# Enterprise PeopleFinder: Combining Evidences from Web Pages and Corporate Data

*Alistair McLean, Anne-Marie Vercoustre, MingFang Wu*

CSIRO – ICT Center
Clayton, Victoria, Australia
(+61) 03 9545 8461

*{firstname}.{surname}@csiro.au*

## Abstract

*The main source of competitive advantage of an organization is its knowledge assets and its learning capacity. We argue that finding knowledge in an organization is not just about finding documents, it is also about finding the right person. In this paper we present PeopleFinder, a Web based system which automatically identifies experts in an area, based on the documents already published on an organization's intranet or part of the corporate data. 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 leverages on organizational data, such the project descriptions and membership to infer which documents can be used as best evidences of expertise. The paper also includes preliminary evaluations*

**Keywords:** Information Retrieval, People Finder, Corporate data.

## 1. Introduction

The main source of competitive advantage of an organization is its knowledge assets and its learning capacity [1]. Traditional document centric knowledge management approaches have mostly focused on capturing relevant corporate documents and making them available through search or more proactive delivery mechanisms. These approaches have failed in making the knowledge in the corporate memory operationally available (eg making the information relevant to the task at hand and converting it into effective actions).

**Proceedings of the 8th Australasian Document** Computing Symposium,

Canberra, Australia, December 15, 2003.

We argue that finding knowledge in an organization is not just about finding documents, it is also about finding the right person.

In the corporate world, needs for finding experts include:

- For workers to identify people to ask for expert opinions and perhaps collaborate;
- To be able to get a good picture of capabilities and areas of expertise;
- To be able to identify appropriate people to build teams.

We are developing PeopleFinder, a tool for finding people in an enterprise based on their expertise. Unlike some commercially available tools our approach does not rely on a profile database but on a combination of the documents that people create and publish during their actual work and organizational knowledge encoded as an organization's structure.

## 2. Related Work

Existing people finding work such as Peer Helper and more recently I-Help [6] use a knowledge profile of a helper which is organized to particular tasks a user may be engaged in. The profile indicates, among other things, ability, willingness and availability for that person to help with a task (or step in a task). Other systems such as expert marketplace (www.experts-exchange.com) rely purely on a scoring mechanism whereby the person asking for help scores the person who answers his question. Neither of these approaches is particularly suitable to an organization where minimal effort is required to keep the assessment of knowledge up to date and where we are trying to capture the expertise of people who might not actively take part in voluntary “market places”.

One approach often used in organizations is to assess and record peoples’ skills in a skills database (eg Microsoft’s SPUD [4] or SAGE’s PeopleFinder at http://sage.fiu.edu). However, in large organizations, particularly those spread over multiple offices, undergoing organizational change or experiencing high staff turnover, it can be difficult to keep track of employee expertise. Furthermore a fixed set of terms are used to describe skills which leads to terminological mismatches. We argue that database approach is limited because: 1) the database has to be created and moreover continually updated and 2) a user profile is not going to capture all the implicit expertise that a person is acquiring in their regular work.

A people finding system based on the electronic documents that people publish, own and access is more relevant, is automatically updated as new documents appear and provides a rich data environment for which term disambiguation can be computed. Work such as MITRE’s ExpertFinder [5] works in this manner. It combines evidence from documents that are published by the person with evidence from documents that mention that person’s name. In the latter case, the system uses heuristics to try to determine the page type and then uses type-specific heuristics to extract the relevant text associated with the name.

Recognizing that the exploitation of heterogeneous expertise indicators is essential for good expert finding systems [7] proposes a general architecture that combines centralized expertise models with decentralized expertise indicator source gathering.

Other systems (eg Autonomy [2]) that monitor the files that people read/publish etc do not differ between the type of the data nor do they take advantage of the implicit organizational structure.

## 3. PeopleFinder Architecture

CSIRO’s original prototype, P@NOPTIC Expert [3] is a web based system which automatically identifies experts in an area, based on the documents already published on an organization's intranet. Like a standard web search engine the system takes a subject query and 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.

The limitations of our initial prototype included, in some instances, low quality results due to poor quality documents being used as expertise evidence. In our current work we refine the computation of a person’s expertise by (a) accepting more documents as evidence of expertise using inference based on the corporate structure, and (b) we do not rate all documents as equal for evidence.

