ADVANCED TECHNOLOGY SUPPORT FOR INFORMATION MANAGEMENT AT FRIENDS OF THE EARTH

C.D.Clack, S.J.Gould, P.R.Lidwell, J.T.McDonnell
Department of Computer Science, University College London

INTRODUCTION

We report early results from a project to study the application of advanced technology to enhance information management in a medium sized enterprise where the collection, analysis and dissemination of information are key business processes. Our two-year TCD-funded project is a collaboration between University College London (UCL) and Friends of the Earth (FOE), a research and campaigning organisation with 65 full time employees and a turnover of about 3.5 million pounds. We explain our strategy for re-engineering information management at FOE and present three example projects which demonstrate the application of innovative IT solutions to problems associated with fundamental working practices.

BACKGROUND

FOE's UK activities are geographically distributed and it has formal links with FOE International offices across the world. The information requirements of the head office, the regional offices, and the international offices, are diverse and a large variety of information must be efficiently exchanged and shared between them. FOE responds to thousands of enquiries each week from the public, press, teachers, academics, industry and government. Timeliness and accuracy of information is essential to FOE's strategic objectives.

Until recently FOE has used mainly paper-based information management systems. This is becoming unmanageable: research documentation alone currently totals about a million pages and is expected to double in the next three years. To manage this growth FOE plans to introduce a more efficient information management system involving innovative use of IT to provide a framework for sustainable long-term expansion; the involvement of UCL via the Teaching Company scheme is an essential part of this plan.

Innovative information management practices have been introduced in three areas of interaction:

With the outside world FOE was one of the first environmental campaigning groups to establish a presence on the Internet (FOEnet[1]). This enables FOE to present itself as an information resource, and to be able to advertise and to encourage public involvement with the organisation directly via interactive electronic communication (e.g. on-line subscriptions and ordering and as well as comments and information contributions). Internet connectivity also presents FOE with the means to enhance its research through the gathering and intelligent filtering of information from electronic sources.

At head office Early study of working practices has lead to the identification of some key information resources and processes. On-line indexing and cataloguing

[1] FOEnet is at http://www.foe.co.uk/

allows for greater sharing of information (using electronic mail and document access), though it requires some changes in working practices.

At regional offices Advanced IT facilities established at head office have been extended to all regional offices which enables remote access to FOE resources. Research data can be transferred electronically direct from the field; all staff can have immediate access to the results published electronically.

A STRATEGY FOR CHANGE

Advanced information technology provides the support platform for the exploitation of information. The main properties of advanced IT which affect organisations are those for facilitating communication (rapidly, easily, reliably, cheaply, across large distances, among selected large or small groups and among individuals) and those for assisting with decision making namely by storing huge amounts of data automatically, retrieving data selectively, combining and reorganising data, making inferences, running simulations, presenting information in a variety of understandable forms and structures. Facilitating communication and aiding decision making in these ways increase the *availability* of information. Some of the consequences of this increased availability will be intended and planned effects, but others, which cannot be predicted in advance, will arise inevitably. It follows therefore that the use of advanced IT not only increases the effectiveness and efficiency of an organisation when applied appropriately (the "information as resource" view) but is also likely to affect the organisation's nature and objectives (what it knows, what it can do, how it is organised).

Mintzberg et al. (1) classify strategies into two perspectives "deliberate formulation through systematic analysis, and emergent formation through the interactive learning of people". Our framework for radical change in the way information is managed at FOE is strongly based on the latter perspective. The activities which are undertaken and the applications of advanced IT supporting specific information management needs which are being developed can be understood collectively from this view. Some of the practical consequences of taking this perspective are now briefly outlined.

We focus on short projects which are selected to deliver rapidly something useful in areas where staff perceive that IT will have a high impact. Early results increase the visibility of the project as a whole and naturally lead to increasingly informed participation by a growing proportion of the staff. The approach is closer to evolution than revolution; it encourages a culture shift towards a working environment which can sustain continual improvement through exploitation of new IT and away from the idea that improved information management is some fixed goal to be achieved after an interval of upheaval.

