dissertation describes the technology of the eSIGs and evaluates the extent to which they have delivered on their potential.

The development of a World Wide Web based tool which was created in response to a demand for detailed information on university computing courses in Europe is also described. The tool uses a simple user interface based on maps and allows users quick access to computing faculty web sites in European universities.

Finally, a set of recommendations for further development of the eSIGs and the Web site is discussed.

Chapter One - Technology of eSIGs

The CTC Annual Conference of 1994 identified eight distinct areas of computing established an electronic Special Interest Group for each.. The areas identified were:-

- Artificial Intelligence
- Databases
- Formal Methods
- Multimedia and Authoring Tools
- Programming Languages
- Computer Architecture and Operating Systems
- Human Computer Interaction/Multimedia
- Software Engineering

1. Background Information

The concept of an electronic discussion group is similar to that of any other discussion group where a number of people with similar interests meet and discuss topics of interest. What sets an electronic discussion group apart from any other is that discussions take place via electronic means, often utilising the rapidly growing world wide network known as the Internet, to 'talk' via electronic messaging or email systems. The Internet consists of many types of computer connected together and communicating in a common way at high speed all around the world. It is often referred to as the Information Superhighway or Cyberspace, but whatever it is called, this international network has evolved significantly since its inception in the 1960's [GUL-95] .

The Internet was conceived as an experimental network to support military research for the U.S. Defence Department. The project was initiated by the Dept.'s Advanced Research Projects Agency (ARPA) and one of the primary goals of the ARPAnet was to build a network that could withstand partial failures of equipment and still function. In other words, ARPA wanted to build a *fault tolerant* network. Funds were granted to computer science departments in many US universities and to a

few private corporations to stimulate research into computer networks, and on using distributed networking to protect a network from destruction.

As a result of this research, a technology known as packet switching evolved that has the capability to allow communications to be re-routed should communication lines be damaged. This common communications 'language' was called TCP/IP¹. With the introduction of this technology and the development of additional communications software, more and more networks were linked together in the early '80's. Regardless of what type of computer equipment was being used by government, educational and commercial institutions all machines could communicate with each other using TCP/IP.

The increased ease with which it was possible to transfer information has resulted in nearly 150,000 new users per month accessing the Internet, and to date has resulted in almost 7 million computers have been connected to the Internet. This allows between 40 and 50 million users access to this Information Superhighway.

The Internet is not merely a world-wide network of computers. The Internet is the data available on it and the ease with which this data can be accessed. Electronic mail (email), FTP (file transfer protocol), Newsgroups (public discussion groups), World Wide Web and Telnet are more readily available to make Internet easy and worthwhile to use.

Definitions

Email and File Transfer: Internet users can send mail messages, data files and software programs to other Internet users.

Telnet: Users on one system can access other host systems across the network with just a handful of commands. A user with an account at one Internet site can log in to check for mail messages or files or other electronically save data from anywhere else on the network.

News: Internet is the home of Usenet, a giant bulletin board system that contains thousands of ongoing teleconferences on almost every topic imaginable. Millions of readers check into one or more Usenet groups regularly.

¹Transmission Control Protocol/Internet Protocol

World Wide Web: A graphical interface for browsing documents that have been made publicly available on the Internet.

1.1 Newsgroups

A newsgroup is a communication medium that allows people with Internet access to discuss issues of mutual interest. Readers interested in a particular topic can read and respond to messages posted to a particular newsgroup by other readers. Readers have access to experts on a variety of topics. People can have discussions about everything and anything with other people on a world-wide basis. These discussions can be used for work or simply for recreation and enjoyment. Newsgroups are ideal for browsing and do not require a lot of commitment. If you're only marginally interested in some topic, you can 'drop in' and read up on the latest discussions once a month, or once every 6 months, or only once a year. It is not necessary to subscribe to a newsgroup and read all the postings to it everyday. If you are extremely interested, then you can check up the latest happening on a frequent basis.

There are two types of newsgroups - *moderated* and *unmoderated*.

'moderated' - Individuals are not allowed to post directly to the newsgroup. Postings go to a 'moderator' who decides whether to pass it on or not.

'unmoderated' - A reader is allowed to post directly to the newsgroup

The CTC eSIGs are very similar to these newsgroups. Details on how they operate follow.

1.2 How the eSIGs operate

The eSIGs are slightly different to the newsgroups outlined in section 1.1, in that interested parties must subscribe to the eSIGs. One can access the eSIGs via email and in fact, an eSIG can be considered to be an email alias that has multiple

destinations. The eSIGs, which will be discussed in more detail below are maintained on a server in the UK known as MAILBASE.

1.2.1 MAILBASE Server

MAILBASE is the UK's major electronic mailing list service that enables groups to manage their own discussion topics and associated files. The CTC eSIGs are unmoderated, allowing postings (messages) to go directly to members of the lists. To join a particular eSIG, one simply emails the group and joins up. After this a member can send a message to that group, check on previous messages sent to the group or leave the group (see Appendix A for more information on MAILBASE)

1.3 Objectives of the CTC eSIGs

An eSIG is intended to consist of a community of like people who seek to promote excellence in research and good practice in their specialist area. This community would communicate electronically via the MAILBASE server as outlined previously in this chapter and as a result the members of the 8 eSIGs will

- maintain bulletin boards
- help organise seminars and workshops
- contribute to the annual conference
- help the CTC draft an educational policy
- draw up recommended syllabi.

These objectives are consistent with the overall aim of the CTC.

1.4 Management of the eSIGs

Management involves deciding what is to be done, deploying the resources required to do it and ensuring the task is completed [LUP-83]. Having decided to

use eSIGs as a forum for discussion between computing academics, the CTC has employed the resources supplied by MAILBASE to facilitate this electronic collaboration. The main management concern at present is that of task completion. This entails ensuring that the eSIGs are being used for their intended purpose and that the objectives set out by the CTC are being met.

Preliminary research into managing this Information Resource for the CTC shows that Information Filtering techniques would be the best way to analyse performance. Using these techniques, a system could be designed and developed to monitor the eSIG activity and generate reports that the CTC would subsequently use to carry out qualitative evaluations. Chapter Two describes Information Filtering and details current techniques and systems in use. Chapter Three outlines how these could be useful in managing the eSIGs and details are given on the success of eSIGs.

Chapter Two - Information Filtering

2. Introduction

The filtering of information is not a new concept. In fact, since the invention of the written word, people have been continually filtering information. People only buy certain magazines or papers that contain articles on a particular subject and skim through the articles to find the interesting ones. This extraction of relevant information is referred to as filtering. Figure 1 shows a simple model of information filtering.



Figure 2-1 - Simple Filter Architecture

Up to the early '80's the attention was on generating information, but since then more focus is being placed on receiving information - the process of controlling and filtering information that reaches the people who must use it [DEN-82] . For over 40 years, research in Information Retrieval has been focused on ensuring that users only receive information that is relevant to them. Information Filtering also concerns the presentation of data that will satisfy an information requirement and is in essence a specialised form of Information Retrieval.

Research in the area of Information Filtering is endeavouring to keep up with the explosion of information available in machine readable (electronic) form. A prime example of the growing rate of information traffic is the Internet, which up to 1988, had only a few thousand users. In 1994, Krol reported in his book, 'The Whole Internet' that since then the number of Internet users had increased a thousand fold [KRO-94] and consists of 31,000 networks, with over 2 million computers connected

to it. Over 20 million people could be reached by email² and have access to the resources on the Internet. These figures have increased even further since 1994. It has been estimated that now over 7 million computers are connected to the Internet, resulting in over 40 million people having access to email. Thousands of new users are being added daily to the Internet, the world's largest computer network that provides an indescribable wealth of information. Several tools have been developed to access some of this data scattered across the network. These tools try to be user-friendly and help in searching a variety of on-line resources and access servers such as Gopher, World Wide Web (WWW) and Wide Area Information Servers (WAIS). Gopher allows the user to navigate through the Internet by selecting resources from menus. WWW is another menu-based service for browsing hypertext³ documents with Internet links to other documents dispersed over the Internet. WAIS helps the user search indexed material on the Internet. Search engines have been developed which also help people find information on the WWW. These include Altavista developed by Digital, Lycos developed at Carnegie Mellon University and others such as InfoSeek, Yahoo, Excite, WebCrawler and Opentext. These tools search all web pages which have been made publicly available on the Internet and return a list of those which best match your query. Other techniques have been put in place to help people utilising the Internet - news groups, mailing lists and specialist information sites. Thousands of mailing lists exist, covering a wide variety of topics. The user subscribes to mailing lists of interest and receives messages on the topic via email.

With access to these information sources becoming easier, and the numbers of users growing rapidly, the need for automated information filtering is becoming critical.

2.1 What is Information Filtering?

Filtering is often understood to mean the removal of information from a stream of data. However, generically the term Information Filtering is used to refer to selecting relevant information (positive filtering) and eliminating that which is non-

² electronic means of sending information from one user to one or more users in the form of a message

³ text containing words that provide access to other documents on the Internet

relevant (negative filtering). One of the earliest forms of electronic information filtering came from research into how to minimise scientists' efforts in keeping up to date with new developments in their areas of specialisation. This research known as Selective Dissemination of Information (SDI) [ROW-78] involves the user expressing interests via a number of long-term, continuously evaluated queries, called profiles. The user will then passively receive documents filtered according to the profiles.

Ever since the invention of computers, research into how to store and access the electronic information available has been ongoing. Storage and accessing relevant information is now referred to as Information Retrieval. As most current information filtering systems use techniques and methods developed as a result of research undertaken in information retrieval, information filtering could be described as a specialised application of information retrieval.

Information Filtering has more recently been defined as 'a name used to describe a variety of processes involving the delivery of information to people who need it' [BEL-92]. Many information filtering systems:

- are designed for unstructured or semi-structured data
- deal primarily with textual information
- involve large amounts of data
- involve streams of incoming data
- are based on descriptions of individual/group information preferences

Many of these features are also found in other text-based information systems. Text routing systems involve sending relevant incoming data to individuals or groups. Categorisation systems are designed to attach one or more predefined categories to incoming objects [LEW-92]. The major difference between these text categorisation systems and information filtering systems is the static nature of the categories, as compared with profiles. Extraction systems involve extracting facts from the text of incoming objects [SUN-91]. Determining which of the incoming objects are relevant is secondary in text extraction. Information Retrieval systems share most of the features outlined above and much of the research in this area is directly relevant to Information Filtering.

2.2 Information Retrieval

Information Retrieval can be defined as the process of searching for specific information from a number of stored information items. It is concerned with the representation, storage, organisation and accessing of information items [SAL-83]. Information Retrieval (IR) research involves developing the most effective and efficient methods of accessing natural language information using computers. It is also concerned with finding representations and methods of comparison that will discriminate between relevant and non-relevant documents.

Since the 1940's the problem of information storage and retrieval has attracted increasing attention [VAN-79]. With the enormous advances in networking and the rapid expansion of the Internet, considerable attention is now being paid to Information Retrieval. The Internet, the world's largest computer network is like a sea of information and with so many information sources to wade through, it is difficult to sort out the useful from the uninteresting without spending inordinate amounts of time and money. The primary objective of an IR system is to retrieve as many relevant documents given a particular query while at the same time avoiding non-relevant documents. This reduces the time and effort invested by a user in looking for useful information. With the current fast pace of information generation, the huge growth of the Internet and the increasing popularity of the World Wide Web this goal is even more important than ever.

In order to achieve this goal, much of the research and development in information retrieval is geared towards improving the effectiveness and efficiency of IR systems. The efficiency of a system is usually measured in terms of the computer resources, such as CPU time and storage space, which are used in responding to a particular query. The effectiveness of a given IR system is usually measured in terms of precision and recall [VAN-79].

- *Precision* refers to the proportion of retrieved documents which are relevant
- *recall* refers to the proportion of relevant documents which are retrieved.

IR systems can be geared towards high *precision* or high *recall* depending on the application.

An IR system consists of a set of information items (DOCS), a set of requests (REQS) and a mechanism (SIMILAR) for determining which of the information items meets the requirements of the requests [SAL-83]. The information items are usually natural language documents and the requests are referred to as queries. These queries can also be natural language (usually using a restricted vocabulary) or part of a more formal query language (using an artificial syntax and vocabulary). The documents and queries are generally converted to a special form using an indexing language (LANG). (See Figure 2) This process of assigning suitable terms and identifiers which can represent the content of the documents is an important operation in the retrieval system and is generally referred to as indexing.

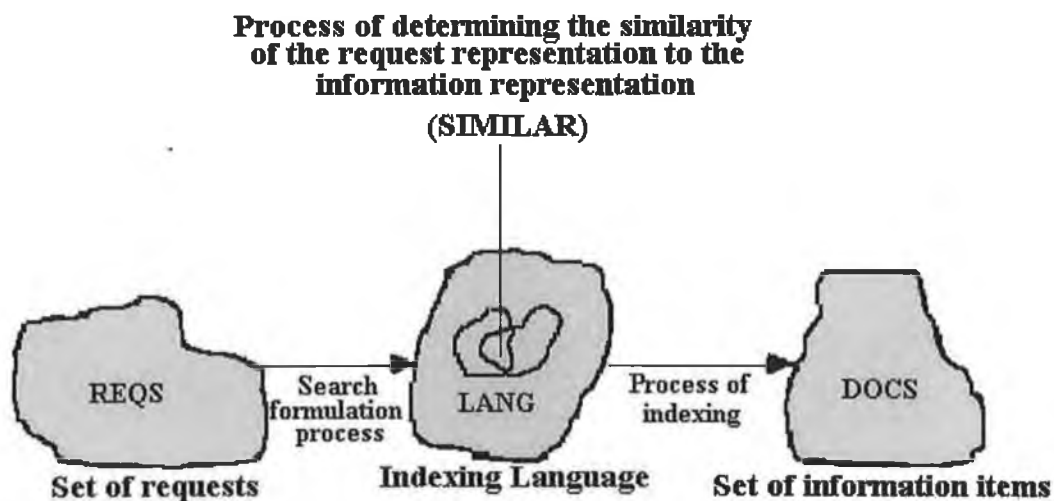


Figure 2-2 - Overview of Information Retrieval

2.2.1 Automatic Indexing

In all modern IR systems, indexing is carried out by computer and is divided into 2 operations - (1) assignment of the actual terms to represent the content of the information item and (2) assignment of a weight to each of these terms which reflects the importance of each term in the document.

2.2.1.1 Word Frequency

Assigning terms which will actually represent the item has been largely based on the work of Luhn, who introduced the idea of using the frequency of word occurrences as a basis for determining good index terms. According to Luhn “measuring word significance by use-frequency is based on the fact that a writer normally repeats certain words as he advances or varies his argument and as he elaborates on an aspect of a subject. This means of emphasis is taken as an indicator of significance...” [LUH-58].

Zipf uses this idea in his rank-frequency law, which states that the product of frequency of use of words and the rank order is approximately constant [ZIP-49].

$$\text{Frequency} \cdot \text{rank} \cong \text{constant}$$

The frequency of a given word multiplied by the rank order, that is, the order of its frequency of occurrence in the overall document, will be approximately equal to the frequency of another word multiplied by its rank.

The “principle of least effort” is used to explain Zipf’s law - it is easier for a writer to repeat certain words than to use new ones. This accounts for the fact that the words with the highest frequency in most documents tend to be short words such as “and”, “of”, “but”, “the”. The frequency is denoted as FREQ_{ik} where k is the term and i the document.

Once the frequency of all the words (terms) has been established, all the high and low frequency terms are removed. This is usually done by picking some high and low threshold values and any terms with frequencies above or below these values will be removed. A ‘stop-list’ made up of a few hundred high frequency words is usually used to eliminate those above the threshold. The remaining medium frequency terms have ‘resolving power’ - the ability to discriminate content and are used as index terms.

2.2.1.2 Suffix removal

The next step in representing the content of the information item is suffix removal. This process serves to broaden the scope of the index terms by reducing the

term to stem form. Thus, the words computational, computing, computer would be reduced to 'comput', or effective, effectiveness, effectual would all be reduced to 'effect'.

There are some issues in suffix removal which must be handled with care.

1. It is desirable to remove certain endings but not others for e.g. the ending 'ing'. It would not be correct to remove this from 'ring' or another example is the ending 'ual', it would be desirable to remove this from 'factual' but not 'equal'. To avoid this problem, a minimum stem length is specified which must remain after the suffix removal.
2. Some words are changed when an ending is added for e.g. relief to relieving, or absorb to absorption. Rules to account for such transformations are included in the suffix removal process.
3. For words with multiple suffixes, such as 'effectiveness', the process can be applied recursively several times. First 'ness' removed then 'ive'. Alternatively a suffix dictionary can be used which contains all the multiple suffix entries and the longer suffix is removed in preference to the shorter one.

In practice, most IR systems use a suffix list combined with a few contextual rules applicable to particular suffixes. A list of transformations or a suffix dictionary to recode some of the generated stems is also used.

Now that stems instead of full words are to be used as index terms, a greater number of potentially relevant items can be identified. Some stems may be more important than others. This leads to the next stage in the representation of the document which is the weighting of each stem to reflect its usefulness in the whole collection of documents.

2.2.1.3 Term Weighting

Good index terms are those which have high frequencies for some individual documents but with a relatively low total frequency - $TOTFREQ_k = \sum_i FREQ_{ik}$. This ensures that good levels of recall and precision will be achieved. Recall refers to the proportion of relevant documents retrieved and precision refers to the proportion of

retrieved documents which are relevant. Most of the research into the effectiveness and efficiency of IR systems concentrates on improving both recall and precision.

The *inverse document frequency (idf)* weight is often used and assumes that the term importance is proportional to the standard occurrence frequency of each term k in document I ($FREQ_{ik}$) and inversely proportional to the total number of documents to which each term is assigned ($DOCFREQ_k$)

$$\log_2 \frac{n}{DOCFREQ_k} + 1 = \log_2(n) - \log_2(DOCFREQ_k) + 1$$

Under this weighting scheme, high weights are assigned to those stems occurring in only a few documents of the collection.

Another method measures the degree to which the use of the term will aid in distinguishing the documents from each other. This measure is known as **Term Discrimination Value**. The density of the document space is calculated initially. This is the degree to which documents are 'bunched up' in the space of documents.

$$AVERAGE-SIMILARITY = CONSTANT \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n SIMILAR(D_i, D_j)$$

$SIMILAR(D_i \text{ and } D_j)$ - is a similarity function used to represent the similarity between documents D_i and D_j in a collection

The space density can be computed more efficiently by constructing an artificial 'average' document \bar{D} as the *centroid*, in which the terms exhibit average frequency characteristics. The AVGSIM is then the sum of the similarities of each document with the centroid

$$AVGSIM = CONSTANT \sum_{i=1}^n SIMILAR(\bar{D}, D_i)$$

The discrimination value $DISCVALUE_k$ can now be computed for each term k as

$$DISCVALUE_k = (AVGSIM)_k - AVGSIM$$

Terms can now be ranked according to their discrimination values. It is generally medium frequency terms that are good discriminators since they have a positive DISCVALUES. For each term, a weighting is now assigned by combining the term frequency factor with the DISCVALUE.

$$\text{WEIGHT}_{ik} = \text{FREQ}_{ik} \cdot \text{DISCVALUE}_k$$

2.2.1.4 Similarity Measures⁴

A similarity coefficient measures the relationship between two individuals, given the values of a set of p variates common to both. In general similarity coefficients take the values in the range 0 to 1 [EVE-80]. In Information Retrieval they are used to show the similarities between the index terms for each pair of documents in a collection. It does this by comparing pairs of rows of a document matrix [SAL-83]. The rows of this matrix represent the document vectors for the whole collection and the columns show the terms in each document.

	T_1	T_2	T_t
D_1	d_{11}	d_{12}	...	d_{1t}
D_2	d_{21}	d_{22}	...	d_{2t}
...
...
...
D_n	d_n	d_{nz}	...	d_{nt}

Figure 2-3 -Matrix of Document vectors

The similarity between two objects (documents/queries) is normally computed as a function of the properties that are assigned to both objects. The number of properties that are jointly absent may also be taken into consideration.

Dice Coefficients

$$\text{SIM}(\text{DOC}_i, \text{DOC}_j) = \frac{2 \left[\sum_{k=1}^t (\text{TERM}_{ik} \cdot \text{TERM}_{jk}) \right]}{\sum_{k=1}^t \text{TERM}_{ik} + \sum_{k=1}^t \text{TERM}_{jk}}$$

⁴ often called **similarity coefficients** or **association coefficients**

The numerator represents the sum of the products of corresponding term weights for two given vectors - DOC_i and DOC_j . The denominator consists of the sum of the weights of all properties for the two given vectors.

Jaccard Coefficients

$$SIM(DOC_i, DOC_j) = \frac{\sum_{k=1}^t (TERM_{ik} \cdot TERM_{jk})}{\sum_{k=1}^t TERM_{ik} + \sum_{k=1}^t TERM_{jk} - \sum_{k=1}^t (TERM_{ik} \cdot TERM_{jk})}$$

Cosine Coefficient

$$COSINE(DOC_i, DOC_j) = \frac{\sum_{k=1}^t (TERM_{ik} \cdot TERM_{jk})}{\sqrt{\sum_{k=1}^t (TERM_{ik})^2 \cdot \sum_{k=1}^t (TERM_{jk})^2}}$$

This is the most commonly used measure and takes into account the length of documents and assigns a weight to each term and calculates the cosine of the angular distance between vectors. As the numerator is divided by the product of the lengths of the document vectors, long documents with many terms are quite long and usually produce small cosine similarities.

