The Value of Taxonomies, Thesauri and Metadata in Enterprise Search

Martin White

Intranet Focus Ltd. , 12 Allcard Close Horsham West Sussex, United Kingdom of Great Britain and Northern Ireland, RH12 5AJ, <martin.white@intranetfocus.com>

Martin White has been in the information business for over forty years. He is Managing Director of Intranet Focus Ltd. He consults in intranet and information management strategy and implementation, and also in enterprise search implementation. Martin has been a visiting professor at the Information School, University of Sheffield since 2002 and is the author of a number of books, including the *Intranet Management Handbook* (Facet Publishing 2011) and *Enterprise Search* (O'Reilly Media 2015). Martin is an Honorary Fellow of Chartered Institute of Library and Information Professionals, a Fellow of the Royal Society of Chemistry and a member of the British Computer Society.

White, Martin. 2016. "The Value of Taxonomies, Thesauri and Metadata in Enterprise Search." *Knowledge Organization* 43 no. 3: 184-192. 28 references.

Abstract: Although the technical, mathematical and linguistic principles of search date back to the early 1960s and enterprise search applications have been commercially available since the 1980s; it is only since the launch of Microsoft SharePoint 2010 and the integration of the Apache Lucene and Solr projects in 2010 that there has been a wider adoption of enterprise search applications. Surveys carried out over the last five years indicate that although enterprises accept that search applications are essential in locating information, there has not been any significant investment in search teams to support these applications. Where taxonomies, thesauri and metadata have been used to improve the search user interface and enhance the search experience, the indications are that levels of search satisfaction are significantly higher. The challenges faced by search managers in developing and maintaining these tools include a lack of published research on the use of these tools and difficulty in recruiting search team members with the requisite skills and experience. There would seem to be an important and immediate opportunity to bring together the research, knowledge organization and enterprise search communities to explore how good practice in the use of taxonomies, thesauri and metadata in enterprise search can be established, enhanced and promoted.

Received: 18 January 2016; Revised 11 February 2016; Accepted 16 February 2016

Keywords: enterprise search, information, applications, metadata, taxonomies, thesauri

1.0 Introduction

This paper examines the current state and future potential for taxonomies, thesauri and metadata in enterprise search based on the outcomes of recent surveys of enterprise search implementation and management, on presentations at the only two conferences specifically for enterprise search (Enterprise Search Summit, New York and Enterprise Search Europe, London) over the last five years and the authors' own consulting practice in this area dating back to 1998. Many of the statements made in this paper are, therefore, on the basis of oral communications and cannot be referenced for reasons of commercial confidentiality.

The objectives of this paper are to illustrate the challenges that enterprise search managers face in improving search performance and the extent to which they are in a position to take advantage of the potential benefits of taxonomies, thesauri and metadata.

The technical, mathematical and linguistic principles of search date back to the early 1960s, and by the early 1970s, the use of computer bureau services (such as Lockheed Dialog and ESA-IRS) for real-time search of primary and secondary scientific and business databases was widespread among information professionals. They often worked alongside research scientists to create high-quality queries and assess the search results. (Bourne and Hahn 2003). Many of the databases were enriched with metadata tags, sometimes derived from thesauri, controlled term lists and classifications. The use by Predicasts of the SIC industry codes is an example (Hass 1977).

It was not until the late 1980s that the concept of searching internal corporate databases using "enterprise search" software applications such as Inktomi, Ultraseek and Verity started to be more widely adopted, initially mainly in the United States. With the notable exceptions of Recommind and Autonomy (White 2015), enterprise

search vendors have continued to develop applications that do not differ significantly in terms of technical concepts from the bureau services of the 1970s. Despite this long period of evolution, global surveys undertaken since 2010 by Findwise (2015), the Association for Information and Image Management (Association for Information and Image Management) (2014) and NetStrategy/JMC (MacConell 2015) indicate a high degree of user dissatisfaction with enterprise search applications. This seems to be a result of an assumption by information technology (IT) departments holding the budget for enterprise search (Association for Information and Image Management 2014) that technology alone is the solution to search.

In the context of this paper, it is important to highlight that there is comparatively little communication or joint research between the information retrieval community and the enterprise search community. Although there is a substantial volume of research into the value of taxonomies and thesauri in information retrieval, the test collections are almost always relatively small, comparatively homogenous and usually in the public domain. By comparison, enterprise collections are often very large (many millions of documents), heterogeneous (one major pharmaceutical company has over 400 intranets and eleven languages stretching over a period of 10 years) and most certainly not in the public domain.

