C: Proposal Description

Gauguin:

New evaluation approaches for information seeking environments

1a. Research Topic

The proposed research training network will adapt multiple methods drawn from several disciplines and apply them to evaluating information seeking environments (ISEs): that is, IR (information retrieval) systems in the widest sense, including new digital library technologies. It will furthermore attempt to develop an evaluation framework capable of relating and combining the different relevant approaches and techniques. Appropriate evaluation is important to developing ISEs that are both effective and usable. New technical developments require new approaches to evaluation, while retrieval systems are becoming ever more important not just in libraries and archives, but in the internet and the intranets now becoming central to the way many companies and organisations operate. This research seeks to carry forward both the topic and the partnerships developed in the MIRO and Mira working groups (BRA 6576 and working group 20039 respectively).

The IR field has a long history mainly concerned with techniques for text document retrieval based on a complete indexing of all the words (actually, word stems) in all the documents in a given collection. Along with the focus on this retrieval technology has been a strong focus on a single evaluation method based on test collections, which measure the performance of retrieval engines in terms of precision and recall by comparing the documents each piece of software retrieves given a standard set of queries against the judgements about relevance by a set of human experts. Although the test collections and their stored human judgements are expensive to create, software can then be tested in the lab. without further involvement of human users.

The technical basis of IR has changed enormously in recent years as computing has advanced. Important features of this include:

- Interactive software: a typical retrieval nowadays will involve one user and many retrieval cycles, with the query being repeatedly modified by the human. The performance of the machine on one cycle is much less important than the overall success of the session.

- The WWW (world wide web): by far the most important collection, but huge, dynamic, and unsearchable by any current engine. (Current engines admit they cannot cover the whole web.)

- "Documents" in media other than text are now important: images, video, sound, etc.

- As well as queries in the same medium, cross-media retrieval is important e.g. giving a text query to retrieve images; (and cross-language retrieval e.g. a French query to retrieve German text documents).

- Multi-collection and multi-engine retrieval tasks (e.g. one query applied to several document collections and/or several retrieval engines).

- Hypertext (hypermedia) e.g. the links in a WWW document: explicit authored links between documents, not just similarity based on textual content or predefined data structures.

- The use of IR as part of collaborative work contexts e.g. CSCW (Computer Supported Collaborative Work) applications.

Each of these developments entail their own evaluation problems. For instance, observing ordinary users (since nearly everyone now uses IR, specially trained users are of less importance) shows that often they do not open a document unless the keyword they expect is in the surrogate (the list entry representing that document), and sometimes they do not even scroll the window to show more than the top three items on the list. Clearly changing the user interface not the underlying IR engine is what is needed, but traditional IR evaluation cannot address this. Another example is that retrieving images of paintings by a text query dealing with the painter, date, and so on is exactly what is wanted by some users (e.g. art historians), but is of little use to another kind of user who wants a picture that looks "similar" in theme, mood, or colour to another one: instead another kind of retrieval engine (that computes picture to picture similarity, and takes images as queries) will be required. It is pointless to ask which kind of engine is better: it is a case of matching the software to the type of task and user. Again, this is beyond traditional evaluation approaches.

Thus technical developments require matching developments in evaluation methods, which however have not advanced so rapidly. However methods for studying individual users and human-human interactions in the workplace, drawn originally from the social sciences but already applied in parts of the field of Human Computer Interaction (HCI), are likely to lead to substantial progress here.

IR is being used by an explosively expanding user population (e.g. WWW and intranets). Furthermore, its scope is being hugely and rapidly expanded by its application to new media, and by other technical advances. Since the many new technologies involved in digital libraries will introduce many difficult technical challenges, the danger is that approaches to evaluation will slip backwards to an emphasis on the technical testing of software if user-centered approaches are not vigorously promoted at the same time. Should that happen, digital libraries could develop technically, yet in practice only be usable by librarians and a few other information specialists, and not by the much larger number of the users of the information itself. For developments to be more than mere technological novelties, and to be adapted to be of real use to human users, depends upon the development and application of new and improved evaluation techniques.

1b. Project (research training network) objectives

There is thus a wide ranging need to meet the new challenges for IR evaluation posed by new technologies. This will require importing techniques from other disciplines, building up experience of their use in IR, and attempting to understand how they might fit together in a comprehensive approach.

Our specific objectives, then, are:

Research objectives

1. The first class of objective is to apply, adapt, and develop different methods of evaluation in a set of studies. Examples appear below in the projects offered by individual sites. The range of methods to be applied will include traditional IR evaluations using precision and recall measures, and test collections; HCI style studies, with human users in a lab. setting, using observation and thinkaloud protocols; and workplace field studies, using ethnographic approaches to observation.

2. Secondly, we will perform IR evaluation studies of various of the new technologies e.g. large image collections, multi-collection searching. Although these will use the methods mentioned, the focus of this objective is to investigate the distinctive issues brought out by new technologies.

3. Similarly, we will perform studies of various work domain applications, thus identifying types or groups of actual IR users, and types of task important to those user groups.

4. Finally the network as a whole will seek to develop a general framework for IR evaluation that can relate the methods, technologies, and issues identified, and may eventually specify the set of methods to be applied to a given problem. Where practicable, we shall evaluate a particular ISE by several methods, as comparing the results allows the evaluation methods themselves to be assessed for validity, cost, etc.

Training objectives

5. Train a group of young IR researchers in a range of techniques and approaches for IR evaluation which they would not normally receive in their home institutions.

6. Transfer expertise between the research groups involved, as they currently have a wide range of different expertise from each other.

2. Scientific originality

The field of information retrieval (IR) began in the late 1960s, addressing the problem of retrieving text documents from large collections, by computer, based on full-text indexing of words. It has always been characterised by a strong focus on evaluation: on methods of measuring retrieval effectiveness, traditionally just the performance of the software engines. This dominates how most research is now done and reported, but probably stems from the peculiar problem of not being able to judge the quality of any retrieval by simple inspection of the results: you can only judge them if you have extensive knowledge of what might have been retrieved by that query on that collection (using a perfect engine) and that is inherently expensive information to acquire,

requiring a much more systematic and formal approach to evaluating test results. The emergence into widespread use both of multimedia rather than only text documents, and of interactive user interfaces has extensive implications for this field and the evaluation methods on which it depends which are far from being worked out.

How can we evaluate how good an engine is at finding documents which the user considers relevant? The traditional measures used are *precision* and *recall*. *Precision* measures the fraction of the retrieved documents which are in fact relevant to the user's information need, and *recall* measures the fraction of those documents the user would consider relevant which were actually retrieved. Obviously, the practical importance of each of these measures varies widely with the type of task being done (e.g. contrast "find any two examples of…" with "find all papers on…"). Furthermore, to calculate recall-precision figures an experimenter must have a collection of documents to search, a collection of queries to find documents for, and a set of independently made decisions as to which documents are "really" relevant to which queries. Constructing such a "test collection" of documents, queries, and relevance judgements is very time consuming and open to criticisms of bias in the judgements. Considerable efforts have been invested in building large standard test collections (e.g. Cranfield (1) and TREC (2)). By creating a standard set of queries and relevance judgements, the test collection approach has removed the end users from the evaluation loop, representing them by the queries and judgements stored in the test collections. This may be acceptable when the search techniques are non-interactive and it allows fast experimentation, but it also makes it extremely hard to evaluate the worth of interactive techniques such as user relevance feedback, user query expansion and changes to the interface.

Information retrieval systems were initially designed to be used by intermediates in a library setting. These trained searchers would interview a user to build up a model of their information needs and then carry out searches at a later date — often specifying very clearly the information (or topic) that the user was looking for. This is not how people search the Internet — users often have only a loosely formed notion of what they are looking for when they start a session and often have very little idea of what the collection will contain on that topic. This stark difference in user population from the traditional models of IR is one of the challenges facing modern IR researchers. As well as Internet searching, the widespread use of encyclopaedias on CD, large volume hard disks and cheap, very fast personal computers has led to many end users with no computing training using search engines on fairly large collections of text. Furthermore, the speed of the engines, the spread of mouse and window user interfaces, together with non-specialist users has made repeated exploratory retrievals the normal procedure, rather than single carefully designed queries. The net effectiveness of a session, or at least a set of retrieval attempts, has become much more relevant than the performance of a single retrieval cycle. Users typically find interesting information at many different steps in a session which not only is used to modify their query formulations but may also modify their goals and relevance judgements (a point made as early as 1973 (3)). This means that to study, measure, and optimise the useful work done with an IR program, we must measure the retrieval done by an interactive user over a set of retrieval cycles. This will depend partly upon the software, but also partly upon what the user does. The "system" being studied is not the function computed by one call on the retrieval engine, but the combined human-computer interaction over as many cycles as the user is observed to initiate in the course of one task. This redefinition of "system" affects how evaluation must be done, what measures can be used to compare designs, and of course the designs themselves – for example, features of the user interface may prompt users to formulate better or poorer queries or to try more/fewer cycles.