Overlap Measure

$$SIM(DOC_i, DOC_j) = \frac{\sum_{k=1}^t (TERM_{ik} \cdot TERM_{jk})}{\min \left(\sum_{k=1}^t TERM_{ik}, \sum_{k=1}^t TERM_{jk} \right)}$$

The denominator in this case consists of the lower-weighted terms from the two documents. As a query normally contains low-weighted terms, the query-document correlations using this measure are usually larger in magnitude than those of the Cosine Coefficient.

2.2.1.5 Phrase formation and Thesaurus Classes

To increase the chances of retrieving more relevant documents terms with high frequencies but low weight can be combined to form term phrases and those terms with low frequencies can be joined together to form thesaurus classes.

A phrase is a pair of 2 stems which occur in the same sentence within a document text and at least one of the stems has document frequency in the collection exceeding a given threshold. Forming phrases from stems improves the level of precision as the specificity of the document and query content identifiers are increased.

A thesaurus combines groupings of similar or related terms into term classes. By adding thesaurus classes, the range of index terms is broadened and this enhances the level of recall in retrieval.

2.2.2 File Structures

Having chosen the terms to represent the information items, the next step is to organise these into a file structure which ensures efficient operations such as updating and searching. Various file structures exist ranging from no organisation to complex relationships between the terms. A very simple file structure is called a linear list, which is basically an unordered collection of items. A file can be ordered sequentially according to the values of one of its fields used to obtain access to the stored records⁵. Although addition of a new record requires making room at a specific location, the improvement in searching for a record over the linear list more than compensates for such a drawback.

2.2.2.1 Indexed files

An even more complex file structure is an indexed file where an index provides access to particular segments of a file. Searching is improved even more

⁵ these fields which are used for accessing stored records are known as **keys**

than the ordered sequential file as now only a search of the index and that part of the file specified by the index is required.

2.2.2.2 Signature Files

Another structure further down the spectrum is a signature file which is a probabilistic method for indexing text. Each document has an associated *signature* in which every indexed term is used to generate several hash values and the bits of the signature corresponding to those hash values are set to one. To expand on this, a signature file is usually a fixed length bit string within which bits are set to represent the contents of an information item. A hash function is applied to each of the index terms chosen to characterise the document resulting in one or more bits in the bit stream to be switched on (for each operation on a term), thus creating the bit string. A signature file could be defined as a collection of word fingerprints. When these word fingerprints are combined, they form an abstract of the original text file [ZEZ-91] . In this case, the fingerprint is the fixed length bit-stream created by the hashing function and it uniquely identifies a particular term in the original document.

Once the signature file has been created, the retrieval works by comparing the query signature with the document signatures. Less storage space is needed to store the signatures and matching operations can be implemented more efficiently as a result of the bit-parallel, word-serial nature of conventional processors. Using the signature file approach, results in improved performance in the retrieval as there is less data being transferred from storage to the processor because the signatures are a great deal smaller than the original documents from which they were generated.

2.2.2.3 Inverted Files

The most widely used file structure in IR systems is an inverted file. This file contains a *postings list* for each term in the index's lexicon. Each record in the postings file stores a list of pointers to all occurrences of a given term in the document

corpus. Each postings list entry is, in effect, a unique identifier for a document in which the term appears [WIT-94]. Using this type of file eliminates the need to search the entire collection of documents as only the postings list entries for each query term need to be checked.

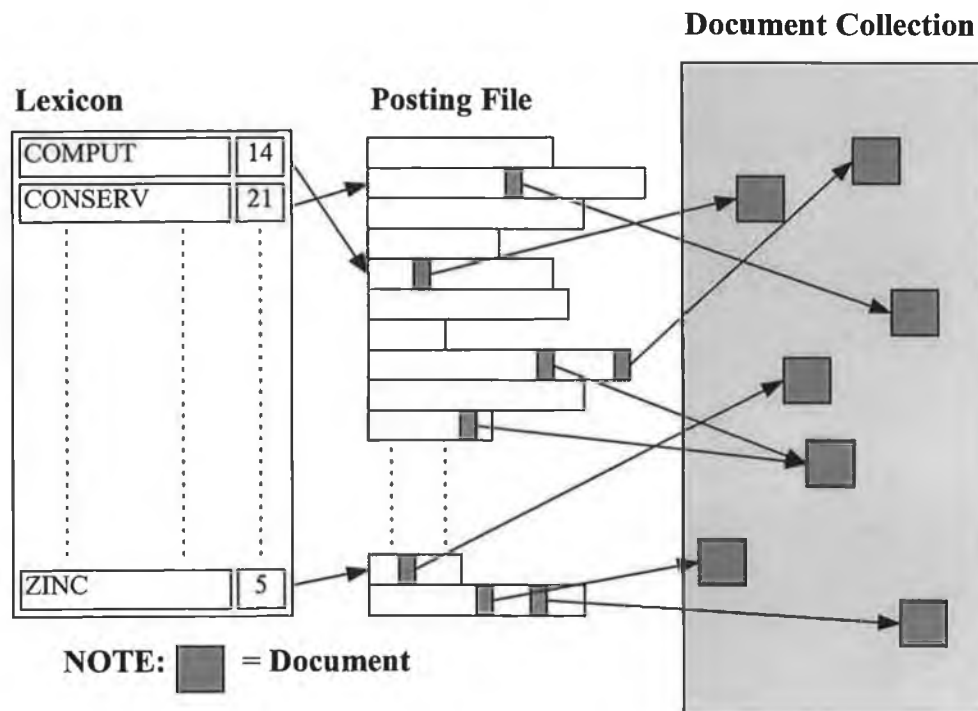


Figure 2-4 - General Structure of an Inverted File

2.2.2.4 Clustered Files

The traditional approach using inverted index files has a disadvantage in that the information regarding a particular document is scattered among many different inverted-term lists. In addition, information pertaining to different documents with similar term assignments is not in close proximity under this file system [SAL-89]. It is useful to organise the stored information in a manner which simplifies file access and manipulation. Clustering is one method of classifying a collection of stored records and involves grouping similar documents into common classes. Using clustered files, documents in a particular class can be stored in adjacent locations allowing a single file access to reach an entire class of documents [SAL-83]. This

cluster search process results in an increase in efficiency as the search is confined to those document clusters which seem to be most similar to the queries made. Effectiveness may also be improved under the *Cluster Hypothesis* which states that closely associated documents tend to be relevant to the same queries [VAN-79]. The cluster search process is implemented using a **cluster centroid**. As with the ‘average’ document described earlier, the cluster centroid is typically the average vector of all the documents in a particular cluster. That is, the weight of a centroid term is calculated as the average of the weights for that term in the clustered documents.

$$CTERM_k = \frac{1}{m} \sum_{i=1}^m TERM_{ik}$$

$CTERM_k$ is the k th term in the centroid, $TERM_{ik}$ is the k th term in the i th document in the cluster and there are m documents contained in the cluster. The cluster search proceeds as follows -:

```

for (i = 1 to number_of_clusters)
{
    csim = compare(query_vector, centroid[i])
    if (csim > threshold)
    {
        for (j = 0, j < docs_in_cluster[i])
        {
            dsim = compare(query, doc[i][j])
            if (dsim > threshold)
                retrieve doc[i][j]
        }
    }
}

```

The query is first compared with the various centroid vectors by calculating the centroid-query similarities. For the clusters whose centroids show a sufficiently high similarity with the query the individual documents in the cluster are then compared to the query and those exhibiting high enough similarities with the query are retrieved for the user.

2.2.3 Information Retrieval Models

The goal of an IR system is to retrieve all the relevant documents that match the users query while retrieving as few non-relevant documents as possible [VAN-79]. Many IR systems incorporate query reformulation, known as “relevance feedback”, whereby the query is reformulated to improve the retrieval of relevant documents based on the relevance assessments supplied by the user. The aim is to construct new queries which show greater similarity with the documents previously deemed relevant by the user than the original queries [SAL-83]. At the same time, the new queries are supposed to be less similar to those documents identified as non-relevant than the originals. The assumption is that the newly reformulated requests will retrieve more information items resembling the relevant documents previously returned, and fewer items resembling the non-relevant ones. Figure 3 shows a general model of IR which describes the basic entities and processes in the IR situation.

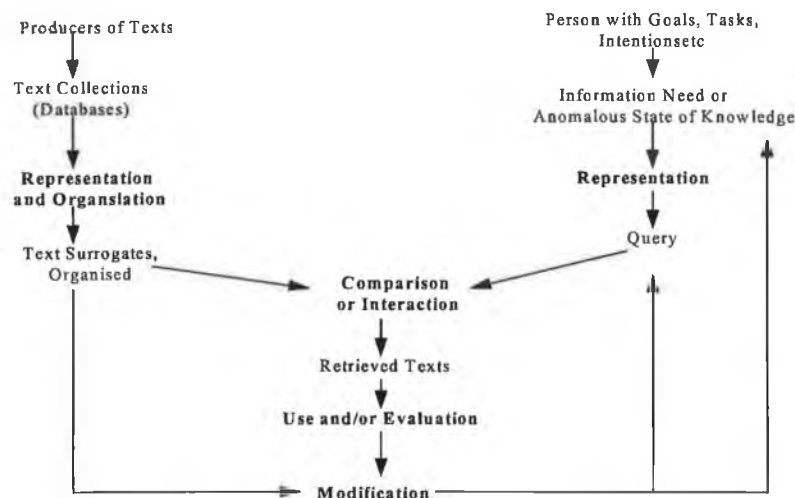


Figure 2-5 - General model of Information Retrieval

2.2.3.1 Boolean Retrieval

This is the retrieval model currently used in the vast majority of IR systems and is the simplest retrieval model in operation today. It utilises the inverted file approach, outlined previously and is based on the 'exact match' principle. The Boolean operators AND, OR and NOT are used in conjunction with the query of terms in order to specify what type of match is required for retrieval to take place. This is a relatively straightforward task which involves the following steps -: the documents are searched for each term in the query; each inverted file entry is retrieved and decoded and the entries are merged, taking the union, complement, intersection, as appropriate. Then the documents which have been indexed in this way are retrieved and displayed to the user [WIT-94]. To clarify the procedure this example is given -:

Sample Boolean Query

(computer AND applications) OR programming

- The posting list for 'computer' is accessed and decoded. This activates all documents containing the word 'computer'.
- The posting list for 'applications' is then accessed and decoded. As the Boolean query specifies an AND operator between 'computer' and 'applications', the two postings lists are combined using a set operation. This results in a new set of active documents which contain 'computer' and 'applications'.
- The next step involves in accessing the posting list for 'programming' and decoding it. This is then combined with the set from the previous step using a set union operation.
- The net result is an active set of documents containing 'computer' and 'applications' or 'programming'.

In addition to specifying that the existence of query terms within a document is sufficient to warrant its retrieval, it is also possible to impose conditions so that the terms must appear in the same sentence of a document before it will be retrieved. A

major problem with this model is that it does not allow for any ranking of the retrieved documents.

2.2.3.2 Vector Space Model

This is based on the concept of ‘best match’ and is a response to the problems of the exact match model. In this model documents are represented by **term vectors** in the form $DOC_i(TERM_{i1}, TERM_{i2}, \dots, TERM_{it})$ with the coefficient $TERM_{ij}$ representing the value of term j in document i and the query is represented by the vector $QTERM_j(QTERM_{j1}, QTERM_{j2}, \dots, QTERM_{jt})$ with $QTERM_{jk}$ representing the importance of term k in query j . These values are usually set to either 0 or 1 to indicate presence or otherwise of a term in the document and likewise for the query. Documents are retrieved by performing similarity computations between the query vector and documents vectors. The assumption is that the more similar a vector representing a text is a query vector, the more likely that the text is relevant to that particular query. The most common similarity measure used is the cosine of the angular distance between vectors outlined earlier.

$$COSINE(DOC_i, QUERY_j) = \frac{\sum_{k=1}^t (TERM_{ik} \cdot QTERM_{jk})}{\sqrt{\sum_{k=1}^t (TERM_{ik})^2 \sum_{k=1}^t (QTERM_{jk})^2}}$$

The document and query terms can be weighted to account for their importance, contributing to greater numbers of relevant documents being retrieved.

The documents retrieved can be ranked in decreasing order of the query-document similarity. Alternatively, a variable amount of documents can be retrieved as requested by the user.

2.2.3.3 Probabilistic Retrieval

Probabilistic Retrieval is also a ‘best match’ model and is based on the Probability Ranking Principle [ROB-77] . This states that the function of an IR

system is to rank the documents in the collection in the order of their probability of relevance to the query, given all the evidence available. This principle accounts for the uncertainty in the representation of the information need and the documents in the database. Under this model, a numeric weight is assigned to each retrieved document as an estimate of the probability that it is relevant to the query. As the relevance of a document with respect to a query is a matter of degree, it is assumed that when the document and query vectors are sufficiently similar, the corresponding probability of relevance is large enough to make it reasonable to retrieve the document in answer to a query [SAL-83]. Documents are presented to the user in weight order, so those considered to be the 'best match' come top of the list. This newer method of retrieval can include term dependencies and relationships and major parameters such as weighted query terms.

2.2.3.4 Latent Semantic Indexing

This is an extension of the vector space retrieval method which has improved performance over the traditional vector space techniques. The associations between terms and documents are calculated and exploited in the retrieval process. This is achieved by organising information into a semantic structure that takes advantage of the implicit higher-order associations of terms with documents. The structure created reflects the major associative patterns in the data while ignoring some of the smaller variations which may be due to idiosyncrasies in the word usage of individual documents [FOL-90]. The assumption in LSI is that there is a 'latent' or underlying structure in the pattern of word usage across documents and that statistical techniques can be used to estimate this latent structure. Using the singular-value decomposition technique, a term-document matrix is formed, from the collection of documents, in which each entry consists of an integer representing the number of occurrences of a specific term in a particular document [DEE-90]. The Singular Value Decomposition (SVD) of this matrix is then computed and any small singular values are eliminated. These small singular values are judged to be practically irrelevant when computing document similarities and including them would reduce the accuracy of the relevance judgements. The resulting singular vector and matrices are then used to map term

frequency vectors for queries and documents into a subspace in which semantic relationships from the term-document matrix are preserved while the 'noise' that results from term usage variations is suppressed. As the LSI method results in a subspace based on the underlying semantic similarity, documents with similar content will tend to cluster into this space. In determining whether a particular document is relevant to a query, the document is firstly folded into this subspace on the basis of the terms contained therein. If it appears close to documents already judged to be interesting in the space, then it is considered to be of interest and is subsequently retrieved for the user.

2.3 Information Filtering and Information Retrieval

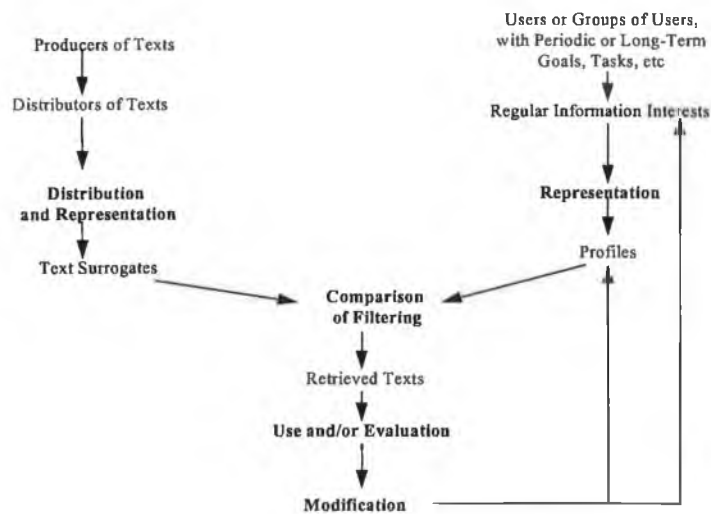


Figure 2-6 - Information Filtering Model

In the IR model a person with an information need represents this need as a query and submits it to the system. On the other side, documents are grouped into collections and are indexed (into text surrogates) for easy manipulation. The query is compared to the text surrogate which leads to a selection of possibly relevant documents. These retrieved documents can be evaluated which may lead to some sort of modification to the query known as relevance feedback.

In the Information Filtering model, a person has a fairly stable, long-term information need and represents this need in the form of a profile which is then put to the system. On the other side, the producers of the documents distribute them to make them available to those interested. The documents are represented as text surrogates and compared to the profiles. Various documents are then brought to the attention of the user, who may or may not evaluate how these satisfy the information need and interests represented in the profile. This evaluation may lead to some modification of the profile.

By comparing the two models, it seems that there is a close relationship between them. According to Belkin and Croft, IR and IF are 'two sides of the same coin' [BEL-92]. Both are concerned with presenting information to people who need it and both are deal with practically the same kind of information.

However, while comparing the models shows that the entities and processes in both are almost identical, there are differences:

Information Filtering	Information Retrieval
Repeated uses of the system by a person with long-term interests	Single uses of the system by a person with a one-time query
Assumes profiles are correct specifications of information interests	Recognises inadequacy of queries as representations of information needs
Distributes texts to person/persons	Collection and organisation of texts
Selecting texts from a dynamic stream of data	Selects texts from a relatively static database
Concerned with long-term changes over a series of queries	Responds to user's interaction within a single query
Timeliness of a text and privacy significant	Timeliness not an issue and privacy not important
Undefined user groups with highly varied domains	Well-defined user groups in specific domains

Figure 2-7 - Information Filtering/Retrieval Comparison

2.4 Current Information Filtering Systems

Most of the information filtering systems being used today, employ IR techniques such as indexing, Boolean retrieval, expert/rule based systems, vector space techniques and clustering that have been developed in the field of IR. In the past the most common approach using these techniques is *content-based filtering*. This type of filtering recommends items for the user's attention based on similarities or correlations between the content of the items and the user's preferences. Keyword filtering and latent semantic indexing are examples of content-based filtering. There are some limitations associated with this approach. The items must be in some machine parsable form or attributes must have been assigned to the items by hand. With current technology, media such as sound, photographs, art or video items cannot be analysed automatically for relevant attribute information. Often it is not possible or practical to assign attributes by hand due to a lack of resources. There is no inherent method for generating accidentally satisfying finds as the system recommends more of what the user has already seen. Also, systems based on content-based filtering cannot filter items based on quality, style or point-of-view. For instance, they cannot distinguish between a well written and a badly written article if both articles happen to have the same terms. In recent times another filtering approach is being investigated as a complementary method to the content-based filtering. This is known as *social filtering*. Social filtering basically automates the process of 'word-of-mouth' recommendations. Using this type of filtering, systems may suggest items to the user which are different (content-wise) from what the user preferred previously. In addition, the recommendations are based on the quality of the information items, rather than the more objective properties of the items themselves. *Collaborative filtering* is a form of social filtering that involves people working together to help with filtering various documents by making their opinions on the documents available to everyone.

There are many information filtering systems being developed today. We have broken these down into two basic categories, namely personal filters and group filters. There are many information filtering systems available and being developed,

so we have chosen to describe only a portion of these. However, the systems we describe are typical of the filters in the two categories⁶.

2.4.1 Personal Filters

Personal Filters include all filtering systems which are used by individuals who are not part of any particular group. Although several people may use a particular filter, a person who does so is not affected by or has no contact with anyone else employing the same filter. These personal filtering systems include electronic mail filters such as Elm, Procmal and Mailfilt. Other systems included under the Personal Filter category are Borges, SIFT and InfoScan which all utilise a 'user profile'.

2.4.1.1 Elm - Email Filter

This filter is built into a mail tool and was first developed by Taylor in 1986 [TAY-86] to deal with electronic junk mail and store messages that, while interesting and important could be filed and read at a later time avoiding a cluttered mailbox. It employs content-based filtering allowing rules to be defined by matching certain keywords within the header field or message body. The rules can then filter out unwanted messages, sort message types into mail boxes or folders or perform more complex commands. Since Taylor first floated this idea back in 1986, quite a bit of work has gone into this filter system and it is now designed to run on almost any UNIX system and because it is screen-oriented it is easier to use than earlier mail systems.

2.4.1.2 Borges

The Borges project was initiated as part of the European Union's Libraries Programme. The partners who worked on the project are Digital Equipment Corporation, (The Netherlands), Dublin City University and University of Barcelona.

⁶ The University of Maryland WWW site - <http://www.ee.umd.edu/medlab/filter/software.html> is an excellent source of information on current information filtering systems

The resulting Borges Filtering Tool is an information filtering service which uses keyword matching between a user's profile or query and the text of Usenet News articles or newly found WWW pages [SME-96] . The underlying retrieval engine employed is SMART which uses vector space information retrieval techniques outlined in section 1.2.3.2. Users can access the Borges Tool via Netscape or some other WWW based browser and having logged on to the system can define any number of profiles which will then be matched against the incoming documents on a regular basis. At present this filtering is carried out 5 times per week. When the user logs on to check out what matched the profile, a list, called a digest, of the top 50 ranked documents for each profile is presented, with HTML links to the actual documents.

