People and Expertise Search

June 2017

Martin White

INTRANETFOCUS

Contents

Exploiting corporate knowledge

Finding information about people and finding people with the expertise that would help solve a business problem are often given as a core business case for enterprise search. As a result it is quite common to find two search boxes on an intranet, one of which is for 'people search' even though it may well not use the search application.

Although searching for a named person and searching for someone with specific expertise could be seen as a single search use case, in practice they could not be more different. A search for a person is, by definition, an exact match search. This can be a significant challenge because often the 'correct' spelling of the name of a person is not known. The role of the search application is to help in identifying the 'correct' name as in most cases a person search will be carried out against a database which only permits exact matching.

In addition, there could be a requirement to find references to a person in a document or other content, and in this case the search application is using entity recognition to match a name against a query. The outcome will be a very high precision search. If the search fails to locate the named person then there will be other ways of tracking them down. It may take time but in the end the named person should be able to be found.

The issue of how to define 'expertise' is highly complex and books have been written about it, notably The Cambridge Handbook of Expertise and Expert Performance, published in 2006 and running to almost 1000 pages. The searcher looking for expertise and the person describing their own expertise may use different words and concepts. Just matching the query to the expert profile is very challenging. Seeking expert help usually only comes when other channels, such as a search through documents or talking to colleagues, have failed to provide a solution to a problem.

Finding someone in the organisation then becomes not only very important but marks the end of the discovery road. If no expert can be found then the problem cannot be solved and the organisation is now at risk of not meeting an objective. If the query is difficult to formulate and the expertise is inevitably not well defined the outcome is very likely to be a list of possible experts. Then comes the decision about which expert to contact. How can expertise be ranked, and is a search application able to do this in a way that meets the expectations of the searcher and the expert?

The development of expertise search applications dates back around twenty years. It is only over the last few years, and in particular the last few months, that software solutions promising the ability to find people with particular expertise have become highly visible in the market. This report examines the technologies behind both people search and expertise search and identifies some of the wider issues about the sharing of expertise inside an organisation.

Data privacy implications

All EU Member States and many other countries around the world have legislation to ensure that the data privacy rights of individuals are adequately protected. In the case of the EU, new General Data Protection Regulations come into force in May 2018, with potential fines for organisations in breach of these Regulations at the level of 4% of gross revenues. Any initiatives to enhance any database application which records personal data should be signed off by the appropriate manager as being compliant with the Regulations. This is especially important where this information will be transferred to locations outside of the EU.

Defining the requirements

There are very few large-scale studies of the way in which employees search for people inside an enterprise. That in itself is interesting. Many search and social software vendors make claims about how important it is to locate people by expertise but present no research information. A study carried out within IBM in 2011 showed the following distribution of the information being sought.
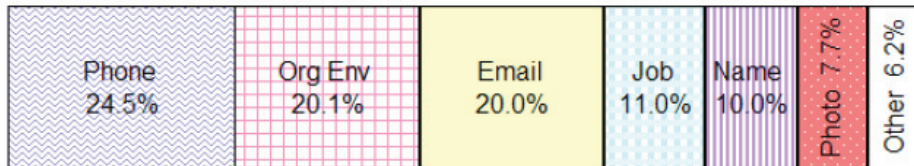
From 'Best *Faces* Forward: A Large-Scale Study of People Search in the Enterprise'. https://www.research.ibm.com/haifa/dept/imt/papers/guyCHI12.pdf

This would suggest that locating expertise was not a core requirement. The research also indicates that there are multiple entry points into profile information. For example, an employee may know the name of the head of a functional area and use an organisational structure to locate a direct report in a specific country. Having located the person they may then look at their experience and expertise to find a common interest that would provide a basis to facilitate the initial contact. A search use case that is often overlooked is the requirement to find the name of a person with a specific role. "Who is the intranet manager in Sweden?" The problem here is that the person managing the intranet may not have the title Intranet Manager, and even if they do the title might be in Swedish and not English.

Defining requirements for expertise search is discussed in more detail later in this report but for people search the critical issue is whether the search is just against an internal HR-owned database or against any occurrence of the name in a document. This raises the question of whether there should be separate search boxes for Search and People. In my opinion there should just be one search box with the search application then offering options to limit the search to an HR database or to all occurrences of the name.