Every opportunity is taken to encourage staff to develop a more informed appreciation of what IT can do to increase their efficiency and effectiveness. The project leads by example. Firstly, all questionnaires associated with the work have been produced, disseminated, filled-in, collated, analysed, and the results published to interested parties electronically, allowing full use to be made of the facilities for presenting information which are offered by hypertext based media. Secondly, all information related to the project is made immediately available by means of internal use of the World Wide Web e.g. meetings agendas, minutes, progress reports, prototypes of all kinds, questionnaires and other survey results and findings. The visibility of the project is high: many different staff groups have been involved in the development work and have benefitted from this exposure to advanced IT.

EXAMPLE PROJECTS

To illustrate our approach further we now briefly outline three examples of innovative solutions to problems associated with some of the fundamental activities at FOE.

Remote Publishing

FOE's network of local groups are widely distributed across England, Wales and Northern Ireland. There is a need for these groups to be able to make information available to other groups, regional offices and head office. Conventionally this has been done by means of a local group newsletter, produced by head office, which assimilates information sent in by the local groups. The problem with this is the time required by staff to assemble and publish information and the associated problems with turn-round times and restrictions on quantity of information.

A solution is required that will allow a local group to publish information to other local groups and that is not only simple to access and simple to publish, but also secure, low cost, and has a fast turn-round time. Many of these aims can be achieved through the use of the Internet for electronic publishing; the World Wide Web (WWW) has revolutionised the Internet because of its ease of use, and was our choice of information delivery mechanism. However there are no obvious mechanisms for updating the content of WWW pages from a remote site.

Several WWW-based commercial solutions such as Symposia[2], Marmot[3], Frontpage[4] and NaviPress[5] are at present being developed and will in the future provide a full solution to the problem of remote publishing. However the need for a low cost solution led us to use technology that groups already own. The three principle means by which local groups can currently send information electronically to head office to automatically update a WWW page are electronic mail, file transfer (ftp), and WWW using *forms*. We chose WWW using *forms* as our prototype solution, as this gives the user immediate feedback on what their edits look like (unlike electronic mail and ftp), and requires little additional training (unlike ftp). We have produced a prototype system which is simple to use and requires no additional software or equipment for either the client (local group) or the WWW server (head office). It uses a standard *forms*-capable browser (e.g. Netscape[6]), and two CGI[7] scripts running on the WWW server[8].

The format of an electronic newsletter, or any other remote-authored text, is determined at head office. Each WWW page may contain areas that may be remotely updated with text; these areas are implemented using an extension to the HTML (HyperText Mark-up Language)tags mechanism. The text may be viewed using a standard Web browser and an authorised user can edit each Web page after just two mouse clicks. When a page is presented for editing, it appears as normal except for the regions marked as updatable which appear in text entry boxes. After the client has finished editing the page, a single click updates the page on the server and the updated page is then displayed.

Opening FOE's WWW pages to editing raises a whole host of security implications, for example the possibility of unauthorised users editing pages, creating new

[2] Information about Symposia can be found at http://symposia.iria.fr/symposia
[3] Information about Marmot can be found at http://bbs.pnl.gov:2080/marmot
[4] Information about Frontpage can be found at http://www.microsoft.com/msoffice/frontpage/
[5] Information about Navipress can be found at http://www.navisoft.com/
[6] Information about netscape can be found at http://www.netscape.com/
[7] Information about the Common Gateway interface (CGI) can be found at http://www.w3.org/pub/www/CGI/
[8] The CGI scripts require that CGI is activated on the WWW server and the PERL programmming language is installed on the server machine.Information about Perl can be found at http://www.perl.com/perl/

pages and corrupting other pages. Our prototype currently addresses these issues on several levels. Firstly, the editor can inspect the Internet address of the user requesting edit privileges and restrict access accordingly. Secondly, the CGI scripts on the server will only allow updates to a fixed collection of named pages. Thirdly, the pages can be protected by password, and finally, if the information is sensitive, using a secure server ensures all information passed between the client and the server is encrypted.