The challenges of undertaking research studies of enterprise applications were discussed at the Second Strategic Workshop on Information Retrieval 2012 (Allen et al. 2012), but in a discussion at the Enterprise Search Europe 2015 conference in London (chaired by the author), it became clear that not only had no progress been made but that there seemed to be no mechanisms in place to consider how the two communities could work more closely together.

There are a number of important reasons why very little information retrieval (IR) research has been carried out inside the applied search environment of organisations:

- The confidentiality of the information on these internal repositories;
- The scale of the search solutions, indexing perhaps 500,000 plus documents;
- The difficulty of creating test collections to use in a controlled assessment of search performance;
- The need to be fully conversant with the business and technical languages of the organisation; and,
- The difficulty of persuading an organisation to allow a paper to be published, which perhaps casts concerns about the quality of the internal search applications.

There is an innovative large-scale, cross-organisational IR-enterprise search project currently being conducted by Cleverley, Burdett and Muir (2015). This paper provides a valuable insight into the ways in which exploratory search is used within the oil and gas production sector but is not specifically addressing issues around taxonomy and metadata use. Other enterprise studies, which may not be representative of other enterprises, are usually undertaken within a specific company, often in the IT sector (Guy et al. 2012).

It does not help that the research literature is invisible to most enterprise search managers, because it is behind a subscription firewall. This should not be taken as a criticism of the subscription model *per se* but as an observation that tracking research in information retrieval to assess its potential value in improving enterprise search performance requires skills and time that are very rarely available.

2.0 Enterprise search—definitions

Although the phrase "enterprise search" is widely used, there seem to be two different interpretations. Many vendors of search applications position their products as though they were the sole search application for an enterprise. The website of Coveo is just one example of what might be regarded as a sales-driven approach (http://www.coveo.com/en/solutions/enterprise-search):

> Enterprise search technology connects your employees with the information and expertise they need to be proficient, productive and effective. Coveo securely indexes each and every content source and application across your enterprise—both on-premise and in the cloud. It builds and constantly updates a unified index of this diverse content, and analyzes every document and record to understand precisely what it is about, and how it relates with the rest.

As this paper will consider, the reality is that the investment in staff to support a search application will probably be larger than the technology investment. Moreover, without a search support team, the promise of the technology cannot be delivered.

Hawking (2011) interprets enterprise search as search of digital textual materials owned by an organisation, including search of their external website, company intranet and any other electronic text that they hold such as email, database records and shared documents. The issue becomes whether this is carried out by a single enterprise search software application or whether it should be regarded as a strategic approach, which may involve more than one application. Larger organisations in particular may already have a range of enterprise applications that

incorporate search functionality, including those being used for document management intranets, records management, customer relationship management and enterprise resource planning.

There is a blurred boundary between enterprise search and "federated search," where a user is able to search multiple applications and repositories from a single search interface. This can be accomplished either by the search application sending out a query to each of the other applications and then integrating the results in some way or through indexing the content of all the applications and running the query against the master index, sometimes with the ability to restrict the search to one or more specific repositories. Search vendors have a tendency to use "enterprise search" and "federated search" almost as synonyms, which is confusing to potential purchasers. In this paper, the definition by Hawking is used. It should be added that in this paper the term "enterprise" is used as a generic description of any organisation and not as a term to describe a large multinational company, and that "document" is used as a generic description of a content item.

3.0 The renaissance of enterprise search

The technology for retrieving digital information dates back to the late 1950s and the establishment by Engelbart in 1959 of the Augmented Human Intellect Program at Stanford Research Institute (SRI). By 1963, SRI was able to demonstrate remote online searching of both bibliographic records and the full text of documents. Over the next two decades, three requirements for information access dominated the initial development of search technology. The first was to be able to search scientific literature, notably to support the U.S. space missions and the work of the Environmental Protection Agency. The second was to be able to find internal information in the context of anti-trust actions taken by the U.S. Government. One of the outcomes was the development of the proto-typical IBM STAIRS retrieval application. The third was the requirement to be able to search U.S. legal statutes and case law where judgements depended on being able to find appropriate precedents (Bourne and Hahn 2003).