Name search

Name search is very easy for a user to evaluate. All they have to do is search for someone they know. From the moment they find that the search application does not immediately find this person they are unlikely to trust the search engine again. Almost certainly they will also use their own name as a search term and then be either very surprised or very concerned about the amount of information they find!

To illustrate the challenges a case study may be helpful. In a recent project for a client with over 100,000 employees the Intranet Focus team had been asked to interview Carsten Hollinger. A search for this person returned no results. We then discovered that the correct spelling of the manager's name was Karsten Hollinger, but the search application was not set up to manage pre-stemming. Of greater concern was that there was only one person in 100,000 with the family name of Hollinger. The search application should have proposed Karsten Hollinger as a suggestion based just on the family name with a "Did you mean?" option.

Even if the nominal business language of an organisation is English the issues of language quickly appear in name searching. Many employees will have family or given names which reflect the culture and heritage of their family and not the language that they may speak. Even working out which is the family name and which is the given name can be very difficult. In Chinese, as in many other Eastern languages, the family name precedes the given name, but it could be that the person concerned has inverted the structure for use in a Western culture.

Another common challenge is when a name requires the use of an extended alphabet. The Swedish given name Åsa is not the same as Asa and in an alphabetical list comes after Z. It may well have been transcribed differently, especially if the HR database cannot cope with extended character sets.

To gain an appreciation of the complexity of searching for a name it is advisable to read the range of briefing papers from Basis Technology. The text of the section that follows is based on a Basis Technology white paper "The Name Matching You Need – A Comparison of Name Matching Technologies" published on 29 February 2012.

Typographical errors
A slip of the finger at the keyboard causes transposition of characters, missed characters or other similar errors. (e.g., "Htomas" or "Elizbeth")

Phonetic spelling variations
Some names simply sound alike, but are spelled differently (e.g. "Christian" and "Kristian"). Neglecting to confirm spelling produces errors. (e.g., "Cairns" vs. "Kearns" vs. "Kerns"; or "Smith" vs. "Smyth").

Transliteration spelling differences
Multiple transliteration standards or "approximate" transliterations from a non-Latin script to English lead to multiple spelling variations. In the case of Arabic to English, Arabic has many consonant sounds which might be written with the same English letter, and Arabic vowels may be expressed more than one way in English, giving rise to many spelling variations. (e.g., "Abdul Rasheed" vs. "Abd-al-Rasheed" vs. "Abd Ar-Rashid")

Initials

Sometimes all name components are spelled out, other times initials are used. (e.g.,"Mary A. Hall" vs. "Mary Alice Hall" vs. "M.A. Hall")

Nicknames

In some cultures, nicknames are numerous and may be often used in place of a person's formal name (e.g., "Elizabeth", "Beth", "Liz", and "Lisbeth")

Re-ordered name components

The order of family name and given name may appear swapped due to database format or ignorance of cultural naming convention. (e.g., "JohnHenry" vs. "Henry, John"; or "Tanaka Kentaro" vs. "Kentaro Tanaka")

Missing name components

Sometimes a middle name or patronymic (personal name derived from ancestor's name - e.g., Olafsson = "son of Olaf") may be absent. (e.g., "Abdullah Al-Ashqar" vs. "Abdullah Bin Hassan Al-Ashqar"; or "Philip Charles Carr" vs. "Philip Carr".

Missing spaces

Some names are commonly written with spaces in different places, both in common English names (e.g., "Mary Ellen", "Maryellen", and "Mary-Ellen") and those less common in English (e.g., "Zhang Jing Quan" and "Zhang Jingquan").

Names in different languages

Names from languages using different writing systems can be notoriously difficult to match against English representations of the names. Here is just one name spelled in English, Russian, simplified Chinese, and traditional Chinese, respectively: "Mao Zedong" "Цзэдун", "毛泽东", or "毛澤東").

Writing personal names in forms other than their native languages is essentially a type of translation. The process of translating words between languages can be divided into three categories.

- zero translation, where the usage of a word in one language is directly adopted into another
- phonetic translation, where the spelling or pronunciation of the source language (SL) is expressed in the alphabet of the target language (TL)
- meaning translation, where a word of the SL is substituted for a word from the TL with an identical meaning. That would render my name as Martin Weiss in German!