At first sight, this redefinition of the system to be designed and measured might not seem to require much change to the evaluation method. Simply set the user the retrieval task, take what they finally select at the end of a session as the result, and again consult the stored "answers" in a test collection in order to measure the combined performance. In addition, direct observation of the users (for instance, by think aloud protocols) would yield useful formative information about how user interface features affect performance and could be improved. However, things are not so simple (3, 4, 5, 6, 7, 8) and have begun to be addressed by the interactive track of TREC (9) and Mira (10). The first problem is that of how to "set the user the retrieval task". In test collections, these are often specified by the query that would be typed directly into the software. But the formulation of that query, given a goal in the user's mind, is one of the major steps in the overall task, and it strongly affects the outcome. Consequently in many cases test collections would have to be rebuilt with search tasks specified in more realistic ways; and furthermore should be backed by other studies of what tasks occur in actual work places, and in what forms. Borlund (11) is currently researching the use of "simulated work tasks" as a way of addressing this problem while Reid (12) is investigating including task aspects in test collections. However this would still only address those tasks where the user begins with a definite and articulated retrieval task. But only a little observation of real users shows that a lot of retrieval concerns browsing, not just as a method but as a type of goal where the user just looks for something "interesting", not something definitely known in advance. There is a growing need to increase the usage of HCI type evaluations to the IR field. There have been a number of user studies of information retrieval software (e.g. 13, 14), often published in the HCI rather than IR literature. These need to become a standard component of IR research. Rasmussen and his collaborators have developed a comprehensive framework of the issues in human-machine-work interactions that can be used to guide evaluation (15, 16, 17). It covers a wide spectrum of evaluation from low level issues such as the user noticing and understanding the output from an interaction through to measuring how the IR system has helped them achieve their work goals.. However, a difference in the evaluation of IR systems, which has started to be addressed at Risø, is the linking of interface design evaluation with the performance of the underlying IR engine, whereas the design of the interface of a word processor, say, may require a lot of improvements but there is little need to measure the accuracy of the word storage facilities. In contrast to this complex framework, Harper & Hendry presented the notion of Evaluation Light (18): concentrating on using very focused small experiments to answer constrained questions concerning users' interaction with IR systems (similar in spirit to Andrew Monk's work on lightweight HCI evaluation techniques (19)). Another lightweight technique in IR is to use limited user modelling combined with the test collection approach (20).

Analysing retrieval from non-textual collections, such as collections of paintings or photographs provides an insight into the problems of evaluating state-of-the-art IR systems. One approach is to use traditional meta-tag indexing to provide access to the image by attributes such as photographer, date of photograph and similar external information. To access the content of the images, we could add a set of keywords ("meta-tags") to each image. In this way, textual queries are used to retrieve non-textual documents: cross-medium retrieval. A second use of associated textual descriptions is to index text which already exists and is, somehow, related to the image (e.g. 21, 22, 23) as found in, say, web-based art collections (24). This approach avoids the problem of a human having to create text just to make the images indexable. The alternative to using text in any form is to analyse the content of the image but this leads to a multitude of problems: high level attributes are very hard to extract, low level attributes may bear little resemblance to items users would wish to search for and there is a much wider set of possible relevance relationships for images than for texts. Most image search systems currently use techniques such as colour histogram and texture matching (25, 26, 27) between query and document images, possibly in combination with main object shape detection. While these approaches have shown considerable success in finding images which are visually similar to each other, it is extremely hard to move away from this visual similarity to a more semantic matching: there are only a limited number of tasks in which you are looking for an image and know the texture and colours of the matching set. It is extremely difficult to devise evaluation approaches which categorically answer questions such as which of these techniques is most suitable for which users.

Field studies of how IR is used in real work are particularly important as we face the problems posed by the new horizons in IR. For instance, a study of a commercial image bureau showed, among other things, that in this business at least, image retrieval is done by text queries not because that is the only thing current technology supports but because that is how the customer specifies and thinks about what they want. Similarly, as mentioned above, they can uncover kinds of relevance that IR engines, so far, have almost no way of representing. As new approaches to IR evaluation worry about what kinds of user task really exist and matter in practice, workplace studies can collect them. Studying new classes of IR user, for instance WWW users or school children (28), shows how these users do not come with any prior search skills at all: success of IR software here will depend either on having the interface communicate such skills or else by avoiding the need for them altogether. However, workplace studies are expensive to do, as they absorb many hours of investigator time (although they are invaluable for the above reasons). Their expense however means that they will not replace other kinds of study. Thus HCI studies in which participants representing users are invited in to use software will retain a place in IR evaluation e.g. for rapid improvement of the user interface. It is likely too that the benchmark style of study using recall and precision will retain some place. Combinations are likely to become more important: for instance, inviting participants into the lab not to use the whole program but to test a small part of it against benchmark measures.

Future work will be characterised by attempts to explore basic tensions in direction. One is the tension between highly standardised workbench tests using precision and recall and no human users (fast and highly comparable with the work of others thus good for competitions, but with doubtful relevance to any real work applications) vs. workplace studies (highly valid, but expensive and of doubtful comparability with each other). Another is the tension between comprehensive evaluation using the Rasmussen et al. framework exhaustively vs. lightweight techniques that are more often affordable in practice. One of the main directions to address is that of characterising, and measuring performance with respect to, other types of user task than those specified by explicit pre-given queries. If evaluation is to correspond to large amounts of current retrieval in practice, it must find a way of measuring how well a retrieval session went with respect, not just to concrete goals like "Rembrandt's last painting", but to "browsing" goals of just looking for something interesting, and also to explicit but vague goals such as looking for a "nice" or "novel" or "beautiful" picture.

Our proposal is, using the diverse skills across the network, to adapt and apply a wide range of evaluation methods to ISEs: methods originating from psychology and sociology as well as computer science, and already often applied in areas such as HCI and CSCW as well as in workplace studies done for other reasons. We will apply them to old and also to some of the new technological opportunities in IR (e.g. large image collections). Finally in addition to developing methods and accumulating cases of their application, we will work on developing a framework to combine them in a unified approach.

Our first impulse had been to develop a multimedia test collection. However it is now apparent that this is much too ambitious, and also in some ways backwards looking. Firstly, the effort just of collecting and organising the documents is a major research direction of its own (now often called "digital libraries"). Secondly, assembling expert judgements of relevance for the whole of such collections seems likely to be beyond reach, given that technology can now support such big collections. Thirdly and most importantly, however, is that many of the objections to test collections that were worrying for text documents now seem overwhelming in the context of new technology: could anyone really identify a "representative task" for all users and all media? Instead, it is clear on the one hand that new methods are already proving valuable and need to be explored widely in ISEs, and on the other hand that much exploration is needed to identify what the important issues are in the new technological contexts. Only then can we begin to formulate the questions that evaluation of ISEs should address, and the methods by which each question might be answered. These are our aims, and the usefulness of the methods in other areas and of some studies in the IR field allow us to expect substantial progress.

References for section 2 (scientific originality)

(1)   CLEVERDON, C.W., MILLS, L., AND KEEN, M., Factors Determining the Performance of Indexing Systems, ASLIB, Cranfield Project, Cranfield. 1966.
(2)   HARMAN, D.K. "The TREC Conferences". In R. Kuhlen and M. Rittberger (Eds), Hypertext - Information Retrieval - Multimedia: Proceedings of HIM 95, Konstanz, Germany. 1995.

(3)   COOPER, W.S. "On selecting a measure of retrieval effectiveness". Journal of the American Society of Information Science, **24**, pp. 87-100, 1973

(4)   SARACEVIC, T, "Relevance: A review and a framework for the thinking on the notion in information science". Journal of the American Society of Information Science, **26**, pp. 321-343.1975

(5)   BELKIN, N.J., AND VICKERY, A. Interaction in information systems: a review of research from document retrieval to knowledge-based systems, British Library – Library and information research report 35, 1985.

(6)   BORGMAN, C.L. "All users of information retrieval systems are not created equal: an exploration into individual differences", Information Processing and Management, **25**(3), pp. 237-252,1989.

(7)   BATES, M.J. "Where should the person stop and the information search interface start?", Information Processing and Management, **26**(5), pp. 575-591, 1990.

