

P@NOPTIC Expert: Searching for Experts not just for Documents

Nick Craswell, David Hawking, Anne-Marie Vercoustre, Peter Wilkins
CSIRO

Mathematical and Information science [HREF1]
723 Swantson St., Carlton, Victoria, 3053, Australia.

[[@cmis.csiro.au](mailto:Nick.Craswell, Peter.Wilkins, Anne-Marie.Vercoustre, David.Hawking)]

Abstract

In large organisations, particularly those spread over multiple offices, undergoing organisational change or experiencing high staff turnover, it can be difficult to keep track of employee expertise. P@NOPTIC Expert is a Web based system which automatically identifies experts in an area, based on the documents already published on an organisation's intranet. The system can be queried like a standard Web search engine, but instead of returning documents it returns a list of experts. Each expert listing includes contact details and supporting evidence. The prototype shows the benefit of integrating Web data with structured data (a list all of employees) to deliver more valuable corporate information.

Introduction

Documents on a corporate intranet may include reports, meeting minutes, project/product pages, employee pages and descriptions of working groups. Although such documents are a valuable resource, for many questions it is necessary to find the right *person* rather than the right document. As reported in [1], people search for documents to find people and search for people to get documents and to get information quickly. For example, a user might wish to find members of a particular project, employees who deal with a particular client or programmers with experience in a particular programming language.

Such information might be found by searching the available intranet documents, using a standard document search system. Looking for Java programmers, the query "Java" might return a number of Java-related documents. Some of these might contain the names of Java programmers. Some of those programmers would still be working for the organisation and it would be possible to look up programmer contact details. However, in the process, the user must sort through many Java-related pages, then spend time checking each programmer's employment status and contact details.

P@NOPTIC Expert solves these problems. It analyses all intranet documents and using a list of all current employees, it finds those who are most often mentioned in the context of Java. It then presents employees found, along with their contact details and a list of matching intranet documents as supporting evidence.

Some expert finding services are already available. Example are [HREF5], [HREF6], where an expert can register under a particular predefined subject and users can search for experts. [HREF7] also asks experts to register, but for the purpose of answering specific user questions. Questions and their answers are available on the site, and answers can be rated by anyone. In all cases experts register for specific expertise, either from a predefined list of categories (very broad), or by free terms. Although providing for specific answers from the web is very attractive, it is not aimed at support direct contact between people. Tacit [HREF8] offers a product that can

be added to a portal and will build Expert profiles from any document that has been published on the portal. It is probably the closest example of what we have achieved with P@NOPTIC Expert.

In the remaining of the paper we first describe P@NOPTIC Expert from the user point of view, then the architecture and implementation. We will finish with a brief evaluation and a description of likely future developments.

Looking for Experts with *P@NOPTIC Expert*

Our prototype searches for experts employed by CSIRO Mathematical and Information science (CMIS). The user can type any query. For example, in CMIS the user query might describe a topic (image analysis), a group name (TED), an internal project name (ISOLDE), a product name (P@NOPTIC), a technology (XML), a customer name (NRMA) or any other free text query. At a less official level it could even be used to find a fellow employee from whom a tennis racquet might be borrowed.

Figure 1. gives an example of P@NOPTIC expert search results. The current system lists up to 10 experts. The top ranked expert is presented with detailed contact information, a photo and a list of supporting documents. Other experts are listed with basic contact details and a link to supporting documents.



Fig.1. P@NOPTIC Expert results for the query "MPEG maaate", including details of the top ranked expert, a list of four experts and a list of supporting documents.

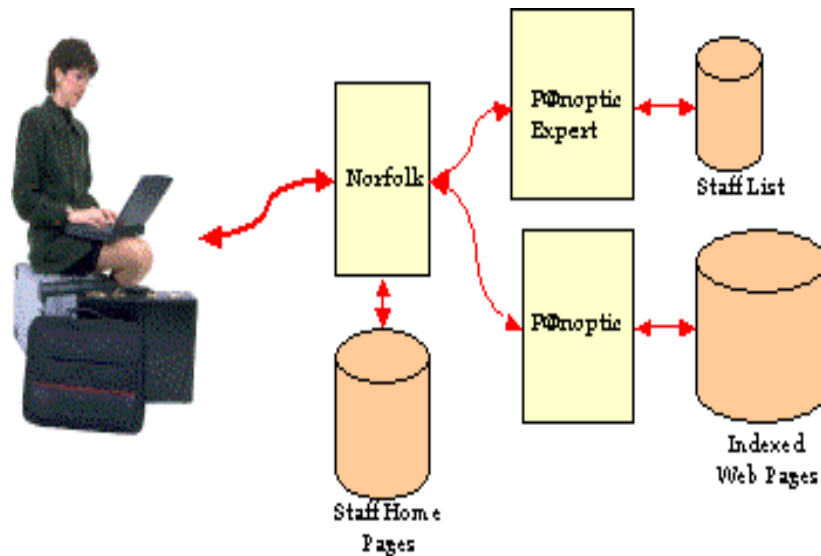
Architecture of the system

The architecture of the system is described in Figure.2. The user interface is generated by Norfolk [2]

Our virtual document generator Norfolk gathers information from the staff Home page and the results returned by the P@noptic Expert engine and the P@noptic search engine [HREF9]. P@noptic Expert access a staff list that should be up to date and include employee contact details and home page URLs. First Norfolk queries a special P@NOPTIC expert index, which is described below, retrieving a list of experts.

This list of experts is displayed on the results page. However, Norfolk adds extra details on the top ranked expert. It extracts a photo and additional contact information from the expert's home page. It also presents a list of supporting documents, which were used in finding the expert. The

latter comes from a standard P@NOPTIC index of the intranet documents.