Chinese, Japanese and Korean names present very substantial challenges. Korean names have a single syllable given name and a two-syllable family name. Western names are uniformly spaced between given name, middle name, and surname. By comparison, the three syllables of a Korean name can be written as all attached or spaced. Inconsistencies in separating the two syllables of the given name then leads to difficulties in Anglicised name identification.

A further complication arises from the fact that in the case of Korea (and the situation is similar in China) there are comparatively few family names. In the UK there are well over one million family names but there are only 286 Korean family names listed in the 2010 South Korean census. The large number of family names in most European languages means that it is unusual for there to be more than two or three people with identical names even in a large European company.

A recent challenge in language management has been the arrival of Arabizi, which is also referred to as Arabish or Araby. This is a form of Arabic used for text messaging, blogs and microblogs and for communicating in Arabic when only a QWERTY keyboard is available on a smartphone or tablet.

Wikipedia has many good entries on issues around naming conventions, including

Arabic	http://en.wikipedia.org/wiki/Arabic_names
Chinese	http://en.wikipedia.org/wiki/Chinese_names
German	http://en.wikipedia.org/wiki/German_names
Japanese	http://en.wikipedia.org/wiki/Japanese_names
Portuguese	http://en.wikipedia.org/wiki/Portuguese_names
Spanish	http://en.wikipedia.org/wiki/Spanish_names

Core information

Decisions need to be made about the core information that should be associated with each employee, such as role, reporting lines, email address, telephone numbers etc. In a multinational organisation, some of this information could be in a local language and it is easy to miss off international dialling codes. Consistency and currency are important. There is nothing so frustrating as finding out-of-date contact information for someone you are trying to contact.

Evaluating people search functionality

The user satisfaction barrier with people search is very high indeed. The number of times people search is needed may be infrequent but when there is a need the solution needs to deliver. Rarely is people search performance evaluated at Proof of Concept or User Acceptance Testing stages but it should be. It is not a function of the number of countries that an organisation operates in. Within the EU there could be people from any number of the 27 Member States (and other countries) working in any one single country and all present challenges in name finding.

Capitalising on expertise

Although it is unclear who at Hewlett Packard (HP) said "If only HP knew what HP knows", the company was very much in the vanguard of the development of knowledge management in the mid-1990s. One of the applications that was developed was Connex, which was among the first systematic attempts to create an expertise database of a large multinational organisation. Around the same time Microsoft launched its Systems Planning and Development (SPuD) application to link developer expertise around its global research centres. This was based around 137 implicit and 200 explicit competencies derived from a list of up to 60 competencies that an employee could use to describe their expertise. This selection was agreed in a discussion with their manager. Twenty years later we are seeing a renaissance of interest in expertise databases, driven by a technical capability to build profiles of the expertise of employees by tracking their work activities.

In principle the benefits of being able to identify expertise are clear but the processes of doing so are more complex than often appreciated. In a professional services company the core business case is selling expertise to clients. In a manufacturing company the expertise is bundled into the product and so is far less visible. Another factor is whether being recognised as an expert is going to result in more rapid promotion through the organisation. If this is the case then there are some tangible benefits for someone taking the time to make their expertise widely available. When it comes to diversified organisations, typically large multinational companies, the challenge becomes one not only of scale but of identifying expertise gained in one area that could be applicable to others.

There are six primary mechanisms through which personal expertise can be integrated into business processes.

Tracking

Employees track the progress of projects and other business activities and are able to offer their expertise to solve problems that become evident through social media and email messages. Whether they do so is entirely their own decision.

Sharing

A member of a group may not only offer their own expertise but link the group to an expert in another group, usually making the introduction. This process is often referred to as 'boundary spanning'.

Promoting

Profiles of experts are linked to specific search terms and appear as promoted content in search results.

Responding

In many organisations individuals and groups can post requirements for expertise on one or more social platforms, in anticipation that experts within the organisation will track these platforms and respond if they choose to do so. There are many external examples of these Community Question Answering applications, such as Quora, Stack Overflow and Mathematics Stack Exchange.

Identifying

A search is carried out for documents on a topic and the names of people who have authored the document or who are mentioned in the document are presented as a facet of the search user interface. One of the problems with this approach is that the author or contributor may have left the organisation.

Profiling

A profile of each employee is prepared by tracking the actions they have taken in writing their own profiles, writing corporate and external documents, contributing to discussion groups, attending meetings and working on projects. Jeremy Thake (Hyperfish) has published a good summary of the way in which Office 365 creates profiles.