(8)   MIZZARO, S., "How many relevances in information retrieval?", Interacting with Computers, **10** no.3 pp.305-322, 1998.

(9)   BEAULIEU, M., ROBERTSON, S. AND RASMUSSEN, E. "Evaluating interactive systems in TREC". Journal of the American Society of Information Science, **47**(1), pp. 85-94, January1996.

(10)   DRAPER, SW, DUNLOP, MD, RUTHVEN, I, AND VAN RIJSBERGEN, CJ, *Proceedings of Mira 99 Conference*, Electronic workshops in Computing, 1999.

(11)   BORLAND, P., AND INGWERSEN, P., "The development of a method for the evaluation of interactive information retrieval systems", Journal of Documentation, **53**(3) pp 225-250, June 1997.

(12)   REID, J, "A task-oriented, non-interactive evaluation methodology for information retrieval systems", *Journal of Information Retrieval*, in press 1999.

(13)   KOENNEMANN, J, AND BELKIN, N.J. "A case for interaction: a study of interactive information retrieval behavior and effectiveness", CHI96 Conference Proceedings, (Edited by M.J. Tauber, V. Bellotti, R. Jeffries, J.D. Mackinlay, and J. Nielsen), pp. 205-212, 1996.

(14)   PIROLLI, P., SCHANK, P., HEARST, M., AND DIEHL, C. "Scatter/gather browsing communicates the topic structure of a very collection", CHI96 Conference Proceedings, (Edited by M.J. Tauber, V. Bellotti, R. Jeffries, J.D. Mackinlay, and J. Nielsen), pp. 213-220, 1996.

(15)   RASMUSSEN, J., PEJTERSEN, A.M. AND GOODSTEIN, L.P. Cognitive systems engineering (Wiley: New York) 1994.

(16)   PEJTERSEN, A.M. "Emperical work place evaluation of complex systems", Proceedings of the 1st International Conference on Applied Ergonomics. (ICAE'96), pp21-24, Istanbul, Turkey, May 1996.

(17)   PEJTERSEN, A.M., AND FIDEL, R. *A framework for work centred evaluation and design: a case study of IR on the web*, Glasgow University Mira Research Report, TR-1999-35, 1999.

(18)   HARPER, D. AND HENDRY, D. "Evaluation light", Proceedings of the Second Mira Workshop (Edited by M.D. Dunlop), Glasgow University Research Report, http://www.dcs.gla.ac.uk/mira/workshops/padua_procs,1996.

(19)   MONK, A. F. "Lightweight techniques to encourage innovative user interface design". In L. Wood (Ed.), User interface design: Bridging the gap from user requirements to design , CRC Press. pp. 109-129. 1998.

(20)   DUNLOP, M. D. "Time Relevance and Interaction Modelling for Information Retrieval", Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR97) (Edited by N.J. Belkin, A.D. Narasimhalu and P. Willet), Philadelphia, pp. 206-213, 1997.

(21)   DUNLOP, M. D. Multimedia Information Retrieval, PhD Thesis, Glasgow University Computing Science Research Report 1991/ R21, October 1991.

 (22)   FRANKEL C., SWAIN M. J., AND ATHITSOS V. WebSeer:An Image Search Engine for the World Wide Web, University of Chicago Technical Report TR-96-14, 1996.

 (23)   SMEATON, A.F. AND QUIGLEY, I. "Experiments on Using Semantic Distances Between Words in Image Caption retrieval", Proceedings of the 19th International Conference on Research and Development in Information Retrieval (SIGIR96), Zurich, pp.174-180, August 1996,

(24)   HARMANDAS, V., SANDERSON, M., AND DUNLOP, M.D. "Image retrieval by hypertext links", Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR97) (Edited by N.J. Belkin, A.D. Narasimhalu and P. Willet), Philadelphia, pp. 296-213, 1997.

(25)   FLICKNER, M., SAWNHEY, H., NIBLACK, W., ET AL. "Query by image and video content: the QBIC system", IEEE Computer, **28**(9), pp. 23-30, September 1995.

(26)   EAKINS, J.P., HARPER, D.J., AND JOSE, J. Proceedings of The Challange of Image Retrieval (Edited by J.P. Eakins, D.J. Harper and J. Jose), Electronic Workshops in Computing, to appear 1998.

(27)   FOUNTAIN, S., AND TAN, T. "Content based annotation and retrieval in RAIDER", Proceedings of IRSG98 (Edited by M.D. Dunlop), Electronic Workshops in Computing, to appear 1998.

(28)   FIDEL, R., AND CRANDALL, M. "User's perception of the performance of a filtering system", Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR97) (Edited by N.J. Belkin, A.D. Narasimhalu and P. Willet), Philadelphia, pp. 198-205, 1997.

3. Research method

As a research training network, the basic approach is organised around individual projects suitable for young researchers. Sites have proposed at least two examples each of such projects (see below for sample projects), although these may change according to developments within the network, and to suit the researchers employed. The important criterion for any project undertaken is that it directly address one or more of objectives 1,2,3. Our aim is thus to build up substantial collective experience in the form of these cases of evaluation methods, retrieval technologies, and workplace studies.

The network will additionally hold joint workshops at least once a year, and these will be the locus for pooling this experience and making progress on objective 4: the creation of a general IR evaluation framework. These will continue and intensify the Mira workshops whose success gives us confidence that this method will be productive. Reports from some of the Mira workshops can be seen at: http://www.dcs.gla.ac.uk/mira/workshops/

The training objectives will be met by hiring young researchers with a prior training different from the host site, and training them in the approach that site specialises in. In addition, provision has been made for individual inter-site visits by the young researchers. Thus training objectives will be met by the selection of researchers and their mobility, while the scientific objectives will be met by their individual projects, and by the network workshops.

Projects for young researchers at each site

Due to space limits, only one example project per site is given here. A longer list is available at
http://www.psy.gla.ac.uk/~steve/gauguin/projects.html

Collaborative filtering using logs (Glasgow)

An evaluation of the collaborative use of logs of information-seeking activity to filter and recommend multimedia information, both in competition and in combination with existing content-based retrieval systems.

Usability of still image and video retrieval systems (CLIPS IMAG)

Several methods exist for the extraction of the still image and video documents. For still images: colour, texture, shape extraction, and assignation to elements of the indexing vocabulary. For videos, analysis of the audio track, the image track are also a source of information for the content representation. The problem addressed here is the study of the adequacy of the fusion of all these heterogeneous sources of information according to the usability of the systems, according to query sessions and also to browsing. We will also focus on the need for additional sources of information (for instance textual descriptions, scripts for videos) to enhance the usability of such still image and video retrieval systems.

Smart Media (GMD)

Internet-based "ezines", electronic newspapers on demand, and other new information services revolutionised the mass media. The next step will integrate them into the users' personal information environment. GMD-IPSI participates in the development of Internet-based information-on-demand systems which organise new items according to user interests enriching them with more background information if requested. This entails the use of conceptual information retrieval and collaborative filtering techniques. The young researchers will develop adequate evaluation methods.

Analysing engineers during co-operative work (Risø)

A young researcher will be involved in the project on Ecological Information Systems that addresses the development of principles for design and evaluation of multimedia systems that support information seeking in complex co-operative work environments. This is a cornerstone activity of the Centre for Human Machine Interaction and it investigates the information seeking practices of engineers during co-operative work, the information they need, the heterogeneous sources they use and the strategies they apply, such as similarity searching.

Collective group information seeking (Robert Gordon)

Extend and develop existing information seeking tools (e.g. SketchTrieve) for collective use by (small) groups of users, and experiment with the resultant environment in an actual work setting, e.g. assisting librarians offering WWW search assistance to remote user clients.

Intranet information seeking evaluation (Ubilab)

Ubilab aim to develop and evaluate innovative tools based on results from our current Informia project in collaboration with real user communities within our organisation. In particular, tools that take into consideration user preferences, context, and task will

be considered in combination with traditional techniques such as text/web indexing and retrieval systems and newer developments such as information mediation systems. In this project the young-researcher will evaluate tools & techniques for information seeking in a large real-world intranet environment in the financial domain, involving analysis of domain, documents, tools, tasks, users, and information needs.

Information retrieval models (Dortmund)

We want to consider user-oriented data (e. g. relevance feedback, profile, interactivity) in information retrieval models. Our probabilistic approaches allow already to incorporate various facets of the information retrieval process such as uncertain document representation and vague querying. Visiting researchers could add the user facets to our approaches.

