Discovering Areas of Expertise from Publication Data

Meredith Taylor and Debbie Richards

Computing Department, Maquarie University
North Ryde, Australia, 2109
richards@ics.mq.edu.au

Abstract. Expertise recommender systems are a valuable tool for keeping track of who has expertise and in what areas within an organization. The key problem is acquiring validated knowledge of expertise and keeping that information up to date. In research organizations, publications are one source of evidence of expertise which can be used to identify who knows about what. In this paper we focus on evaluating the feasibility of a simple technique for uncovering expertise used as the foundation and starting point of maintaining a profile of validated expertise within an organization.

Keywords: Expertise recommender system, knowledge acquisition, knowledge validation.

1 Introduction

Recommender systems are electronic (often web-based) systems that recommend objects of interest to searchers based on search queries (e.g. Amazon.com recommends books and other products, imdb.com recommends films). Once a searcher has selected a recommendation, the system will often recommend other things that it thinks the searcher will like based on the preferences of other searchers who also chose that recommendation. This is called Collaborative Filtering and works on the assumption that if person a likes one thing (or several things) that person b likes, it's likely that person a will also like other things that person b likes.

The term Recommender system was first introduced by Resnick and Varian [12] [11]; however, the first recommender system was Tapestry [5]. Tapestry was an electronic mail system which filtered the mail sent to searchers, returning only those messages that were of interest. Searchers specified how they wanted their mail filtered by providing search queries that the system ran over the set of documents. In addition to content filtering, Tapestry also used (and first coined the phrase) collaborative filtering. A searcher could pick out messages of interest, which could then be sent on to other searchers on the same mailing list who would have received those messages, but were not sure if they wanted to read them. Recommender systems can also be used to recommend experts. Expert Recommender Systems (ERS) are the focus of this project.

1.1 Expert Recommender Systems Issues

Expert recommender systems (systems that recommend experts) are a useful and convenient tool for finding experts [1] without having to spend time combing the

D. Richards and B.-H. Kang (Eds.): PKAW 2008, LNAI 5465, pp. 218–230, 2009.
© Springer-Verlag Berlin Heidelberg 2009

internet, staff web pages, or publication repositories. As in any system that stores data, collecting the appropriate data can be hard and time consuming. Collecting the knowledge held in people's heads on who is an expert and in what area is harder still. Even more problematic is ensuring that the data obtained are complete, consistent and current so that the system can provide accurate recommendations. If a user is not certain that the recommendations provided by the system are valid, they will have little reason to trust the system, and thus the system will not be used.

To handle data and knowledge acquisition, two main approaches are used in expert recommender systems:

1. Manual: Experts are required to register their own areas of expertise with the system by filling in surveys or entering keywords that can be matched with search queries; and
2. Automatic: Data mining and other information retrieval techniques are used to search through sources that may hold evidence of peoples' expertise (such as web pages, publication repositories, citation indexes, and conference proceedings) to determine if someone is an expert in a certain field and to what extent they are an expert in that field (see for instance [2], [10], and [11]).

The first type of system is fairly easy to implement on a technical level as the experts themselves have to do most of the work by entering their own areas of expertise. The system developer will only need to develop appropriate forms for data entry and retrieval, and a simple query matching algorithm. Thus when a searcher searches the system, all that needs to be done is to match their search terms with the expert's keywords. This technique is often referred to as a yellow-pages approach to finding an expert as that is the way people usually find a plumber, lawyer or doctor. It is a simple and yet effective method for finding people who have certain skills. It uses the assumption that only someone who is actually an expert in a certain field would list themselves as such and since they went to the trouble of registering their areas of expertise, they are probably interested in being contacted. This assumption is not always valid. An expert may initially enter their data into the system, but may not perform any regular updates due to a lack of time or interest. As a result, people using the system can never be certain that a recommended expert's expertise is current, or that they are still willing to be contacted. It is even often the case that the expert will leave the organisation and their data will still be in the system. There are also the issues of experts failing to find the time to enter their data in the first place or entering incorrect data that does not reflect their true levels of expertise. However, this is a less likely occurrence than someone simply not entering their details initially since most members of an organisation would have a fairly realistic view of their level of expertise and would not wish to be contacted by someone if they are not confident that they would be able to help them.