The prototype is currently being used by a network of local groups to maintain a common diary. Any local group can add an event to the diary and this is then immediately available to all other local groups. The use of this technique can be expanded to any situation where a page of information has a consistent format but its content changes regularly. The prototype is so easy to use that it has provided an efficient and effective solution not only for pages that are updated remotely but also pages updated by staff at head office.

Profiling

In our preliminary analysis of information flow at FOE we have noted a governing theme, sometimes explicit, sometimes implicit, which is that of the *information profile*. The profile exists as a set of terms which indicates the specialist areas of interest and the current priorities of an individual, team or department. We extend the idea of a profile to include classification systems for document storage and quality and distribution standards at the editorial and publication stage. Thus the profile has three facets and represents the information user in three directions, towards the inflow, towards the outflow and towards the transient flows of filing and archiving. The importance of this is that it allows the information users to focus on constructing successful profiles rather than on keeping up with the constantly changing and expanding environment of remote locations, databases and proprietary services to which it pertains. The profile can reflect, and be directly linked to, the organisation's or team's strategic aims because it represents purpose-driven categories such as daily news, enquiries or scientific sources reflecting the strategic aims of responding quickly to environmental events, informing the public, engendering support, and constructing well supported arguments.

We have started work on the profile in the area of the information flow from the external environment. The profile is presented to each team as a page on a web browser (accessible only within the organisation). A master profile is constructed in two stages: firstly by selecting one of a set of top level terms which represent the “immutable issues” upon which FOE campaigns and then by adding the particular aspects of interest under each main term. We aim to use this list of immutables as a consistent structuring device across the organisation's information resources. Elements from this master profile can then be used to select and filter information from different classes of information resources. For the case of the Internet (sub-)profile, the following sequence of events then occurs. Each night, a program, written in Perl, collects the profile elements and uses them to query Internet search engines[9]. The findings from this query are then compared against an index of previously returned information files to prevent unnecessary retrievals. Any newly discovered files are then downloaded onto the local system and a list, in most-recently-modified order, is linked in to the profile page and also sent by electronic mail to the users with brief abstracts of the new files found.

This feedback presents a guide to the user of the success of the search. A user can indicate those files which are not considered to be of interest (these will be

[9] We use the freely distributed libwww-perl package by Roy Fielding of the University of California to do the routine, intensive Internet searching (see http://www.ics.uci.edu/dir/grad/Software/fielding).

ignored on the next search) and modify the terms in the profile to broaden or narrow the search. The best results from Internet searching can be built up gradually and the best formulation of queries learnt incrementally. There are also great practical benefits to this system in the saving of time spent searching on the Internet during the working day and in the freeing up of network resources. Early results show this system to be popular and especially attractive to people who have otherwise been frustrated by spending time waiting for Internet files to download from remote sites (with no guarantee of quality or relevance). We aim to extend the idea to the gathering of press cuttings and to the routing of enquiries from the public to those best qualified to deal with them.

We see the profile as being a control panel for each team to tune its own information flow. As such it contains three kinds of devices. Modification devices, as described above, allow the users to control and refine their own information flow and also promote a consistent use of terminology. Display devices which provide instant and daily feedback on the “state” of information flow (at present these consist of electronic mail and hypertext). Finally, sensing devices, to operate as a management information tool to provide feedback, for example, on which files are being accessed most and, as described under Learning Algorithms, operating as programs to supplement the profile with intelligently filtered information.

