



# Searching for people and for expertise

Intranet Focus Ltd Research Note 03/13

June 2013

Intranet Focus Ltd  
12 Allcard Close  
Horsham RH12 5AJ, UK  
+44 1403 267030  
[www.intranetfocus.com](http://www.intranetfocus.com)

# Searching for people and expertise

## Summary

One of the most important uses of a search application is to be able to find people working for the organization or to identify people with specific skills and expertise. In the quest for 'relevance' most of the limited effort available to optimize a search experience is directed at document search, and little attention is paid to these two aspects of people search. Language and culture play an important role in people search. Even if the organisation's corporate language is English employees will come from many different countries and cultures, each with a different and distinctive way of spelling and setting out their name.

Searching for expertise is equally challenging. An approach that assumes employees will write down all that they know is simplistic and ineffective. Many organisations, notably HP and IBM, are developing search applications which can blend together a range of expertise characteristics that are acquired passively through membership of communities of practice and contributions to the technical literature.

Searching for people and for expertise should be a core element of a search strategy for any organisation that wishes to gain the best from its employees, in particular those who have recently joined the organisation.

The Research Note concludes with a set of eight recommendations.

## Contents

<b>1. Introduction</b>	<b>3</b>
<b>2. Name search</b>	<b>3</b>
<b>3. Solutions to name matching</b>	<b>5</b>
<b>4. Entity extraction</b>	<b>5</b>
<b>5. Implementing name search</b>	<b>6</b>
<b>6. Expertise search</b>	<b>7</b>
<b>7. Strategy implications</b>	<b>9</b>

## Research Notes

*This is one in a series of Research Notes published by Intranet Focus Ltd. since January 2012. For further information see <http://www.intranetfocus.com/resources/downloads>. Previous Research Notes covered enterprise mobile strategy development, enterprise search team management, digital workplaces, virtual teams and legal issues for intranet managers, the future of enterprise search, and stakeholder management.*

# Searching for people and expertise

## 1. Introduction

One of the most important uses of a search application is to be able to find people working for the organization or to identify people with specific skills and expertise. The importance of employees being able to find other employees by name or by expertise is of crucial importance in taking full advantage of the investment that the organization has made in its workforce over a period of many years. Much is made of the benefits of collaborative working but this style of working has to start with creating the best possible team.

The US IT company Autodesk has created a visualization of how the [organizational structure](#) of the company changed between May 2007 and June 2011. For each of 1498 days the entire hierarchy of the company is constructed as a tree with each employee represented by a circle, and a line connecting each employee with his or her manager. Larger circles represent managers with more employees working under them. The tree is then laid out using a force-directed layout algorithm.

From day to day, there are three types of changes that are possible:

- Employees join the company
- Employees leave the company
- Employees change managers

The point that this visualization makes is that networks are constantly forming and (no matter how hard people try) breaking as an organization develops. Relying on personal contacts is not good enough to be sure that the best available people are working together on a project or task.

In the quest for 'relevance' most of the limited effort available to optimize a search experience is directed at document search, and little attention is paid to these two aspects of people search. It is important to appreciate that a document search may be carried out in order to find people with specific expertise. In such a situation the most relevant document may not be the one with the most useful information but the document that identifies one or more people to turn to for assistance.

In this Research Note some of the issues around searching for employees by name and by area of expertise are discussed.

## 2. 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 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!

Even if the nominal business language of an organization is English the issues of language quickly appear in name searching as 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. Indeed 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. However 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.

## Searching for people and expertise

### ***Typographical errors***

A slip of the finger at the keyboard causes transposition on 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, or 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”, “Мао Цзэдуи”, “毛泽东”, or “毛澤東”).

© Basis Technology 2012

Chinese, Japanese and Korean names present very substantial challenges. For example 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 Anglicized name identification.

## Searching for people and expertise

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.

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 smart phone or tablet.