The second type of system is less reliant on the time and interest of the expert. In some cases an automated system may provide a less biased profile of someone's expertise, but this depends on the appropriateness and range of sources available and information extraction techniques used. However, these systems are more difficult to implement as they require a large amount of data to be available for each expert. Additionally, expertise could be identified from many different sources that will vary across individuals and organisations making it difficult to have predefined sources for the system to search through. Sources that are used in existing systems to locate experts include email [3, 6], bulletin boards [7], web pages [4, 9], program code [8, 13], and technical reports [2].

To date we have developed a prototype system known as “Who Knows?” We have implemented and tested selected components of our proposed solution as initial proofs of concept. In the remainder of the paper we present our results to evaluate the feasibility of capturing initial data from artefacts and having experts validate the results. In the larger framework, our approach will encompass more sophisticated automated methods using a range of inputs such as individual web pages, project/grant repositories, citation indexes (e.g. CiteSeer - http://citeseer.ist.psu.edu/) and publications databases. In the longer term we intend to create a toolkit or workbench (like WEKA) which draws together the body of disparate work in this area by incorporating many algorithms and automated techniques such as those cited above. In the shorter term for proof of concept we have used a simple text analysis approach and an internal data source comprising a collection of all publications, grants and impact factors of individuals within our university.

2 Evaluating Automatic Expertise Acquisition

The Research office (RO) at Macquarie university runs and maintains IRIS -Integrated Research Information System (http://www.research.mq.edu.au/ researchers/iris) in which staff are required to enter information about all their publications from 2001 (the year the system was first put into use) onwards. The system also stores information about each staff member’s research projects and grants (accepted and rejected) in the profile for the staff member. In their profiles, staff members are able to nominate RFCD (Research Fields, Courses and Disciplines) codes that correspond to their areas of expertise as well as the percentage of expertise they have in each area (Fig. 1).

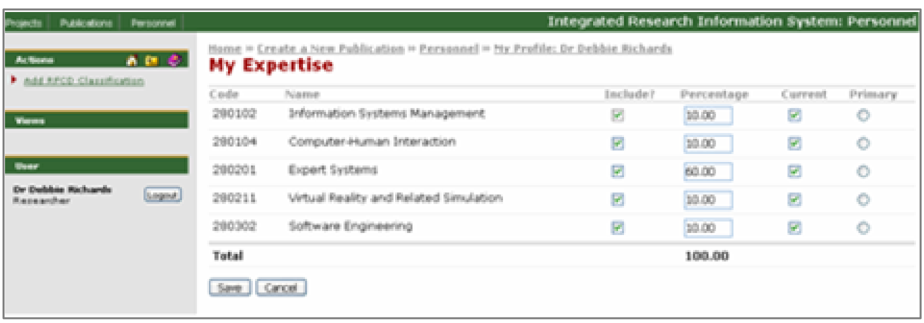

Fig. 1. Areas of expertise in IRIS

RFCD codes are issued by the Australian Research Council (ARC[1]) in order to categorise research and development activity and other activity within the higher education sector in a uniform manner. They are split into divisions (for example:

[1] see http://www.arc.gov.au/applicants/codes.htm

250000 - Chemical Sciences, 260000 - Earth Sciences, 420000 - Language and Culture, and 280000 - Information, Computing and Communication Sciences) which are then split into subdivision (such as 280101 - Information Systems Organisation and 280102 - Information Systems Management, which are subdivisions of Information, Computing and Communication Sciences).