Integration of novel information processing technologies (Sheffield)

Using simple automatic natural language extraction and summarisation techniques to complement statistical IR approaches to support interactive searching. This project will include laboratory-based retrieval tests incorporating user involvement.

Multilingual IR (Tampere)

Work is required on application of IR and HCI evaluation techniques to multilingual IR. The researcher will gain experience in working with multilingual document collections, query-construction approaches and word normalisation for IR.

4. Work plan

The network will last for four years, while individual young researchers will be employed for between one and three years. This allows flexibility for each site in recruiting the best available young researchers. Each young researcher will carry out an individual project, of the kind described in the previous sections, and will produce a report on it by the end of their employment.

The network will hold joint workshops at least once a year. A report based on each workshop will be produced within a month after each workshop, and sent to the commission. The last workshop will be held near the end of the funding period. These reports will thus provide approximately annual reports to the commission (but their exact times will depend upon the workshop dates rather than the calendar); and the final workshop will lead to the final report.

As specified further in section 7 below, these reports will give progress on the joint objective 4, the progress reported by each young researcher, and allow monitoring of each site's progress in employing young researchers. Thus progress may be assessed from each of these reports by: consulting the activity report from each site to check whether they have recruited researchers (and their plans for this); for each researcher already employed for a year or more, checking on their reports which should be included; and finally looking at the report for the section on progress towards an overall framework (objective 4).

Professional research effort on the network project			
Participant	Young researchers to be financed by the contract (person months) (a)	Researchers to be financed from other sources (person months) (b)	Researchers likely to contribute to the project [(a) and (b) financed] (number of individuals) (c)
1: Glasgow	36	12	4
2: CLIPS-IMAG	24	12	6
3: GMD-IPSI	23	12	3
4: Risø	36	12	4
5: Robert Gordon	33	12	4
6: Ubilab	30	12	4
7: Dortmund	26	12	6
8: Sheffield	36	12	6
9: Tampere	34	12	4
Totals	278	108	41

N.B. The network lasts for four years, and the site staff contributions (column (b)) are expected to last throughout that period at all sites, even though the young researchers are likely to be employed for only part of that period at each site.

5. Collective expertise

This network has a diverse collection of sites which will be grouped here under the themes: workplace evaluation; laboratory based evaluation of novel IR systems; and application of HCI, IR and IS (information science) theory to evaluation of information seeking environments. This section first describes these themes, followed by an overview of existing linkages between these sites, application domains to be targeted by Gauguin through the industrial connections of each site, and finally details of each individual site. However, such a thematic grouping of sites is only part of the story – one of the underlying aims of Gauguin is to balance the knowledge and experience at each site through the exchange of young researchers, and this is motivated by a strong interest from sites in all aspects of the project that has already been demonstrated in Mira.

Participant roles

Tampere, Risø, Robert Gordon, Sheffield and Ubilab will look at workplace evaluation of information seeking environments. Mainly through the REGIS group, Tampere has been involved in studies of task-embedded information searching in various work environments as well as on the use of electronic networks by professionals and lay persons. The project at Risø will be associated with the Ecological Information Systems project there that is addressing the development of principles for design and evaluation of multimedia systems that support information seeking in complex co-operative work environments. Recently, Robert Gordon has arranged with the Macaulay Land Use Institute (Aberdeen) to conduct workplace trials of WebCluster using information provided by CAB International, a large supplier of scientific agricultural information based in England. At Ubilab, the IT innovation laboratory of UBS, the aim within Gauguin is to develop and evaluate innovative tools based on results from their Informia project in collaboration with real user communities within UBS. Sheffield is currently involved in a major project with Glaxo Wellcome Research and Development Ltd which will investigate user information needs assessments of different user groups and their information seeking behaviour for the design of corporate information systems in the pharmaceutical sector. Tampere and Risø both have portable usability laboratories that can be exploited for workplace evaluation, in addition Risø has eye tracking equipment which can be used for closer monitoring of end-users' interactions.

A major strand of work within Gauguin will be the application of laboratory evaluation techniques to novel information retrieval systems: either novel interfaces and information access methods to traditional collections or interfaces to state-of-the-art information seeking systems. The Robert Gordon University have been involved in three main projects on this theme: WebCluster which is a set of tools that allow information sources to be automatically structured according to semantic themes, and these structured collections can then be used to mediate access to the WWW (via search engines); Flair is an object-oriented framework for constructing IR servers, an example of which is their EPIC photograph retrieval system; SketchTrieve is an information seeking environment, which enables a user to co-ordinate searching over multiple search engines and sources via a 2D searching "canvas". Extensive end-user evaluations are planned for both the SketchTrieve tool and the WebCluster client. The University of Dortmund focus on probabilistic information retrieval models, integration of database and retrieval systems, hypermedia retrieval, network retrieval, and digital libraries. Within Gauguin young researchers are expected to exploit document collections (large textual, image and video collection), related representations of the semantics of the documents which are kept in database management systems such as Postgres and Oracle and various search engines which are hosted at Dortmund. Members of the Sheffield Group have been long-standing participants in the TREC Interactive track which focuses on user involvement in laboratory-based evaluation. They have also carried out the testing of the Okapi advanced probabilistic IR system in operational library settings which has also included the evaluation of highly interactive interface environments for query expansion. Together with Dortmund and Tampere they are interested in running user experiments to test other complex search interaction environments for different interactive retrieval tasks such as multilingual retrieval (a technology that is just reaching end-user availability, but on which there has been little evaluation work). CLIPS-IMAG and Glasgow University are also interested in developing user evaluation techniques which will assess the worth of content-based image and video retrieval and collaborative filtering / recommender approaches. CLIPS-IMAG has been involved in two main projects in the area of content-based image retrieval, FERMI with Glasgow and other partners (see below), and DIVA with the National University of Singapore and Kent Ridge Digital Laboratories of Singapore on indexing and retrieval of home videos plus the individual work of Georges Quénot on video segmentation. Novel interfaces being developed at Glasgow include path models for aiding browsing, collaborative filtering systems and image retrieval systems for both 2D and 3D images. GMD has been involved in the HERMES project (Esprit 9141) on developing methodologies for indexing, presentation, and retrieval of multimedia information, especially continuous data provision aimed at scientific applications and novel information services. The research at GMD within the context of Gauguin will focus on the problems of content-oriented filtering, indexing, and retrieving structured multimedia objects. The approach combines methods from logic-based probabilistic information retrieval (IR) and object-oriented database theory. As the users and their behaviour - as individuals and groups - essentially determine the features and uses of a digital library, the research may embrace the issues of designing problem-specific ways of providing access to digital libraries, such as pull and push services,

information brokering, recommender systems, and embedded applications, such training and teaching environments, virtual companies etc.

The Universities of Dortmund, Glasgow and Sheffield plus CLIPS-IMAG, GMD-IPSI and Risø have strong theoretical and practical backgrounds in the domains of HCI and IS. This will be the main thrust behind work on applying results from these fields to the application domains of Gauguin and to the problems of evaluating dynamic interactive information seeking environments.

The Robert Gordon University, CLIPS-IMAG, Risø and Ubilab have done previous work on frameworks for evaluation which will underpin the network's objective 4: the development of a multimedia evaluation framework.

Existing linkages

As mentioned above, the proposed network stems from the Mira working group [Esprit 20039]. Glasgow, which led Mira (and its predecessor MIRO [6576]), was strongly linked with twelve other institutions, five of which are participants of this proposed network: Dortmund University (Germany), GMD IPSI (Germany), Robert Gordon University (UK), UBS Ubilab (Switzerland), and University of Tampere (Finland). The University of Sheffield's contribution to Gauguin is effectively a continuing association, since participants from City and Glasgow universities have recently moved to Sheffield. Similarly, a Glasgow member of staff (Mark Dunlop) is in the process of moving to Risø: another case of the active collaboration already existing between the sites of this proposed network. Annelise Mark Pejtersen at Risø was also heavily involved in the Mira working group events although not a formal member of the consortium.

In addition CLIPS-IMAG, Dortmund and Glasgow were also involved in the FERMI project [8134] together with CNR in Pisa. Along with other sites, CLIPS-IMAG, GMD-IPSI and Glasgow were involved in the IDOMENEUS Network of Excellence [6606]. At a more individual level Robert Gordon University and Ubilab have run joint research projects, Tampere and Sheffield are currently establishing other funded joint research, and short staff exchanges have taken place between Glasgow and CLIPS-IMAG.