When defining a profile a user is allowed to specify phrases or word collocations as part of the profile and these phrases are matched against the same phrase occurring in documents. This leads to a higher quality of output in the digests with better matching between profile and news article. In addition, WordNet⁷ is used to disambiguate polysemous words (words with multiple semantic meanings). In a user's profile these words are highlighted and the user is asked to choose between alternative "senses". These disambiguated word senses are then expanded by the inclusion of related terms into the user's profile but this is done transparently to the user. For example, if the user had the term "bank" in a profile and chose the "financial institution" sense of that word then the terms "finance", "money", "loan", "borrowing", "financial institution", etc. would possibly be added to the profile. Related terms are taken from the WordNet lexical database developed at Princeton University.

2.4.1.3 SIFT

The Stanford Information Filtering Tool (SIFT) is a tool for information providers to perform large-scale information dissemination [YAN-94] . Similar to Tapestry, SIFT deals with NetNews articles. It supports full-text filtering using well-known information retrieval models and the filtering engine used implements novel

⁷ WordNet is a lexical database developed at Princeton University

indexing techniques. The engine is capable of processing large volumes of information against a large number of profiles.

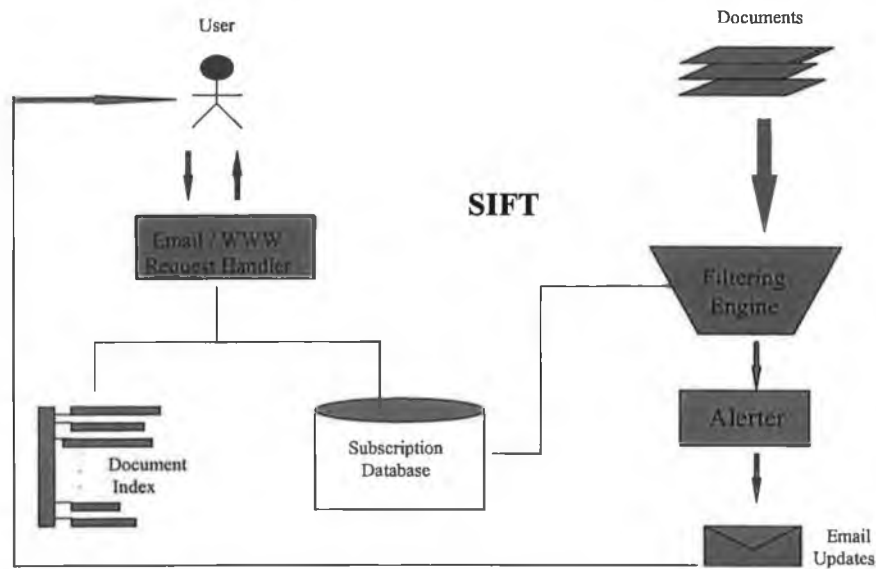


Figure 2-8 - Overview of SIFT

A user subscribes to a SIFT server with a subscription for each topic of interest. This subscription includes -:

- a profile which can be expressed in one of two IR models : Boolean or vector space model [SAL-89]
- additional parameters to control the frequency of updates
- the amount of information to receive - the user specifies a *relevance threshold*, which is the minimum score that a document must have against a profile for it to be delivered and
- the length of the subscription. A subscription is identified by the email address to the user and a subscription identifier.

Unlike the traditional IR approach to processing queries against a collection of documents, the SIFT filtering engine builds an inverted index of profiles and the documents are treated as queries. Each profile has an entry in two arrays - Threshold (relevance threshold outlined earlier) and Score (the sum of all the term-term matches

after the document is matched against the profile). If the Score value is greater or equal to the Threshold value, an excerpt (length of this specified by the user) of the document is sent out in an email message to the user. After reading the excerpt the user can access the SIFT server to retrieve the entire document if so desired.

SIFT is currently used for selectively disseminating NetNews articles and there is also a SIFT server for disseminating Computer Science Technical Reports.

2.4.1.4 Connectionist Filtering System

McElligott and Sorensen [McE-93] presented a design for filtering news articles on the Internet which draws on theories and techniques in artificial intelligence and cognitive psychology. Many of the current approaches used in Information Filtering (Boolean and vector space models, automated keywording, expert/rule based systems, thesauri), are dependent on the user being expert in describing what is of interest. This is seen as being a major flaw, therefore it was decided to build a filtering mechanism which would determine when the right terms appearing in an article also appear in the right context. To build such a filtering mechanism, it is necessary to have previous experience of just what the right context is. A design is detailed which constructs a user profile based on previous articles that the user deemed interesting. The user profile acts as an intelligent user agent for the filtering of subsequent articles, while adapting to user feedback regarding its success in filtering those articles.

In light of the fact that a user's profile is based on articles already read by the user, the authors suggest designing a dynamic memory system. This dynamic system would facilitate the formation and storage of a user profile based on the underlying principle of episodic memory. An outline is given of how a user profile might be designed based on the script theory of Schank [SCH-77]. In an effort to overcome the problem of relevance, a document is viewed as a society where each word is an agent [ROW-91]. These agents co-operate with each other to form phrases and sentences. The field of neural networks aids in ascertaining how the society (document) can pick out the important phrases. An adapted version of anti-Hebbian Learning is used -

“The strength of the link between one word and another depends on how often these words occur in sequence in interesting articles. This link is strengthened each time the sequence occurs and possibly weakened when it does not. Each word has a maximum strength which the sum of its links cannot exceed.”

As the success of this approach largely hinges on the quality of the articles that the user presents to teach the system, the authors make use of biological evolution theory, using the following analogies -:

document (society)	an organism
un/interesting articles	life-threatening/enhancing substances
the quality of results from the system	life expectancy

Genetic algorithms are employed to ensure the system evolves to a point where 70% or more of the articles selected are relevant. This approach was designed to filter news articles on the Internet.

2.4.2 Group Filters

Group Filters are those filtering systems which are designed to cater for groups, where there is communication via the system between all the people using the filter. This category includes the Information Lens System, Tapestry and GroupLens.

2.4.2.1 Information Lens System

Malone et al 1987 [MAL-87] established an intelligent information sharing system in their efforts to reduce information overload in an organisation. The Information Lens system is designed to help people share and filter information communicated by computer-based messaging systems. The goal of the system is to increase the amount of useful information that can be exchanged electronically without leading to information overload. They identified three categories of filtering techniques - cognitive, social and economic based on the sources the techniques draw on in order to predict a user's reaction to an article.

The filtering mechanism is constructed out of a set of rules that is then applied to each incoming message/article - an expert/rule based system.

The systems provide categorisation and filtering of semi-structured messages such as email. The user defines rules for filtering, and the processing is carried out at the user site. The user interface is seen as being the key to efficient information processing in this system. It is built on top of an existing electronic mail system and provides four important optional capabilities:

1. Structured message templates are available for message composition
2. Receivers can specify rules to automatically filter and classify messages arriving in their mailbox.
3. Senders can include a special mailbox (as an addressee of a message) to indicate that the sender is willing to have this message automatically redistributed to interested parties.
4. Receivers can specify rules that find and show messages that they would otherwise have not seen (via this special mailbox).

The system exploits concepts from artificial intelligence such as frames, production rules and inheritance network. It avoids the problems of understanding natural language by providing users with a rich set of semi-structured message templates. So the system works as follows -:

- at the time of authoring a message, the user selects a message type that best conveys the goal of the message
- the message is then distributed and people receiving the message subject it to a series of rules that filter the messages
- the rules allow for actions to be performed automatically on the messages and also messages can be sent around the organisation and 'grabbed' by other members of the organisation using special filters.
- this method allows readers to handle messages in a customisable manner and to find shared messages that relate to subjects of particular interest to them.

This system relies primarily on **cognitive filtering**, or content-based filtering where the content of a message and the information needs of the recipient are characterised and then these representations are used to intelligently match messages to recipients.

To a lesser extent **social filtering** and **economic filtering** are employed. Social Filtering complements cognitive filtering by focusing on the personal and organisational interrelationships between sender and receiver. It involves including, in the automatic filter mentioned above, rules concerning who sent the message and who has received a copy. Economic filtering is based on a cost-benefit assessment of a message, such as length of a message or for example if it is sent to one person only (one to one), it would be deemed more important than if it is sent to a distribution list. The system also allows prioritising of messages.

2.4.2.2 Tapestry

The Tapestry mail system [GOL-92] was developed at the Xerox Palo Alto Research Centre to address the growing problem of users being inundated by a huge stream of incoming documents yet at the same time feeling insufficiently informed. A user could just subscribe to certain mailing lists of particular interest to overcome these problems. However, items of interest don't generally fall neatly into existing mailing lists. The Tapestry system facilitates the filtering of all mailing lists, by selecting interesting items regardless of which mailing list they originated from. It processes information in two steps - firstly it filters a large continuous stream of electronic documents into a stream of smaller stream, satisfying a users interests and secondly organises and prioritises these selected documents. The basic idea behind the Tapestry system is that more effective filtering can be achieved by involving people in the process. To this end it employs collaborative filtering in addition to content-based filtering. Collaborative filtering is based on the premise that people looking for information should be able to make use of what others have already found and evaluated. The Tapestry system was the first filtering system to use the collaboration between users of the system. This collaboration involved recording users reactions to documents they have read. Readers mark those documents they find particularly interesting or relevant and these annotations are collected and made available to other readers.

The system is more than just a mail system as it is designed to receive, filter, file and browse electronic documents that arrive in a continuous stream [TER-93] . It consists of five main components -

1. **the database** which is built on Sybase (a relational database management system). Over 5000 new messages per day are added to this database. The current database contains approx. 3 weeks worth of NetNews (a bulletin board system on the Internet which is organised into discussion groups covering a wide variety of topics) and other news feeds.
2. **the clipping service** which runs filter queries continuously over this changing database and mails the results to users,
3. **the browser** for running ad-hoc queries using TQL (Tapestry Query Language),
4. **the shell** provides for manipulating a user's set of filters, including commands for prioritising messages and adding, deleting, modifying and listing filters and finally
5. **the mail reader** which displays and organises documents into folders.

The developers of Tapestry have detailed work to be undertaken on the system in the future -:

- accumulate more user experience with Tapestry to improve the design and cater for larger numbers of users at distributed sites
- include security into the system.

2.4.2.3 GroupLens Filtering System

GroupLens is a system which also employs collaborative filtering, in conjunction with user profiles which reflect the user's preferences. It is used to filter Usenet NetNews articles and helps people find items of interest in the huge stream of available articles [RES-94] . Usenet, is a distributed bulletin board system encompassing government agencies, universities, commercial organisations of all sizes and home computer users of all descriptions. It was first implemented in 1979 and has rapidly grown to become international in scale and is now probably the largest decentralised information utility in existence. It hosts well over 1200 newsgroups and an average of 40 megabytes of new technical articles, news, discussion, and chatter

are added every day. GroupLens is a distributed system for gathering, disseminating and using ratings from some users to predict other users' interest in articles. GroupLens is based on the idea that people who agreed in their evaluation of previous articles are likely to agree again in the future. Users assign ratings to the articles they read on a numerical scale. The system then determines the correlations between the ratings different users have given the articles. By ascertaining which of these user ratings are most alike, the system can then predict what new articles will be interesting to the users based on ratings from similar users. The GroupLens system is very similar to Tapestry. It is in fact an extension of Tapestry - it caters for sharing evaluations between sites and implements aggregate ratings from several evaluators. Tapestry, on the other hand, only allows sharing of evaluations within a single site and does not include any aggregate ratings.

2.4.2.4 Active Collaborative Filtering System

Another approach to collaborative filtering is outlined by Maltz and Ehrlich [MAL-95], whereby a person evaluates a particular document/article and shares this knowledge by sending 'pointers' to those documents to colleagues. The previous two systems could be described as employing 'passive' collaborative filtering because there is no direct connection between a person evaluating and rating an article or document and the other people who then filter these articles based on the ratings. The system adopted by Maltz was designed to support the informal practice of people telling friends and colleagues of interesting documents. As there is a definite intent on the part of the person who finds something interesting to share this knowledge with certain people, the system is referred to as being an 'active' filtering system. The basic concept in the system is the 'pointer', which consists of: a hypertext link to the document of interest, contextual information such as the date of document, title and optional comments by the sender. The added contextual information and any additional comments by the sender serve to help the recipient determine the interest and relevance of the document prior to retrieving and looking at it.

The filter was implemented inside a Lotus Notes environment. Lotus Notes is a commercial product from Lotus Development Corporation which consists of a group of application programs that allow organisations to share documents and exchange email messages. The documents which are stored in Notes are stored in databases and can be created by users within the Notes system or can be from external information sources such as NetNews, WWW or clipping services. Once a pointer to a particular document has been created, it can be saved or sent to one of these databases. A pointer can also be emailed or posted on a distribution list. This means of distribution is similar to the Tapestry system.

The greatest scope for improving information filtering systems lies with this collaborative filtering. Features of documents and articles such as quality and readability are easily evaluated by people. However, such features would be difficult for a computer to detect automatically. By including this 'human' aspect, information filtering systems can prove to be even more useful in the future as more and more information is made available on-line.

2.5 Summary

In this chapter, the area of Information Filtering was introduced. Details were given on Information Retrieval which is directly relevant to Information Filtering as most current filtering systems rely on IR techniques in presenting data to users which satisfies their information needs. Models of IR and IF were shown to have close ties but differences between the two were also outlined. Finally the chapter gave details about current information filtering systems.

Chapter Three - Monitoring the eSIGs

3. Overview

As outlined in Chapter 1, the eSIGs and the information they generate need to be monitored by the CTC. It was decided to try and develop some form of automatic monitoring system which would allow the CTC staff to periodically review eSIG activity. Ideally this system would be able to inform the CTC of what topics were being discussed and the regularity of communication on the eSIGs. By reviewing this data regularly, informed decisions can be made on whether or not the eSIGs are of benefit to the CTC members.

3.1 Monitoring the eSIGs using IF techniques

Also in Chapter 1, we referred to the findings of preliminary research which indicated that Information Filtering techniques were the best way to tackle the issue of monitoring the eSIGs. This is due to the nature of the data contained in the eSIGs which is textual and semi-structured data. In addition, we anticipate large flows of data as a result of activity on all the eSIGs. As can be seen from section 2.1, Information Filtering is designed to deal with this type of information.

In most IF systems, the person/people using the system know what kind of information they want to receive. Help is provided by the system when defining the request for information. For example, in the BORGES [SME-96] project phrases can be used and the specific sense of a word is also included to improve the filtering out of undesirable items and the retrieval of relevant information.

Having chosen IF as being ideal for monitoring the eSIGs, a closer look at current IF systems was undertaken. These systems were described in section 4 of Chapter 2. Content-based filtering and Social Filtering are the most common approaches being employed by these systems.

3.1.1 Content-based versus Social Filtering

Content-based filtering checks for correlations between the content of the items and the user's request. The limitations of this type of filtering would not seem to affect a CTC filter. Electronic mail messages are in machine parsable form, so no formatting is necessary to use this approach. In addition, the limitation of missing serendipitous finds is not an issue, nor is the disadvantage of being unable to distinguish between well or badly written articles, or determining the quality of a particular article/document.

Social Filtering appears to be unsuitable for the eSIG system as this type of filtering is used in filters falling under the Group Filter Category. As the CTC is the only body involved in monitoring the discussion groups, the filter designed falls in the Personal Filter Category.

Having looked at the filtering techniques, an in-depth analysis of the current status of the eSIGs was then taken before deciding on a specific eSIG filtering system. By doing so, we were able to check the number of members and the traffic between members to get a rough picture of what volume of traffic to expect and perhaps gain an early insight into whether or not the eSIGs were delivering on their potential.

3.2 Preliminary Appraisal of eSIGs

The primary objectives of the eSIGs are worth reviewing at this point -:

- To establish a database of European recommended syllabi for the area of Computing
- To develop courseware written to the specifications of these syllabi

It was anticipated by the CTC that the provision of this type of electronic forum would facilitate a higher level of communication between academics within their areas

of expertise. This in turn would result in the objectives outlined above being achieved.

When the CTC was being established, one of the main activities to be undertaken was the organisation of an annual conference on 'Computing Curriculum Development and Delivery' [CTC-94]. This has been a successful activity since the inaugural conference in September 1993. The conference is designed to strike a balance between the traditional lecture presentations, demonstrations and interactive workshops. Reports are presented on the various other CTC activities that have taken place over the year and the work programme for the next year is agreed between the delegates. The emphasis throughout the conference each year is on co-operation between delegates and sharing of ideas and experiences is strongly encouraged during the workshop sessions and indeed during the presentations. It was during these interactive meetings that the idea to set up an electronic forum to continue and support the work of the annual conference was first suggested. Because of the enthusiasm surrounding the establishment of the eSIGs, it was expected that there would indeed be discussion between the members who joined. It was also expected that the discussion taking place would involve curriculum and related topics and perhaps some information about seminars, workshops, conferences and calls for papers.

3.2.1 Members and Traffic on the eSIGs

Membership of the CTC groups was investigated (see Figure 3-1). Due to the large number of academics who had become members, it was anticipated that the expectations outlined would indeed be met.

Members are mainly from the UK and Ireland, but others hail from Canada, France, Greece, America, Poland, Italy, Germany, Tunisia, Thailand, Portugal, Belgium, Brazil, Singapore, Austria, Croatia and Russia.

Current Status - August 1995

eSIG	Members	No. of msgs
cti-comp-ai	95	6
cti-comp-caos	49	3
cti-comp-dbms	70	2
cti-comp-fm	42	4
cti-comp-hci	70	13
cti-comp-mat	129	7
cti-comp-pl	99	4
cti-comp-se	142	14
Totals	597	53⁸

Figure 3-1 - Current Status of eSIGs

As can be seen from Figure 3-1, the Software Engineering and Human Computer Interaction eSIGs have been the most active and they account for a third of eSIG members.

Most messages sent to the eSIGs concerned conference announcements, details of upcoming workshops and seminars and requests for information and replies to these. Others included details of new postgraduate courses, a new journal and frequently asked questions for a particular area.

As can be seen from Figure 3-1, the number of messages was not as high as anticipated. There was very little traffic and hardly any discussion about curriculum or syllabus issues. It was decided at this point to investigate why this was the case. To this end some research was undertaken to ascertain why the problem was occurring and ultimately help with improving the situation. The position is very similar to any company looking for information pertaining to a particular item in a product range which may not be performing as predicted. This information can help when developing a strategy to right the situation. Any planned and organised effort to gather new facts and new knowledge in order to help make better market decisions is referred to as Marketing Research [BRE-82]. Having formulated the problem at hand, namely that there is not enough communication between members of the eSIGs, the next step involves selecting an appropriate data collection method.

⁸ not all unique (i.e. same msg sent to several eSIGs). Total no. of unique msgs - 32

3.3 Data Collection

The main data collection methods include the following -: Mail and telephone surveys, secondary sources and qualitative interviewing [LOC-94]. In marketing research, the secondary sources collection method involves searching through internal company sources such as files, records and accounts and also external sources such as published material including statistics and reports issued by governments, trade associations or other organisations. Many advertising agencies and research companies frequently circulate facts and figures about products, consumers and market trends which can also be rich sources of information. In this particular situation, the secondary sources avenue has already been followed, namely establishing the facts and figures about the eSIGs from MAILBASE⁹ sources. The other methods generally employ a questionnaire. A questionnaire is a method of obtaining specific information about a defined problem so that the data, after analysis and interpretation, result in a better appreciation of the problem [CHI-82]. Qualitative interviewing involves administering the questionnaire in person to either an individuals or a group of people. Panel research is another method of qualitative interviewing where a questionnaire is administered to certain 'panels' of regarding their buying behaviour. Such interviewing was deemed unsuitable in this situation due to cost and time constraints. Mailing or telephoning a sample of the members was also ruled out due to these constraints. However, there was another possible data collection method open to us. It was decided to use the existing infrastructure available to us and put a questionnaire on the CTC World Wide Web site. Chapter Five will discuss the World Wide Web further. A request to all members to complete the questionnaire was made by posting a message to all the eSIGs.

3.4 Questionnaire Design

Building a questionnaire is more of an art than a science. If several different research experts were asked to build a questionnaire given a specific problem it is

⁹ See Appendix A for more information about the services supported by MAILBASE

almost certain that several different questionnaires would be produced. Although this is the case, it is advocated that the introduction to a questionnaire should be simple and to the point. It should explain the nature of the survey and invite the respondent's co-operation [CHI-82]. This advice was followed for our introduction which simply states that the CTC are collecting information about the operation of the eSIGs and would appreciate members' co-operation in filling in the questionnaire. Appendix B includes the actual questionnaire. The questions used in the body of a questionnaire fall into 2 main categories. Open-ended questions and closed questions. Open-ended questions allow the respondent to answer freely. No possible answers are suggested, rather the respondent is allowed to structure the reply as seen fit. There is another form of the open-ended question which directs the respondent to some item which may be of specific interest to the researcher. Such questions are known as direct open-ended questions and allow you to limit the replies given, preventing answers that stray off the point. Closed questions involve giving the respondent a selection of answers to choose from. There are two types of closed questions - the binary or simple alternative question and the multiple-choice question. The former gives the respondent only two choices of reply while the multiple-choice question allows an answer or answers to be taken from a list of possible answers. Most questionnaires contain both types of questions just described. The CTC questionnaire follows the same format with a mixture of open-ended and closed questions. The first 2 questions are included to provide us with some basic data about the respondent - name and membership of which eSIG(s). The rest of the questionnaire is a mixture of open-ended questions (Q5), direct open-ended questions (Q3, 6b, 7b, 8b, 9b, 10, 11) and closed simple alternative questions (Q4, 6, 7, 8, 9, 12). It was intended that using such questions would gather the information necessary to allow the CTC to appreciate the eSIG situation better.