The availability of enterprise search solutions that could be implemented without significant in-house development effort dates back to the late 1980s when Verity was spun out of Advanced Decision Systems (ADS) to exploit the Topic search application developed by ADS. Commercial adoption of this application was slow and it was not until 1999 that Verity achieved profitability. By this time, the Ultraseek enterprise search application had been developed by Go.com. This was acquired by Inktomi in 2000 to complement its website search application. In 2002, Inktomi sold Ultraseek to Verity. This gave Verity a departmental-level application to add to its enterprise-level application. In the same year, Google launched its Google Search Appliance (GSA) for enterprise search. In 2005, Verity was acquired by the UK search vendor Autonomy.

This situation has changed significantly since the launch of the Enterprise version of SharePoint 2010 which included a very powerful search application based on the FAST Search and Transfer software acquired by Microsoft in 2008. Many organisations discovered the benefits of enterprise search for the first time as they began implementing SharePoint 2010 to support information management and collaboration requirements. Another factor in raising the profile of enterprise search was the integration of the Apache Lucene and Apache Solr open source search software projects in 2010. Until 2010-2011, major IT vendors had paid little attention to search applications. Such was the urgency of the situation that these vendors decided that the business case for acquisition was stronger than that for internal development.

Within the period from 2011 to 2012, the Vivisimo search application was acquired by IBM, Endeca was acquired by Oracle, Isys Search was acquired by Lexmark and Autonomy was acquired by Hewlett Packard. All these acquisitions indicate that the major information technology (IT) vendors had gaps in their product range for search applications that had to be filled quickly to meet customer requirements (White and Nikolov 2013; White 2015).

There remain around 80 companies specialising in enterprise search software (White 2015). Most of these companies have revenues of less than $50,000,000.00 (White and Nikolov 2013) and do not have the local support needed to meet the requirements of multinational companies. They also face significant competition from the wide-spread use of Microsoft SharePoint (both the SP2010 and SP2013 releases have feature-rich search applications) and the emerging use of open source search applications in the enterprise.

4.0 Assessing enterprise search performance

The evidence from academic IR research is that search performance can be improved using taxonomies and thesauri to populate metadata schemas as well as adding non-topical metadata. Although search managers intuitively feel that these techniques could be of benefit, apart from the Findwise and Association for Information and Image Management (AIIM) surveys there is no published research to support the wider adoption of taxonomies and thesauri. In addition, without being able to benchmark search performance on a rigorous and longitudinal basis it is not possible for an organisation to assess with certainty whether the effort involved in making wider use

of taxonomies and thesauri is in fact justified by improvements in performance. A major challenge in enterprise search management is the scope and scale of search, making performance management very difficult to accomplish. It is of note that one of the standard texts on IR performance evaluation (Buttcher, Clarke and Cormack 2010) runs to 600 pages but provides no guidance on assessing enterprise search performance.

The primary metric of performance of most enterprise applications is that designated work-flow processes can be carried out as quickly as required to complete a business transaction. Time to completion of a process is a very important success metric. User-created content is usually limited to adding notes about a transaction. Search is limited to a "find" command which will search specified fields for a defined parameter, such as Account Number or Address.

However, there has also been a significant amount of research into aspects of user satisfaction with these systems, a metric that is often used for enterprise search applications. This research dates back to seminal work by James Lewis (1995) working at IBM, in which he set out a computer usability satisfaction questionnaire. The three factors that were assessed as influencing user satisfaction were system usefulness, information quality and interface quality.

The level of trust in a search application is arguably a very important metric, as the information found may have to be included in a decision with significant impact on the enterprise and perhaps even on the personal career of the decision maker. However, a survey that asks whether users trust a search application may raise more questions than it answers. Although "trust" and "satisfaction" are not synonyms, gaining a sense of the level of satisfaction with a search application will give a good indication of overall trust in the application and its content.

Although the IBM work referred to above took user satisfaction into account, the number of users of most enterprise solutions (for example, a treasury management application or a human resources (HR) application is quite small. An enterprise search application is almost certainly being used directly or indirectly by every employee in an organisation. Indirect use takes place where a manager may use the search application for information that is then passed on down the line of management. However, the information provided by a search application will usually only be one element of the information required by a user. Evaluations by the academic IR community will usually focus on immediate assessments of relevance by the study participants, or explore task completion. It might also not be for perhaps several days, if not weeks, that the user is able to judge whether the information gained from the intranet is relevant and of value.