P@NOPTIC expert index

The P@NOPTIC expert index is based on special "employee documents". Given a list of 150 employees, there would be 150 employee documents. In each one is text associated with that employee, taken from intranet pages. So if Fred Nerk appears in 12 intranet documents, the Fred Nerk employee document would contain text from those 12. In the current implementation, the employee document is simply the concatenated text of the 12 documents. In future systems, text might be extracted from the 12 more selectively, to extract only the text appearing near Fred's name.

P@NOPTIC then indexes and processes queries over the employee documents. If Fred is a Java expert, his employee document is likely to mention Java. So given the query "Java", Fred's document is a match, identifying him as a potential expert. The expert results list the ten best matching employee documents.

Norfolk Wrapper and page generator

Norfolk is a system that allows for extracting and gathering information from various structured and semi-structured sources, including HTML and XML pages. The result can be displayed as new sets of XML or HTML pages. Norfolk offers a tree language for extracting information from sources and creating new trees. A complete description of the language can be found in [2], and its application to the creation of Web pages in [3].

In P@NOPTIC Expert, Norfolk has been used 1) to extract one expert's details from his/her corporate home page, 2) to compact the list of experts returned by the P@NOPTIC Expert engine [4], 3) to send a query to P@NOPTIC search engine to get the list of evidence for a given expert and, 4) to assemble all the results into a new page with links to other expert pages. These later pages are created on demand when a link is activated.

Benefits of P@NOPTIC Expert

Unlike conventional directories or search engines P@NOPTIC Expert provides automatic and rapid identification of experts:

- Without manual maintenance of employee expertise lists
- Without any specialised user training
- With contact details and photos of the experts found
- With supporting evidence so the degree of expertise can be judged

Decisions are based on the documents that staff load onto the corporate Intranet; this automatically tracks expertise as staff develop.

Evaluation and further work

The prototype is operational and can be accessed from [\[HREF2\]](#). We have not carried out formal evaluation yet, but informal testing with a variety of queries are very encouraging. More precise evaluation is needed in two directions: 1. Precision in finding the right experts and ranking them properly, and 2. Comparing user experience using expert search to their experience using a standard search tool and staff list, to see if expert search helps in finding better information more quickly..

From the informal evaluation we can make a couple of remarks and suggestions. It seems that we get better results for finding experts in our research group than for other groups. Since we are involved in Web technology, we put a lot of documents on the Web, such as project descriptions, minutes of internal meetings, publications, reports, etc. Such documents usually mention the names of our staff members, so can be used in expert search. In other groups, the same type of information is not put online, so it is not possible to exploit corporate memory to the same extent.

Further study might identify the type of documents which hinder proper expert search. For example, news pages list a large number of employees along with various potential query keywords. The present system would associate all listed employees with all those keywords, leading to noise in results. This could be avoided by eliminating such pages from the expert index. Another possibility would be allowing users to have some manual input in describing areas in which they are or are not involved, although care would need to be taken not to compromise the useful automatic nature of the system.

Another issue is in finding new employees, particularly those taking over from an expert who has left the organisation. As soon as the old expert is removed from the staff list, they will disappear from search results. In some situations it would be appropriate, at least until the new expert is represented in intranet documents, to forward searchers from the old expert to the new.

In conclusion, we believe that P@NOPTIC expert demonstrates the benefit of integrating semi-structured information, such as Web documents and home pages, with more structured information such as staff lists and phone directories. It illustrates the type of added-value service which can be built using existing corporate memory resources. Finding relevant documents is not enough.

References

[1] Hertzum, M. and Pejtersen, A. M (2000). "The information-seeking practices of engineers: searching for documents as well as for people", in *Journal of Information Processing and Management*, Vol(36),pp.761-778, 2000.

[2] Vercoustre, A-M. and Paradis, P., "A Descriptive Language for Information Object Reuse through Virtual Documents", in *4th International Conference on Object-Oriented Information*

Systems (OOIS'97), Brisbane, Australia, pp299-311, 10-12 November, 1997. [[HREF3](#)].

[3] Paradis, P. and Vercoustre, A-M, "A Language for Publishing Virtual Documents on the Web", in *International Workshop on the Web and Databases*, Valencia, Spain, 27-28 March, 1998. [[HREF4](#)].

[4] Hawking, D., Baily, P. and Craswell, N., *Efficient and Flexible Search Using Text and Metadata*, CSIRO Mathematical and Information Sciences, Technical Report TR2000-83, 2000, [[HREF10](#)].

Hypertext References

HREF1

<http://www.cmis.csiro.au/>

HREF2

<http://www.ted.cmis.csiro.au/proj/yellow/>

HREF3

<http://www.cmis.csiro.au/TIM/staff/Francois.Paradis/papers/OOIS97/>

HREF4

<http://www.cmis.csiro.au/TIM/publications/1998/paradis98b.ps>

HREF5

<http://www.allexperts.com>

HREF6

<http://www.expertisearch.com>

HREF7

<http://www.askme.com>

HREF8

<http://www.tacit.com>

HREF9

<http://www.panopticsearch.com/index.html>

HREF10

<http://www.ted.cmis.csiro.au/~dave/TR2000-83.ps.gz>

Copyright

© Copyright 1997-2001, CSIRO Australia. The authors assign to Southern Cross University and other educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Southern Cross University to publish this document in full on the World Wide Web and on CD-ROM and in printed form with the conference papers and for the document to be published on mirrors on the World Wide Web. No Rights to Research Data is given. CSIRO and the Author/s remain free to use their own research data including tables, formulae, diagrams and the outputs of scientific instruments.