The questionnaire was posted on the CTC World Wide Web site using the Hypertext Markup Language (see Section 5 for more details) and members were contacted via email asking them to complete and submit it back to the CTC. The CTC Home Web page also included a request and encouragement to fill in the

questionnaire. Appendix B contains the plain text version and the WWW version (with the HTML source code) of the questionnaire.

3.5 Results of Questionnaire

A total of 40 members completed and submitted the questionnaire. This amounts to less than 10% of the total number of members for all the eSIGs. The following results were obtained -:

Members join :

- To keep up to date with current developments in their area, latest web pages, developments in material for supporting teaching and course delivery
- To receive information on conferences, workshops, seminars, technology in teaching, other information not specifically related to teaching
- To share experiences, knowledge, materials
- To contribute to discussions and offer comments in specialised areas
- To keep up contacts made elsewhere and to find out what others are doing in the area.

These answers coincide with the CTC reasons for establishing the eSIGs.

When asked about membership of other mailing lists, most members indicated that they are also members of various other mailing lists.

These other lists are found useful for the following reasons :

- More active - more information coming through regularly
- There are many more contributors doing modern work
- Access to a larger research community
- Receive early postings of sources of information - research papers, book publications
- Information varied - not always specifically teaching issues

- Similar to electronic newsletter

It appears that the main reason for finding these other lists useful is that there is actually discussion between the members resulting in a regular flow of information being available.

The answers regarding what a discussion forum for academics would provide and the objectives of setting up the eSIGs (question 9), begin to give an indication into why very few members chose to communicate with other academics via the CTC eSIGs. For example :

- Objectives are reasonable, but perhaps we don't have sufficient time to want to participate in these activities, through eSIGs
- They are a very ambitious set of aims which require a great deal of altruism on the part of members
- Objectives listed are already served by other lists on the Internet - hardly worthwhile replicating these services for a small community
- Too much to expect
- Noble objectives, but with increased teaching and marking load of recent years, who has time to manage bulletin boards
- Too ambitious - don't think you can draw up policies and syllabi via email. Need a more formal arrangement for that.

The number of respondents was lower than expected which is indicative of the poor participation levels of eSIGs members. It was clear that some further measures were necessary to support discussion on the eSIGs.

At this point in our research, the CTC became involved in a European Union project initiated to increase the level of co-operation between universities in Europe in the field of computing. One of the tasks of the project was to examine the success of the ERASMUS¹⁰ programme introduced in 1987 by the European Commission. The resulting report [RYA-96] identified that the scheme had been very successful in promoting the exchange of students and teachers between European universities. The

¹⁰ European Community Action Scheme for the Mobility of University Students

report also indicated that the main area for improvement was the joint development of joint curricula by participating institutions. It was decided that developing a tool which would allow people quick access to information on the area of computing in universities across Europe could help address this issue and at the same time prove to be a useful support measure in improving discussion on the CTC eSIGs. As a result, the next step in our research was investigating what format this tool would take and what user interface would be suitable for such a tool. The next chapter will discuss 'User Interface Design' issues and outline the format and user interface chosen for the CTC tool mentioned above.

Chapter Four - User Interface Design

4. Overview

Previously, the emphasis in system design was on solving technical problems. Such an approach focused on the computer hardware and software and the limitations of current technology faced by system designers. Due to the continuing increase in the power of computers, improvements in reliability and flexibility and recognition of how poorly designed many systems were, there is less importance being placed on the technology side of system design. More emphasis is now being placed on the human-computer interface and has led to a more user-centric approach being employed in system design. Indeed there is enormous interests in improving all aspects of communicating between systems and people as computers are being used for many purposes from industrial and commercial systems used in offices and factories to life-critical applications such as air traffic systems and medical systems in hospitals to home and entertainment applications.

4.1 User Interface Design

Work on improving human-machine interaction is known as User Interface Design. User interfaces are those parts of computing systems that allow the person using the computer access to the facilities offered by the computer. It is this user interface which acts as an information channel that conveys information between user and computer [THI-90]. The needs and characteristics of humans with respect to designing interactive systems represents a wide field of human knowledge. For this reason user interface design combines psychology and computer science. The psychology component is concerned with people and the computer science deals with computer machinery and the mechanics of how it has to be adapted to suit the users.

4.2 Design Issues

The following principles are offered as general guidance in designing a user interface. These principles represent a condensation of the many general principles detailed in the literature on user-interface design.

- **Maintain Consistency:**

Sequences of actions should be consistent in similar situations and identical terminology should be used in prompts, menus and help screens. Commands should be also be consistent throughout the whole system.

- **Provide Shortcuts and Flexibility:**

As the frequency of use increases, so does the need to reduce the number of interactions necessary to complete a specific task and also increase the pace of interaction. Abbreviations, special keys and hidden commands can be built in for the experienced user.

- **Present Informative Feedback**

For every action some form of system feedback should be provided. This feedback will differ depending on the importance of the action.

- **Aid Orientation and Navigation**

Provide orientation aids and instructions to help users maintain a sense of where they are in the system, what they can do, and how they can get out. Sequences of actions should be organised into groups with a beginning, middle and end. The provision of some informative feedback at the completion of such a sequence ensures that the user can proceed on to the next set of actions, satisfied that the previous task has been completed.

- **Design a simple error handling procedure**

As will be discussed below, as much as is possible, design should prevent the user from making serious errors. If errors are made, simple, easy to understand instructions for handling the error should be available to the user.

- **Allow simple action reversal**

Where possible actions should be reversible. Users can be sure that any actions can be undone, providing users with the confidence to explore new sequences of actions, satisfied that no irreparable errors will occur.

- **Support the feeling of control**

Operators and users of systems like to feel in control. If it is tedious and difficult to obtain necessary information, then a feeling of incompetence may arise and lead to total dissatisfaction with the complete system.

- **Reduce the short term memory demand**

In order to cater for the limitations of the human information processing in short term memory, displays should be kept nice and simple. Reducing the frequency of window motion and providing ample training time for sequences of actions can help to see that this principle is enforced [SHN-87]

These principles have to be interpreted and adapted for different environments. When dealing with the other design issues outlined below, these principles should be kept in mind at all times. Doing so will ensure the provision of an interactive system that is highly productive.

There are other issues associated with the design process. These fall into two categories, namely primary and secondary or human factor goals. While the growth in the use of computers stimulates further interest in the human factor issues associated with user interface design, the most important issue for interactive computer systems is no different from early computer systems, interactive and otherwise. They simply have to work. The following section discusses the steps necessary to ensure that the system first and foremost fulfils this basic requirement

4.2.1 Primary Design Goals

Functionality: The first step is to determine the necessary functionality. Task analysis is central, because systems with inadequate functionality frustrate the user and are often rejected or under-utilised [BAI-82]. If the functionality is inadequate, it doesn't matter how well the human interface is designed. Equally, excessive functionality is also a danger, as too much clutter and complexity makes usage more difficult [SHN-87].

Compatibility: Any differences among computer systems may lead to instability and cause user errors. High levels of incompatibility may require significant retraining of users.

Reliability: The next step is ensuring proper system reliability. It is vital that the hardware support and software architecture provide high availability and ease of maintenance. The system must also be secure, preventing unauthorised access.

Schedules: Ensuring delivery of a system on time based on the planned schedule and within budget is the final step in the primary stage of design.

4.2.2 Secondary Design Goals

If the primary goals have been achieved, attention can then turn to secondary issues in the design process. These secondary goals concern the human factors which must be considered when designing user interfaces for computer systems. Human Factors (ergonomics) is the study of how people and machines interact. It is a technology for creating designs that work well in *human* terms [RUB-84]. An understanding of the intended user(s) is very important at this point in user interface design. Knowledge of what systems and user interfaces are currently being employed by the users will help in the production of a user interface that will meet users needs. In addition, any insight into how users have learned to use their current systems can assist when devising what training may be required for users to enable them to exploit the full capabilities of the system. Generally users fall into 3 categories - novice, knowledgeable intermittent and frequent users [SHN-87]. If multiple usage classes have to be accommodated in one system, then there are two possible strategies to take. The first permits some form of structured learning, whereas the second allows the user to control the amount of informative feedback that the system provides.

The human factor goals outlined below are central to designing a system with an appropriate user interface which reduces the time and effort required of users to learn about it. While all of these goals may be desirable, many applications may require a trade off. For example, if the error rate is to be kept to a minimum, then the speed of performance may have to be sacrificed.

- Ease of learning:

How long does it take for a typical user to learn how to use the interface and carry the appropriate commands to manipulate the system. The time taken for users to learn how to interact with the user interface and carry out the required set of tasks can dictate how successful the whole system will be. The goal here should be to provide short learning times. Nelson's 'Ten-Minute' rule suggests that a novice user should be able to learn how to use a system in just ten minutes[NEL-80]. This may not be applicable in many situations, but can act as a guideline for designers when developing a 'user-friendly' user interface. The robustness of the underlying system will also affect ease of use and learning curve for users. Robustness refers to the system's tolerance of difficulties and its protection of the user from them [RUB-84].

- Speed of performance:

How fast can a set of tasks be carried out. The speed of the system should allow rapid task performance. Reducing the time waiting for a response will contribute to ensuring that the user will be satisfied and high productivity can be achieved.

- Error rates:

How many errors and what type of errors are made in carrying out a set of tasks. The rate of errors should ideally be kept as low as possible. The reliability of the interface can help to reduce user error. Reliability refers to the absence of malfunctions. When a system is reliable, a user will be more confident and less unforced errors will result.

- Level of satisfaction:

How much do users like using the system. It is desirable to aim for high satisfaction rates, as ultimately it is this which will affect the acceptance of any system.

- Retention:

How well do users retain knowledge. This is closely linked to the first goal. The more often the system is used the greater the amount of information users retain about how the interface operates. This can affect how long it takes to learn how to carry out the specified tasks.

Once these issues have been taken on board, the next step is to draw up the tasks which need to be done. The appropriate sets of actions are devised for each particular task. When the tasks have been determined, choosing an interactive style for the user interface is the next phase in the design process.

4.2.3 Interactive Styles

There are 3 basic interactive styles of control in human interfaces - command, menu and direct manipulation.

- Command language- the user is required to create a statement using some predefined language describing what task the computer is to implement.
- Menu selection - the user is presented with selection among a number of choices. The key objectives of menu design are to minimise the user's search-and-selection time and to optimise the user's ability to navigate through the menu structure (
- Direct manipulation - involves representing objects to the user which can be directly manipulated. Graphical representations are often much easier to retain and manipulate than are textual or numeric representations. In addition, the results of actions are continuously visible. Techniques of direct manipulation include open/close, minimise/maximise, drag-and-drop, resize, and scroll. Many direct-manipulation tasks are initiated by the user's action of pointing and clicking on screen-control graphics or icons.

Once one of these interactive styles has been chosen, it is important to ensure that enough information is transferred between user and computer for the user to be able to make effective use of the system.

4.2.4 Data Display

It is advisable to present data simply and usefully. It is easy to provide the user with information that is not informative, and paradoxically, providing more information may make it less informative [THI-90]. Smith and Mosier put forward 5 high-level objectives for data display which can avoid clutter and make the user interface more informative [SMI-86]

1. Maintain Consistency in Look and Feel

Interface characteristics should be uniform and familiar, with consistent sequences of actions in similar situations. Terminology must be used consistently to avoid confusing the user.

2. Provide for efficient information assimilation

The format and terminology of the displayed data should be familiar to the user. The data displayed should be relevant to the tasks required to be performed.

3. Reduce Memory Demands

Users should not be required to remember information from one particular text box, screen or window to another. Long complicated sequences of actions should be avoided so that only a few commands will result in task completion.

4. Compatible Data Display with Data Entry

There should be a clear link with the manner in which data is entered and displayed.

5. Provide User Control of Data Display

Based on their knowledge, skill, and experience, users should have the flexibility to change or structure a system to suit their particular requirements. They should be allowed to get the information they need in a form which is most suitable for the task at hand.

Screen designs are a key component for most interactive systems. Screens are often more difficult to read than paper. In general cluttered displays should be avoided. They can be particularly annoying and do not exhibit consistent location, structure or terminology which as outlined above are important features that should be included in interactive systems. Colour is fast becoming a major technique for supporting the user's quick detection and discrimination of data items or groupings of data displayed on a screen. Other visual-coding techniques include variations in brightness, flashing, line style, symbols, sizes, and shapes. It is important not to misuse or overuse these coding techniques. A basic rule of thumb is to follow is to use visual coding for functional purposes only.

The next section will discuss the CTC tool and the reasons for developing the chosen user interface.

4.3 Developing the CTC Tool

As outlined in Chapter Three, it was decided to develop a tool that would permit quick and easy access to information about computing in universities in Europe. Some form of data collation would be necessary to gather all the necessary information together so that the tool could be developed. The source of this information lies in various textual publications, for example 'The World of Learning' [EUR-95], a large tome, published each October, which gives details of universities world wide, other government and European publications and also online sources such as the World Wide Web (WWW). As this will be referred to frequently it is worth giving a brief description of the WWW at this point.

4.3.1 World Wide Web

The WWW [WWW-95] was first conceived at the European Laboratory for Particle Physics (CERN) in 1989 and first implemented in late 1990. The official definition given by CERN is

‘A wide area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents’ .

The WWW uses hypertext and hypermedia as its method to interact with users. Hypertext is similar to ordinary text, but it contains connections within the text to other documents¹¹. Hypermedia is hypertext with a difference - it contains links not only to pieces of text but also to other forms of media - sound, images, movies. As described in the previous chapter, these Web pages are created using HTML. The connections, called *links*, to other pages are created in HTML using what is known as a ‘Uniform Resource Locator’ (URL). Basically this is the address of a specific page/resource on the Internet. The URL consists of the *protocol identifier* (see <http>

¹¹ Web documents are often referred to as **pages**

below), a *domain name* - the exact Internet address of the computer that the page resides on and a *resource location* - the directory and filename for the HTML file. For example the URL or address for the Geographical Interface Tool developed as part of this research is <http://www.compapp.dcu.ie/~cryan/intro.html>. The domain name in this example is [compapp.dcu.ie](http://www.compapp.dcu.ie/~cryan/intro.html) and the resource location, namely the directory where the HTML file, [intro.html](http://www.compapp.dcu.ie/~cryan/intro.html), is stored is [~cryan](http://www.compapp.dcu.ie/~cryan/intro.html)¹². To create a link to a specific URL, the *anchor* tag, `<A>`, is used. For example to create a link to the Introduction Page, 'intro.html', the following code would be used :

```
<A HREF="http://www.compapp.dcu.ie/~cryan/intro.html">Introduction  
Page</A>
```

In order to read WWW documents, a connection to the Internet is needed plus your computer must have a program called a browser. Current browsers include Netscape Navigator [NET-95] and NCSA's Mosaic [NCS-95], Microsoft's Internet Explorer [MS-96] . A browser acts as an interpreter, converting HTML into meaningful text and graphics. The browser will also allow one to view the HTML source code used in a Web page. Using the 'View' option on the menu bar, and choosing the 'Document Source' will bring up a separate window containing the code.

The WWW has been established on the client-server paradigm. The client is the browser software on a computer which interfaces with the server computer and requests documents the user wants to see. The server shows or 'serves' these documents to the browser when requested to do so. The communication between browser and server is based on the Hypertext Transfer Protocol (http). This protocol which is still undergoing development allows object-oriented hypertext and multimedia operations to be performed over great distances at high speeds.

Initially Internet access was confined to government and academic institutions, but today anyone with a computer and modem can connect anytime, anywhere and thereby gain access to the information on the WWW. Due to the millions of people who do exactly that the amount of pages now on the WWW is very large and growing very rapidly, largely due to the ease with which the Web technology allows people to present, organise and display information on the Internet.

¹² The tilde (~) indicates that this is owned by cryan and is the directory which contains all the HTML files that cryan wishes to be publicly accessible

As mentioned earlier, almost every third level college in Europe has Internet access. The vast majority of these institutions have pages on the Web with information about the courses on offer, student numbers, facilities, research centres and more. As this is the case, it is clearly evident, that the Web will be an invaluable source of information when developing the CTC tool.

4.3.2 Geographical Web Tool

Given that the information on computing at universities across Europe is textual in nature, either published in paper form or available online, it was decided to try and create a tool that would logically group all the data and represent it in a different format. A tool that would display the data dynamically would be more satisfying to use and easy to follow and at the same time provide quick responses to user requirements. As outlined earlier, it is wise to present data in a simple and useful format, avoiding clutter using a format and terminology that is familiar to the users. In addition, it is advised to reduce the short term memory load on the users. To this end, it was decided to represent the data visually using a geographical context. Specifically, the data would be available via a map of Europe. The graphical representation would be much easier to retain than a textual one. Most people are familiar with reading through an atlas where countries are grouped together into one big map (for example, Africa, Asia, Europe) and more detailed maps of each country are also included. This map would allow users to access information on computing in third level institutions in a similar fashion. The map could be manipulated by the user with a simple pointing or clicking action, bringing them to a further more detailed map of the chosen country. Again by pointing, clicking this map could be manipulated to bring up the actual information on computing in the various colleges for that particular city, town, region in that country.

Having determined the functionality, it is clear that the best interactive style to represent the 'map of Europe' tool is the method which allows objects to be directly manipulated. The alternatives open to us at this stage were to build a new custom designed interface or use an existing interface. For reasons which will be discussed

below, the second option was taken and we decided to build use the World Wide Web.

4.3.2.1 Reasons for using the World Wide Web

- **Functionality:**

All Web browsers are already capable of displaying images easily and provides a context sensitive point and click interface which is ideally suitable for the application we want for our tool. Time and effort can be saved instead of having to develop some X-windows application for example which would have to be custom written to allow the same functionality just described.

- **Compatibility:**

The language used to publish information on the Web is universal, it is supported on the vast majority of operating systems such as Solaris, NT, and Macintosh. Writing some new specially designed interface would involve producing a version for the various operating systems. Omitting to produce a version for one particular operating system or platform would instantly rule out some people who might be among the intended users.

- **Reliability:**

As outlined earlier, virtually all academic organisations and indeed commercial, government and other institutions have access to the Internet and have Web software available on their local networks. The user does not have to take an active in maintaining this software. Keeping the CTC Web tool up to date would be relatively simple and need only be done once. If a custom designed interface had been used, updates would have to be distributed to users regularly.

- **Schedule and budget:**

It would have taken considerably longer and more expensive in terms of effort and money had a Web interface not been chosen. No distribution costs were incurred as no installation disks had to be distributed to the users.

- **Easy to learn:**

Users will not have to be given significant training to be capable of using the tool on the Web. Even people who are not highly computer literate are already familiar with using Web browsers. The intended users, academics in the field of computing, students, people in the computing industry and European Union officials, cover the three categories of user described in section 4.2.2. The Web is so easy to use that the people falling into the novice category would find it extremely easy to learn how to operate this tool and while not being as quick as the more experienced user, would soon be navigating around computing faculty Web sites in Europe comfortably.

- **Consistency:**

The actions involved in navigating through the map are consistent as is it a simple point and click interface the whole way through.

- **Shortcuts and Flexibility:**

As can be seen in Figure 5-2, the frame gives users a shortcut back to either the map of Europe or a specific country, avoiding the need to navigate back from a particular university Web site which has been just been visited. The in-built features of Web browsers also allows users to return home to a some predefined page of their choosing.

The Web solution is a write once, run anywhere solution. It is a logical extension to use the Web as much of the information we are trying to provide is already available via the Web. The next chapter goes on to describe how the geographical based tool for the Web was designed and implemented using the Hypertext Markup Language.

Chapter Five - CTC World Wide Web Tool

Before going on to describe the development of the tool, it is worth giving details of the programming language used to create the pages for the World Wide Web and the CTC Web tool.

5. HTML

World Wide Web documents¹³ are created using Hypertext Markup Language (HTML), which specifies to the Web software how to display a document and create the links to other sites on the Internet. Document markup is the process of adding codes to a document to identify the structure of a document or the format in which it is to appear. The need for longevity of markup and the need for marking documents for electronic database storage led to the development of the international Standard Generalised Markup Language (SGML) in the mid eighties [WAT-92]. SGML is used to define HTML which is used to format the documents which we see on the Web everyday.

HTML defines the *parts* of a document, such as titles, headings and body text. These parts are known as *elements*. To define these elements, *tags* are used. Tags take the form of angle brackets surrounding letters, numbers and other information, for example ****. Tags usually come in pairs, with one tag at the beginning of text which is to be formatted in a particular way and another at the end. For example the expression “**The early bird catches the worm**” would make those words appear in boldface in a Web browser. Using these HTML tags it is possible to enable user feedback. The **<FORM>** tag allows you to add interactive features to documents. The following line is taken from the Web Questionnaire - it tells the browser that when a user clicks the *submit* button, the information input by the user should be sent to a CGI script, which is defined in the HTML code and stored on the

¹³ Web documents are often referred to as **pages** or **resources**

server computer (see Section 4.3.1 for more information on the browser/client relationship).

```
<FORM METHOD="POST"
```

```
ACTION="http://www.compapp.dcu.ie/cgi-bin/carmel/quest.pl">
```

The script, in our case, a PERL file then processes the incoming information and performs the requested action (see Appendix B for the full source code). This script is stored in a special directory which provides communication between external sources of information and the local server. This special directory contains programs which specify the standard (CGI - Common Gateway Interface) for running external programs (see APPENDIX B for more information on HTML).