5.0 Some challenges for enterprise search implementation

A major challenge for enterprise search implementations is that enterprises speak multiple business languages. To give an example from a major hospital, clinicians will use the terms "oncology," "renal," "paediatric" and "phlebotomy," whereas support staff without a medical background will use "cancer," "kidney," "children" and "blood tests," partially driven by a need to relate to patients and their carers who will use what might be regarded as colloquial usage. Many clinicians, for example psychologists, may also refer to their patients as clients. These language issues present a challenge to a hospital intranet which has to enable staff to whichever terms they are most familiar with and lead them to the information they need to make what may well be life-changing decisions.

It is not just technical usage that might vary across an organisation. In the same organisation, the department titles could include Personnel Department, HR Department, Human Resources, Employee Services and Personalabteilung. Staff with similar titles may also have different job responsibilities. In the London office of a major law firm, there might well be a Chief Billing Officer, but in a small country office in Germany, client billing might be the responsibility of the General Manager.

One of the many differences between enterprise search and searching the World Wide Web is that there is no "Plan B" with enterprise search. Via the Internet, the requirement to find the arrival time of a flight at Heathrow airport could be met by a search of the Heathrow website, the airline website, or one of the many third-party flight information services. Web search applications are also able to use past search histories to help in search optimisation and in recommending other potential sources of information. When using enterprise search, the user is limited to (in effect) a single website and does not know whether a failure to find the required information is because:

- The information is not held by the enterprise;
- The information is held by the enterprise but has not been indexed;
- The information has been indexed but the concepts associated with the information are not conveyed by the index terms;
- The information has been indexed and tagged with additional metadata but the access permissions are such that the employee is not able to see it;
- The information has been indexed and tagged with additional metadata but the tagging is no longer fit for purpose; and,

– The information, when found, is not of a quality that enables it to be used for the purpose intended.

In 2008, there was considerable public concern about the loss in transit by the UK tax authority of sensitive private information on 10 million people. A report was commissioned into the reasons for this loss. This highlighted that the appropriate secure handling procedures were difficult to find, and therefore gave rise to an inappropriate method of transit being used. The procedures were set out in the Departmental Security Standards Manual (DSSM) (Poynter 2008). The report noted:

> The primary dissemination method for information security policy in HMRC is via its intranet. However, almost all interviewees contacted in my team's investigations expressed a lack of knowledge as to exactly where on the intranet, security policy is to be found. In addition, staff have noted that the intranet search function is unhelpful in generating relevant results for search terms such as "DSSM."

Another challenge for enterprise search implementation is that users have a requirement for both high precision and high recall depending on the nature of the query. The requirement for high precision is usually to find a known document, person or application. This often arises because of problems with the information architecture of an intranet and so the fall-back is search. High recall is typically required when the organisation has to make a business-critical decision and needs to reduce as far as possible the risks of making an incorrect decision.

The overall status of the latest enterprise search implementation surveys is that users are now starting to expect a much higher level of satisfaction with enterprise search as the volumes of internal enterprise information continue to grow at very rapid rates.

There is also a better understanding by organisations of the importance of search as an enterprise-wide application. The results of the 2015 survey (Findwise 2015) show the steady rise in the number of organisations that have a strategy for search and findability from twenty percent in 2012 to nearly fifty percent in 2015. User requirements for improved search performance cannot be achieved by technology alone. Attention has to be paid to content quality, the adoption of taxonomies and thesauri as a key component in providing consistent and appropriate metadata tags for content. What seems to be evident from the AIIM and Findwise surveys is that where an organisation invests in addng taxonomies and structured metadata into the search implementation then the quality of the search performance, measured in terms of user satisfaction, is markedly higher.

6.0 Enhancing the search experience

Search requires a dialogue to be established between the search application and the user. A significant amount of research has been carried out on information-seeking models and information-seeking behaviour for several decades, building on the work of Wilson (Ford 2015). The challenge for search user-interface designers is to provide an interface that supports a range of information seeking models. It could be argued that an enterprise search application is not a single process, but that it has to be implemented in a way that each user feels confident in using it for a range of different purposes.

Employees at all levels now have multiple roles. A manager may have line responsibility for a group of employees, be a member of a number of different project teams and communities of practice and have to undertake continuous professional development. Search is important to all these different roles, yet one may need to offer high precision and another excellent exploratory search. In this respect, it is interesting to compare the search interfaces and facets of Google web search and Google Scholar. In the web search, the date facet is relative (e.g., "last year"), but in Scholar, it is absolute (e.g., "2014, 2015"). The result display is also different, with Scholar showing alternate locations for research content.