Classifying experts with RFCD codes gives an indication of their general areas of expertise and would be a good addition to an expert's profile in an expert recommender system. Since few staff members have entered this data it is not possible to obtain this information directly from IRIS. If it was possible to automate the process of locating the RFCD codes for experts, it would not only provide a useful addition to each expert's profile in our prototype system, but it would also provide each expert with a more realistic view (that is an evidence-based view) of what their expertise areas actually are. Therefore we chose to use the publication data within IRIS to classify each publication with an RFCD code and then to assign RFCD codes to each staff member in IRIS based on the RFCD codes for each staff member's publications.

The publication information contained in IRIS includes details such as the name of the publication, the name of the publication it belongs to (in the case of journal and conference papers, for example), the author's name or list of authors' names, the primary department the publication belongs to, and the year of publication. It does not include paper abstracts, relevant keywords, or any online locations of papers. If we had this additional information we expect we could achieve better results and the effort of incorporating alternative and advanced algorithms would be more appropriate.

2.1 Methodology

The tasks involved in classifying experts with RFCD codes are the following:

1. Match RFCD codes with paper titles and publication titles using a simple string matching algorithm that checks to see if a keyword (or several keywords) in an RFCD code title occurs in the title of a publication or paper. (In this study this was done only for publications from the Computing Department).
2. Classify each staff member with the major RFCD codes found. (In this study we classified on the smaller division level, rather than the subdivision level or the major division level.
3. Check against self reported codes. It has been necessary to request assistance from members of the Computing Department for this exercise by asking them to classify their areas of interests with RFCD codes.
4. Record the percentage of experts that agreed with their automatically found codes.

Several Python scripts were written to complete these tasks which will now be described in more detail below.

2.2 Matching RFCD Codes with Paper and Publication Titles

This was done in several stages. The first stage was to collect the relevant data from the XML file that held the IRIS publication data. For the purposes of this experiment, only papers written by people from the Computing Department were considered. From these entries, the title of the paper (or book) was extracted, as well as the title of any accompanying publication (such as the journal or conference a paper was published in) and the list of authors.

The second stage was to collect the names of the staff currently in the Computing Department at Macquarie University. When this list of names was compiled, those staff members who did not have any publications in the list created from the IRIS data were eliminated. Similarly, publications in the list of IRIS data that were not authored by at least one person from the list of staff members were eliminated. The result of this process was a Python dictionary associating staff members with the publications that they had authored, co-authored, or edited.

The third stage involved gathering the relevant RFCD codes from the Australian Bureau of Statistics Website (http://www.abs.gov.au/). This collection was done prior to the new RFCD codes being released, thus the division used for matching was 280000 - Information, Computing and Communication Sciences. While some members of the Computing Department do have publications written in other domains, we felt that restricting the classification to one domain would simplify the process and show any significant results fairly quickly.

One goal of this experiment was to test what information from the IRIS publication data would provide the most accurate and predictable classifications. To this end, we classified each staff member's documents 3 times, once only using the paper (or book) titles, once using the containing publication titles (if applicable) and once using both paper and publication titles.

Matching RFCD codes to paper or publication titles was a fairly simple task. Each RFCD code was split into words. Then, each word was tested against the title in question using a simple string search. If the word was found, then that RFCD code was counted as a match. The only exception to this rule was the word 'computer' which is a common word to use in the domain and would have yielded too many false matches.

Minor tweaking of the string matching process was also performed to match words that share the same root (to make 280504 - Data Encryption match a paper with 'cryptography' in the title, for example). This was achieved by creating an ontology of terms found in the RFCD codes along with several words that share the same root and seem likely to appear in a publication title (Fig. 2).

```
"simulation": ["simulating", "simulate", "simulations"],
"analysis": ["analyse", "analysing"],
"representations": ["representing", "representative"],
"encryption": ["encrypting", "cryptography", "encoding",
"decryption",    "decoding", "cryptology"],
"security": ["secure", "unsecure", "secret"],
```

Fig. 2. Snippet of code from the ontology of terms, written as a dictionary in Python

The ontology also matched terms in the RFCD codes with words that referred to similar concepts. For instance the concept of a knowledge-based system is the same as for an expert system (RFCD 280201 Expert Systems). Thus the term 'expert' in the dictionary was matched with the term ‘knowledge-based’. While this would not be a realistic task if we wished to classify staff from all disciplines, it was fairly simple to implement for only one discipline, and serves to show the possibility of such a task.