5.1 Information Sources

The World of Learning [EUR-95] was used to come up with an initial list of third level institutions offering computing courses. The World Wide Web was very useful for checking that these colleges had 'Computing' Web sites. Additional searches were conducted using keywords such as 'universities in Europe' and 'computing courses' using the search engines Altavista and Yahoo¹⁴. Two very useful sites were also used in an effort to obtain a complete list of college 'Computing' Web sites. One is maintained by a UK Publishing company known as Datalake [WES-96] and the other by the University of Birmingham [UBH-96] and both contain links to university Web sites.

5.2 Designing the Web site

One of the first things to be done when setting about designing a Web site, is to check that it is not already on the Web [PFA-96]. When searching for a complete list of college Web sites containing information on computing courses in Europe no sites were found that give or point to such information. Some sites were found with links

¹⁴ The search engines - Altavista and Yahoo were used - URL's: <http://www.altavista.digital.com> & <http://www.yahoo.com>

to university web sites, but either these did not have direct links to computing faculty/school/department pages or were not exhaustive. For example, both the sites mentioned in the previous section, while useful did not contain the links we were looking for. The University of Birmingham site included links to British university home pages only. The other site had a more extensive set of links to universities world-wide, but no link directly to computing information.

Another issue to be considered in the design stage of a Web site, is the intended audience. This Web site is first and foremost intended to provide information for CTC eSIG members, who are mainly computing lecturers, but we also expect that the following audience would benefit from having such easy access to information about computing at third level in Europe -:

- Students
- People in the Computing Industry
- European Union officials
- Non-academic staff in computing faculties throughout colleges in Europe

Students who may wish to study away from their home country can find out information about other university computing courses that are on offer around Europe. Contacts can be made with students studying similar courses in different universities in their home country or indeed further afield.

Computing Industry professionals can use this tool to find information about what research is being undertaken in computing faculties and if interested in becoming involved or initiating research projects, companies can use this tool to find contact information. In addition, industry people can get computing course information and ascertain the type of graduate which will be entering the workforce. It is a long term goal of the CTC that there would be more input from industry into curriculum subjects. Industry people know what kind of people they are and will be looking for and can make a significant contribution towards improving the availability of high quality computer personnel to service the needs of the rapidly changing computing industry in Europe.

European Union officials can make use of this tool also. When looking for information on what universities in Europe actually offer computing courses, it was discovered that there is no list available with such information [EUR-95] . The process of finding these universities will be outlined later in this chapter. This site can now be used to collate such information.

There are other practical issues to be considered when designing a Web site - the page layout, size and amount of inline images and how much information to put on a single page. Practically all publications which deal with designing Web sites advocate that a Web page layout should be attractive and easy to read [SIE-96] [PFA-96]. Inline images should be limited and if possible kept small. Long pages are not in vogue as the reader may get overwhelmed and they usually involve excessive scrolling. Organisation and intelligent linking also play an important role in the look and feel of a site. The next section deals with the development of the tool and will detail how these matters were addressed when establishing the site.

5.3 Development of the Tool

When developing the Web site, we wanted to incorporate the features outlined above. We knew that there would be a large number of universities offering computing courses across Europe. If these were to be presented in a linear fashion, we would have very long pages which, as mentioned earlier, would involve excessive scrolling and a long list of text would not be particularly attractive to users. It would have been the easiest option, once the full list of colleges was collated, to list all the countries and then present a list of all those universities teaching computing in those countries. The Datalake site, mentioned in section 5.1, takes this approach and provides a long list of countries and universities. By steering away from this presentation format, a more interesting site could be developed. By using an image of the map of Europe instead of a list of European countries and country maps with the links to the various colleges, instead of a long list of universities, not only would the user find it less boring and more attractive to read, but using the maps gives those who

don't know their geography too well, information on exactly where in a particular country a university is situated.

Images with links are known as *imagemaps*. Section 4.2.1 describes *imagemaps* in more detail. Using these *imagemaps* would allow us to subdivide the information by country and by region and present it in a clear concise manner.

However, Web site designers are specifically advised that inline images should be limited and kept small. The reason for doing so is to prevent prolonged periods of, waiting for a page to appear on the screen due to an image being downloaded. In addition, images can clutter up a page unnecessarily making text hard to read. This was a major concern as we intended to base our site on country maps. When creating the maps, it was decided to try and keep them as simple as possible and create them using the same format - size and colouring. The first map developed was the map of Europe (see section 4.2.1 for more details). When completed, the size of the image was 14Kilobytes. It was included in a test page and only took a few seconds to download. In addition, a favourable feature of Web browsers, is that once an image has been downloaded by a user, the next time that page with the inline image is requested, download time is much quicker. This occurs because the image is stored in the computer's cache. The Cache is used to keep copies of frequently accessed Web pages, which reduces the time connected to the network¹⁵.

The test image used was not interactive, links had to be added so that users could actually navigate around Europe. The following section gives more information on these interactive maps, known as *imagemaps*.

5.3.1 Imagemaps

An *imagemap* is a means of graphically representing the links on a Web page. It is a standard graphic file (gif) with hotspots defined in a map file on the server. It is a "clickable" image with links (hotspots) that function as normal HTML links.

There are three steps involved in creating an *imagemap* [NCS-95]:

¹⁵ The Reload button in a Web browser, will always compare the cached page to the network page and show the most recent one. The Cache usually resides in the same directory as the browser and its associated files

1. Design the image (picture file) to be used as the imagemap.
2. Create a data file containing the "hot spot" co-ordinates on the image and the URL associated with them.
3. Write the HTML code that refers to the image and its associated datafile.

5.3.1.1 The Picture File

This can be any GIF formatted picture that you want to use for an imagemap. Pictures can be scanned in and saved with scanning software such as Ofoto and Photoshop. Graphics and paint programs may also be used to create pictures (images) and save them in **.GIF** format. The images for this project were created using downloaded images from various sources on the Web. These images were then edited extensively using PaintShopPro and Paintbrush - two paint programs widely available for Personal Computers.

5.3.1.2 The Map File

Once an image is ready to be turned into an imagemap, the hotspots must be defined. These are defined in a map file on the server as pixel co-ordinate bounded shapes or regions. This map file could be created using any text-word processor, but it would be necessary to accurately know the pixel co-ordinates of every point in an image or count them. This procedure can be extremely time consuming, but there are imagemap file creators which simplify this task. There are many of these programs available on the Internet. Mapedit [BOU-96] was the program employed to create the various imagemaps needed for our tool. Mapedit and other imagemap programs such as MAPTHIS [MAP-96] and MapMarker [MUR-96], allow the image file to be opened, hotspot shapes to be drawn and links to other URL's to be made. By saving the file, a text file is actually created with the proper dimensions for the hotspot followed by the link. This is very convenient as it can be edited directly if a link but not the hotspot needs to be changed.

So the map file is basically a text file that contains the co-ordinate information on where to go when different parts of the picture are clicked with the mouse. Each line of text in the **.map** file defines a shape (circle, rectangle, or polygon) and gives the destination URL for a click within that shape. The map file defines clickable spots on the image by identifying locations on the grid with x,y co-ordinates.

The format for each line of the file is:

shape where-to-go.html co-ordinates

- shape** This gives the type of shape that will be defined. The valid values are circ, rect, point, or poly.
- where-to-go.html** This gives the URL for the destination site, document, image, video, audio, or text file to connect to when a user clicks in the section defined by this shape.
- co-ordinates** This defines the co-ordinates of the rectangle, circle, point, or polygon. The co-ordinate system is defined in pixels and begins at the upper left corner and ends at the lower right corner.

In very non-technical terms the imagemap has a grid of x,y co-ordinates that are associated with each hotspot. When the hotspots are clicked, a map file of X,Y co-ordinates is queried to find the associated URL. The connection is then made to that Web page. In other words, the map file tells where the click should take the user.

The general map file has the following format:

default URL	#anything after a # is a comment
circle x y r URL	#circle -given by centre and a radius
rectangle x0 y0 x1 y1 URL	#square -two points making a diagonal
polygon x0 y0 x1 y1 x2 y2.... URL	#polygon -series of points
point x y URL	#a point

Default: This line of the file tells where to go if the place where the user clicked was not defined as a section of the map.

The following lines are taken from the imagemap file of Europe used in the WWW Geographical tool - the full file can be read in Appendix C.

```
default http://www.compapp.dcu.ie/~cryan/error.html
rect http://www.compapp.dcu.ie/~cryan/iceland.html 4,66 106,115
poly http://www.compapp.dcu.ie/~cryan/irl.html 108,216 62,251
62,251 91,281 91,281 135,268 135,268 144,247 144,247 119,212
119,212 107,215
```

The default URL used in all our imagemaps is **error.html** which contains a simple message telling users the wrong area has been clicked and to go back to the map and try again.

The second line in English is "A mouseclick anywhere between 4,666 (x,y of the upper-left corner of the rectangle, in pixels) and 106,115 (lower-right corner) will take you to the URL <http://www.compapp.dcu.ie/~cryan/iceland.html>".

In creating the Europe imagemap, the polygon shape was the most widely used as it allowed the exact shape (more or less) of each country to be outlined, however, in the individual country imagemaps, the rectangle shape was also used for creating the links to the universities and colleges in the various cities and towns in each country.

5.3.1.3 Writing the HTML Code

The web page from which the viewer will see and click on the imagemap needs to contain a link to the location where the imagemap is stored and then a link to the actual image. The following code appears in the web page displaying the map of Europe (see Appendix C for the full source code).

```
<A HREF=
"http://simpr1.compapp.dcu.ie/cgi-bin/imagemap/europe">
<IMG SRC="http://www.compapp.dcu.ie/~cryan/europe.gif" ISMAP></A>
```

The first line calls a CGI script to perform the image mapping function (see Appendix C for the source code of this script - `imagemap.conf`). The ISMAP part of the IMG tag proclaims this picture to be an `imagemap`.

5.3.1.4 Making the Imagemap Work

In order for the `imagemap` to work, the server must know where to find it. In the server's directory structure there should be a `cgi-bin` directory and within it a file called `imagemap.conf`. This is the file which contains information on all `imagemaps` served by a particular server. It contains one line for each `imagemap` served by the server. When adding a new `imagemap`, a line similar to the following example can be used:

```
scotland :/users/capg/cryan/public_html/scotland.map
```

See Appendix C for the full `imagemap.conf` file which references the `imagemaps` used in our WWW tool.

5.3.2 Usage of Frames

Continuing our efforts to present the information in a non-linear and more visually appealing manner, it was decided to use HTML *frames* in the Web site. Frames [NET-96] were first introduced by Netscape as an added feature to their Web Browser - Netscape Navigator. Frames are graphical and logical subdivisions of a single Web page into two or more sections and allow a site to be more 'browseable' [BIG-96]. In addition, as visitors to the CTC site would be navigating through the various countries of Europe, it was decided that it would be useful to simply have an index of these countries readily available avoiding the necessity to go back through all the links visited to get back to the main map of Europe. The CTC Web site was divided into two frames with the first frame being the site navigation index and the other frame holding the selected pages. Figure 5-1 shows how the frames appear in the CTC Home Page Web site.

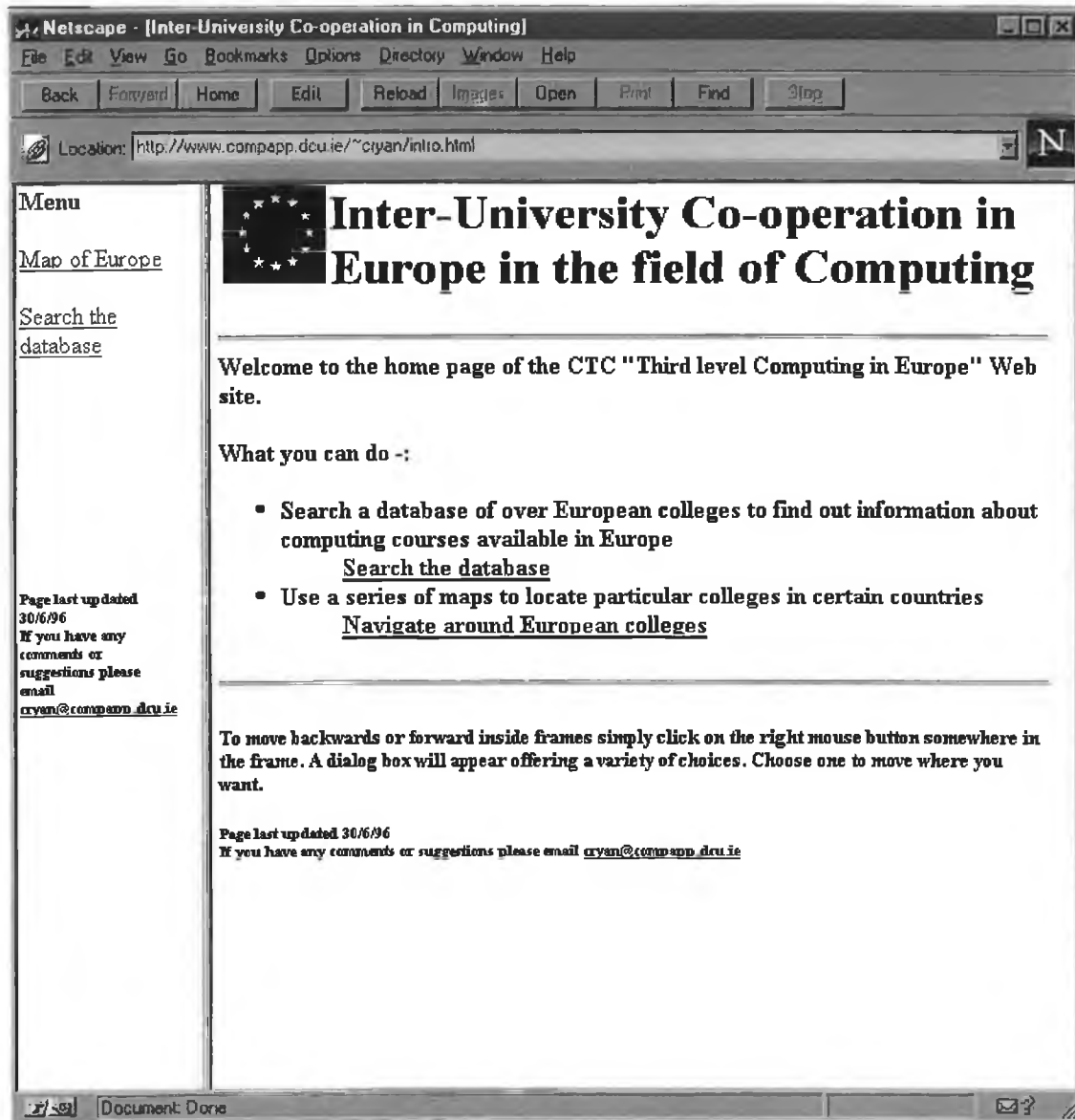
The page is divided into 2 frames, each having its own unique URL and scroll bars. Employing the frames in the site gives visitors two ways of accessing links, either by using the side frame links or by using the imagemaps which appear in the main frame.

To make frames the `<FRAMESET>` tag is used in the HTML code. The width and name of each frame is included in the `<FRAMESET>` tag, as is the URL of each page to be displayed in the frames. If the URL's are not included the frame would appear blank or empty. The following HTML source code is taken from the page seen in Figure 5-1 and shows how the frames are defined.

```
<html>
<head><title>Inter-University Co-operation in Computing</title></head>
<frameset cols="130,*">
<frame src=side.html name="side" scrolling=auto>
<frame src=euhome.html name="display" scrolling=auto>
</frameset>
</html>
```

The `'cols="130,*'` indicates that the first frame is to be 130 pixels wide and the second is to take up the remaining space. In this page the first frame is given the name `'side'` and the main frame called `'display'`. The URL's corresponding to both frames are `'side.html'` and `'euhome.html'` respectively. The `'auto scrolling'` means that the browser will determine whether or not scrollbars are necessary in a frame and shows or hides them as appropriate

Figure 5-1 - Home Page of Geographical Interface Tool



In order to make a page appear in a particular frame, the link must be altered. It is the 'TARGET' attribute in the *anchor* tag (described in section 4.3.1) which allows a name to be assigned to a specific window and pages can then be targeted to appear in the window bearing the matching name. Normally the *anchor* tag, `<A>`, specifies a link to be loaded when an active item is clicked. Adding in the TARGET

attribute forces a page to be loaded into the targeted window. The format for the anchor tag would then be as follows :

```
<A HREF="URL" TARGET="window-name"> Targeted Anchor </A>
```

To illustrate how this TARGET tag was used in the CTC pages, the following code is taken from the Display frame of the Home Page and allows the page containing the interactive map of Europe to be displayed in the Display frame :

```
<A HREF="eurmap.html" TARGET=display>Navigate around  
European colleges</A>16
```

See Appendix C for the full listing of the HTML source code.

5.3.3 Search Facility

In addition to having the interactive maps for finding information on third level computing in Europe, a search facility was also included in the site. All the Web pages from college Web sites that are accessible via the imagemaps are stored in a database locally. The search page employs the <FORM> tag (described in section 5) to accept a text query from a user. A CGI-script file is then called to process the query. The search engine used was developed as part of a Ph.D. project here in DCU. It employs vector space modelling techniques(see Section 2.2.3.2) and uses a *tf*idf* weighting scheme (see Section 2.2.1.3) to return a list of ranked documents which match the user query. The Search page and its HTML code for this page plus the PERL code are included in Appendix C. Two sample query pages are also included.

¹⁶ There is no need to write the full "http://www....." path as the file 'eurmap.html' is stored in the same directory as the current Web page

5.3.4 Using the WWW Tool

Having described how the imagemaps were developed and how frames were used, the following pages show how the tool would work for a user upon selecting the option to Navigate around European colleges from the Home Page depicted in Figure 5-1. The user will initially be presented with a list of countries and a map of Europe (Figure 5-2). The sample figures following Figure 4-3 and Figure 5-4) show the pages a user would be presented with if the country France and then Compiegne were chosen.