It is important to appreciate that in the case of Tracking, Sharing, Promoting and Responding the expert takes the decision whether or not they wish to share their expertise. In the case of Identifying and Profiling, both of which are search-based, it is the requestor who makes the contact. The fundamental questions are then whether or not the expert is available and is willing to share their expertise.

Group expertise

Little attention seems to be paid to the fact that it is not always an individual that has the expertise needed to solve a problem. It could well be a group. This is likely to be the case in high risk or disaster management situations. It might well take too long to identify all the people in the organisation that would be able to respond to the requirement and then bring them together in a team. Instead there is a requirement to identify a group or team with the collective skills, and this is something that is invariably not addressed with technology-based expertise finding.

Barriers to sharing expertise

Quite a large number of studies have been carried out into the barriers to expertise sharing once the seeker has located an expert. The potential barriers include the following, which are based on a review paper by Morten Hertzum in which he reviews the outcomes of 72 papers on expertise finding and provides an excellent introduction to the topic.

Context
- Company size and culture does not support expertise seeking and sharing
- National management cultural issues inhibit direct access to expertise
- There is no incentive for the sharing of expertise
- There is an incentive to refrain from sharing to ensure status as the 'go-to' person

Seeker
- Time-consuming to get a response from the expert
- Expert is not willing to commit to a time when they can be available
- Expert is not willing to commit to a time by when they will reply
- Seeker does not personally know the expert and it is difficult to build a relationship
- Expert wants to know why their expertise is required before agreeing to share
- Seeker feels they are losing face by revealing uncertainty and lack of knowledge
- Seeker cannot formulate question because of a lack of expertise
- Expert only willing to give an oral answer, either because they do not have the time to write a reply or do not want to be quoted as the source of the expertise

Expert
- Expert's knowledge turns out to be incomplete or unreliable
- Expert cannot be immediately located – an issue with experts often travelling
- Expert is perceived as unapproachable or unwilling to help
- Expert is biased in their reply but the bias will not be evident to the non-expert
- Expert has concerns about sharing confidential information
- Expert may have expertise which is not visible as it was gained on a confidential project
- Expert's credibility difficult to assess
- Expert is not up-to-date with developments
- Expertise required is owned by a group rather than an individual
- Expert is external to organisation
- Expert works in a different first language than the seeker which inhibits a dialogue

The fundamental question for any organisation planning to use search-supported Identification and Profiling approaches to locate expertise is the extent to which any of these barriers may apply, especially as every expert may be behind one of more of these barriers.

It should also be noted that most expertise database applications take into account US/Western European attitudes to team working and the sharing of knowledge. The situation is quite different in countries to the east of the Mediterranean Sea where hierarchical management is much more rigorously imposed. For example, a Japanese employee might be embarrassed to be asked direct for their expertise – they would expect the approach to come via their line manager and for the response to go back through their line manager.

In addition, Japanese companies invest significant time and resources in training to develop employees as generalists. They rotate individuals around the business so they gain broad experience and so the experts that do exist tend only to be in research and development departments.

If these barriers are significant (and this may only be apparent once wide-scale implementation of a technology 'solution' has taken place) then the investment in the technology will be wasted. It does not help that there will usually be no reliable information as to the extent and value of Tracking, Sharing and Responding.

Making the business case for expertise finding applications

The question to answer before considering any expertise profiling solution is why these existing channels are not being effective. Is it because of fundamental issues about sharing expertise or because employees are not aware of these channels? Solving the barriers to expertise sharing and raising the visibility of existing channels may result in both substantial improvements to expertise sharing and a better definition of requirements for where technology solutions can add additional value.

To make the best use of the technology there should be an expertise location strategy that is linked into HR processes, knowledge management, training, job appraisals and social media development. Finding people with expertise is not a 'search problem'. Good search tools can certainly help but without attention being paid to profile quality (even if other types of content are being searched) and a commitment by employees to share their knowledge expertise, discovery will not be as successful as anticipated or required.

The development of expertise finding

The challenges of locating expertise inside an organisation have in the past been the concern primarily of knowledge managers. The core problem has been working out how to capture expertise in a document format. This was the basis for My Site profiles in SharePoint 2007. Employees could include a profile of their expertise in a My Site and this could be searched as a defined field by the SharePoint application. IBM had been carrying out a substantial amount of research in parallel as a component of the development of Domino and subsequently IBM Connections.