Learning Algorithms

FOE's activities at head office are distributed between several groups each specialising in a different research area. Each group organises its own information and uses its own cataloguing and indexing procedures. Whilst this structure is essential to support the development and application of specialist knowledge and skills, there are occasions that require the sharing of information. The strong culture of mutual support at FOE means that the different groups are highly co-operative in sharing information, yet the process is very labour-intensive. Furthermore, the present system is highly reliant on individual knowledge of what information is available and where and how it is stored. If a person is ill or leaves, that knowledge is lost. The existence of information is not always clear and duplication (including the associated unnecessary costs) is a potential problem. The initial perceived need is for information management systems that are not dependent on individual knowledge; however, the system must continue to support the fundamental need for each research group to determine the relevance or otherwise of any given document in a different way.

Our approach to this problem is to evolve new working practices whereby data will become accessible electronically, with a standard system for filing and cataloguing (such as is used in a public library) but with that data being individually assessed electronically and automatically in different ways by different groups. This supports the observation that each group has a different perspective on the shared information which itself will shift over time; an automated assessment can be used to characterise one or more documents that have been archived by someone who has no knowledge of how these documents may be relevant to others. Thus, it is clear from the start that we cannot rely on technology such as rule-based systems (2) to encapsulate the human knowledge of how the assessment process works.

Automated individual assessment is however not easy to achieve. It is human nature to combine some degree of standardisation with ad-hoc mechanisms to cope with the many cases of documents that do not fit into a rigid system. This mixture of the standard and the ad-hoc can work remarkably well, yet it can be very difficult to explain the system to another person (for example, in the hand-over period between one person leaving and another starting work with FOE).

The solution that we are pursuing is the use of learning algorithms (3) (in particular, genetic programming (4)) which will, over a period of time, learn to assess information in the same way as the human researcher. Each group will have its own, separate, learning algorithm which can be used to search for and assess data that is relevant to the group.

We have produced a first prototype Genetic Programming information assessor which learns to classify documents by giving them a percentage "relevance score" according to the presence or absence of words and other word relationships such as inter-word proximity. Our prototype uses a single internal Genetic Program (GP), but in the future we intend to use multiple GPs internally in order to characterise the requirements of a research group; this will provide multiple relevance scores and it is the *profile* of these scores which is the ultimate determinant of relevance.

A very simple prototype has been tested on tiny test files, taking about 6.5 minutes to learn to distinguish with 95% accuracy between test cases, and less than a second to provide a relevance score when applied to a larger file of about 6 kilobytes (all timings carried out on a modest Sun Sparcstation IPC). These initial results are very encouraging and we are therefore pursuing the technique further; in particular, we need to pay attention to reducing the internal complexity of the algorithms in order to deal with the extremely high numbers of large documents used every day at FOE. Our final aim is for a system that may take many days to learn to assess documents adequately, but that will (after the learning process has finished) characterise any given document in a matter of seconds.

CONCLUSIONS

The project is already delivering real improvements in the way some key information-intensive activities are carried out at FOE. We believe that the approach we are taking will lead FOE to the emergent formation of an information strategy which can continuously exploit new advances in information technology. More importantly, for the longer term, the staff at FOE are rapidly becoming more informed about advanced IT and consequently better able to see where new information management practices can improve their productivity. We expect that the information strategy which emerges will be consistent with the organisation's strategic aims, the style of management which characterises the organisation, and the demands of the volatile environment in which FOE operates.

REFERENCES

(1) Mintzberg,H., Quinn,J.B., and Ghoshal,S., 1995, "The Strategy Process", Prentice Hall.
(2) Hayes-Roth,F., Waterman,D.A. and Lenat,D.B., 1983, Building Expert Systems", Addison-Wesley.
(3) Michalewicz,Z., 1995, "A Pesrpective on Evolutionary Computation" in Progress in Evolutionary Computation ed. X.Yao, LNAI 956,Springer,73-89.
(4) Koza,J.R., 1994, "Introduction to Genetic Programming" in Advances in Genetic Programming ed. K.E. Kinnear Jr., MIT Press.