In general, we calculate a person’s expertise by analysing the documents that contain that person’s name and documents that may not contain their name but may occur “close” to other important documents in some organizational structure. For example, on a corporate intranet we may designate a project page as highly relevant evidence of the expertise of the project members. Documents close to this project page in the intranet structure may also be counted as relevant even if they do not explicitly mention project members’ names. Moreover a home page or a project page will be weighted higher as evidence than, for example, a news page or a document page.

The architecture of our PeopleFinder system is shown in Figure 1. First, we create the evidence that is used to determine a person’s expertise in an offline manner. Corporate data is used extensively to help determine what documents to use as evidence for expertise, to help determine how to rank the returned list and how to display the results. For this data, we currently use a manually created XML document that represents the hierarchy of organizational units, existing projects, current staff and the relationships between them (see Table 1).

```
<unit id="ted" type="team">
  <details>
    <title>Technologies for Electronic Documents
    </title>
    <descriptionurls>
      <url>
        http://www.cmis.csiro.au/ted/about.html
      </url>
    </descriptionurls>
    <member personID="p1">
      <role roleID="tl"/></member>
    <member personID="p2">
      <role roleID="bdm"/></member>
    <project projectID="expert_search" />
      …
  </details>
</unit>
```

**Table 1: Snippet of structured data used to represent corporate data.**

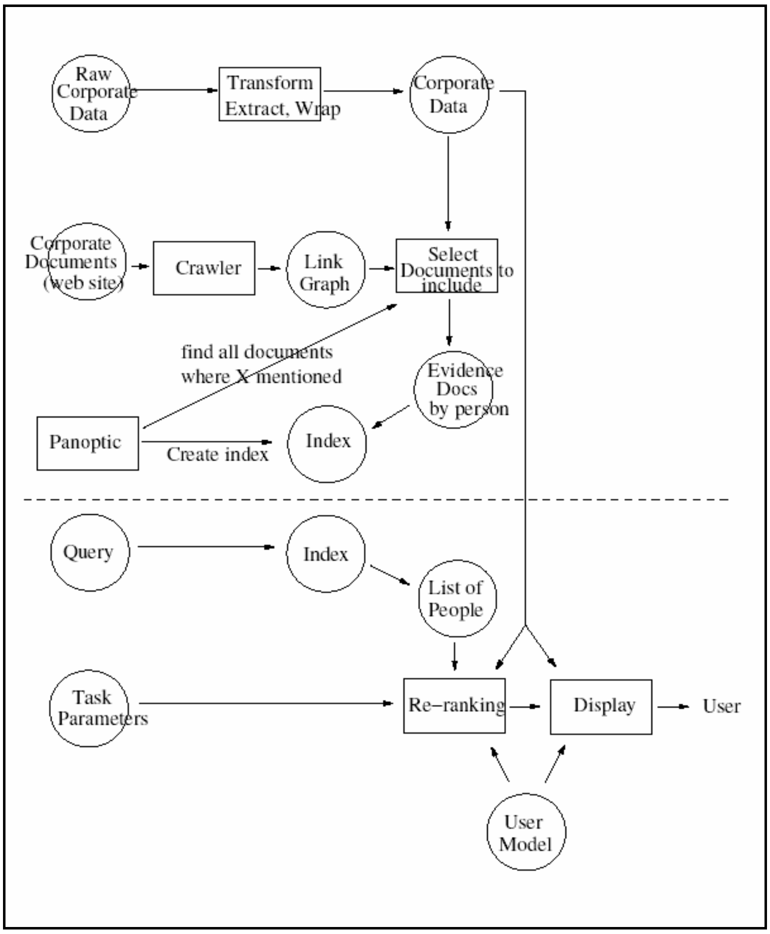


Figure 1: Data flow architecture diagram.

We also use the organization's web graph to provide us with some information structure and to allow us to identify related information to certain seed points. From this combined set of information we construct evidence fragments for each person. This sub-collection is then indexed.

At run-time, a user may be involved in a task and we represent this as a set of parameters that feed into our PeopleFinder system. The idea is to be able to perform different ranking and presentation of the results according to the task so that, for example, if someone is trying to put together a team of people to write a tender on a specific subject, then the system would return sets of people, organized as teams (with required roles and skills specified). Currently, however, we have implemented the "find an expert to ask for information" task where a required role-type can be requested.

Once again, the structured corporate information can be used to assist in the ranking and organization of the results and the user's profile can be employed to assist in both the ranking and presentation. For example, a scientist may want to see other scientists first whereas a business client may want to see group leaders first. Figure 2 shows an example of results presented for the query "electronic records and government", that includes the ranked list of people and evidence of expertise. Clicking on a button near the name of the first person has opened a small window showing the position of the person in the organization chart (the TED team in the ECT group), and the projects he is involved in.