This work led to one of the IBM knowledge management team, Dave Snowden publishing a blog in 2008 in which he highlighted some fundamental issues with capturing knowledge/expertise.
http://cognitive-edge.com/blog/rendering-knowledge/

His observations include
- Knowledge can only be volunteered. It cannot be conscripted.
- You can't require people to share their knowledge, because you can never measure if they have.
- We only know what we need to know when we need to know it.
- The way we know things is not the way we report we know things.
- We can always know more than we can say and we will always say more than we can write down.

The implication of these five principles is that no matter how much effort is put into persuading employees to write down what they know and what their skills are, it will only be a very partial and probably biased commentary. Searching through these self-completed profiles is not a complete solution to identifying expertise and knowledge because of the inevitable inconsistencies between profiles. Then comes the challenge of keeping these profiles current. If there is no overall policy on profile management, supported by managers, then there is no incentive for people to spend time on this process.

The concept of machine-compiled profiles is certainly not new. Sopheon was just one of the companies working on this technology in the late 1990s, as was a team developing P@noptic Expert at CSIRO in 2001. In 2006 MITRE Corporation published a report on the principles of expert profiling and included descriptions of the applications from TACIT, AskMe, Autonomy IDOL K2, Endeca 25, Recommind 30, Triviumsoft's SEE-K and Entopia Expertise Location. Although this report is now ten years old it provides still-relevant advice on specifying and evaluating expertise finding applications.

After its experience with Connex in the 1990s, Hewlett Packard put a significant amount of effort into its Enterprise Collective application from 2010-2012. The innovation in this application was the deep profiling of the use being made by employees of work products. There is an interesting Research Paper dating from 2012 from HP Labs which gives a substantial amount of detail into the processes behind the application but it seems not to have been developed as a commercial product.

Over the last few years there has been a steady stream of expertise search applications from search vendors (e.g. Attivio, BAInsight and Sinequa) and from specialist application vendors (e.g. Profinda and ThingMap). Both technology and market development has been stimulated by profiling software developed to track terrorists by national and international security organisations.

The technology of expertise profiling

The technology (more specifically the algorithms) behind expertise search is very complex and this short section cannot do justice to the range of approaches that are being used at present, and are in development.

Document-centric profiling identifies the relationships between people and topics in collections of documents, using a classic Vector Space Model to rank the outputs. Because people know when they have written or been cited in documents they will have some sense of the potential relationships and will not be surprised to be contacted by another employee that starts the conversation by referring to a specific document. Ranking of expertise is primarily a function of the rank of the document and not specifically the expertise of the contributors

Relationship-centric profiling takes into account a wide range of sources, which might include:
- Participation in projects, teams, and groups
- Comments in blogs, microblogs, and discussion forums
- External papers and reports (e.g., papers in peer-reviewed journals and patents)
- Internal papers, patents and reports
- Being cited in internal papers and reports (i.e., not just as the author)
- Internal and external presentations
- Structured CVs

One of the objectives of this approach, especially with respect to crawling graph networks, is to provide an internal solution to the page link algorithms used by Google. One of the potential implications of this approach is that the profile information created could be used to enhance enterprise search performance by using these links in a similar way to that adopted by Google for web search.

It is important to appreciate that the algorithms used to crawl and interrogate artefacts of expertise are not the same as those used to match a query to one or more experts. Query management for experts has to take account of the fact that if the searcher knew exactly what they were looking for then they would probably already have the solution to the problem. In reality the queries will be exploratory, perhaps initially with just one or two words, as an initial test of the potential value of the system.

The mathematical modelling of the probability that a profile satisfies a query is very complex

$$p(q|\theta_{ca}) = \prod_{t \in q} \{(1 - \lambda) \sum_{d} p(t|d) f(d, \theta_{ca}) + \lambda p(t)\}^{n(t,q)}.$$

This equation is only to show that there are multiple factors that affect the probability and that makes ranking of expertise a very challenging issue. Not only is there the inherent multifactor characteristic but expertise can change with time. A given expert may not have been in a position to continue to enhance their expertise, and being involved in a current project (the scope of which may be on a need-to-know basis) could have substantially enhanced their expertise.

The next question is to consider what the user is seeking.