Figure 5-2 - Map of Europe Web Page

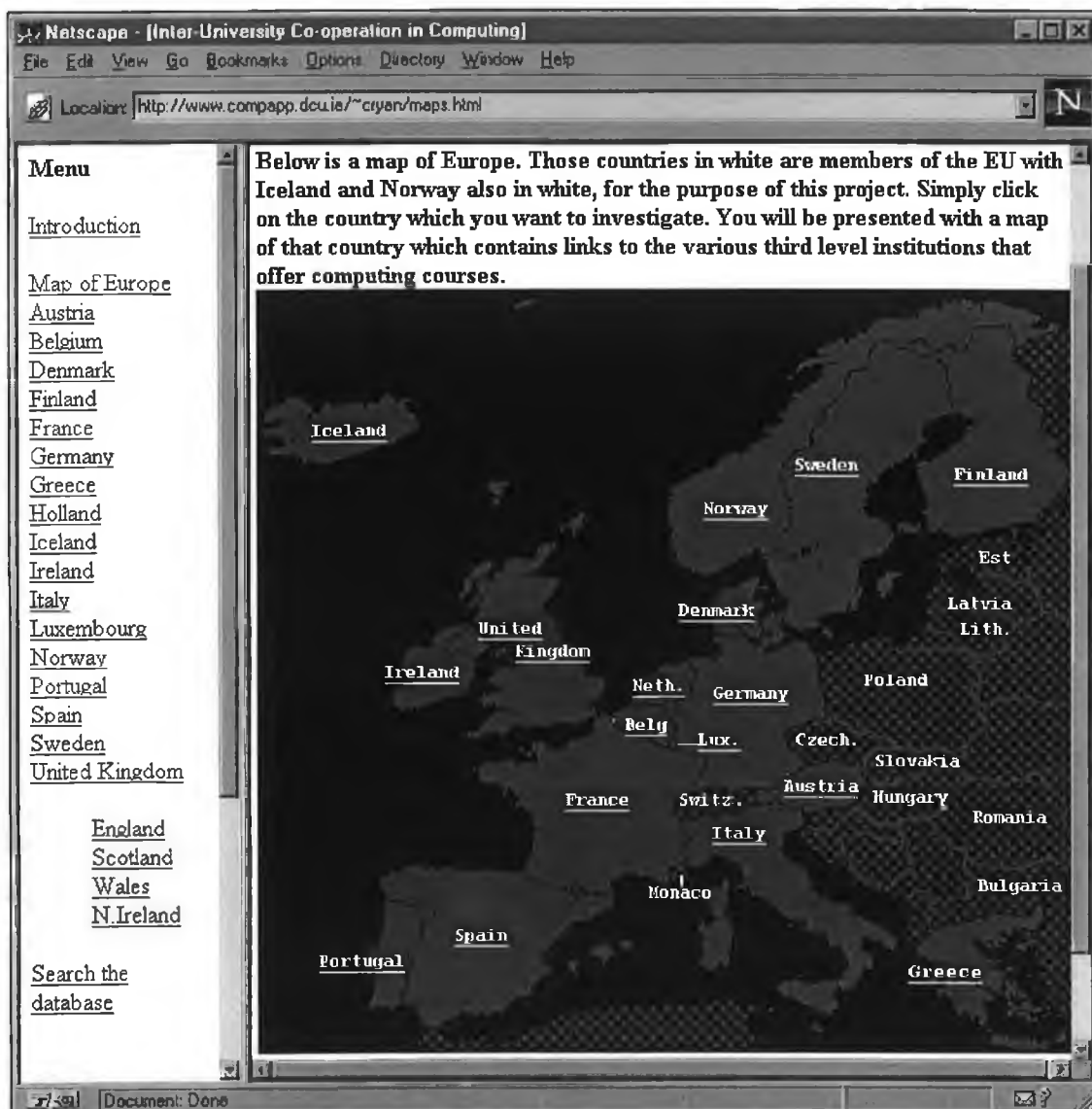


Figure 5-3 - Map of France Page

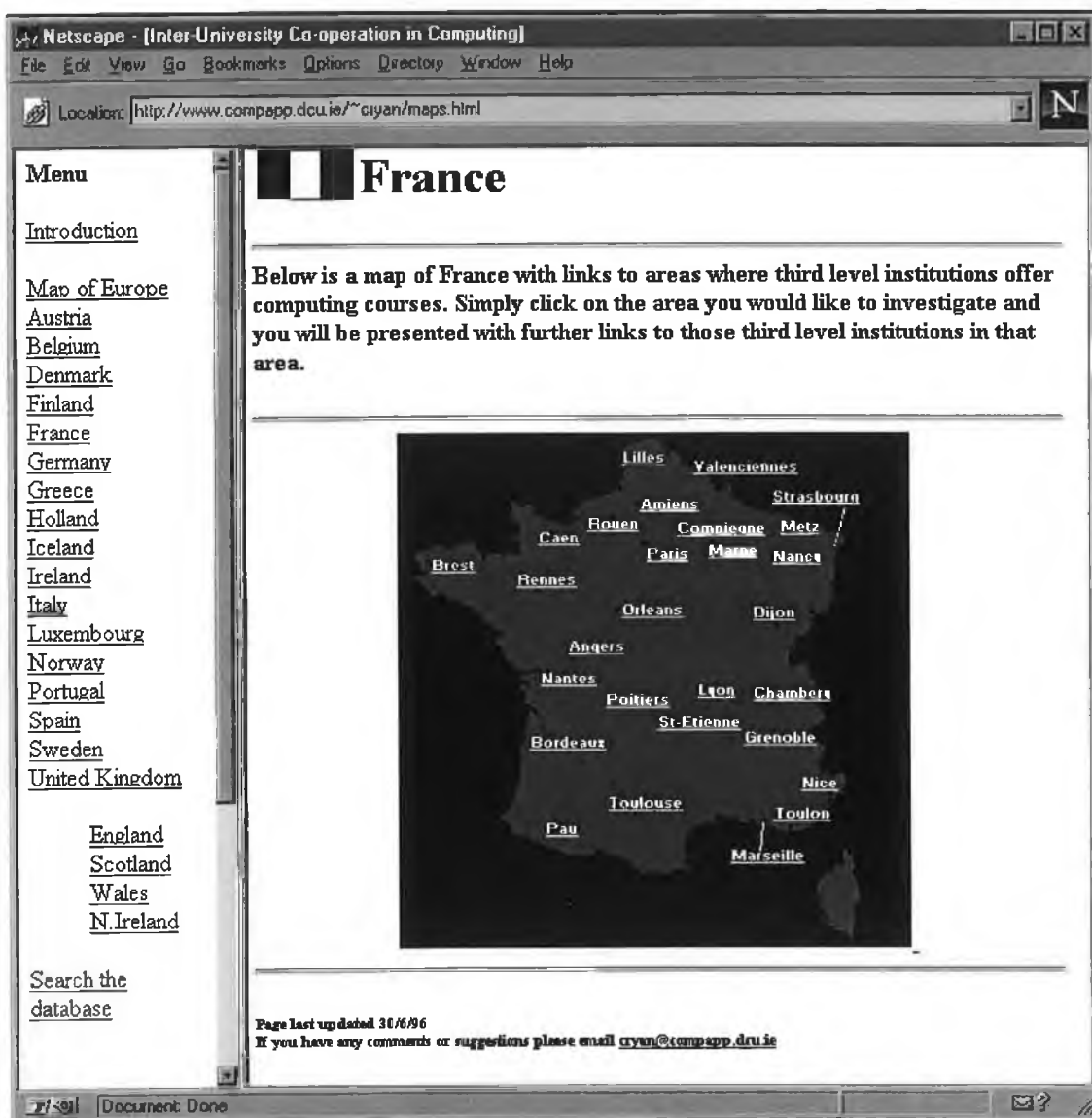
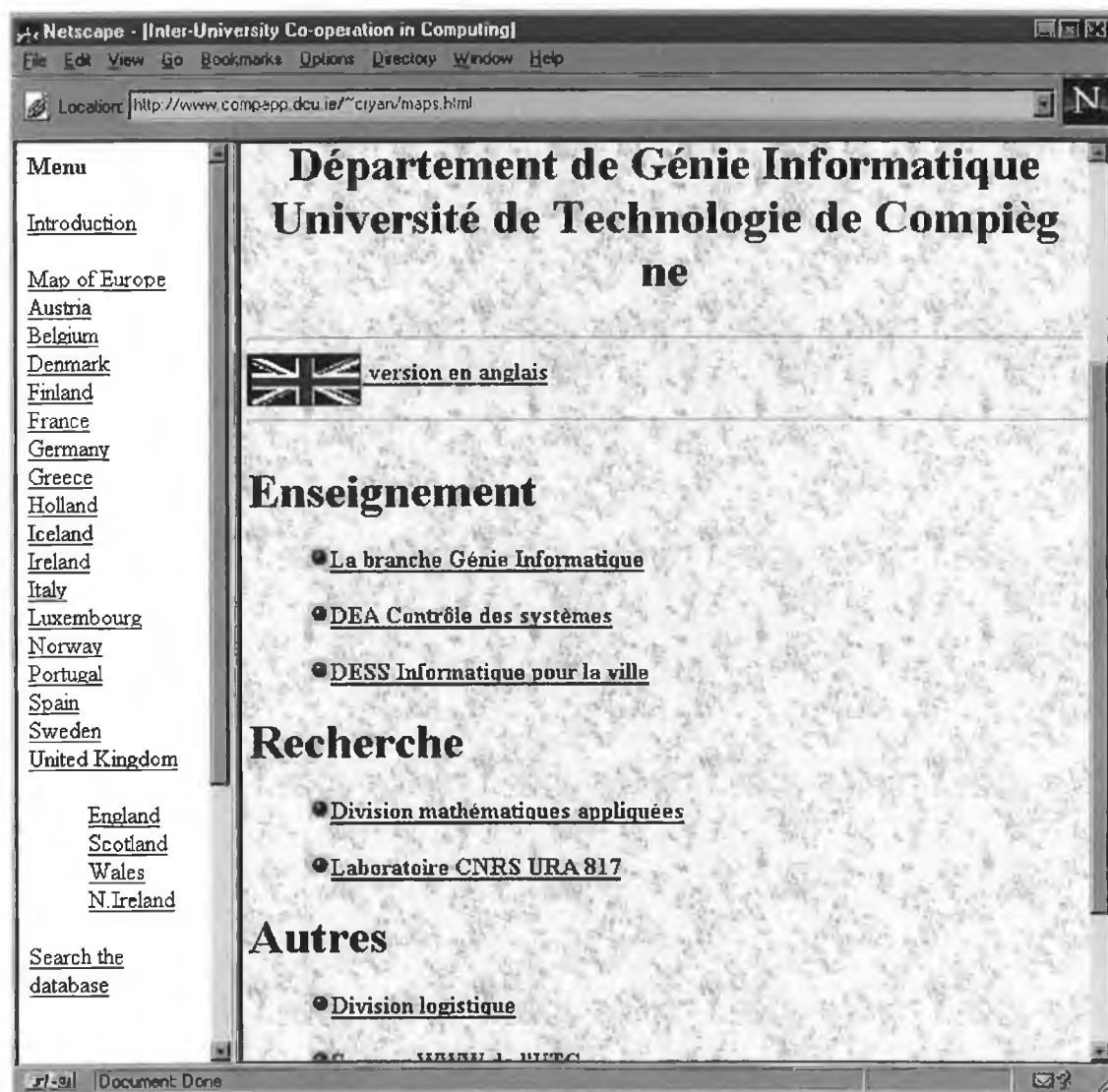


Figure 5-4 - University of Compiègne (Computing Dept)



The CTC Geographical Interface Tool was launched at the Evaluation Conference on Inter-University Co-operation in Europe in the Field of Computing. A successful 'live' demonstration was held for the conference delegates which involved visiting a few computing sites (including the example shown above) from European universities. It is currently under evaluation and is accessible through the Centre for Teaching Computing Web site in DCU.

Chapter Six - Conclusions

6. Introduction

As outlined in the introduction, the previous chapters have described how and why the CTC eSIGs were established. The research undertaken focused on finding a method to monitor the eSIGs to ensure they were being used for their intended purpose. To this end, Information Filtering techniques were investigated and found suitable for this application.

6.1 Conclusions

The primary conclusion of this research is that the eSIGs are not yet delivering on their potential. As they were not being used to the extent we expected, further investment of time and effort in developing the CTC filter was put on hold. The situation was disappointing as there had been great enthusiasm for their establishment at the CTC annual conference. In addition, the number of people who joined within a very short time period was very encouraging. It was anticipated that it would not be long before a skeleton database of recommended syllabi would be drawn up as a result of communication between all these members. This, however, has not happened yet.

In our opinion the reason for this was that there were no support measures put in place at the time of establishment to get the eSIGs up and running and it was felt that until such measures were implemented, the eSIG situation would continue in the same way.

Since its establishment the CTC has been involved in projects funded by the European Union. At this time, the CTC became involved in one such European project with similar aims. This particular project had as its primary aim the promotion of co-operation between third level institutions in the field of computing in Europe. As a result of involvement in this project and the need to get the eSIGs up and running a graphical Web-based tool was developed. The tool was designed to assist in the finding of information quickly and easily about computing at third level.

An additional support measure was also put in place to assist in improving the activity on the eSIGs. Each eSIG was assigned a 'mentor' who would endeavour to initiate communication on the particular eSIG for which they have responsibility. The mentors were appointed by the CTC board based on their expertise in the area and their keen interest in developing a database of recommended syllabi in their area. The specific methods to be used by each mentor were not discussed in fine detail, but would proceed along the lines of inviting members to become involved in an important curricular issue pertinent to that area of computing. For example on the Programming Language eSIG - the issue of which programming language to teach first year Computing students might be raised by the mentor.

6.2 Future Work

As the traffic on the eSIGs increases, work will proceed once again on developing the actual filter to assist the CTC in managing the groups and evaluating their success.

Additional eSIGs may be established as each area originally identified, can be broken down even further. For example, the Programming Languages eSIG could be sub-categorised into Object-Oriented Languages, 4GLs, and 3GLs.

The Web tool will be updated regularly to include the following:

- New pages being added to sites (currently accessible)
- New Web sites being created by universities (currently not accessible)
- New countries joining the EU (or other countries as specified by CTC)

There are new products now available to create Web sites and imagemaps. These products will be investigated and if suitable will be used to carry out the above steps involved in maintaining the CTC Web Tool up to date.

Development of the Search engine is still continuing here in DCU and as it is expanded, the new versions will be incorporated into the Web Tool.

6.3 Summary

It is our opinion that the eSIGs are a good idea and will result in the establishment of a European database of recommended syllabi. The support measures have now been put in place and we feel that these will help to get members communicating on the eSIGs.

BIBLIOGRAPHY

- [BAI-82] "Human Performance Engineering: A Guide for System Designers". Bailey, R.W., Prentice Hall, 1982
- [BEL-92] "Information Filtering and Retrieval : Two sides of the Same Coin", Belkin, N.J., Croft, W.B., *Communications of the ACM*, Vol. 35, No.12, December 1992
- [BIG-96] "HTML Authoring and WWW Programming", Biggs, T. D., Harris, A., World Wide Web document URL:
<http://klingon.cs.iupui.edu/~aharris/220html/ht7.html>
- [BOU-96] Boutell.Com, Inc., World Wide Web document
URL: <http://www.boutell.com/mapedit/>, Mapedit Information Page
- [BRE -82] "Do-It-Yourself Marketing Research", Breen G.E., Blankenship, A.B., McGrawHill, 1982
- [CHI -82] "Marketing Research Analysis and Measurement", Chisnall, P. M., McGraw-Hill, 1982
- [CTC-94] 'Prospectus', Centre for Teaching Computing, November 1994
- [DEN-82] "Electronic Junk", Denning, P., *Communications of the ACM*, 163-165, March, 1982
- [EUR-95] "The World of Learning 1996", Europa Publications Ltd., 1995
- [EVE-80] "Cluster Analysis", Everitt, B., Halstad Press, 1980
- [FOL-90] "Using Latent Semantic Indexing for Information Filtering", Foltz, P., *Proceedings of the Conference on Office Information Systems*, Cambridge, 1990
- [GOL-92] "Using Collaborative Filtering to Weave an Information Tapestry", Goldberg, G., Nichols, D., Oki, B.M., Douglas, T, *Communications of the ACM*, Vol. 35, No. 12, December 1992
- [GUL-95] GulfNet Professional Internet Service, World Wide Web document
URL: <http://www.tgn.net/whatsit>
- [KRO-94] "The Whole Internet", Krol, E., O'Reilly & Associates, 1994
- [KRO-94] "The Whole Internet", Krol, E., O'Reilly & Associates, 1994

- [LEW-92] "An evaluation of phrasal and clustered representations on a text categorisation task", Lewis, D. D., In *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1992
- [LOC-92] "Principles of Marketing Research", Lockhart, D.C., Russo, J.R., Chapter 4 'Mail and Telephone Surveys in Marketing Research: ', Blackwell Business, 1980
- [LUH-58] "The Automatic Creation of Literature Abstracts", Luhn, H.P., *IBM Journal of Research and Development*, Massachusetts, 1949
- [LUP-83] 'Management and the Social Sciences', Lupton, T., Penguin, 1983
- [MAL-87] "Intelligent Information Sharing Systems", Malone, T.W., Grant, K.R., Turbak, F.A. et al, *Communications of the ACM*, Vol. 30, No. 5, May 1987
- [MAL-95] "Pointing the way: Active Collaborative Filtering", Maltz, D., Ehrlich, K., *Proceedings of Conference on Human Factors in Computing Systems*, May 1995
- [MAP-96] MAPTHIS, World Wide Web document
URL: <http://galadriel.ecaetc.ohio-state.edu/mt/>
- [MCE-93] "An Emergent Approach to Information Filtering", McElligott, M.S., Sorensen, H., *UCC Computer Science Journal*, Vol. 1, Issue 4, December 1993
- [MS-96] Microsoft Corporation, World Wide Web document
URL: <http://www.microsoft.com/ie/ie.htm>, Microsoft Internet Explorer Page
- [MUR-96] Murray-Rust, P., World Wide Web document,
URL: <http://gserv.dl.ac.uk/CBMT/mapmarker/HOME.html>,
MapMarker Home Page
- [NCS-95] National Centre for Super Computing Applications, WWW documents
URL's:
<http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/mosaic-docs.html>
<http://hoohoo.ncsa.uiuc.edu/docs/tutorials/imagmapping.html>
<http://www.ncsa.uicu.edu/General/Internet/WWW/HTMLPrimer.html>
- [NEL-80] "Interactive Systems and the Design of Virtuality", Nelson, T., 'Creative Computing', Nov/Dec 1980
- [NET-95] Netscape Communications Corporation, World Wide Web document
URL: <http://www.netscape.com>, Netscape Home Page

- [NET-96] Netscape Communications Corporation, World Wide Web document
URL: http://home.netscape.com/comprod/products/navigator/version_3.0/layout/frames/index.html, Frame Page
- [PFA-96] "Publish it on the Web", Pfaffenberger, B., AP Professional, 1996
- [RES-94] "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", Resnick, P. et al, *Proceedings of ACM 1994 Conference on Computer Supported Co-operative Work*, 175-186, 1994
- [ROB-77] "The probability ranking principle in IR", Robertson, S.E., *Journal of Documentation*, 33, 294-304, 1977
- [ROW-78] "The Dissemination of Information", Rowley, J.E., Turner, C.M.D., Andre Deutsch, 1978
- [ROW-91] "An Emergent Computation Approach to Natural Language Processing", Rowe, J., McKevitt, P., *Paper presented at AICS Conference*, Cork, September 1991
- [RUB-84] "The Human Factor: Designing Computer Systems for People", Rubinstein, R., Hersh, H., Ledgard, F., Digital Press, 1984
- [SAL-83] "Introduction to Modern Information Retrieval", Salton, G., McGill M., McGraw-Hill, 1983
- [SAL-89] "Automatic Text Processing", Salton, G., Addison Wesley, 1989
- [SCH-77] "Scripts, Plans, Goals and Understanding", Schank, R.C., Abelson, H., Lawrence Erlbaum, 1977
- [SHN-82] "The Future of Interactive Systems and the Emergence of Direct Manipulation", Shneiderman, B., *"Behaviour and Information Technology"*, Nov. 1978 p 1005-1008
- [SHN-87] "Designing the User Interface: Strategies for Effective Human-Computer Interaction", Shneiderman, B., Addison Wesley, 1987
- [SIE-96] "Tips for Designers and Writers", Siegel, D, World Wide document,
URL: <http://www.dsiegel.com/tips/>
- [SME-96] "Filtering News and WWW Pages using the BORGES Information Filtering Tool", Smeaton, A. F., *Proceedings of the Workshop on Practical Applications of Information Filtering*, October 1996
- [SUN-91] *Proceedings of the Third Message Understanding Evaluation Conference*, Sundheim, B., 1991
- [TER-93] "A Tour through Tapestry", Terry, D., *Technical Report*, Computer Science Laboratory, Xerox Palo Alto Research Centre, CA, 1003

- [THI-90] "User Interface Design", Thimbleby H., ACM Press, 1990
- [UBH-96] University of Birmingham, World Wide Web document,
URL: <http://www.bham.ac.uk/documents/ukuwww.html>
- [VAN-79] "Information Retrieval", van Rijsbergen, C.J., Butterworths, 1979
- [WAT-92] 'Brief History of Document Markup', Watson, G., *University Circular 1086*, University of Florida, November 1992
- [WES-96] Westlake Publishing Co. Ltd., World Wide Web document,
URL: <http://www.westlake.co.uk/>
- [WIT-94] "Managing gigabytes : compressing and indexing documents and images", Witten, I. H., Moffat, A., Bell, T., Van Nostrand Reinhold, 1994
- [WUR-89] "Information Anxiety", Wurman, R. S., Doubleday, 1989
- [WWW-95] World Wide Web document URL:
<http://www.w3.org/hypertext/WWW/TheProject>
- [YAN-94] "SIFT - A Tool for Wide-Area Information Dissemination", Yan, T., Garcia-Molina, H., *Technical Report*, Dept, of Computer Science, Stanford University, 1994
- [ZEZ-91] "Dynamic Partitioning of Signature Filters", Zezula, P., Rabitti, F., *ACM Transactions on Information Systems*, Vol. 9, No. 4, Pages 336-369, October 1991
- [ZIP-49] "Human Behaviour and the Principle of Least Effort", Zipf, G.K., Addison Wesley Publishing, 1949

APPENDIX A

MAILBASE Server

MAILBASE is the UK's major electronic mailing list service that enables groups to manage their own discussion topics and associated files. The service is run by a dedicated team based at Newcastle University and offers two main facilities to its users via electronic mail :

1. Discussion of topics using electronic mailing lists 'managed' by MAILBASE
2. Commands for retrieving files associated with a discussion list and for obtaining information about other lists or about MAILBASE in general.

In addition to the email service the following facilities allow users to browse through some of the publicly available material held on MAILBASE :

- MAILBASE On-Line Service
- MAILBASE Gopher Service
- MAILBASE World Wide Web Service

MAILBASE Discussion Lists

MAILBASE is a computer program that manages electronic discussion lists. Each list has four major elements :

1. **An electronic mail address for the list** - This uniquely identifies the list and is the address to which members of the list send contributions. Contributions are made in the form of email messages. Each message sent to that address is then distributed to every member of that particular list. For example, the **eSIG on Formal Methods** has the following address : **cti-comp-fm**.
2. **Members of the discussion list** - A list has associated with it a set of email addresses, each of which identifies a member of the list. Every item of mail sent to a list is distributed to each individual member's address.
3. **A list administrator, or owner** - Each list has one or more owners who carry out administrative functions, and who provide a first point of contact for problems or

queries related to the list. The List Owner is identified by an email address that always follows the pattern : <listname>-request@mailbase.ac.uk.

4. The list attributes - The level of public access to a MAILBASE list is determined by the attributes set the by the List Owner. Consequently a list may be “open” or “closed” for non-members. The List Owner can decide whether a non-member may:

- join the list automatically
- send mail to the list
- review the list membership
- retrieve those files associated with the list

A List Owner may also decide whether:

- to store list contributions in the monthly archive files
- replies to list messages are normally directed to the message originator or the whole list
- to moderate the list whereby all messages are sent directly to a List Moderator who will decide if they are suitable for broadcast to the whole list.