## 4. Evaluation

A first version of the PeopleFinder prototype has been implemented and we have started its evaluation. The objective of the evaluation is to be able to answer the following questions:

Does our approach work? i.e. can the system find staff with relevant expertise, rank them sensibly and deliver an useful answer what amount and type of corporate data make a difference? In some organizations corporate data may not be easily available, or costly to configure.

what is the optimal performance of the algorithms? Is this performance query-dependant or is it independent of the type of queries?

To answer these questions we have designed an evaluation framework that includes two distinct parts: 1) *expert* evaluations that follows the standard evaluation method in the Information Retrieval community, and 2) evaluations that involve users' participation. In this paper we report only on preliminary *expert* evaluations using our test collection.

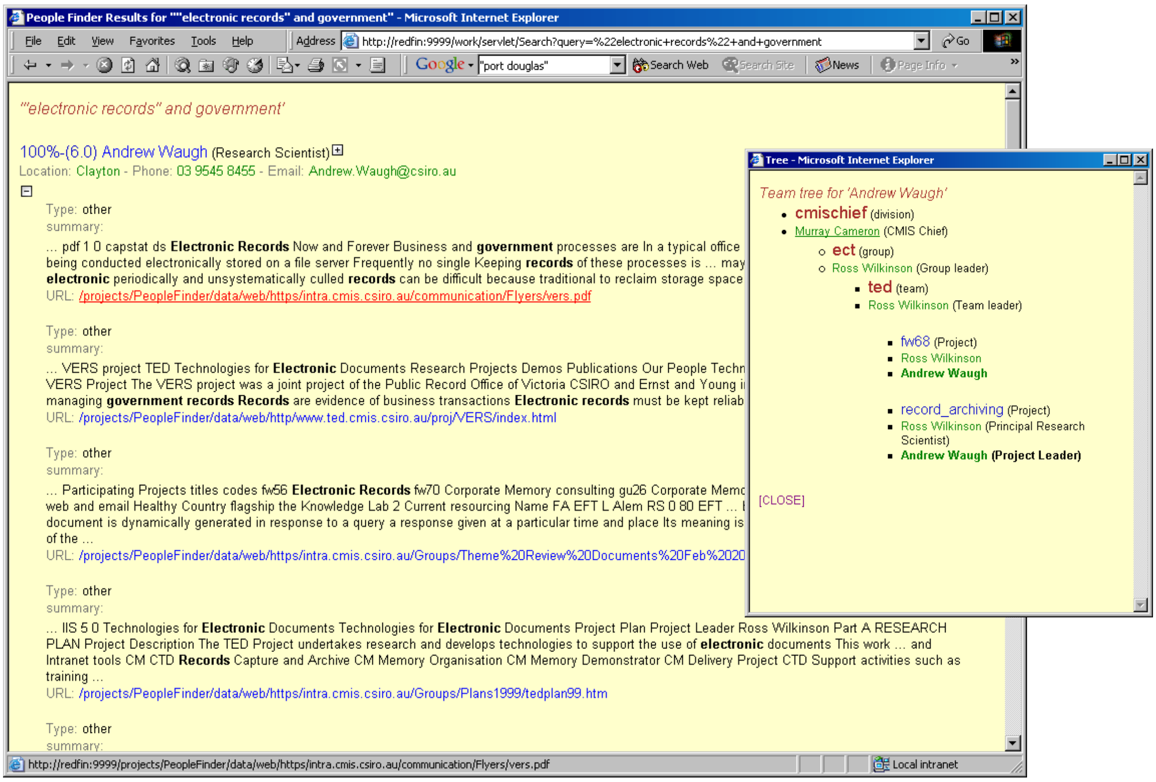


Figure 2. User interface showing the results.

## 4.1 The Test Collection

Part of our effort is to create a test collection of documents, queries and answers. The advantage of a test collection is that it can be used over time to compare various algorithms or parameters. The document collection includes:

1. the web pages from our division's (CMIS) Extranet and Intranet, collected by our in-house Panoptic crawler, on 20 March 2003; (this collection will be referred as "web collection".)
2. the CMIS staff list on 20 March 2003;
3. project descriptions and associated members manually extracted from the extranet;
4. current project list and members from our corporate project database, for the month of February 2003, as well as current project description from the project plans
5. publication list from the CMIS publication page and groups' publication pages;
6. business development contact database (mails)

Data 2-6 have been manually or semi-automatically created and converted into XML documents. (Thus data 2-6 will be referred as "xml data".)