- Anyone with at least some of the expertise that could be appropriate
- At least one person that matches all the required expertise
- A list of all the people that match some of the expertise requirement
- A list of all the people that match all the required expertise
- The most expert person in the organisation
- A ranked list of the people that match the expertise requirement.

All of these options have implications for the way in which the query process is managed. Much will depend on the weighting of the elements of the profile. It is not unusual for individuals to ask for their profiles to be weighted so that they have a higher profile. However, this process could have unforeseen implications for other profiles and the way in which they match a query.

In summary, using search to locate expertise is not going to produce a definitive list of "people who know." It may provide a starting list, but the moment a user finds a reference to someone whose skills they have little confidence in, the level of trust in the application will drop to zero and will not bounce back.

From research to product

As a very rough estimate around 1000 research papers a year are currently being published on various aspects of expertise profiling and expertise finding. As a topic, it presents some very interesting challenges in working out how to match poorly defined enquiries against complex collections of data and information and then produce a ranked list that is of value to the user. These papers exclude any reference to the challenges of using expertise profiling in an organisation as this was not in the scope of the research.

A particular challenge arises in validating the algorithms. Access to corporate collections is not possible but there are test collections, such as those developed for the TREC Expertise Track and the LExR collection for academic expertise. The extent to which these collections are representative of the requirements of organisations is always a major concern for research teams.

Of greater concern is that vendors offering software applications also make no reference to wider organisational and cultural issues, especially with regard to data privacy issues, the ethics of tracking the actions of employees and the value of expertise search against other ways of integrating expertise into decision making.

Impact of incomplete profiles

No matter how good the technology of the application there are many situations in which the query may not be matched by the profiles of individual experts.

These include

a. People working on projects where the security clearance is limited to a need-to-know access list. The paradox here is that these could be among the most expert in the organisation as they are working on the most sensitive and innovative projects.

b. People joining the firm may not be in a position to disclose what they have been working on for (perhaps) a competitor. The paradox here is that these people have been specifically hired for their expertise. Since on average around 10% of an organisation's employees leave each year and 10% join the question is how long it will take for the new joiners to be regarded as having equivalent expertise to colleagues who have been with the organisation for many years.

c. People working for contractors and advisors who are retained specifically for their expertise but as they are not employees they do not show up in an expertise search.

d. People who do not want to be bothered with others asking for advice. They might well game the system by (for example) not writing blogs or not being named as the lead author on a report.

e. People with soft skills in areas such as mentoring, training and team management. These skills can be very valuable to an organisation but may well not show up on a profile where the emphasis is on technical skills.

f. People with hidden skills who may have joined the organisation from a completely different business area, perhaps as a deliberate change of career. Clearly the organisation sees their work as important but their experience to date may not be reflected in the profile.

g. People who have cultivated a range of external social media and professional networks that may not be included in an internally-sourced profile but which could be important indicators of expertise.

h. People working in a language that is not included in the indexing of the crawls and therefore do not show up in the responses to a query.

In all these cases the expertise of these employees may not be ranked in a way that reflects their expertise.

Taxonomy development for expertise

There has been a substantial amount of interest in the development of taxonomies to enhance the performance of expertise databases. As with the addition of taxonomy-based metadata to documents, the discussion always centres around the balance between the effort to develop a taxonomy and to apply the terms against the benefit in retrieval performance. Increasingly semantic and cognitive models are being applied to support taxonomy development and the paper by Delroy Cameron and his colleagues is a good example of this approach.

Profile management and data privacy

A very sensitive area in the creation of profiles is the extent to which the person being profiled is able to see and edit their auto-created profile. Since this profile may contain personal information, in the EU and in other countries where there is data privacy legislation the employee has the right to ensure that this information is correct.

As they review their profiles it could be that they will become concerned about the level of detail that is being included in the profile and whether it accurately reflects their expertise. As indicated above it is not just the level of detail but the rating that is applied. They may well wish to know who decided on the rating of the expertise and perhaps even wish to challenge the rating. They may also be concerned that their future in the organisation may be judged, even if only partially, on the ranking of their expertise against their peers.

Expertise finding

In selecting an expertise finding application it is important to appreciate that the technology used for profiling and the technology used to support finding need to be in total harmony. If the profiling is excellent but the finding is not effective then the investment has been a waste of time and money. If the profiling is poor but the find interface is very powerful then the list of experts presented may be not only incomplete but worse may be misleading. A recent paper from IBM illustrates well both the complexities of expertise finding and the areas which still require significant research.