This last option is intended to filter out unwanted or duplicated material which may be of no use to the majority of list members. Currently however, most lists maintained on the MAILBASE server are not moderated.

MAILBASE offers the facility for storing file associates with a particular list, sot that the users may retrieve them when required. There are different types of file that may be associated with a list :

- Monthly archive files
- Files contributed by List Owners and members
- Introduction file

The team at Newcastle University also maintain a helpline to assist list members and various files are available that detail all commands including those for retrieving files, checking list membership, leaving a list among others.

The current number of lists maintained on the server including the CTC eSIGs is 755 with the number of distinct members being 41,600 approx.

How to join a CTC eSIG and make contributions

To join one of the CTC's eSIGs an email message must be sent to the following general address : **mailbase@mailbase.ac.uk**. The subject field may be left blank but the text of the message must read as follows :

join <listname><firstname(s)><lastname> and then end the message as normal.

For example

To: mailbase@mailbase.ac.uk

Subject: test (may be left blank)

Body of message: join cti-comp-fm Carmel Ryan
stop

MAILBASE Commands

MAILBASE is a computer program that automatically deals with the routine administrative tasks such as joining a list or retrieving files. All commands should be sent to the general address mentioned above. Commands sent to MAILBASE are executed in sequence until a terminator or an error is found. A mail message reporting success or failure is then returned to the sender. More than one command may appear in any one mail message to MAILBASE, but each command must be on a separate line. Except when stated otherwise, commands may appear in any order and can be upper or lower or mixed case.

The MAILBASE commands are:

- find lists <word>*
- help [<command>][†]
- index [<listname>]
- join <listname><firstname(s)><lastname>
- leave { <listname> | all }[‡]
- line limit <number>
- list me

* items in angle brackets <> are to be replaced appropriately

[†] square brackets [and] enclose optional items

[‡] braces {and} enclose alternative items one of which must be given, a vertical bar means "or"

lists [full]
resume mail {<listname> | all}
review <listname>
send <listname><filename>
statistics [command | list | <listname>]
stop
suspend mail {<listname> | all}

It is possible to get information from MAILBASE on the various commands by sending a help message to the general address.

For example

To: **mailbase@mailbase.ac.uk.**

Subject: **help** (may be left blank)

Body of text: **help index**

stop

This will result in a description of the index command being returned to the sender in an email message.

MAILBASE also provides a document of frequently asked questions that is particularly useful for new users as many users have the same type of problems which can generally be solved in quite easily. If there is anything still not clear the List Owner can be contacted at the address -: **listname-request@mailbase.ac.uk.**

Contributing

To contribute a message to the list simply send your message to

listname @mailbase.ac.uk, all the list members will receive this message.

For example

To: **cti-comp-fm@mailbase.ac.uk**

Subject: **First message** (may be left blank)

Body of message: **I have found an excellent paper on**

stop

In order to check what previous communication has gone on in a particular list, archive files can be checked. Messages sent to a particular list are normally collected together and placed in a monthly archive file for retrieval. Archive files, containing all list messages circulated, are generated automatically. These files - one for each month that the list has been in existence, can be checked by using the MAILBASE **index** and **send** commands. Documents requested by email are normally sent in plain text, but most of the documentation is available in "RTF" format. This format allows some users to obtain the information in the original formatted style and print it at their own sites. To obtain an "RTF" version it is necessary to append the suffix **.rtf** to the email request.

For example

Body of message: send mailbase gopher-tutorial.rtf

If membership of a list is no longer required then sending the command **leave listname** will remove your name from the specified MAILBASE list, or if the **leave all** option is used, from all those lists where your membership address corresponds to your current email address. To temporarily suspend mail from a list (or from every list if the "all" option is used), the command **suspend mail cti-comp-fm** is used.

APPENDIX B

Hypertext Markup Language

As outlined in Chapter Three, HTML is an application conforming to International Standard ISO 8879 - Standard Generalised Markup Language (SGML). SGML is a system for defining structured document types and markup languages to represent instances of those document types.

HTML, invented by Tim Berners-Lee of CERN [NCS-95] is a collection of platform independent styles (indicated by markup tags) that define the various components of a WWW document.. It is the HTML that tells the Web browser to view information in a certain way.

The language is composed of formatting commands called *tags*. Tags define the start and end of headings, paragraphs, lists, character highlighting and links. Tags work in pairs to format the text and graphics in an HTML document. Start tags are delimited by < and > and end tags are delimited by </ and >.

For example -:

```
<H1> This is a heading</H1>
<P>This is a paragraph.
```

There are certain standard HTML tags which should be contained in every HTML document. Required elements are shown in this simple document -:

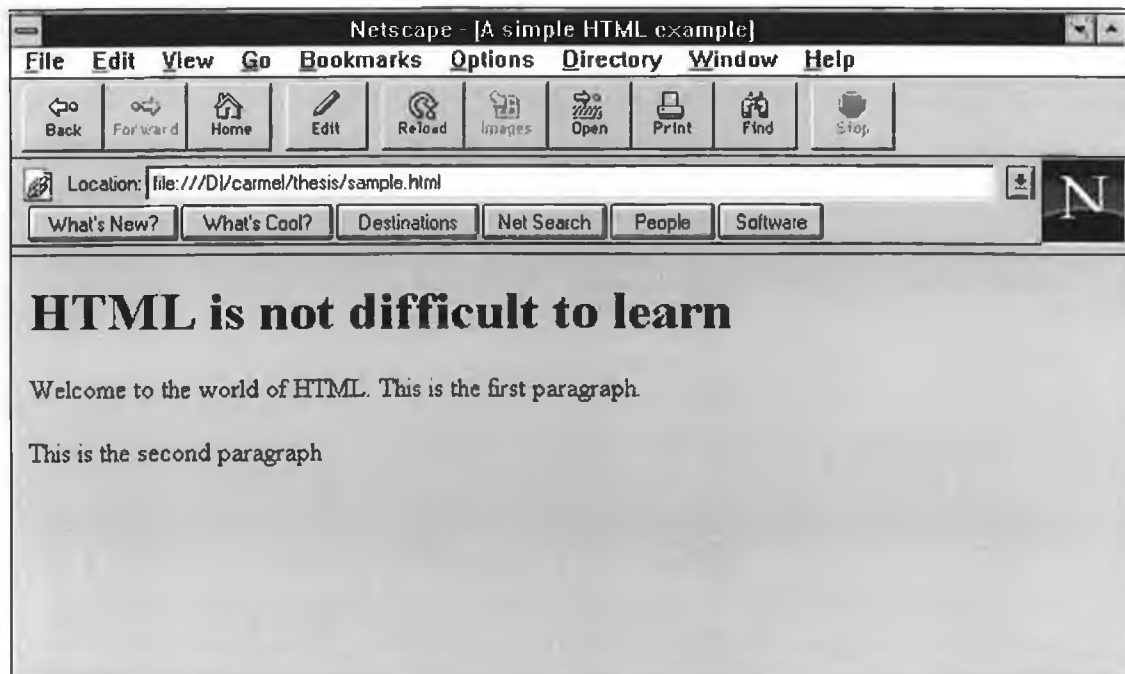
```
<HTML>
<HEAD>
<TITLE>A simple HTML example</TITLE>
</HEAD>
<BODY>
<H1> HTML is not difficult to learn</H1>
<P> Welcome to the world of HTML. This is the first paragraph.
<P> This is the second paragraph
</BODY>
</HTML>
```

HTML:

This element tells the Web browser that the file contains HTML-coded information.

HEAD:

This head element identifies the first part of the document and contains the title. The title included in the document is shown as part of the Web browser's window as shown in this example -:



TITLE

The title element contains your document title and identifies its content in a global context. It is usually displayed at the top of the browser window. It is better to choose a short but descriptive title as it is this that will be displayed in a user's bookmark list.

BODY:

The largest part of your HTML document is the body which contains the content of the document that is displayed in the text area of the browser window.

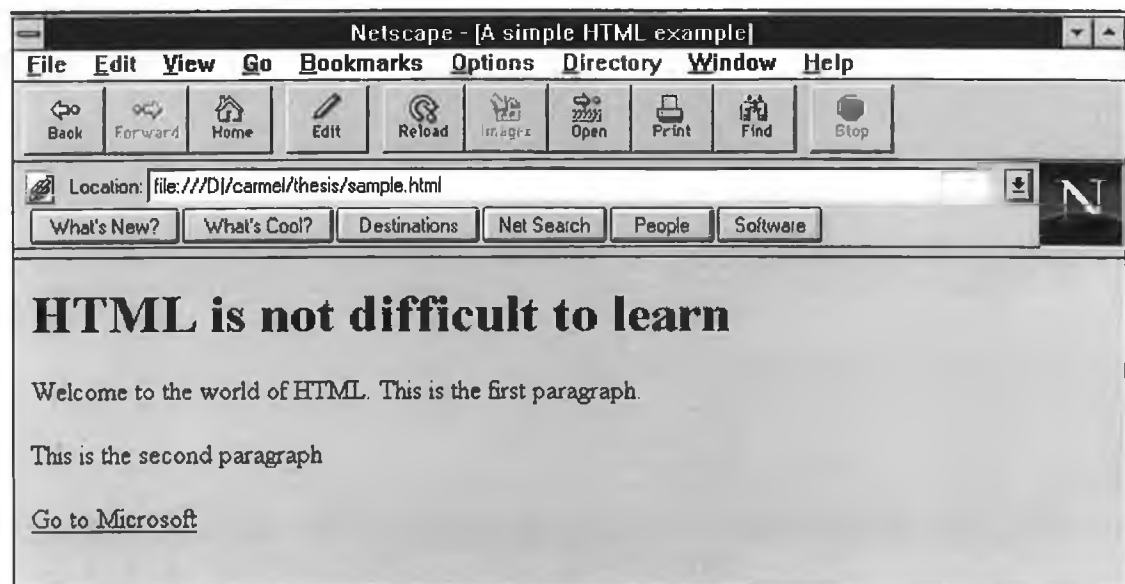
The other tags which make up HTML documents include Heading, Paragraph and List tags. Preformatted text using the `<PRE>` generates text in a fixed-width font. It is

useful for program code among other things. Different colours for background or text can be chosen using <BODY BACKGROUND> or <BODY BGCOLOR> tags.

The main power of HTML documents lies in the ability to link text and/or an image to another document or section of a document. A browser highlights the identified text or image with colour and/or underlining to indicate that it is a hypertext link. The *anchor* tag is used to create the link. For example in the example above the following line could be included in the body of the document to create a hypertext link to the home page of Microsoft -:

```
<A HREF="http://www.microsoft.com">Go to Microsoft</A>
```

The page would then appear in the browser window as follows :



Inline images can also be used to link to another Web document. The *img src* tag is used to display an image on a page. By including this in the *anchor* tag, the image becomes active and can be used to access another page with the click of a mouse button. For example adding the following line in the sample Web document will result in an image called 'welcome.gif' appearing on the page with a black line around it to indicate that it is also a link to another Web page, namely the introduction page of the Web tool described in Chapter Four.

```
<A HREF="http://www.compapp.dcu.ie/~cryan/intro.html"></A>
```

HTML also includes an interactive feature that enables users to submit feedback to the owner/creator of the Web document. This accomplished using the <FORM> tag. The tag has 2 attributes METHOD and ACTION which must be used in order to ensure the HTML form will work correctly. In the example given in section **Error! Reference source not found.** the METHOD = POST, which instructs the server to send the data from the form to the program which will process it. The ACTION attribute gives the URL of this program. This program (a PERL program, called 'quest.pl') is stored in the cgi-bin directory on the server.

There are numerous features of HTML which are not detailed here. To learn about HTML and these features, please see [NCS-95].

CTC eSIG QUESTIONNAIRE

Please check the boxes and fill in the blanks as appropriate

1. Please enter your name: _____

2. Please indicate which eSIGs you are a member of

1. cti-comp-se	5. cti-comp-pl
2. cti-comp-fm	6. cti-comp-ai
3. cti-comp-mat	7. cti-comp-dbms
4. cti-comp-hci	8. cti-comp-caos

3. If you have sent any messages to the eSIGs, please enter the name of the eSIG and the number of messages you sent : _____

4. Did you receive any response to these messages? Yes ____ No ____

5. Why did you join the eSIGs? _____

6. Do you find membership of the eSIGs useful? Yes ____ No ____

6b. Why? _____

7. Are you a member of any other email groups/lists similar to the CTC eSIGs?

Yes ____ No ____

7b. If yes how many? _____

8. Do you find membership of the other eSIG(s) useful? Yes ____ No ____

8b. Why? _____

9. The eSIGs were established to enable collaboration and discussion between computing academics. The provision of such a forum for academics to discuss topics of mutual interest, is intended to provide :

- pointers to new developments/resources on the Internet

- access to evaluations of public domain and commercial products
- access to lists of conferences, software reviews, reports and good practice as drawn up by the group
- a means of making informal acquaintances with other computing academics
- a facility to make informal comments and exchange views on their specialisms
- access to centres of excellence throughout the world

It is also intended that the members of the eSIGs will:

- maintain bulletin boards
- help organise seminars/workshops
- contribute to the annual conference
- help the CTC draft an educational policy
- draw up recommended syllabi

Do you think these objectives are reasonable? Yes ____ No ____

9b. Why?

10. Please list any other objectives you think should be included _____

11. What additional facilities would you like to see in place? _____

12. A message coding scheme when mailing the eSIGs would aid the process of monitoring the eSIGs to ensure the objectives outlined are being met.

For example: in the SUBJECT field of the mail message

A denotes curriculum issues

B denotes exam questions

C denotes research issues etc..

Would you be prepared to use such a coding scheme when sending messages?


Yes ____ No ____

- *Thank you for your co-operation*

CTC eSIG Questionnaire Web Page (1)

Netscape - [CTC eSIG Questionnaire]

File Edit View Go Bookmarks Options Directory Window Help



CTC eSIG Questionnaire

The CTC is collecting information about the operation of the eSIGs. Your co-operation is greatly appreciated.
Please click on the boxes and fill in the blanks as appropriate

1. Please enter your name here :

2. Please indicate which eSIGs you are a member of

<input type="checkbox"/> 1.cti-comp-se	<input type="checkbox"/> 1.cti-comp-pl
<input type="checkbox"/> 1.cti-comp-fm	<input type="checkbox"/> 1.cti-comp-ai
<input type="checkbox"/> 1.cti-comp-mat	<input type="checkbox"/> 1.cti-comp-dkms
<input type="checkbox"/> 1.cti-comp-hoi	<input type="checkbox"/> 1.cti-comp-caos

3. If you have sent any messages to the eSIGs, please enter the name of the eSIG and the number of messages you sent to it :

4. Did you receive any responses to these messages? ☐ YES ☐ NO

5. Why did you join the eSIGs? :

6a. Do you find membership of the eSIGs useful? ☐ YES ☐ NO

Document: Done

CTC eSIG Questionnaire Web Page (2)

Netscape - [CTC eSIG Questionnaire]

File Edit View Go Bookmarks Options Directory Window Help

6b. Why?

7a. Are you a member of any other email groups/lists similar to the oti-comp eSIGs?

☐ YES ☐ NO

7b. If yes how many? :

8a. Do you find membership of the other eSIG(s) useful? ☐ YES ☐ NO

8b. Why?

9a. The eSIGs were established to enable collaboration and discussion between computing academics. The provision of such a forum for academics to discuss topics of mutual interest, is intended to provide :

- pointers to new developments/resources on the internet
- access to evaluations of public domain and commercial products
- access to lists of conferences, software reviews, reports and good practice as drawn up by the group
- a means of making informal acquaintances with other computing academics
- a facility to make informal comments and exchange views on their specialisms
- access to centres of excellence throughout the world.

It is also intended that the members of the eSIGs will

- maintain bulletin boards
- help organise seminars/workshops
- contribute to the annual conference
- help the CTC draft an educational policy and
- draw up recommended syllabi

7/99 Document Done

HTML Source Code for CTC Web questionnaire

```
<HTML>
<HEAD>
<TITLE>CTC eSIG Questionnaire</TITLE>
</HEAD>

<BODY>
<body bgcolor="ffffff" text="DDDDD">
<H1>CTC eSIG Questionnaire</H1><br>
<p>The CTC is collecting information about the operation of the eSIGs. Your co-
operation is greatly appreciated.
<br><Please click on the boxes and fill in the blanks as appropriate</i>
<HR>
<FORM METHOD="POST" ACTION="http://www.compapp.dcu.ie/cgi-
bin/carmel/quest.pl">
<FORM>
<B><PRE>
1.Please enter your name here      : <INPUT TYPE="text"
NAME="username"SIZE="28"><BR>

2.Please indicate which eSIGs you are a member of<br>

<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-se

<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-pl
<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-fm> 1.cti-comp-fm

<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-ai
<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-mat

<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-dbms
<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-hci

<INPUT TYPE=checkbox NAME=eSIG VALUE=cti-comp-se> 1.cti-comp-caos

3.If you have sent any messages to the eSIGs, please enter the name of the eSIG
and the number of messages you sent to it :
<INPUT TYPE="text" NAME="eSIG2" SIZE="20"><p>

4.Did you receive any responses to these messages?
<INPUT TYPE=radio NAME=response VALUE=yes> YES
<INPUT TYPE=radio NAME=response VALUE=no> NO

5. Why did you join the eSIGs?      :
<textarea name="joining" rows="8" cols="40"></textarea><p>
6a. Do you find membership of the eSIGs useful?
```

<INPUT TYPE=radio NAME=membership VALUE=yes> YES
<INPUT TYPE=radio NAME=membership VALUE=no> NO<p>

6b. Why? <textarea name ="member" rows = "4" cols="58"></textarea><p>

7a. Are you a member of any other email groups/lists similar to the CTC eSIGs?

<BLOCKQUOTE>

<INPUT TYPE=radio NAME=othereSIG VALUE=yes> YES

<INPUT TYPE=radio NAME=othereSIG VALUE=no> NO

</BLOCKQUOTE>

7b. If yes how many? : <INPUT TYPE="text" NAME="num_eSIG"SIZE="5"><p>

8a. Do you find membership of the other eSIG(s) useful?

<BLOCKQUOTE>

<INPUT TYPE=radio NAME=mem_useful VALUE=yes> YES

<INPUT TYPE=radio NAME=membership VALUE=no> NO

</BLOCKQUOTE>

8b. Why? <textarea name ="member2" rows = "4" cols="58"></textarea><p>

9a. The eSIGs were established to enable collaboration and discussion between computing academics. The provision of such a forum for academics to discuss topics of mutual interest, is intended to provide :

<BLOCKQUOTE>

- pointers to new developments/resources on the Internet
- access to evaluations of public domain and commercial products
- access to lists of conferences, software reviews, reports and good practice as drawn up by the group
- a means of making informal acquaintances with other computing academics
- a facility to make informal comments and exchange views on their specialisms
- access to centres of excellence throughout the world.

</BLOCKQUOTE>

It is also intended that the members of the eSIGs will

<BLOCKQUOTE>

- maintain bulletin boards
- help organise seminars/workshops
- contribute to the annual conference
- help the CTC draft an educational policy and
- draw up recommended syllabi

</BLOCKQUOTE>

Do you think these objectives are reasonable?

<INPUT TYPE=radio NAME=objectives VALUE=yes> YES

<INPUT TYPE=radio NAME=objectives VALUE=no> NO<p>

9b. Why? <textarea name ="why_reasonable" rows = "4" cols="58"></textarea><p>

10. Please list any other objectives you think should be included

<textarea name="other_objjs" rows="4" cols="67"></textarea><p>

11. What additional eSIG facilities would you like to see in place?

<textarea name="facilities" rows="4" cols="67"></textarea><p>

12. A message coding scheme when mailing the eSIGs would aid the process of monitoring the eSIGs to ensure the objectives outlined are being met.

For e.g. - in the SUBJECT field of the mail message

<BLOCKQUOTE>

A denotes exam question information

B denotes curriculum issues

C denotes research issues

.....

etc.</BLOCKQUOTE>

Would you be prepared to use such a coding scheme when sending messages?