A list of test queries has been manually assembled by looking at terms in research group pages and related terms from the ACM thesaurus, and selecting terms from the query log of our previous prototype. Example of queries are: natural language technologies, mathematical morphology, sampling of minerals, xml protocols, audio analysis, SVG, RDF, data mining, atmospheric science.

The "correct" answers are based on people's evaluation of the expertise of their colleagues. We sent out a questionnaire to our research group leaders and ask them to provide experts on their designated topics/queries. Group leaders were also asked to rank the experts in various level of expertise (high, medium, low, none). Topics for which no assessments of expertise were made, or with no associated expert were eliminated, leaving us with 159 test queries.

## 4.2 Methodology

Traditional IR evaluation uses the recall/precision metrics to determine the quality of a given search engine for a set of test queries. With People Finder we are more interested in finding the right person rather than many experts, so precision[1] is likely to be more important. Usually precision and recall are calculated at the various numbers of retrieved answers; for example, how many relevant answers do you get in the first five answers (noted @5), the first 20 (@20), or the first 100 (@100). In PeopleFinder we are interested in looking only at a few results since the purpose is to contact one person. Also, the list of assessed experts for most topics is small since the research groups are relatively small[2]. So we calculate precision @1, @3, @5 and @10 only.

The next step is to define the notion of relevance for a given expert. In our test collection, various people have been assessed (by themselves or by their group leader) for each test query, as having a **high**, **medium**, **low** level of expertise, or by default **none** expertise. We need to define some function to map expertise into a relevance value. We have defined the following relevance function:

$f_{topic}$(person) := 1 if (expertise = high or medium) or

(expertise = low and there are no high/medium level experts)

0 otherwise

Other functions would be possible, such taking in account only high expertise. Indeed expertise is not quite absolute; it may depend on how you evaluate yourself, or on the environment you are in. A person with a low level of expertise knows at least what the topic is about and can refer you to better experts.

[1] **Precision** is the percentage of relevant documents in relation to the number of documents retrieved. In our case it will be the percentage number of (un-repeated) relevant people out of the retrieved ones.

[2] For technical expertise, such as java, XML, C++, etc., the number of experts is much bigger as they can be found in many different groups. We may consider precision/recall @20 in further experimentations for those topics.

Given the complete set of assessments (expertise of people) and a quantization function for mapping the assessments to a single relevance value we are able to apply evaluation metrics as in standard document retrieval.

# 5. Experiments

We are running a set of experiments along three dimensions: data, topics and algorithms.

**Algorithms:** First we want to compare the new prototype (referred new system) with our basic initial system (referred base system) to see the benefit of building evidence for each person. If the answer is positive, then we will investigate further on the impact of different weighting schemes in the new system for different types of documents.

**Data:** As said in 4.1, various document collections are used. We are interested in selecting a collection(s) with minimum effort to build while achieving best precision. We will investigate, by using the same algorithm, which collection (or combination of collections) will contribute most to the high precision.

**Topics:** The aim is to investigate the dependence of the answers on the type of topics/queries. The topics in the test collection can be classified into the research topics (asking for scientists) and technical topics (asking for software engineers). The evidence documents for these two types of topics vary, for example, scientists tend to have more publications and be mentioned more often in project description.

In this paper, we will report two completed runs. The first run is to compare the new system with the base system over the web collection (web pages from intranet and extranet). The second run is to compare the new system's performance with two collections: web collection vs. web collection and xml data.

# 6. Results

The results of two runs are shown in table 2 and table 3.

Table 2 compares the average precision at different levels of recall between our base system and the new system, using the Web collection. The numbers in brackets are the percentage of the increased precision of the new system over the base system, they are in bold representing the significant improvement.

**Table2. Average precision for base system and new system over the Web collection.** .

| P@ | 1 | 3 | 5 | 10 |
|---|---|---|---|---|
| Base | 0.209 | 0.187 | 0.159 | 0.125 |
| New (%) | 0.405 **(93.8%)** | 0.315 **(68.4%)** | 0.257 **(61.6%)** | 0.172 **(37.6%)** |
| p< | 3.2E-05 | 5.3E-06 | 4.6E-07 | 5.7E-05 |

The last row shows the paired t-test values level that measure the significance of the results[3]. This run shows a significant increase in precision for the new algorithm compared to the base system, at all cut-offs, since all p values are less than 0.05.