2.3 Classification

In the initial process of matching publication titles with RFCD codes, an attempt was made to match each staff member’s publications with one or several RFCD codes. The codes used to match the publications were both subdivisional and divisional codes (e.g. 280100 Information Systems, and 280101 Information Systems Organisation). Thus each publication had on average three lists of codes associated with it: one list of codes matched purely on the title of the paper or book, one list of codes matched on the title of the containing publication, and one list matched on both titles. Because we were interested in classifying staff members’ areas of expertise rather than their publications, we needed to gather the individual results together to provide a general classification of expertise for each staff member.

```
80 papers classified out of 95 by publication only
here are the rfcd codes and their count:
280500 Data Format: 30
280100 Information Systems: 182
280200 Artificial Intelligence and Signal & Image Processing: 53
280400 Computation Theory and Mathematics: 3
280300 Computer Software: 27
__________________________________________________________
72 papers classified out of 95 by paper title
here are the rfcd codes and their count:
280500 Data Format: 49
280100 Information Systems: 132
280300 Computer Software: 13
280400 Computation Theory and Mathematics: 14
280200 Artificial Intelligence and Signal & Image Processing: 36
_____________________________________________________________
90 papers classified out of 95 by paper title and publication
here are the rfcd codes and their count:
280500 Data Format: 79
280100 Information Systems: 314
280300 Computer Software: 40
280400 Computation Theory and Mathematics: 17
280200 Artificial Intelligence and Signal & Image Processing: 89
```

Fig. 3. Example of RFCD classification for a staff member

Since this method of matching would match an RFCD code on only one word in the title, each publication could potentially yield many matches. We decided to simplify the output by classifying staff members’ areas of expertise under the divisional (e.g. 280100 Information Systems) rather than the subdivisional codes (e.g. 280101 Information Systems Organisation). This involved adding up the number of subdivision matches under each division. This was not as straightforward as it seems. A paper with the word "information" in the title, for example, would match once with 280100 Information Systems, 280101 Information Systems Organisation, 280102 Information Systems Management, 280103 Information Storage, Retrieval and Management, and 280112 Information Systems Development Methodologies. Since each of these RFCD codes was counted as a match, the final count for the major division 280100 Information Systems would be 5 for this paper. However, giving such a large weighting to an RFCD code based only on one word would be misleading. Thus, since the same one word was matched from each of these codes, the final count for the major division 280100 Information Systems should be 1 for this paper (if the words 'information' and 'systems' were present, the count would be 2). Thus the number of matches associated with each division was altered to reflect the proportion of words in the title of the paper or publication that yielded the match.

The final output for the classification process is three sets of classifications for each expert: one set of classifications showing the divisional RFCD codes that were matched on the paper titles along with their counts, one showing the codes that were matched on the containing publication titles, and one showing the codes that were matched on both titles (Fig. 3).

2.4 Validating the Results

After automatically classifying each staff member's areas of expertise, we then needed to have staff members view the classifications and accept or reject them. We selected 20 staff members each with more than 10 publications and sent them their results asking them to indicate which they felt was correct and incorrect. We also asked them to select RFCD codes from a list provided to them that they felt most accurately represented their areas of research.

Table 1. Statistics for total number of publications for each staff member

	No. Publications
Mean	15.80702
Stdev	27.41509
Median	6
Mode	1
Max	169

On the 31st of March 2008, a new set of RFCD codes were released to be used from April 1 onwards. When we discovered this, the emails with the old RFCD codes had already been sent to all 20 staff members, and 10 had replied. We decided to classify the remaining 10 staff members with the new RFCD codes and resend them their results. The nature of the algorithm we used was such that it could just as easily classify staff publications using the new codes as it did with the old ones. As the introduction of the new codes was fairly recent, many systems and institutions are still using the old RFCD codes,[2] so we felt that classifying under these codes is still relevant, but only for the short term. Additionally, classifying under both the old and new codes may give us good information about which set of codes more accurately classified the staff members with the algorithm we used.