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

Arabic	<a href="http://en.wikipedia.org/wiki/Arabic_names">http://en.wikipedia.org/wiki/Arabic_names</a>
Chinese	<a href="http://en.wikipedia.org/wiki/Chinese_names">http://en.wikipedia.org/wiki/Chinese_names</a>
German	<a href="http://en.wikipedia.org/wiki/German_names">http://en.wikipedia.org/wiki/German_names</a>
Japanese	<a href="http://en.wikipedia.org/wiki/Japanese_names">http://en.wikipedia.org/wiki/Japanese_names</a>
Portuguese	<a href="http://en.wikipedia.org/wiki/Portuguese_names">http://en.wikipedia.org/wiki/Portuguese_names</a>
Spanish	<a href="http://en.wikipedia.org/wiki/Spanish_names">http://en.wikipedia.org/wiki/Spanish_names</a>

### 3. Solutions to name matching

Four types of methods are most frequently used to score name similarity, with Basis Technologies recommending a hybrid approach as no single method will solve all the potential problems.

#### *Common key*

These methods, such as Soundex, reduce names to a key or code based on their English pronunciation, such that similar sounding names share the same key. Common key methods are fast and produce high recall (finds most of the correct answers) but have generally low precision (i.e., contain many false hits). Precision is yet lower when matching non-Latin script names, which first must be transliterated to Latin characters to use this method.

#### *List-method*

This method attempts to list all possible spelling variations of each name component and then uses the name variation lists to look for matches against the target name. The result can be slow performance if very large lists must be searched. Furthermore, this method will not match name variations not appearing in its lists.

#### *Edit distance*

This approach looks at edit distance, that is, how many character changes it takes to get from one name to another. For example, "Catherine" and "Katherine" have an edit distance of 1 since the "C" is substituted for "K." Edit distance methods work for Latin-to-Latin name comparisons, but precision suffers as each edit is weighted similarly, so a replacement of "c" for "k" is considered equal to a replacement of "z" for "t."

#### *Statistical similarity*

A statistical approach trains a model to recognize what two "similar names" look like so that the model can take two names and assign a probability that the two names match or not. This method produces high precision results, but may be slower than the common key method.

### 4. Entity extraction

Another approach to locating information about people and their expertise is to use named entity extraction. Entity extraction uses a set of rules to identify people, places and many other character sequences that could be of value in searching for information.

This extraction can be accomplished in three different ways, though many search applications will use all three in a blended approach

## Searching for people and expertise

Statistical models provide a means of recognizing never-seen-before names and providing good answers when words can have multiple meanings. Analyzing the correlation with the other words helps identify the correct context such as deciding when the word “Paris” is used as the name of a person or a city.

Telephone numbers and credit cards have standard formats, so as these are indexed a set of rules can be used to determine the category of the entity. This could be extended to other entities such as part numbers. All that is needed is to be able to write a rule that any character string starting with BD and a six digit number is a part number. Dictionaries and gazetteers will support the extraction of places and groups of places, so that a search for EU will also offer a search for all the Member States of the European Union.

A full treatment of named entity extraction is beyond the scope of this Research Note but it does provide a means of identifying people mentioned in the text of documents that could then be used to associate the content scope of the document with one or more employees

### 5. Implementing name search

There are a number of issues to consider when implementing name search. The first problem is to find a complete list of employees of the organization. HR is almost always organized along national lines, with Corporate HR responsible for policies and procedures. Where an organization has grown through acquisition there could be more than one database in a country. Local HR staff will be familiar with using multiple databases in operational circumstances but it complicates the process of name searching.

Another potential source is the Active Directory or similar listing but there may not be a master list and it is not uncommon for there to be differences in spelling between an AD list and the ‘formal’ name used for an HR database.

The next step is to create a test collection of names that show the traits listed above that can be used to ensure that the current application is able to find the names, and also to be used as a means of evaluating a new search application.

A decision then has to be made about whether to offer a ‘subject’ query box and a ‘people’ query box, or to integrate the query process into a single query box. Offering a separate ‘people’ query box will raise the expectations of users that there is a well-grounded search application behind the query box that they can rely on. If this is not the case (and they will be able to test it out very quickly) then either work has to be carried out to provide a dependable search application or the box is best removed.