Site information

1. University of Glasgow

The Computing Science Department is world class in its research, as demonstrated by its achievement of the highest grading (5*) in the last UK Research Assessment Exercise, and is also top rated in its teaching. The research activity in the department demonstrates particular strengths in the theory and mathematical underpinnings of information retrieval, for example the work of van Rijsbergen; in novel approaches to interactive IR, as in the work of Chalmers and Campbell; in exploring multimedia information storage, as in the Revelation project; and in Human-Computer Interaction, which is the central topic of the Glasgow Interactive Systems group (GIST). GIST has a particularly multidisciplinary mix, including members from the Department of Psychology such as Steve Draper, whose past research includes the adoption and application of methods from psychology to HCI. This group will contribute a strong user-centered perspective to the network, and has supported recent work in the IR group that has extended its formal emphasis to work involving empirical user studies.

As a training environment, Glasgow CSD already supports a thriving community of 68 research students. Our Graduate School's growth over recent years cuts against the trend in Scottish universities. Four recent and ongoing PhDs have focused on users and interaction in retrieval. Apart from the staff listed below, the IR group involves nine other members, predominantly postgraduate students, and we also have strong links to other local groups such as that in Strathclyde University's Information Science Department. The collective depth and breadth of expertise is just one contributor to the quality of the environment for young researchers. Glasgow is one of the premier sites for the evaluation of IR software, especially with respect to its utility and usability for individual users. In addition, we can draw upon the administrative expertise built up through the department's involvement in many EC projects over the past years, and also upon a university unit devoted to the support of European projects.

2. CLIPS-IMAG

The MRIM group (Modelling Multimedia Information Retrieval) group is part of the research laboratory CLIPS-IMAG together with several groups all focused on the general research topic of "language based communication and person-system interaction". The group aims at developing models and systems for information retrieval of text, image and video, and for a number of arising application fields of retrieval, namely collective retrieval and Web searching. One of the important aspects of MRIM's work is to carry on experiments to evaluate systems and models.

The MRIM group currently works on the following areas in IR:

- Collective information retrieval including work within the RICOM project on synchronous collective information retrieval and research student work on asynchronous aspects of collective information retrieval;

- Image and video indexing and retrieval including paricipation in the FERMI BRA, the DIVA project with the National University of Singapore and Kent Ridge Digital Laboratories of Singapore on indexing and retrieval of home videos, and work on video segmentation;

- Textual documents indexing and retrieval including the IOTA project which aims to build a full text retrieval system that emphasises precision by automatic natural language processing used in combination with manually built terminological bases (the system is periodically evaluated in the context of "Amaryllis", French TREC-like organisation, and the experiments are to be scaled up to test the impact of huge corpuses such as the Web). Other activities on text and hypertext retrieval involve PhD thesis students who are currently working on the role of hypertextual links for the retrieval of Web pages and the use of datamining techniques to extract knowledge from textual corpuses in order to be used for indexing and/or retrieval

3. GMD-IPSI

GMD - Forschungszentrum Informationstechnik GmbH, the German National Research Centre for Information Technology, conducts research aiming at the development of innovative methods and applications. It co-operates closely with industry and users, thereby increasing the competitiveness of the German and European economies.

The Integrated Publication and Information Systems Institute (IPSI) is one of the eight institutes of GMD. The goal of IPSI is to develop and evaluate concepts, foundations and system prototypes for the next generation of distributed and co-operative multi-media information systems providing tailorable, active, knowledge-based support for the whole range of publication and information activities in a primarily digital information systems environment.

The following are examples of projects carried out at IPSI which are relevant to Gauguin:

- HERMES (Esprit 9141): Development of methodologies for indexing, presentation, and retrieval of multimedia information, especially continuous data (e.g. video), aiming at scientific applications and novel information services.

- ProCORDIS (EU industrial project): Improved user access to the CORDIS databases through probabilistic retrieval, abductive logic, automatic multilingual indexing.

- TREVI (Esprit 23311): Together with leading newswires (eg Reuters) we develop personalised news systems which filters incoming streams of textual data from newswires, enriching them with background information based on user profiles.

- TV ONLINE (Industrial project) Web-based TV guide project incorporating recommender functions and compilation of personalised TV schedule using collaborative filtering.

- AUTOSOFT (Esprit 25762): Indexing and retrieval design for software reuse libraries, including semi-automatic domain generation, user profiles, dynamic user interface definition.

- HAWK (IE 8038) pilots an open knowledge-based publishing model that is meant to enable publishers to better exploit their information assets.

4. Risø National Laboratory

The aim of Centre for Human Machine Interaction is to provide a forum for scientific approaches to the analysis of human behaviour in dynamic, changing, and co-operative work situations, and to use this forum as a platform for the development of novel principles for the design of interfaces that visualise the content of complex work domains in a transparent way. The centre is funded by the Danish National Research Foundation. The research plan is based on scientific approaches that presuppose intensive cross-disciplinary collaboration between centre researchers situation between centre researchers situated at various locations in the country. Among the important objectives for the centre are a tight collaboration and improved synergy among the different approaches to human-machine interaction developed at Risø (Cognitive Systems Engineering) and at the University of Aarhus (Activity theory and computer semiotics) during the last decades. The need for cross-disciplinary research into the analysis and design of Human Machine Interaction has been met by gathering a team of around twenty researchers in computer science, the humanities, and various engineering disciplines.

Risø has cross disciplinary expertise in system design and evaluation of system functionality and interfaces based on the requirements from the work domain. The work place requirements and the work content used as a basis for system evaluation and redesign are related to a framework that structures the dimensions, or categories, of domain information which need to be available for a user. These dimensions include information about the work domain and various related tasks, decision-making activities, division and co-ordination of work, and social organisation. Systems designed to meet these requirements are called ecological information systems. The use of the cognitive systems engineering framework includes field studies at the work place as well as experiments in the laboratory. Risø has experimental facilities to include remote eye-tracking devices, combined digital head and eye-tracking, and a full-scale portable usability laboratory. The facility has been established to strengthen our

experimental research basis on the Human Machine Interaction areas of (1) systems use and analysis of usability test methods and simulations, (2) test and analysis of computer interfaces within information retrieval, and (3) field studies of co-operative work activities. This Human Machine Interaction laboratory aims at facilitating our scientific understanding of usability aspects of people's interaction with computer interfaces based on data gathered both in field studies and laboratory experiments with human subjects.

5. The Robert Gordon University

The School of Computer and Mathematical Sciences at The Robert Gordon University has a developing reputation for high quality research in the areas of interactive systems (including information retrieval) and intelligent systems. There is a thriving community of research-active academics, full-time researchers, and research students, and they are supported by a modern computing environment. In the context of the Network, young researchers will work alongside existing researchers. Additional research training will be provided using relevant modules from the specialist MSc Computing (Information Engineering).

The Multimedia Information Retrieval Research Group at Robert Gordon develops models, tools and systems for content-based retrieval of multimedia information. Recent work of the group has focussed on software architectures for constructing information retrieval (IR) systems, novel information seeking tools for accessing information on the world-wide web, and models and tools for multimedia retrieval of still photographic material (including text, image and attribute data), and latterly video material. Our work is firmly based on the principles of user-centred design, and on rigorous evaluation involving both traditional test collection approaches and the emerging end-user evaluation methods.

Recent projects of the group include the following. WebCluster is a set of tools, which allows information sources to be automatically structured according to semantic themes, and these structured collections can then be used to mediate access to the WWW (via search engines). Flair is an object-oriented framework for constructing IR servers, an example of which is our EPIC photograph retrieval system. SketchTrieve is an information seeking environment, which enables a user to co-ordinate searching over multiple search engines and sources via a 2D searching "canvas". It was constructed using our IR component framework, FireWorks. SketchTrieve was designed by considering the requirements of end users to co-ordinate, organise and plan searches involving multiple search engines and document repositories.

Formative evaluations of the SketchTrieve tool and WebCluster client have been conducted using simulated work environments and tasks. A summative evaluation of the EPIC photographic retrieval tool was conducted using a simulated graphic design task, and real graphic designers, and the effectiveness of spatial querying was established for this task.

6. Ubilab

Ubilab is the IT innovation laboratory of UBS, one of the world's largest financial institutions. Ubilab performs a dual role of keeping tight links to the international research community as well as serving UBS as an effective consultant and partner in the application of new technologies. Ubilab is involved in several research collaborations with external academic and industrial partners, and is also involved in a range of internal projects within the Bank. As an effective bridge between the research community and industry, Ubilab is in a position to rapidly test out and evaluate new research results for industrial application, while at the same time bringing business related problems and experiences into the research community.