[2] See, for instance the ARC website: http://www.arc.gov.au/applicants/codes.htm (last accessed: 9/6/08), and *Find an Expert* at the University of Melbourne: http://www.findanexpert.unimelb.edu.au/ (last accessed: 9/6/08)).

2.5 Results

There were 57 members of the Computing Department who had publication data listed in IRIS. Information about the total number of papers for each staff member is shown in Table 1. Fig. 4 shows the percentage of papers that were able to be classified using only the paper title, only the publication title, and both the publication and paper titles. The Expert IDs are sorted by their total number of publications from smallest to largest. It can be observed that the paper title only method in most cases consistently performed worse than both the publication title only method and the publication and paper title method, while the paper and publication title method consistently performed the best.

On average, the system was able to classify 96.15% of each staff member's papers with the new RFCD codes, and 96.04% with the old RFCD codes using both the paper and publication titles. Additional information about the number of publications classified for the old and new codes is shown in tables 2 and 3, respectively.

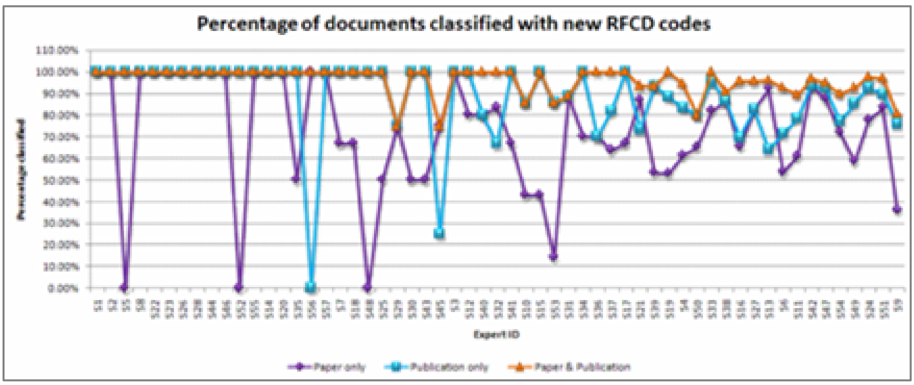

Fig. 4. Percentage of documents classified for each staff member via the three classification methods

Table 2. Percentage of documents classified with Old RFCD codes by classification method

Classification method	Mean	Stdev	Median	Mode	Max	Min
Paper title	70.55%	26.47%	71.79%	100.00%	100.00%	0.00%
Publication title	88.26%	18.35%	95.45%	100.00%	100.00%	0.00%
Paper & pub. title	96.04%	6.59%	100.00%	100.00%	100.00%	75.00%

Table 3. Percentage of documents classified with New RFCD codes by classification method

Classification method	Mean	Stdev	Median	Mode	Max	Min
Paper title	71.96%	26.65%	75.00%	100.00%	100.00%	0.00%
Publication title	88.26%	18.35%	95.45%	100.00%	100.00%	0.00%
Paper & pub. title	96.15%	6.48%	100.00%	100.00%	100.00%	75.00%

As can be observed in Tables 2 and 3, classifying on paper and publication title classified more documents on average than classifying on publication title only, which in turn classified more documents that classifying on paper title only. Table 4 shows the results of a Wilcoxon signed-rank test comparing each of the methods.[3] From this table it can be observed that that the three methods differ significantly in the average number of documents that they are able to classify. This indicates that, at least for publications authored by staff members in the Computing Department, more information about a publication than the title of the paper or book is needed. This is not surprising, as often a certain amount of creative license is taken with the title of a paper or book so that it may not be easily associated with the domain (e.g. "Training for High Risk Situations"). Conferences and journals, on the other hand, will generally contain domain specific keywords in the title (e.g. "Proceedings of Fourth International Joint Conference on Autonomous Agents and Multi Agent Systems").