If the decision is taken to provide two different query boxes then will searching for a person in the ‘subject’ query box find all the same employees as the ‘people’ search box plus additional results found from text entity extraction.

The subject of good practice in employee search is covered in detail in [Volume 6](#) of the Nielsen Norman Group report on Intranet Usability Guidelines.

94. Provide employee search for the organization.
95. On every intranet page, present a search box for searching the employee directory.
96. Do not open the employee search in a new window.
97. Make it clear which type of information can be entered in an open search field or form, such as last name, first name, job title, or location.
98. Be forgiving about formatting and be clear about any format requirements. Ideally, suggest people’s names as users are constructing their queries.
99. Place notes and examples to describe field inputs adjacent to (not as prompts within) employee search fields.
100. Provide a way to search for employees by partial information, including first name, last name, initial, nickname, title, department, or other information.
101. Consider enabling reverse-lookup, so people can search by phone number to find the person it belongs to.

## Searching for people and expertise

102. Consider enabling phonetic name spelling search (especially at global organizations).
103. Provide a clear link to advanced employee search options
104. Provide search results on a new page, not just in a section of the page users searched on.
105. In employee search results, make it easy for users to identify the person they are looking for. Provide the name, location, title, email address, and telephone number so users do not have to open a profile document to find this common information.
106. Make the employee's name an underlined link to the employee's profile document.
107. Write out the full email address in the search results.
108. Clicking the email address link from the search results page should automatically open and address a new email message to the person.
109. Make the employee search results pages as simple as possible by eliminating redundant information.
110. Provide filters and sorting—such as by department or job title—to help users narrow down employee search results.
111. If possible, include employee search results with the general intranet search results even if you also have a directory-only results section.
112. Clearly distinguish employee directory results from general intranet results if presented on the same page.
113. If possible, in addition to the employee search function, provide an alphabetical list of all employees at the organization.

Each of these Guidelines is discussed in detail in the report on pp 137-176 and illustrated with screen shots.

### 6. Expertise search

A search for a document is often the result of needing to find the name of someone in the organisation with specific expertise. In theory it would be good if every employee created a detailed cv but a quick look at Dave Snowden's [Principles of Knowledge Management](#) will show that this is not a realistic objective.

1. Knowledge can only be volunteered it cannot be conscripted. You can't make someone share their knowledge, because you can never measure if they have. You can measure information transfer or process compliance, but you can't determine if a senior partner has truly passed on all their experience or knowledge of a case.
2. We only know what we know when we need to know it. Human knowledge is deeply contextual and requires stimulus for recall. Unlike computers we do not have a list-all function. Small verbal or nonverbal clues can provide those ah-ha moments when a memory or series of memories are suddenly recalled, in context to enable us to act. When we sleep on things we are engaged in a complex organic form of knowledge recall and creation; in contrast a computer would need to be rebooted.
3. In the context of real need few people will withhold their knowledge. A genuine request for help is not often refused unless there is literally no time or a previous history of distrust. On the other hand ask people to codify all that they know in advance of a contextual enquiry and it will be refused (in practice its impossible anyway). Linking and connecting people is more important than storing their artifacts.
4. Everything is fragmented. We evolved to handle unstructured fragmented fine granularity information objects, not highly structured documents. People will spend hours on the internet, or in casual conversation without any incentive or pressure. However creating and using structured documents requires considerably more effort and time. Our brains evolved to handle fragmented patterns not information.

## Searching for people and expertise

5. Tolerated failure imprints learning better than success. When my young son burnt his finger on a match he learnt more about the dangers of fire than any amount of parental instruction could provide. All human cultures have developed forms that allow stories of failure to spread without attribution of blame. Avoidance of failure has greater evolutionary advantage than imitation of success. It follows that attempting to impose best practice systems is flying in the face of over a hundred thousand years of evolution that says it is a bad thing.
6. The way we know things is not the way we report we know things. There is an increasing body of research data which indicates that in the practice of knowledge people use heuristics, past pattern matching and extrapolation to make decisions, coupled with complex blending of ideas and experiences that takes place in nanoseconds. Asked to describe how they made a decision after the event they will tend to provide a more structured process oriented approach which does not match reality. This has major consequences for knowledge management practice.
7. We always know more than we can say, and we will always say more than we can write down. This is probably the most important. The process of taking things from our heads, to our mouths (speaking it) to our hands (writing it down) involves loss of content and context. It is always less than it could have been as it is increasingly codified.