Ubilab has been active in the area of information retrieval since 1994. We were a member of both the MIRO and Mira projects, and in 1996 we co-organised the ACM SIGIR conference on information retrieval in Zurich. Our research over the past years has focused on application frameworks for multimedia information retrieval applications (our FIRE project) and mediator systems for access to heterogeneous and distributed information sources (our Informia project). Apart from our expertise and interests in these areas, we have a strong interest in innovative interactive systems for information seeking in large intranet environments. As a financial institution, the financial domain is naturally of particular interest to us.

7. University of Dortmund

The Information Retrieval Group at the University of Dortmund (IRG UniDo) is active in research on probabilistic information retrieval models, integration of database and retrieval systems, hypermedia retrieval, network retrieval, and digital libraries. The group develops IR systems (freewais-sf, wait, hyspirit, dolores) which are used for operational (freewais-sf is used by several hundreds organisations) and experimental purpose. A rich system and software environment for managing large-scale collections, experiments and evaluation of IR techniques is constantly maintained.

In the past nine years the group was involved in European and German projects: FERMI (Formalisation and Experimentation on the Retrieval of Multimedia Information), Mira (Evaluation of IR Systems), EuroSearch, EuroGatherer (classification and gathering of multilingual web pages), Medoc and Interdoc (digital library initiative of the German Department of Research).

In summer 1999, the Digital Library project Carmen/GlobalInfo will start. The aim of Carmen is to integrate Hypertext and Information Retrieval methods in Digital Library environments, to support high-level search strategies, and to evaluate the usability and effectiveness of Digital Libraries.

8. University of Sheffield

The Department of Information Studies at the University of Sheffield has consistently gained the highest rating as a centre of excellence with an international reputation in each of the Research Assessment Exercises which evaluate research performance of university departments in the UK. The proposed training research network would draw on the expertise of two of its research teams which are concerned with information systems from different perspectives. The Computational Information Systems Research Group focuses on the development and evaluation of novel techniques for the representation, searching and retrieval of textual, biological and chemical information. The Information Management Group is concerned with the impact of information systems and technology in organisations and the information seeking behaviour of user communities. Over a period of more than a decade, both groups have made substantive contributions to information research through the development of theoretical models for information seeking as well as empirical studies on information users and the design of information retrieval systems.

Research activities over the duration of the proposed Research Network Project will focus on a number of externally funded research projects including: the representation of chemical structures for drug discovery, the integration of natural language processing and information retrieval techniques for corporate information services, data fusion for multimedia digital libraries, user interface design and human-computer interaction issues for retrieval systems, learning styles and Internet searching, and cognitive models for information seeking. A common feature which links this range of projects, is that they all raise methodological issues for evaluation. These will provide a rich environment for young researchers to be exposed to different aspects of evaluation on a single site. The research groups at Sheffield are particularly keen to explore the integration of user-centred and systems performance approaches in the design of evaluative experiments across the different projects.

In addition to the unique position of having scientific staff within the same Department who have extensive experience in different evaluation paradigms, including quantitative and qualitative methods associated with laboratory and field testing, Sheffield will further develop its involvement in collaborative research based in operational settings. A current industrial partnership is concerned with the development of corporate information services for a large pharmaceutical international company. The project involves the design and evaluation of prototype systems based on real user needs with end user participation in the whole of the iterative design and development cycles. This initiative is part of the Department's research strategy to explore information rich work domains of which Health Informatics is a prime example. Funding is currently being sought to establish a Health Informatics Collaboratory to provide an appropriate infrastructure and environment for collaborative interdisciplinary research involving a range of participants across the health sector.

9. University of Tampere

The Department of Information Studies at the University of Tampere has a very good international reputation in both research areas information seeking and information retrieval. Research groups on information retrieval (FIRE – The Finnish Information Retrieval Expert Group) and information seeking (REGIS – Research Group on Information Seeking) were established in the beginning of 1990s to strengthen research activities. The present aim is to pursue further research in these areas and integrate them toward task-based information searching research.

The FIRE group has made a major contribution in developing laboratory based evaluation frameworks for text and image retrieval research and on applying these to evaluate thesaurus based query expansion, computer-linguistic methods in IR, query structures in Boolean and probabilistic IR, and content based image retrieval algorithms. REGIS has concentrated on studies on task-embedded information searching in various work environments as well as on the use of electronic networks by professionals and lay persons. Models of information searching in task context has been developed.

The department has expertise in several areas relevant for the Gauguin network:

- Laboratory-based IR research: mono-lingual and cross-lingual text retrieval as well as image retrieval; beginning in 1999 we are starting to work with digital video.

- A task-based test database for evaluating image retrieval has been currently developed. As far as we know it is the first of its kind and could serve as a test bed for the relevant subprojects in the proposed network.

- Information seeking research: IS in organisational task settings, IS by citizen in the WWW, IS and IR by journalists in their task settings, longitudinal studies on relations between problem stage in task performance and information searching and relevance assessments. These field studies use both quantitative and qualitative techniques of data collection and analysis.

- Information retrieval learning environments.

The department has good facilities for laboratory-based IR research through several large document collections, retrieval systems, NLP software and electronic monolingual and translation dictionaries. A task-based test database for image retrieval is in use. The department also provides basic equipment for qualitative field studies such as tape recorders, video camera and equipment aiding in transcribing tape-recorded interviews into written protocols. The department also has close connections to the HCI laboratory of the Department of Computer Science.

Staff time & recent publications from each site

This section gives an overview of staff at each site who will be involved in Gauguin, their positions, how much of their time will be given to research directly in line with Gauguin, and two recent publications related to Gauguin.

1. University of Glasgow

Dr. Matthew Chalmers, Computer Science Department, Lecturer, 10%

Dr. Stephen Draper, Psychology Department, Senior Lecturer, 10%

Prof. Keith van Rijsbergen, Computer Science Department, Professor, 5%

Crestani, F, Lalmas, M, Van Rijsbergen, C J, and Campbell, I. "Is this document relevant?...probably": a survey of probabilistic models in information retrieval', *ACM Computing Surveys*, Volume 30, No. 4 (Dec. 1998), pp. 528-552.

Draper, S W, Dunlop, M D, Ruthven, I, and Van Rijsbergen, C J. (eds) *Proc. Mira Conference*, Glasgow, April 1999. To be published by Kluwer in 1999, and also in electronic form as part of the *British Computer Society 'electronic Workshops in Computing' series*.

2. CLIPS-IMAG

Prof. Marie-France Bruandet, Head of MRIM Group, 5 %

Dr. Nathalie Denos, Lecturer, 5 %

Dr. Georges Quénot, Researcher, 5 %

Dr. Jean-Pierre Chevallet, Lecturer, 5 %

Prof. Yves Chiaramella, Head of CLIPS Laboratory, 5%

Chevallet, J.P., and Chiaramella, Y. Experiences in IR Modeling using Structured Formalisms and Modal Logic, in *Information Retrieval, Uncertainty and Logics - Advanced models for the representation and retrieval of information*, Fabio Crestani, Mounia Lalmas, Cornelis Jost "Keith" van Rijbergen, University of Glasgow Scotland, Kluwer Academic Publisher, october 1998.

Ounis, I, and Pasca, M. RELIEF: Combining expressiveness and rapidity into a single system, in *21st International ACM SIGIR Conference on Research and Development in Information Retrieval*, W.B. Croft and A. Moffat and C.J. van Rijsbergen and R. Wilkinson and J. Zobel Eds., ACM Press, Melbourne, Australia, August 24-28, pp266-274, 1998.

3. GMD-IPSI

Dr. Adelheit Stein, 20%

Dr. Ulrich Thiel, head IPSI's Digital Libraries Division, 5%

Gulla, Jon Atle; van der Vos, A.J. & Thiel, Ulrich. (1997). An Abductive, Linguistic Approach to Model Retrieval. *Data and Knowledge Engineering* 23(1): 17-31.

Stein, Adelheit; Gulla, Jon Atle & Thiel, Ulrich (1999). User-Tailored Planning of Mixed Initiative Information-Seeking Dialogues. *User Modeling and User-Adapted Interaction, Special Issue on Mixed-Initiative Interaction*, 8(1-2): (In press)

4. Risø National Laboratory

Dr. Annelise Mark Pejtersen, centre leader, 5%

Dr. Mark Dunlop, senior scientist, 10%

Dr. Morten Hertzum, senior scientist, 10%

Pejtersen, A. M. and Rasmussen, J. (1997): Effectiveness testing of complex systems. In: *Handbook of Human factors and Ergonomics*. Ed. by G. Salvendy, Wiley.