Table 4. Wilcoxon signed rank test comparing percentage of documents classified by method x against percentage of documents classified by method

Method of classification 1 (x)	**Method of classification 2 (y)**	**W**	**ns/r**	**P(1-tail)**	**P(2-tail)**	**z**
Paper title	Publication title	-481	36	0.0001	0.0002	-3.77
Paper title	Paper & Pub. titles	-780	39	<.0001	<.0001	-5.44
Publication title	Paper & Pub. titles	-276	23	<.0001	<.0001	-4.19

2.6 Testing the Dictionary of Similar Words

We also wanted to test if our dictionary of words in the RFCD codes and words that share the same root, or refer to similar concepts, was able to classify more documents than if we hadn't used it. Fig. 5 shows the percentage of papers classified by paper title and publication with the old codes using the similar word dictionary versus the percentage classified without the similar word dictionary with the experts sorted on total number of papers. We can see that in most cases using the similar word dictionary will classify more documents than not using it, and never less. In fact, using the similar word dictionary classified on average 32.24% more papers with the old RFCD codes than not using it in our algorithm.

Fig. 6 shows the percentage of papers classified by paper title and publication with the new RFCD codes using the similar word dictionary versus the percentage classified without the similar word dictionary with the experts again sorted on total number of papers. We can see that in most cases the two methods classified an equal or very similar number of documents, even when the number of documents got quite large. In fact, using the similar word dictionary will classify on average only 2.83% more papers with the new RFCD codes than not using it with our algorithm.

This indicates that the new RFCD codes (at least in the division that we used) are more suited to our classification task than the old with regards to the number of classifications made. It is also pleasing that the changes by the ARC to the codes are more

[3] The Wilcoxon signed-rank test is a non-parametric alternative to a paired t-test. This type of test was used instead of a t-test, as the population could not be assumed to be normally distributed.

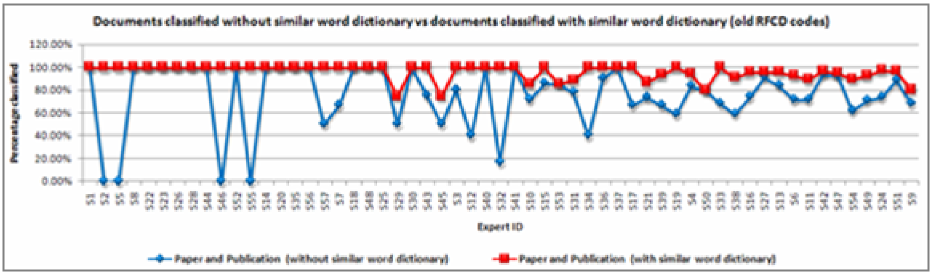

Fig.5. Percentage of documents classified with old RFCD codes for each staff member with and without the similar word dictionary for the paper and publication title method of classification

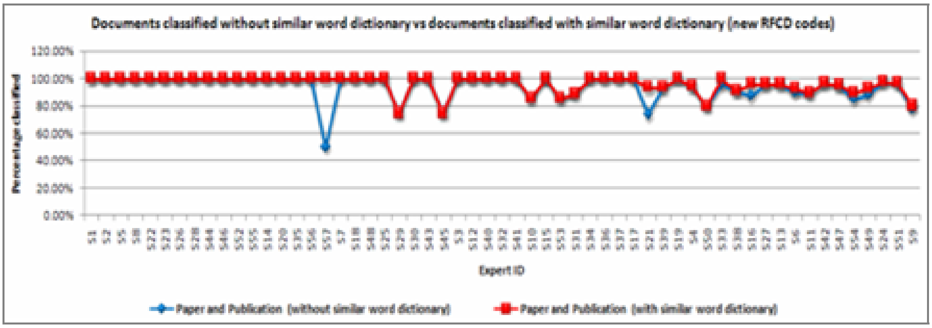

Fig. 6. Percentage of documents classified for each staff member with and without the similar word dictionary for the paper and publication title method of classification (new RFCD codes)

reflective of current research activity. Also very relevant is that there are 102 new RFCD codes as compared to 46 old RFCD classification codes.