**Table 3. Comparison between two collections by using the new system.**

| Average precision | @1 | @3 | @5 | @10 |
|---|---|---|---|---|
| Web | 0.405 | 0.315 | 0.257 | 0.172 |
| Web + xml data | 0.412 (1.7%) | 0.335 (6.3%) | 0.273 (6.2%) | 0.207 **(20.3%)** |
| p< | (0.84) | (0.31) | (0.23) | (0.0003) |

Table 3 compares the average precision for the new system when using the Web collection only with a dataset consisting of the Web collection and the other XML data. This table shows that the precision improves at all cut-offs when adding the XML documents, but significant improvement exists only at cut-off 10.

The increase using the corporate database data as well as Web documents is less significant. This could be explained by the fact that these databases were small compare to the Web collection, and that some data (e.g. the contact report) might not be relevant for the type of queries in the test set.

We looked at the results in more detail and found that the average score had a power law distribution with the rank of the result (see Figure 3) so that the top candidates have very high scores. We then examined some individual results and found that the addition of XML data allowed candidates who were previously ranked poorly to replace candidates who were ranked at a "medium" level. Thus we see that the XML data is capturing important information for some candidates that was not captured by the Web data. However, with the quantity and chosen weights, this was not enough to replace the candidates that scored well with the Web data. Thus we see a slight improvement @1, @3 and @5 but significant improvement @10.

As we have said before the relevance function is somewhat arbitrary. It should also be noted that our relevance assessments are likely to be incomplete since we did not send all the topics for assessment to all the group leaders, and so some experts in a topic may belong to groups that were not assessed for this topic. .However, since we are only interested in comparing systems rather than the absolute value of the precision, all experiments have the same bias.

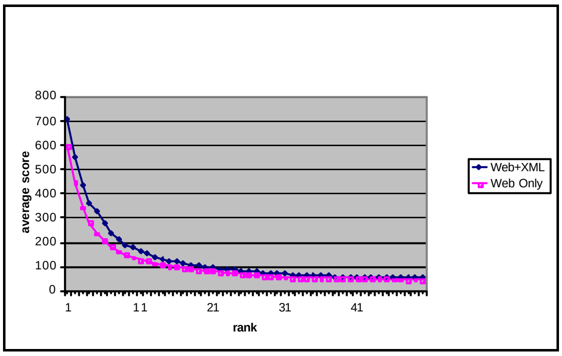


**Figure 3: Graph showing average score distribution**

For user-based evaluation, it would be important to get a more complete and consistent set of assessments.

## 7. Conclusions

In this paper we have argued that finding experts in an organization is best achieved using an automatic, evidence based approach and that structured organizational information can be leveraged to improve the precision of the results. Our initial results are promising and indeed show that the use of structured corporate information improves the precision of finding experts.

We are planning to conduct further evaluations of the system including 1) the impact of weighting differently for different types of documents and 2) user evaluations of the system. Finally, we hope to extend the prototype for retrieving teams of people.

[3] The t-test measures the probability that the improvements in the results happens just by chance. A low value means that the results are statistically significant.

## 8. Acknowledgments

We wish to thank Shaminda Samaratunge for implementing the user interface and the evaluation programs .

## 9. References

[1] Argyris C. and Schon C. “Organizational Learning: a theory of action Perspectives”. Reading Massachusetts, Addison Wesley, 1978.

[2] Autonomy Technology White Paper. www.autonomy.com/Extranet/Marketing/Autonomy White Papers/Autonomy Technology WP.pdf

[3] Craswell N., Hawking D., Vercoustre A-M., and Wilkins P. P@NOPTIC Expert: Searching for Experts not just for Documents in Ausweb 2001, 21-25 April, 2001.

[4] Davenport T.H., Prusak L. “Working Knowledge: How organizations manage what they know”. Harvard Business School Press 1998.

[5] Mattox, D., Smith, K., and Seligman, L. (1998). Software Agents for Data Management. In Thuraisingham, B. Handbook of Data Management, CRC Press: New York. 703-722.

[6] McCalla G., Greer J., R. Vassileva, Deters R., Bull S. and Kettel L. “Lessons Learned in Deploying a Multi-Agent Learning Support System: The I-Help Experience” Proc 10th Int. Conference on AI in education, San Antonio, May 2001, 410-421.

[7] Seid D.Y., Kobsa A. “Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach”. In Beyond Knowledge Management: Sharing Expertise. Cambridge, MA. MIT Press 2002.