Enterprise search managers have to develop a best-fit user interface that can meet all of these requirements. The potential range of user interface options is illustrated by Morville and Callender (2010), and the practical development of these options is discussed by Russell-Rose and Tate (2013). Somewhat surprisingly, enterprise search vendors do not yet offer a range of search interfaces along the lines of the ezDL (http://www.ezdl.de/) library research application.

Most of these options make use of taxonomies, thesauri and metadata and a good introduction to the use of these elements is provided by Rosenfeld, Morville and Arango (2015) who consider in some detail search implementation on websites and intranets. The primary purpose of using these elements is to support the user in building an initial collection of highly relevant results and then further refining these based on a range of criteria. Wetzker et al. (2008) discuss the benefit of tailoring taxonomies for efficient text categorisation and expert finding.

One of the primary means of conducting a search dialogue is through the use of faceted search. This has become a default feature of enterprise search applications. The history and development of faceted search have been described by Tunkelang (2009), who notes that the initial work was undertaken by Shneiderman in the U.S. and by Pollitt in the U.K. This was then the basis for the work undertaken by Hearst in the Flamenco project with its open-

source faceted search system using hierarchical facets. Sacco and Tzitzikas (2009) provide a more detailed assessment of the integration of taxonomies and faceted navigation.

However, faceted search requires documents with faceted metadata. Although there are a range of text mining techniques for enriching unstructured text, these are complex and expensive applications, and so organisations tend to rely either on manually applied metadata or the latent metadata that can be identified from the document record. Using latent metadata can have some significant limitations. One of these is that metadata on (for example) file format can be of very limited value in managing a large result set as the file format (for example PDF) is not semantically linked to the topics of the content.

A very important metadata element in enterprise search is the date of the content. In theory, this should be easy to establish, but in reality, a document could have a number of dates associated with it, including:

– Date of origination;
– Date of authority to publish;
– Date of all or just the latest revision; and,
– Date range of the document (e.g. Q3 sales performance).

To complicate the situation, there are a number of different date formats, notably Day/Month/Year in the U.K. and Month/Day/Year in the U.S. On initial inspection of the document, it might well not be apparent which is the "correct" date, given that "correct" will be a personal construct by the search user. There is also the ISO format of YYYY-MM-DD, but it is perhaps indicative of the challenge of implementing metadata standards in multinational organisations that it seems to be very rarely adopted despite the apparent benefits.

7.0 Exploratory search

The concept of exploratory search was proposed by Marchioni (White and Roth 2009) and describes situations where users may not have a clear requirement that can be used to frame a query, because they are unfamiliar with the technical language of the domain they are exploring and the scope and scale of information that is available. Athukorala et al. (2015) compare information search behaviour for exploratory and lookup tasks.

The support of exploratory search is where taxonomies and thesauri should have very important roles to play. A very comprehensive account of these roles is given by Shiri (2012). In the context of enterprise search across multiple repositories, thesauri should in principle provide a means of mapping the topics and concepts across all of the repositories being searched. However, this is also the role of master data management schemas for the wide range of structured databases in the enterprise. MDM applications enable connections to be made in a relational database. As an example, the MDM will stipulate the format of a customer identification number so that this number is then consistently used on contracts, invoices and payments. It will also stipulate a date format so that, in an international company, there can be no confusion between U.S. and U.K. date formats. In a project conducted by the author for a major law firm the master data schema for the client, matter, billing and finance databases ran to 730 term definitions. Only 33 of these would be of value in searching some 63 million documents in two different document management systems and a global intranet.

Enterprises also often operate in many different countries and may well maintain repositories in multiple languages. At least one major pharmaceutical company has an intranet with content in nine languages, which then present very significant tagging and search issues, especially in the case of Japanese and Chinese, both languages of significance in the pharmaceutical sector.

Both taxonomies and thesauri are of value in providing auto-suggestions to search users along the lines of "do you mean" (Nagy, Pellegrini and Mader 2011). The ability to move through a taxonomy or to use a thesaurus to present related and broad terms is important, but users have to be able to trust these options to lead them in a useful direction. If the search aids fail to meet the expectations of a user, then in a business-critical search, the user is likely to adopt the alternative approach of asking a colleague for the information. This is more time-consuming for both parties concerned and no guarantee that the person asked has the most current and most credible information readily at hand.

8.0 Building, buying and maintaining

The survey undertaken by AIIM in 2014 indicated that among the areas where organisations indicated where they needed the most resource in supporting search were taxonomy management (31%), and correcting, updating or standardising metadata (26%). A major factor in metadata use in an enterprise is maintaining a balance between the value of the metadata being added to the workload on the publisher to do so. Adding metadata can be time consuming and requires an appreciation of how others will seek to find the document, a situation first highlighted by Furnas, Landauer, Gomez and Dumais (2007). The issue around the availability of the skills to develop and implement metadata tagging is considered below.