2.7 Comparing Classifications against Staff Members' Responses

18 out of 20 of the staff members that we contacted responded (90% response rate): 10 with the old RFCD codes and 8 with the new RFCD codes. We asked the experts to indicate which of the RFCD codes they were classified with they felt were correct and which they felt were incorrect. We added the counts of each correct RFCD code together for each classification method (paper title, publication title, paper and publication title) to get an accuracy score for each expert. We then were able to calculate the percentage of accuracy between our classification and what the expert felt was correct.

From analysis of the individual data for each expert we found that all experts except one agreed with their highest ranked RFCD code. The expert who disagreed, S33, said that some of their work could not be classified under the RFCD division 280000 - Information, Computing and Communication Sciences, but rather under 410300 Cinema, Electronic Arts and Multimedia due to their work in computer games. However, they only had one document listed in IRIS that came under the category of computer games (i.e. that had anything to do with computer games in the title). It may be possible that this staff member had written papers on computer games before 2001, and thus these papers were not entered into IRIS, or after 2006, in which case we would not have had access to these papers as we were not provided with any data about publications written after this period. This does, however, raise the issue of the extent to which the system should accept an expert's validation of the data produced by automated searching. If the system simply accepts anything the expert says, despite there being no available evidence for it, it may very well experience some of the same problems faced by systems that rely entirely on self-reporting by experts. However, if the system refuses to accept any changes to the automated searching results unless it is sure that there is evidence somewhere that validates these changes, it may reject experts whose work is either too new, too old, or in a medium or location that the system does not search. While it is clear that a compromise is needed, it is not yet certain how this compromise can be reached or even if such a compromise is possible.

While 17 out of the 18 experts (94.4%) agreed with their highest ranked RFCD code, only 7 out of the 18 (38.9%) agreed with their second highest ranked code. The average percentage of accuracy of the system's codes was 62.05% overall, and 65.86% for the old codes and 57.30% for the new codes (see Table 5). The larger average for the old codes can probably be attributed to two factors: firstly, a couple of staff members who were sent the old RFCD codes responded very generally that they agreed with everything or thought that the classifications were 'good enough'. Secondly, a greater number of codes were assigned to each staff member when using the new codes. This is because the new RFCD codes not only have more categories, but some RFCD codes previously in the division 280000 - Information, Computing and Communication Sciences were moved to other divisions, so it was necessary to include these in the classification algorithm as well. Since the old RFCD codes did not have that many categories under division 280000, the staff members did not have as many codes to choose from when classifying themselves.

Table 5. Percentage of accuracy by RFCD code type

Code Type	Mean	Stdev	Min	Max
Old	65.86%	26.51%	24.42%	100.00%
New	57.30%	22.29%	18.75%	88.98%
Both	62.05%	24.41%	18.75%	100.00%

The fairly low averages for both the old and new RFCD codes could also be attributed to the staff members not having a clear understanding of what the codes meant. As each staff member undoubtedly had a preconceived notion of what their areas of research or expertise were, they may not have considered what each code actually represented, and whether a paper they authored may actually have fallen under a different RFCD code than they expected. On staff member, S51 mentioned that they had thought that most of their work could be classified under 280200 Artificial Intelligence and Signal and Image Processing, but seeing that the system had found their highest ranked code to be 280100 Information Systems, they realised that a lot of their recent work could be classified under this code

In general, this algorithm appears to be quite successful in determining the most prominent RFCD code for staff members in the Computing Department at Macquarie

University. It is not certain how this algorithm would perform for staff members in other departments or at other universities. Further work would need to be done to refine the algorithm to increase its accuracy and to test the algorithm on publications from other departments. However, there the accuracy of this and any other algorithm is limited by only having the titles of the paper and publication as input. Ideally the abstract of the document, or even the document itself should be included for the most accurate results.