The range of retrieval models that have been developed to match a query against an expertise file include (excerpted from Expertise Retrieval – Foundations and Trends in Information Retrieval, Now Publishing, 2012 p36).
http://www.nowpublishers.com/article/BookDetails/INR-024

- Generative probabilistic models estimate associations between query topics and people as the likelihood that the particular topic was generated (i.e., written) by a given candidate (topic generation models) or the other way around, that a probabilistic model based on the query generated the candidate (candidate generation models).

- Discriminative models capture associations between query topics and people by directly estimating the binary conditional probability that a given pair of a query topic and a candidate expert is relevant.

- Voting models generate associations between query topics and people as a voting process that allows documents ranked with respect to a query to vote for candidate experts by different weight schemes.

- Graph-based models determine associations between query models and people by inference on an expertise graph, comprising documents, expert candidates, and different relationships; the graph can be built in a query dependent or query-independent manner.

- Other models use a range of ways of thinking about associations between query topics and people, including modelling people as a distribution of latent variables corresponding to topical themes (author-topic models).

The question for a potential purchaser of an expertise search application is which model is being used, and is it the most appropriate model for the way in which the organisation will search for expertise.

One of the few case studies of expertise retrieval has been carried out by IBM in conjunction with Georgia Institute of Technology. In the summary of this paper Svetlana Yarosh, Tara Matthews and Michelle Zhou note the following

Overall, the ten most useful pieces of information were as follows:

- Company division
- Expertise summary
- Job responsibilities
- Job title
- Geographic location
- Past/current activities & projects
- Past experiences
- Average time to respond to contact
- Top 5 social tags assigned by self and others
- Public documents related to the search term

https://pdfs.semanticscholar.org/3699/09f05141779edb694d0d0834eeb78faac6b2.pdf

This list should be considered cautiously, since there was a great deal of individual difference in the ratings of any particular item (the average of the standard deviations for all items was 1.24). In fact, for every item on the list, at least one person rated it as very useful (5) and at least one person rated it as not useful at all (1), so there was no universal agreement as to which items were absolutely necessary or absolutely useless.

Furthermore, the type of information that was rated most useful for expertise selection varied by task and topic. There were several pieces of information that showed up in the top ten for at least three task/topic types that did not show up in the general top ten:

- Time at current position
- Social networking site current status message
- Reported availability for one-hour meeting
- Approachability rating
- Rating of answer quality for similar searches
- Rating of general answer quality

The overall message of this research project is that the process of finding expertise is not as simple as entering a query into a search box. Behind the search box the query parsing has to be able to match the query terms against the profile information. The problems identified in the IBM study indicate why it is so difficult to rank the expertise of a number of experts and then deliver a dependable list to the search user. In principle it is the same problem as with search relevance. The assessment of relevance is in the mind of the user.

Technology selection and implementation

There are three technology options for expertise search

- Use the expertise profiling capabilities of your current enterprise search provider. Both Attivio and Sinequa are promoting solutions at present.
- Invest in one of the many expertise-specific applications that are now on the market and run it as a separate search application, perhaps with its own search query box.
- Federate the expertise-specific application into the core enterprise search application.

There is no correct answer as there are many factors that must be taken into account. Another consideration should be whether now is the right time to invest. The rate of development of profiling solutions is so great that the functionality and usability of both the profiling and query elements should improve considerably over 2017.

As with any search application undertaking proof of concept tests and user acceptance testing is far from easy. It may be quite straightforward to test some aspects of the user interface with a small group of profiles, but as the entire purpose of the investment will be to locate expertise across the organisation the final tests cannot be carried out without a substantial load of profile information. To do this the application may need to crawl through a significant archive, which presents many technical and procedural problems.

Another consideration is how the performance of the application is going to be assessed. This is not just to ensure that it is working to specification but that it is having a measurable impact on access to expertise over and above the access provided by the other approaches suggested on Page 5 ('Capitalising on expertise). What is almost

certainly going to happen is that people will compare the expertise search with a search through documents using the enterprise search application to see if different experts are identified. If there are there need to be some good explanations if employees (and experts) are to trust the application.

The report from MITRE (see Resources) offers very good advice on selecting a vendor and an expertise-finding application and sets out some lessons learned from implementations.