Rasmussen, J. Pejtersen A.M. and Goodstein, P.L: *Cognitive Systems Engineering*. Wiley, 1994.

5. The Robert Gordon University

Prof. David J. Harper, Research Professor, 5%

Dr. Ayse Goker-Arslan, Lecturer, 10%

Dr. Jihua Cheng, 10%

Hendry, D G, and Harper, D J. An informal information-seeking environment, *Journal of American Society for Information Science*, 48(11), 1036-1048, 1997.

Jose, J M, Furner, J, and Harper, D J. Spatial Querying for Image Retrieval: A User-Oriented Evaluation, In W B Croft, A Moffat, C J van Rijsbergen, R Wilkinson and J Zobel (Editors), *Proceedings of 21st Annual International SIGIR Conference on Research and Development in Informational Retrieval* (Melbourne, Australia, August 1998), ACM Press, pp. 232-240, 1998.

6. Ubilab

Prof. Dr. Hans-Peter Frei, Head of Ubilab, 5%

Dr. Tore Bratvold, Research staff member, 10%

Dr. Hardeep Singh, Research staff member, 10%

Barja, M L, Bratvold, T, Myllymaki, J, Sonnenberger, G: "Informia: a Mediator for Integrated Access to Heterogeneous Information Sources", in *Proceedings of the ACM Conference on Information and Knowledge Management*, ACM Press, 1998, pp 234-241.

Bratvold, T, Sonnenberger, G, Frei, H-P: "A Framework for Developing Information Retrieval Applications", in *Proceedings of the Ubilab Conference 1996*, UV Konstanz 1996, pp 51-64.

7. University of Dortmund

Prof. Dr. Norbert Fuhr, 2.5%

Diplom-Informatiker Norbert Gövert, 10%

Diplom-Informatiker Kai Großjohann, 2.5%

Diplom-Informatiker Claus-Peter Klas, 10%

Fuhr N., Gövert N., Rölleke T.: DOLORES: A System for Logic-based Retrieval of Multimedia Objects *21st ACM SIGIR Conference*, August 1998

Fuhr, N.: A Decision-Theoretic Approach to Database Selection in Networked IR, *TOIS*, 17, 1999 (To appear)

8. University of Sheffield

Prof. Micheline Beaulieu, Head of Department, 5%

Prof. Peter Willett, Director of Computational Systems Research Group 5%

Dr. Mark Sanderson, Lecturer, 5%

Dr. David Ellis, Director of the Information Management Group, 5%

Mr. Nigel Ford, Senior Lecturer, 5%

Beaulieu, M. & Jones, S. Interactive searching and interface issues in the Okapi best match probabilistic retrieval system. *Interacting with Computers*, 10, 1998, 237-248.

Sanderson, M. & Croft, B. Deriving concept hierarchies from text. To appear in Proceedings of the 22nd ACM SIGIR, Berkeley, 1999.

9. University of Tampere

Prof. Pertti Vakkari, leader of a long-term project on task-based IS, 10%

Prof. Kal Jarvelin, Head of Department, leader of the IR research & lab, 5%

Prof. Eero Sormunen, leader of the project on task-based image retrieval, 10%

Pirkola, A. & Keskustalo, H. & Järvelin, K. (1999). The Effects of Translation Method, Conjunction, and Facet Structure on Concept-based Cross-language Retrieval. *Journal of Information Retrieval*. (Accepted)

Vakkari, P. (1999) Task complexity, problem structure and information actions. *Information Processing & Management* 35. (Accepted)

6. Collaboration

Collaboration will occur at two levels: within each site, and between sites.

Within a site, young researchers employed by the network funding will arrive from other countries, often other sites in the network, and be trained in the techniques that that site specialises in (e.g. traditional IR evaluation, workplace studies, etc.). Thus each such project will involve the specialisation of the site being combined with other skills brought by the young researcher, who will be trained by collaboration with the site's personnel.

Between sites, the main vehicle will be workshops for all member sites. At these workshops, the work by the network's young researchers will be reported on. The workshops will furthermore be the main means for developing a new, combined framework for IR evaluation (research objective 4), drawing on and attempting to synthesise the particular specialities of the different sites. The workshops will be held at least once a year throughout the four years of the network. A report based on each workshop will be produced within a month after it takes place. As noted above, these workshops proved very productive ways of collaborating under Mira.

The second inter-site collaboration method will consist of short visits by many of the young researchers and some other participants to other sites than the one employing them, as budgeted for by individual sites.

7. Organisation and management

There will be two important levels of management in the network: at the site level, and for the network as a whole.

As a research training network, the main activity is the employment of young researchers organised by each site from its own budget. Each site is primarily responsible for recruiting its young researchers, and providing training and other support for them. They will ask each funded young researcher to produce reports on both a) the scientific content of their research, and b) on the skills they have learned. These reports will be due at the end of their employment and also at the end of each year of their contracts when these are longer than one year. Inclusion of published papers and any other materials (e.g. software) is expected and encouraged. These reports will be circulated to all the sites including the lead site, and often more widely (see below). In addition, every three months each site will send a simple statement of about five lines to the network co-ordinator summarising the current state of that site's plans and actions regarding the employment and progress of their young researchers.

The network as a whole will be administered by the co-ordinator at Glasgow. The network workshops (at least once a year) will allow for management meetings as well as their main business of pursuing the scientific objectives. A report based on the workshops will be produced within a month after each workshop, and sent to the commission. The last workshop will be held near the end of the funding period. These reports will thus provide approximately annual reports to the commission (but their exact times will depend upon the workshop dates rather than the calendar); and the final workshop will lead to the final report.

These reports will have four sections:

- A summary or copy of any of the research reports from the funded researchers delivered internally since the last report.

- A summary or copy of any of the training reports from the funded researchers delivered internally since the last report. This may include reports on any training sessions or exercises developed by them and run at the workshops.

- A report on our main scientific objective of developing an evaluation framework. This would be based on papers and other activities at the workshops.

- A brief summary of the activity at each site, especially progress in recruiting and employing the young researchers.

Dissemination within the network would mainly consist of the various reports detailed above, and would be primarily by means of the WWW. This method was gradually adopted during the lifetime of the Mira working group (see http://www.dcs.gla.ac.uk/mira/workshops), and proved satisfactory and convenient. Dissemination of the scientific results would be by published academic papers, and by making the scientific parts of the project reports available as printed technical reports, and on the WWW (subject to any copyright restrictions on published papers).

The lead site responsible for co-ordination will be Glasgow, which led the Mira working group which led to this proposal. Many of the organisational arrangements proposed here are based on those used successfully in Mira.

8. Training need

The essential point of the scientific case is that diverse approaches to IR evaluation exist (from technical testing of the software without users, to field studies of users in work settings) and that we propose to explore how they may be fitted together and applied to the new technological challenges emerging. This also means that there is growing awareness that new skills, but still more new combinations of skills, are needed. Thus there is a training need for researchers in the field of IR with a much wider set of evaluation skills than has been required up to now. It is no longer enough to use only technical testing of the software without users as in traditional computer science-based evaluation; but equally, it is not really enough only to employ someone with sociological skills for field studies in work settings. To address the new technologies in IR properly, researchers are needed who are able to perform both kinds of study (and others besides) and to understand when each is appropriate. There are currently no places who produce either graduates or postgraduates with this mix of skills: hence the basic training need. Conversely, if our network produces some researchers of this kind, they will be well placed for jobs in both software companies, information retrieval organisations such as national libraries, and research posts.

The nine sites in the Network represent this range of approaches, for example from the mainly technical emphasis at our German sites to the mainly user emphasis at our Scandinavian sites. This illustrates the European dimension to the diversity of skill, while the interest at all the sites in acquiring skills represented by other sites indicates the European-wide demand in the IR field for multiple types of IR evaluation skill. Moving researchers to work at sites in other countries will give them training and experience at new techniques, while causing them to reflect on the relationship between their new skills and their older skills.

Thus our scientific aims, training young researchers in new skills to equip them with a more multidisciplinary approach, and the diversity of the participating sites all fit together.