3 Summary and Conclusions

This paper has described and implemented a simple method for automatic identification and population of a repository of experts and areas of expertise. It has described an experiment where members of the Computing Department were assigned several RFCD codes relating to their areas of expertise using their publication titles and a fairly simple string-matching algorithm. In summary, the algorithm was quite successful at determining each staff member's most prominent RFCD code, although in most cases it did make some false predictions. After each staff member in the Computing Department who had any documents associated with their names in IRIS had been classified with RFCD codes, 20 staff members with more than 10 documents each were sent their results. Each staff member was asked to indicate if the classifications were accurate and, if not, which RFCD codes they would use to classify their areas of research. A large percentage of staff members responded (90%) and all but one agreed with the RFCD code that was considered most relevant by the system to their areas of expertise. This was considered to be a very promising result from a fairly simple and crude method which, if developed further, could produce even more promising results. Our approach offers a viable alternative to relying on experts to enter and maintain their own data with the outcome being a validated set of recommendations based on integrated machine and human effort.

Acknowledgements. Thanks to the Computing Department staff for their participation in this project and the numerous others who participated in usability studies and interviews.

References

[1] Aïmeur, E., Onana, F., Saleman, A.: HELP: A Recommender System to Locate Expertise In Organizational Memories. In: 2007 ACS/IEEE International Conference on Computer Systems and Applications, AICCSA 2007, Amman, Jordan, pp. 866–874 (2007)

[2] Crowder, R., Hughes, G., Hall, W.: An agent based approach to finding expertise. In: Proceedings of the Fourth International Conference on Practical Aspects of Knowledge Management, Heidelberg, Germany, pp. 179–188 (2002)

[3] Ehrlich, K., Lin, C.-Y., Griffiths-Fisher, V.: Searching for experts in the enterprise: combining text and social network analysis. In: GROUP 2007: Proceedings of the 2007 international ACM conference on Supporting group work, pp. 117–126. ACM, New York (2007)

[4] Foner, L.: Yenta: A Multi-Agent Referral-Based Matchmaking System. In: AGENTS 1997: Proceedings of the 1st International Conference on Autonomous Agents Marina del Rey California, pp. 301–307 (1997)
[5] Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Communications of ACM 35(12), 61–70 (1992)
[6] Kautz, H., Selman, B., Shah, M.: Referral Web: Combining Social Networks and Collaborative Filtering. Communications of ACM 40(3), 63–65 (1997)
[7] Krulwich, B., Burkey, C.: The ContactFinder Agent: Answering Bulletin Board Questions with Referrals. In: Proceedings of the 1996 National Conference on Artificial Intelligence (AAAI 1996), Portland, Oregon, vol. 1, pp. 10–15 (1996)
[8] McDonald, D.W., Ackerman, M.S.: Expertise recommender: a flexible recommendation system and architecture. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW 2000), Philadelphia, Pennsylvania, United States, December 2-5, pp. 231–240 (2000)
[9] Pikrakis, A., Bitsikas, T., Sfakianakis, S., Hatzopoulos, M., DeRoure, D., Reich, S., Hill, G., Stairmand, M.: MEMOIR – Software Agents for Finding Similar Users by Trails. In: Proceedings of the 3rd International Conference on Practical Application of Intelligent Agents and Multi-agents, London, UK, pp. 453–466 (1998)
[10] Prekop, P.: Supporting Knowledge and Expertise Finding within Australia's Defence Science and Technology Organisation. In: Proceedings of the 40th Annual Hawaii International Conference on System Science (HICSS-40), Hawaii, January 3-6 (2007)
[11] Reichling, T., Schubert, K., Wulf, V.: Matching human actors based on their texts: design and evaluation of an instance of the ExpertFinding framework. In: Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, Florida, USA, November 6 - 9, pp. 61–70 (2005)
[12] Resnick, P., Varian, H.R.: Recommender systems. Communications of ACM, special issue 40, 56–58 (1997)
[13] Vivacqua, A.: Agents for Expertise Location. In: Proceedings of AAAI Spring Symposium on Intelligent Agents in Cyberspace, Stanford, CA, pp. 9–13 (1999)