In conclusion

Expertise search has a role to play in making the expertise of the company more widely available. If it is positioned as one of many channels to expertise discovery then search can have a role to play. If the aim is to overcome cultural and organisational barriers to expertise sharing then all that search may do is make them more visible.

Special care is needed to ensure that the expertise of new joiners is available to the company, and in return they have access to the expertise they need to grow their careers and reputation. Expertise search should be jointly owned by the knowledge management team as a primary objective of such a team is to support the exchange of knowledge within the organisation.

Recommendations

Finding out information about other employees in an organisation is regarded as a key role for the search application in an intranet. However, finding the 'correct' spelling of a name is a substantial challenge and the solution needs to be subject to careful testing prior to implementation as a failure to find a known person will result in a significant lack of trust in the search application.

To get the best out of IT applications for expertise finding it is essential that they are selected and implemented within an overall knowledge management/expertise management strategy. This strategy should identify (and where needed resolve) barriers to the exchange of expertise which may vary widely across an organisation.

Special care is needed to ensure that the expertise of new joiners (including employees taking on new and different roles) is available to the company, and in return they have access to the expertise they need to grow their careers and reputation.

It should be recognised that the value of using an IT application for expertise finding through profiling may not emerge for some time after implementation as it will take time for profiles to be developed. They do not represent a 'quick fix' and if fundamental expertise sharing barriers and/or the IT application result in no significant value being gained the overall quality of access to expertise may be irreparably damaged.

The implications of data privacy legislation must be considered, especially where expertise solutions are being implemented in countries which do not have the level of data privacy legislation adopted by the EU and a few other countries around the world.

Acknowledgements

Resources

This is a list of the core books and review papers on expertise finding.

The Cambridge Handbook of Expertise and Expert Performance.

Expertise Retrieval, Krisztian Balog, Yi Fang, Maarten de Rijke, Pavel Serdyukov, and Luo Si, Now Publishing Foundations and Trends in Information Retrieval, 2012
http://www.nowpublishers.com/article/BookDetails/INR-024

MITRE Technical Report MTR 06B000040 Expert Finding Systems, Mark T. Maybury, 2006
https://www.mitre.org/sites/default/files/pdf/06_1115.pdf

A survey on expert finding techniques Shuyi Lin, Wenxing Hong, Dingding Wang and Tao Li Journal of Intelligent Information Systems DOI 10.1007/s10844-016-0440-5

Evaluation of Retrieval Algorithms for Expertise Search, Gaya K. Jayasinghe, Sarvnaz Karimi and Melanie Ayre.
https://www.researchgate.net/publication/311315840_Evaluation_of_Retrieval_Algorithms_forExpertise_Search

On the Assessment of Expertise Profiles, Richard Berendsen, Maarten de Rijke, Krisztian Balog, Toine Bogers and Antal van den Bosch. Journal of the American Society for Information Science and Technology, 2013, 64(10) 2024-2044

The World Is Not Small for Everyone: Inequity in Searching for Knowledge in Organizations Jasjit Singh, Morten T. Hansen and Joel M. Podolny, Management Science 56(9):1415-1438. http://dx.doi.org/10.1287/mnsc.1100.1201

Expertise Seeking: A Review, Morten Hertzum Information Processing & Management, vol. 50, no. 5 (2014), pp. 775-795 http://mortenhertzum.dk/publ/IPM2014_preprint.pdf

Collaborative Information Seeking and Expertise Seeking: Different Discourses about Similar Issues. Hertzum, M. (forthcoming). To appear in Journal of Documentation.
http://www.emeraldinsight.com/journal/jd

The author

Martin White

Martin White is an intranet, enterprise search and information management strategy consultant. He founded Intranet Focus Ltd in 1999.

He has undertaken projects or managed business operations in nearly 40 countries. Martin has been managing global virtual meetings since the mid-1970s. He began his consulting career in 1979 and has participated in several thousand meetings in around 500 projects carried out in 40 countries. Not all have been good meetings but it is from the meetings that don't go to plan that the most valuable lessons are learned. An information scientist by profession, Martin has been a Visiting Professor at the Information School, the University of Sheffield since 2002. He writes a monthly column on enterprise search for CMSWire.

Prior to setting up Intranet Focus he held senior management positions with Reed Publishing, International Data Corporation and Logica plc.