9. Justification of the appointment of young researchers

Young researchers to be financed by the contract				
Participant	Young pre-doctoral researchers to be financed by the contract (person months)	Young post-doctoral researchers to be financed by the contract (person months)	Total (a+b)	Scientific specialities in which training will be provided
	(a)	(b)	(c)	(d)
1: Glasgow	0	36	36	M-10, M-11
2: CLIPS-IMAG	0	24	24	M-10, M-11
3: GMD-IPSI	12	11	23	M-10, M-11
4: Risø	0	36	36	M-10, M-11
5: Robert Gordon	21	12	33	M-09, M-10, M-11
6: Ubilab	12	18	30	M-10, M-11, M-12
7: Dortmund	16	10	26	M-10, M-11, M-12
8: Sheffield	0	36	36	M-09, M-10, M-11
9: Tampere	10	24	34	M-10, M-11
Totals	71	207	278	

N.B. the discipline codes (column (d)) are required, and specified near the end of the Annex to the Guide for Proposers. These proposers' own views on the disciplines and specialities involved are indicated throughout the text and specifically commented upon in section 11.

Each site has picked salaries for the young researchers according to its experience of what is necessary. In this computer science related area, competition from industry often requires higher salaries to be offered than in other academic areas (and so fewer person months for a fixed amount of money). I t also has reduced the difference between pre-doc and post-doc levels.

Each site would be able to train more young researchers than is possible under the funding constraints. This network is a development of the Mira working group. Two of the Mira sites were unable to join as the network funding does not cover the real costs of hosting young researchers (only their salaries, not office space or equipment). Dividing the limit (1.5Meuros) among 8 sites leaves only enough for the person-months shown. However it is clear that even one of the staff at each site could usefully employ and supervise at least two researchers for three years each (72 person months per site). We have a wealth of experienced staff to supervise them, and this research area is enormous. Many sites will be, and in some cases already are, seeking additional funding from other sources.

Vacancies will be published:

a) By announcements within the participant sites. We will use the network email mailing list to distribute these, and agree to publicise each other's announcements locally. The university sites especially produce many students who may wish to enter this area.

b) We will draw up an email list of other sites and site contact people with similar interests, and also send announcements to them. E.g. other members of Mira who were unable to join this network; neighbouring sites with related research groups e.g. at Strathclyde University in Glasgow.

c) By email distribution to international research groupings e.g. IRList, and BCS-HCI. This has become a standard and effective means of advertising jobs in research areas.

We will not have any formal exchange arrangement for employing young researchers from other participant sites, but expect that in practice there will be many cases of movement between the sites. The length of appointments will depend upon each site, but many of the projects are suitable for one year duration.

10. Training programme

As stated in section 8 and elsewhere above, the point of this proposal is to equip young researchers with different skills from those of their original training. Each of the contributing sites typically is most expert in one IR evaluation method only, and the range of methods needed is distributed across countries with none providing the range now believed to be desirable. Collectively, the European-wide network is in a much stronger position than any one country can provide.

The main approach will be on training through pursuing a research project, individual to each young researcher, but using their host site's main methods in collaboration with, and under the personal supervision of, that site's personnel. As is normal for all pre-doctoral and post-doctoral researchers, they will be encouraged and supported in writing, publishing, and presenting scientific papers at the network workshops and at conferences, which are both budgeted for by each site. In addition money is allocated for additional visits by young researchers to other network sites. Some visits of this kind were in fact carried out during the Mira working group (even though it was not a requirement there): for instance a pre-doctoral student Pia Borlund visited Glasgow from her Danish institution and performed some of the experiments there for her PhD thesis.

The network workshops will provide a still more important forum for sharing results and methods, and for the young researchers to practice the communication of these. In fact, at least in the later years of the network, we anticipate that the main activities at these workshops will be performed by the young researchers. In addition, senior staff will put on tutorials at these workshops addressing techniques they know best. (For instance Annelise Pejtersen put on a tutorial exercise on her framework for cognitive work analysis at a Mira workshop, and this can be usefully repeated; and Matthew Chalmers will put on a tutorial on the collaborative filtering technique.)

In addition we will ask our young researchers to provide reports on the skills they have learned (as opposed to the scientific results they have obtained), illustrated by examples and data (a little like the portfolios artists and designers use to illustrate their training and achievements). These are expected to be useful to later researchers in directing their own learning. Furthermore, we will ask the young researchers to produce demonstrations and exercises for use initially in the Network workshops of the techniques they have learned. During Mira we produced several of these, and they were among the most successful activities, which we made a feature of the final conference. Although obviously in one hour's exercise a participant takes only a small first step towards acquiring a new skill, this first step is still much more powerful than reading a paper in allowing a participant to understand by experience what the technique means, and whether they are interested in pursuing it further.

We expect to emphasise the tutorials by senior participants in the early years of the network, possibly in a special "summer school" type workshop; while we hope to emphasise the exercises by young researchers towards the end of the network.

11. Multidisciplinarity in the training programme

As specified in section 8 above, this network proposal concerns adapting and applying multiple methods to evaluating ISEs. Some of these methods originated in other disciplines. For instance field studies in the workplace owe something to ethnographic techniques, which may be said to have come from social anthropology or sociology, but have since been used in HCI and especially in CSCW; and thinkaloud protocols originate in psychology, but have been extensively applied in HCI. On the other hand, the use of test collections in IR is a form of benchmarking, which is used widely in computer science. These methods, however, have all already been used to some extent in computer applications and fields such as HCI. Because of this, recognition and adoption of these methods is becoming fairly common in computer science: although it is still true that most computer science academics have not used such methods, nevertheless it is becoming accepted that HCI (and so its methods) should be part of most computer science courses. Thus they are at the stage where they could be said to be part of computer science, yet actually are still new and applied only in a few areas such as HCI and CSCW, and are currently much less common in the field of IR. Thus, while this network is pleased to contain some people whose research focus is outside computer science (e.g. Steve Draper at Glasgow, those at Tampere), many of the sites see the training they propose as belonging primarily to the information sciences (see the codes in column [d] of the table in section 9). It should be noted that nevertheless some are using methods that originated elsewhere, and that many information scientists do not (yet) use. The main purpose of this network is to adopt and develop this plurality of technique for the single application area of information science. The young researchers will be trained in the methods most used at their host site, which will typically derive from a different discipline than the ones emphasised where they first qualified.

12. Connections with industry in the training programmer

GMD-IPSI's TV-ONLINE project was a co-operation with TV Today (a subsidiary of Bertelsmann) and GMD-IPSI.

Risø has on going collaborations with Kommunedata, a Danish software development company.

Petrotechnics is a SME specialising in knowledge sharing tools for the oil industry. The Robert Gordon University, and more specifically, the IR group, is involved with Petrotechnics in a three year Technical Support Programme, through which we provide expert advice and consultancy on product development. Petrotechnics have developed a Knowledge Sharing Architecture, which enables users to drill down through data associated with a particular "asset", e.g. an oil drilling platform, and to share information concerning parts of the platform. This software would be highly suitable for conducting real end user studies.

Robert Gordon has submitted a proposal to establish a Teaching Company Scheme (TCS) with an small Aberdeen-based Highland "dress" hire company. The purpose of the TCS project is to develop a virtual dress hire shop, which would be assessible via an Intranet by hire wholesale partners in the UK and USA. We are involved in the design and evaluation of the multimedia interface of this system, and this would be an interesting system for the Network to evaluate.

Sheffield has had an ongoing collaboration with the pharmaceutical multinational company Glaxo Wellcome Research and Development Ltd. Previous projects have been concerned with the presentation and exploration of chemical structures for drug discovery and novel applications of Information Extraction in searching biochemical journals in collaboration with the publisher Elsevier. A current project is aimed at further developing existing software for automatic summarisation, information extraction and relevance-based query profiling and at modelling the information seeking behaviour of a range of client groups. Glaxo Wellcome are interested in the methodological development and longer-term significance of needs analysis.

13. Financial information

Financial information on the network project				
Participant	Personnel and mobility costs related to the appointment of young researchers (euro) (A)	Costs linked to networking (euro) (B)	Overheads (euro) (C)	Total (a+b+c) (euro)
1: Glasgow	139,661	52,004	38,334	229,999
2: CLIPS-IMAG	109,800	42,700	30,500	183,000
3: GMD-IPSI	115,105	35,000	30,021	180,126
4: Risø	127,591	22,409	30,000	180,000
5: Robert Gordon	110,000	41,000	30,200	181,200
6: Ubilab (see note)	*110.000*	*33,700*	*28,740*	*172,440 **
7: Dortmund	114,000	36,000	30,000	180,000
8: Sheffield	135,016	14,592	29,921	179,529
9: Tampere	118,000	34,000	30,400	182,400
Totals — EC funded	969,173	277,705	249,376	**1,496,254**
Totals — whole network *	1,079,173	311,405	278,116	1,668,695

The figures were derived by each site to satisfy their local costings, and to conform to the constraints on network funding.

* Ubilab funding will be provided by the Swiss government. Totals for EC funded part of the network and the whole network are shown separately.