<INPUT TYPE=radio NAME=objectives VALUE=yes> YES

<INPUT TYPE=radio NAME=objectives VALUE=no> NO

<INPUT TYPE="submit" VALUE="Submit questionnaire">

<INPUT TYPE="reset" >

</PRE>

</FORM>

<HR>

Back to the CTC Home Page

<p>

<h6>Thank you for your co-operation. If you have any further ideas about developing the eSIGs please do not hesitate to contact us</h6>

</BODY>

</HTML>

CGI Script Source code (PERL) - Questionnaire

```
#!/app1/lang/bin/perl

require "cgi-lib.pl"; # includes a header file containing generic functions

&ReadParse;          # this function included in the header file

# Formatting of the Web page presented after query submitted
print "Content-type: text/html\n\n";
print "<HTML><HEAD><TITLE>Questionnaire submission</TITLE></HEAD>\n";
#print "<body bgcolor='ffffff' text='000000'>\n";
print "<h3>The questionnaire was submitted successfully</h3>\n";
print "<h4>Please return to home page</h4>";
#print "<h4><a href
=\"http://www.compapp.dcu.ie/~cryan/quest.html\">Back</a></h4><hr>";
print "</HTML>";

#inputing all the stuff

$name = $in{'username'};
$ctcesig = $in{'eSIG'};
$msgsig = $in{'eSIG2'};
$response = $in{'response'};
$reason_joined = $in{'joining'};
$membership_useful = $in{'membership'};
$why_useful = $in{'member'};
$otheresig = $in{'otheresig'};
$num_otheresig = $in{'num_eSIG'};
$mem_useful = $in{'mem_useful'};
$why_useful2 = $in{'member2'};
$objectives = $in{'objectives'};
$why_reasonable = $in{'why_reasonable'};
$other_objs = $in{'other_objs'};
$facilities = $in{'facilities'};
$msg_scheme = $in{'msg_scheme'};

#open the file for all the results

open( OUT,">>/users/capg/cryan/perl/results")|| die "cannot open results for writing";

print OUT "NAME $name\n";
print OUT "CTCESIG $ctcesig\n";
print OUT "MSGs_SENT $msgsig\n";
print OUT "RESPONSE $response\n";
print OUT "WHY_JOINED $reason_joined\n";
print OUT "ESIG_USEFUL $membership_useful\n";
```

```
print OUT "WHY_USEFUL $why_useful\n";
print OUT "MEMBER_OF_OTHER_ESIG $otheresig\n";
print OUT "HOW_MANY_OTHER $num_otheresig\n";
print OUT "OTHER_USEFUL $mem_useful\n";
print OUT "WHY_OTHER_USEFUL $why_useful2\n";
print OUT "OBJECTIVES_OK $objectives\n";
print OUT "WHY_OK $why_reasonable\n";
print OUT "OTHER_OBJS $other_objs\n";
print OUT "FACILITIES $facilities\n";
print OUT "USE_MSG_SCHEME $msg_scheme\n";
```

```
close(OUT);
```

APPENDIX C

HTML Source Code for Map-of-Europe Page (Figure 4-1)

```
<html>
<title>Inter-University Co-operation in Europe in Computing</title>
<body bgcolor="ffffff" text="DDDDD">
<h2>Inter-University Co-operation in Europe in the
field of Computing</h2>

<hr><b>
Below is a map of Europe. Those countries in white are members of the EU with
Iceland and Norway also in white, for the purpose of this project. Simply click on the
country which you want to investigate. You will be presented with a map of that
country which contains links to the various third level institutions that offer
computing courses.
</b><center>
<a href = "http://simpr1.compapp.dcu.ie/cgi-bin/imagemap/europe">
<IMG SRC = "http://www.compapp.dcu.ie/~cryan/europe.gif" ISMAP =
"ISMAP"></a></center>
<hr>
<h6>The maps for this web site came from various sources including Virtual Tourist
and National Property Register,UK.<br>
Page last updated 30/6/96
<br>If you have any comments or suggestions please email <a
href="mailto:cryan@compapp.dcu.ie"> cryan@compapp.dcu.ie</a></h6>
</html>
```

HTML Source Code for the Side Frame Page - (Figure 4-1)

```
<html>
<head><title>Inter-University Co-operation in Computing</title></head>
<body bgcolor="ffffff" text="DDDDD">
<h4>Menu</h4>
<p><a href="maps.html" onmouseover="self.status='Maps page';return
true"target=_top>Map of Europe</a>
<p><a href="cr.html" onmouseover="self.status='Search the database';return
true"target=display>Search the database</a>
<br>
<br>
<br>
<br>
<br><h6>Page last updated 30/6/96
```


If you have any comments or suggestions please email cryan@compapp.dcu.ie</h6>
</html>

CGI Imagemap.conf File Source Code

```
ireland : /users/capg/cryan/public_html/ireland.map
austria : /users/capg/cryan/public_html/austria.map
europe : /users/capg/cryan/public_html/europe.map
norway : /users/capg/cryan/public_html/norway.map
sweden : /users/capg/cryan/public_html/sweden.map
iceland : /users/capg/cryan/public_html/iceland.map
lux : /users/capg/cryan/public_html/lux.map
finland : /users/capg/cryan/public_html/finland.map
denmark : /users/capg/cryan/public_html/denmark.map
germany : /users/capg/cryan/public_html/germany.map
holland : /users/capg/cryan/public_html/holland.map
portugal : /users/capg/cryan/public_html/portugal.map
spain : /users/capg/cryan/public_html/spain.map
greece : /users/capg/cryan/public_html/greece.map
italy : /users/capg/cryan/public_html/italy.map
belgium : /users/capg/cryan/public_html/belgium.map
france : /users/capg/cryan/public_html/france.map
uk : /users/capg/cryan/public_html/uk.map
scotland : /users/capg/cryan/public_html/scotland.map
wales : /users/capg/cryan/public_html/wales.map
england : /users/capg/cryan/public_html/england.map
easteng : /users/capg/cryan/public_html/easteng.map
seasteng : /users/capg/cryan/public_html/seasteng.map
swesteng : /users/capg/cryan/public_html/swesteng.map
mideng : /users/capg/cryan/public_html/mideng.map
northeng : /users/capg/cryan/public_html/northeng.map
northirl : /users/capg/cryan/public_html/northirl.map
```


Imagemap File of Europe

default <http://www.compapp.dcu.ie/~cryan/error.html>
rect <http://www.compapp.dcu.ie/~cryan/iceland.html> 4,66 106,115
poly <http://www.compapp.dcu.ie/~cryan/irl.html> 108,216 62,251 62,251 91,281 91,281 135,268 135,268 144,247 144,247 119,212 119,212 107,215
poly <http://www.compapp.dcu.ie/~cryan/norway.html> 469,3 357,30 357,30 336,59 336,59 326,84 326,84 284,116 284,116 260,131 260,131 264,174 263,174 274,188 275,188 301,189 301,188 323,173 323,173 339,150 339,149 337,117 337,117 336,107 336,107 353,96 353,96 362,73 362,73 372,60 372,60 379,52 378,52 384,39 405,36 422,25 443,31 456,31 461,20 474,17
poly <http://www.compapp.dcu.ie/~cryan/sweden.html> 422,40 454,83 435,100 405,130 397,138 416,160 416,160 396,190 390,213 372,217 372,217 335,205 332,178 351,160 353,124 344,118 346,108 359,104 365,100 374,76 390,56 404,49 422,41
poly <http://www.compapp.dcu.ie/~cryan/uk.html> 130,198 177,157 206,200 201,219 228,242 231,263 223,270 191,288 126,299 141,278 143,265 161,249 154,240 127,204 133,197
poly <http://www.compapp.dcu.ie/~cryan/finland.html> 452,44 462,79 447,110 434,116 433,149 461,160 496,151 524,122 505,83 491,45 488,25 474,24 466,42
poly <http://www.compapp.dcu.ie/~cryan/portugal.html> 78,405 102,410 96,427 96,430 88,433 97,442 92,461 90,466 70,470 32,450 34,431 57,432 57,432 78,405
poly <http://www.compapp.dcu.ie/~cryan/spain.html> 100,376 164,378 169,392 225,401 188,427 181,446 187,449 152,475 123,476 116,482 109,484 95,469 104,440 115,414 120,404 81,394 100,376
poly <http://www.compapp.dcu.ie/~cryan/france.html> 178,381 227,392 241,378 257,381 265,366 267,352 260,337 275,320 283,311 225,294 217,294 201,305 181,296 180,309 142,314 168,333 176,380
poly <http://www.compapp.dcu.ie/~cryan/dutch.html> 265,241 239,254 237,268 251,276 268,274 281,268 285,251 265,241
poly <http://www.compapp.dcu.ie/~cryan/belgium.html> 230,274 237,297 269,296 276,283 256,277 228,273
poly <http://www.compapp.dcu.ie/~cryan/lux.html> 277,292 289,288 309,291 318,297 316,303 310,307 295,307 268,298 277,293
poly <http://www.compapp.dcu.ie/~cryan/denmark.html> 324,189 263,203 261,221 311,226 332,212 322,187
poly <http://www.compapp.dcu.ie/~cryan/germany.html> 295,251 308,249 306,236 313,234 327,243 342,240 349,238 360,252 368,277 337,286 327,287 346,310 327,321 290,322 294,311 294,311 324,305 308,287 278,287 275,277 286,266 289,254
poly <http://www.compapp.dcu.ie/~cryan/austria.html> 362,316 340,324 330,330 325,330 325,333 331,333 336,332 343,337 356,340 367,340 384,337 394,335 397,323 390,315 389,313 374,312 361,315
poly <http://www.compapp.dcu.ie/~cryan/italy.html> 293,349 279,354 279,360 276,362 276,365 276,369 281,372 284,374 293,368 297,367 307,371 318,380 332,396 351,411 366,415 377,424 385,435 385,443 381,452 370,456 356,456 344,452 336,459 367,476 376,474 371,465 372,462 379,459 385,458 390,447 396,440 391,433 393,424 410,431 414,428 398,420 381,410 382,405 360,398 336,378 358,395 339,373 337,361 345,352 351,346 330,337 313,347 294,350
poly <http://www.compapp.dcu.ie/~cryan/greece.html> 421,439 443,422 455,420 482,413 501,415 506,407 510,414 507,421 486,422 507,477 512,484 523,493 517,499 473,500 417,452 421,439

HTML Source Code for Map-of-Europe Page (Main Frame)

```
<html>
<title>Inter-University Co-operation in Europe in Computing</title>
<body bgcolor="ffffff" text="DDDDD">
<h2>Inter-University Co-operation in Europe in the
field of Computing</h2>

<hr><b>
Below is a map of Europe. Those countries in white are members of the EU with
Iceland and Norway also in white, for the purpose of this project. Simply click on the
country which you want to investigate. You will be presented with a map of that
country which contains links to the various third level institutions that offer
computing courses.
</b><center>
<a href = "http://simpr1.compapp.dcu.ie/cgi-bin/imagemap/europe">
<IMG SRC = "http://www.compapp.dcu.ie/~cryan/europe.gif"ISMAP =
"ISMAP"></a></center>
<hr>
<h6>The maps for this web site came from various sources including Virtual Tourist
and National Property Register,UK.
Page last updated 30/6/96
<br>If you have any comments or suggestions please email <a
href="mailto:cryan@compapp.dcu.ie"> cryan@compapp.dcu.ie</a></h6>

</html>
```

HTML Source Code for Index Page (Side Frame)

```
<html>
<head>
<title>Inter-University Co-operation in Computing</title>
</head>

<body bgcolor="#FFFFFF" text="DDDDD">
<table width=60>
<tr><td><h4>Menu</h4>
<p><a href="euhome.html" target="display">Introduction</a> </p>
<p><a href="euhome.html" target="display">Map of Europe</a> <br>
<a href="austria.html" target="display">Austria</a> <br>
<a href="belgium.html" target="display">Belgium</a> <br>
<a href="denmark.html" target="display">Denmark</a> <br>
<a href="finland.html" target="display">Finland</a> <br>
<a href="france.html" target="display">France</a> <br>
<a href="germany.html" target="display">Germany</a> <br>
<a href="greece.html" target="display">Greece</a> <br>
<a href="dutch.html" target="display">Holland</a> <br>
<a href="iceland.html" target="display">Iceland</a> <br>
<a href="irl.html" target="display">Ireland</a> <br>
<a href="italy.html" target="display">Italy</a> <br>
<a href="lux.html" target="display">Luxembourg</a> <br>
<a href="norway.html" target="display">Norway</a> <br>
<a href="portugal.html" target="display">Portugal</a> <br>
<a href="spain.html" target="display">Spain</a> <br>
<a href="sweden.html" target="display">Sweden</a> <br>
<a href="uk.html" target="display">United Kingdom</a> </p>
<blockquote>
<p><a href="eng.html" target="display">England</a> <br>
<a href="scotland.html" target="display">Scotland</a> <br>
<a href="wales.html" target="display">Wales</a> <br>
<a href="ni.html" target="display">N.Ireland</a> </p>
</blockquote>
<p><a href="cr.html" target="display">Search the database</a> </p>
<p><h6>Page last updated 30/6/96 <br>
If you have any comments or
suggestions please email
<a href="mailto:cryan@compapp.dcu.ie">cryan@compapp.dcu.ie</a></p>
</td></tr>
</table>
<p><br>
<br>
<br>
</p>
</h6>
</body></html>
```

Search Web Page to access the CTC database

Netscape - [Inter-University Co-operation in Computing]

File Edit View Go Bookmarks Options Directory Window Help

Back Forward Home Edit Reload Images Open Print Find Stop

Location: <http://www.compapp.dcu.ie/~cryan/maps.html>

Menu

[Introduction](#)

[Map of Europe](#)

[Austria](#)

[Belgium](#)

[Denmark](#)

[Finland](#)

[France](#)

[Germany](#)

[Greece](#)

[Holland](#)

[Iceland](#)

[Ireland](#)

[Italy](#)

[Luxembourg](#)

[Norway](#)

[Portugal](#)

[Spain](#)

[Sweden](#)

[United Kingdom](#)

[England](#)

[Scotland](#)

[Wales](#)

[N.Ireland](#)

[Search the database](#)

Search the CTC database

You can search a database, maintained by the CTC, of web pages of European university computing faculties/departments. Simply enter your query in the space provided below and press submit to run the query.

A page will then be presented to you with a list of references to those pages that matched your query.

QUERY:

Page last updated 30/6/96

If you have any comments or suggestions please email ryan@compapp.dcu.ie

Document Done

HTML Source Code for Search Page (Main Frame)

```
<html>

<head>
<title>Search Computing Depts of European Colleges</title>
</head>

<body>
<body bgcolor="#FFFFFF" text="DDDDD">

<table cellpadding=0 cellspacing=0 width=500>
<tr><td width=100%><h3>Search the CTC database </h3>
<hr width=500>
<p>You can search a database, maintained by the CTC, of web pages of European
university computing faculties/departments. Simply enter your query in the space
provided below and press submit to run the query. <br>
A page will then be presented to you with a list of references to those pages that
matched your query. </p>

<form action="http://www.compapp.dcu.ie/cgi-bin/carmel/crsearch.pl"
method="POST">
<p><b>QUERY</b> : <textarea name="query" rows=2 cols=51></textarea> </p>
<hr width=500>
<p><input type=submit name="action" value="Submit Query"> <input type=reset
value="Reset"> </p>
</form>

</td></tr>
</table>
<h6>Page last updated 30/6/96 <br>
If you have any comments or suggestions please email <a
href="mailto:cryan@compapp.dcu.ie">cryan@compapp.dcu.ie</a></h6>
</body>

</html>
```

CGI Script Source code (PERL) - Search Facility

```
#!/app1/lang/bin/perl

require "cgi-lib.pl";          #includes a header file containing generic functions

print &PrintHeader;
&ReadParse;                   #these 2 functions are included in the header file

$request = $ENV{'REQUEST_METHOD'};

$uquery      = $in{query};      #take in the query and rename it uquery

#Formatting of the Web page presented after query submitted
print "<HTML><HEAD><TITLE>Results of searching CTC 'European College
Database'</TITLE></HEAD>\n";
print "<body bgcolor=\"ffffff\" text=\"000000\">\n";
print "<h3>The query submitted
was:</h3><BLOCKQUOTE>$uquery</BLOCKQUOTE><HR>";
print "<h3>Here are the pages which matched the query </h3>";

if( $uquery ne "" )           #if a query is actually entered do the following
{
    chdir( "/users/capg/fkelledy/crs" );  #search engine called, query processed
    system( "/users/capg/fkelledy/crs/search -ws 8 -dn 1 -mda 5000 -I $uquery" );
    system( "/users/capg/fkelledy/crs/get_url" );

    open( F,"/users/capg/fkelledy/crs/results/1.urls" ) || htmdie( "Can't open Url
        file for reading\n" );
    $i = 0;
    while ( <F> )               #processing of results
    {
        $score = substr( $_,0,4 );
        $len = length( $_ );
        $tail = substr( $_,5,($len-5) );
        if( $i < 10 )
        {
            print "$score <A HREF=\"$tail\">$tail</A><BR>";
        }
        $i = $i + 1;
    }
    close( F );

    print "<HR>";
}
else                            #if no query entered
{ print "<H2>query Rejected -- too short</H2>";
}
```

Sample Query #1 Search Page

Netscape - [Inter-University Co-operation in Computing]

File Edit View Go Bookmarks Options Directory Window Help

Back Forward Home Reload Image Open Print Find Stop

Location: <http://www.compapp.dcu.ie/~cryan/intro.html>

What's New? What's Cool? Destinations Net Search People Software

Menu

[Map of Europe](#)

[Search the database](#)

Page last updated 30/6/96
If you have any comments or suggestions please email cryan@compapp.dcu.ie

Search the CTC database

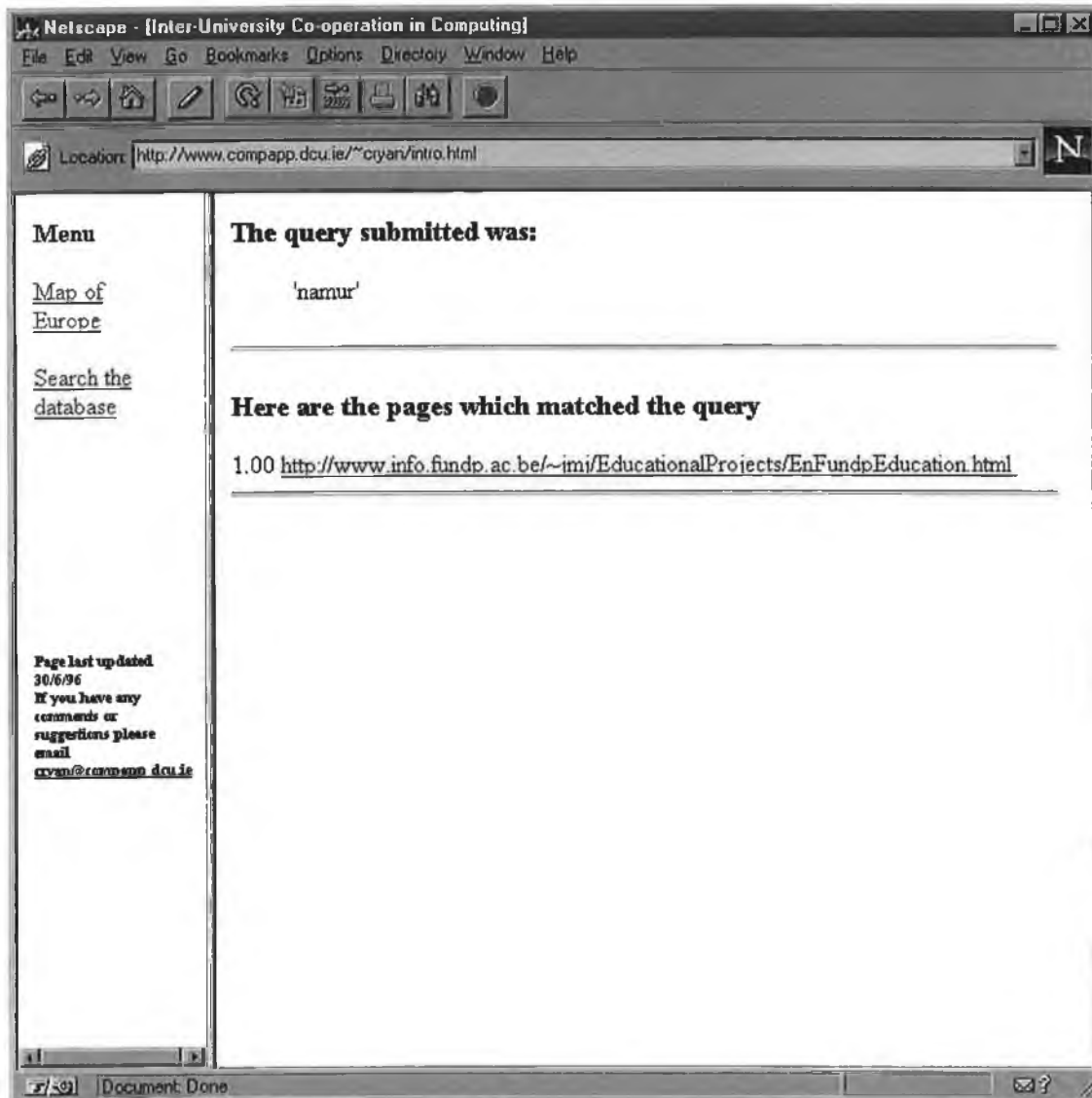
You can search a database, maintained by the CTC, of web pages of European university computing faculties/departments. Simply enter your query in the space provided below and press submit to run the query.
A page will then be presented to you with a list of references to those pages that matched your query.

QUERY:

Page last updated 30/6/96
If you have any comments or suggestions please email cryan@compapp.dcu.ie

Document: Done


Sample Query #1 Results Page



Sample Query #2 - Search Page

Netscape - [Search Computing Depts of European Colleges]

File Edit View Go Bookmarks Options Directory Window Help

Location: 



Search the CTC database

You can search a database, maintained by the CTC, of web pages of European university computing faculties/departments. Simply enter your query in the space provided below and press submit to run the query.

A page will then be presented to you with a list of references to those pages that matched your query.

QUERY:

Page last updated 30/6/96
If you have any comments or suggestions please email crayan@compapp.dcu.ie

 Document: Done 

Sample Query #2 - Results Page