The answer lies in being able to integrate search across multiple applications, each of which may contain some element of information about the expertise of employees. These might include project files, product design databases, blogs, conference papers and patents.

The problems of expertise identification have been the subject of a considerable amount of research over several decades which is well summarized in a recent volume on [Expertise Search](#) in the Foundations and Trends in Information Retrieval series from Now Publishing. There are over 280 references to published research papers.

At the Enterprise Search Summit this year Kas Kasravi (Hewlett Packard) gave a paper on the Enterprise Collective application that he and his colleagues have developed. Details are available in an [HP Technical Report](#) that is well worth downloading. Enterprise Collective is a web application that automatically discovers experts and their expertise via linguistic and semantic analysis of their work products (e.g., e-mails, patents, papers, reports, presentations, and blogs). The system combines multiple data sources, including unstructured work products and structured sources such as Enterprise Directory and document metadata, into one holistic model of organizational knowledge and construct a graph model of organizational knowledge.

The application is built as a combination of three main components:

1. A back-end responsible for connecting to the different content types, people finder, and taxonomy creation. These components create a large graph of entities that represent their relationships with each other.
2. An analytics engine that answers relevance queries on the entity-relation graph.
3. A front-end that allows a user to both view and navigate a knowledge-base represented by the graph, but also to search the graph for relevant and personalized information.

The approach is very comprehensive but there could be a concern about how long it will take until a new member of staff gains enough “HP” expertise to be highlighted in a search. The problems of locating new employee expertise have not been given much attention. One interesting feature of the HP application is that the experts highlighted for a search are rotated through a cohort of experts so that the same two or three experts do not get overloaded with enquiries.

IBM has been working along similar lines with its [Faces project](#), led by Ido Guy and his team at the IBM Research Laboratories in Haifa. One of the design requirements of the Faces project was to be able to return core information in no more than 100ms. University College London have been working on an expertise location application called [KnowDis](#) and Sinequa presented a very interesting paper at Enterprise Search Europe 2012 on the work they were doing on the French consulting company [ATOS](#).

## Searching for people and expertise

### 7. Strategy implications

A strategy for providing and monitoring people and expertise search should be a key element of an overall search strategy. Almost certainly it will involve working closely with HR departments who are collecting employee information who may have concerns about access to departmental databases. There are also some important data privacy implications on the information that is held on each employee, including the inclusion of photographs which could be contrary to data protection legislation in many countries.

Judging from presentations at search conferences the emphasis seems to be on providing employees with all the documents/information they need on the first page of search results. This is a very challenging task but almost certainly less of a challenge than responding to an employee who has looked for information on someone that they know well and been very disappointed with the result, or has not been able to find one or more employees who have expertise that could have made a significant difference to the achievement of a business objective.

We would make the following recommendations

1. Specifically address the requirements of employee name search and expertise search in the search strategy.
2. Set up an advisory group of stakeholders with interests in employee name and expertise search
3. Ensure that there is clarity about which employee directories are being searched and how the search application copes with the range of issues that are inherent in name searching.
4. Pay particular attention to finding details of employees who have only recently joined the organization.
5. Develop a test set of names and core areas of organizational expertise that can be used to evaluate the performance of the search application. This test set should include some recent new employees.
6. Ensure that employee and expertise search applications are compliant with data privacy legislation and corporate policies.
7. Read through the Employee Directory section of the Intranet Usability Guidelines report from the Nielsen Norman Group and benchmark your current employee search implementation.
8. Consider how employee search performance will be assessed from the search logs.