The challenge for these organisations is where to find people with the specialist skills. All the evidence from the AIIM and Findwise surveys indicates that the majority of respondents had one person specifically responsible for supporting the search application despite the awareness across the business of the importance of being able to find information inside the organisation.

These activities require specialist skills that few organisations have available. There are few external consultancies (compared to the scale of the potential requirement) that are able to support taxonomy development. It is certainly possible to purchase taxonomies and there are open source taxonomies and thesauri. In many areas, these tools have been overtaken by progress, especially in the areas of information technology and medicine. Hunink et al. (2010) have highlighted the problems of creating industrial taxonomies in a specific domain. In addition, enterprises have developed ad hoc controlled term lists, master data schema and departmental or process taxonomies and thesauri that may well already have been used to tag content for a number of years. Integrating externally-developed taxonomies and custom-built taxonomies and thesauri is a very considerable challenge for most enterprises, who then face the requirement to assess the performance of these tools through a detailed assessment of search logs and search user satisfaction surveys and to use this information to make revisions to the search application and to the tagging that has been applied.

A report prepared for the European Commission in 2013 (White and Nikolov 2013) identified that a lack of information science and IR skill is becoming a substantial barrier to development of search software applications by commercial vendors, to meeting customer demand for search system integration skills and for building search support teams inside organisations. This research project indicated that there are no undergraduate courses in information retrieval in the European Union and the post-graduate courses are targeted at those wishing to pursue a career in academic research rather than in enterprise search management. There are over sixty information schools world-wide but in their undergraduate courses, the development and management of taxonomies, thesauri and metadata schemes are often taught as one of many specialised options. This is not to be taken as a criticism by the authors of the course design of these schools but only to highlight that the future demand for these skills is likely to be substantially greater than the supply.

9.0 Future imperfect?

From the survey information that is now available and from presentations at the Enterprise Search Summit and Enterprise Search Europe conferences, it is clear that the benefits of using taxonomies, thesauri and metadata to improve search performance are recognised by search managers but barriers to wider adoption are significant, including:

- There are no recent published case studies of the way in which enterprises are using these techniques to improve search performance that would enable good practice to be identified, adopted and further developed;
- Business managers have little experience of specifying and building taxonomies and thesauri;
- Only a minority of enterprises have a dedicated search support team which includes members with experience of the implementation and enhancement of taxonomies, thesauri and metadata schemas;
- Because search teams are so small, the amount of analysis required to assess the benefits of these techniques cannot be carried out on a regular basis;
- It is difficult to find people that combine subject knowledge together with information science and information retrieval skills to join these search teams; and,
- Making a business case to increase the size of search teams, or to invest in external consultants, is very challenging when the results of doing so are not immediately visible.

There would seem to be an important and immediate opportunity for the International Society for Knowledge Organization to bring together the research, knowledge organization and enterprise search communities to explore how good practice in the use of taxonomies, thesauri and metadata in enterprise search can be established, enhanced and promoted.

However, there is a wider issue. There is the lack of awareness of the importance of information management and how information should be treated as an asset of the organisation. An organisation will know very precisely how many hand driers it has, how many employees, the size and value of its offices and factories and the amount of money owed by customers. These and other asset metrics are required for regulatory compliance. These organisations will have no measure of their information assets (Lemming 2015) nor will they have an information management strategy.

Gartner predicts that by 2017, thirty-three percent of Fortune 100 organizations will experience an information crisis, due to their inability to effectively value, govern and trust their enterprise information. In a press release, Andrew White, research vice president at Gartner, stated (Hamilton 2014): "There is an overall lack of maturity

when it comes to governing information as an enterprise asset. It is likely that a number of organizations, unable to organize themselves effectively for 2020, unwilling to focus on capabilities rather than tools, and not ready to revise their information strategy, will suffer the consequences' and goes on to say "Information is becoming the competitive asset to drive business advantage, and it is the critical connection that links the value chain of organizations."

A bottom-up approach of encouraging and educating organisations into the value of taxonomies, thesauri and metadata to support effective enterprise search will only have a significant impact when it is matched by a top-down approach to instil good information management principles and practice.

References

Allan, James, Bruce Croft, Alistair Moffat and Mark Sanderson, eds. 2012. "Frontiers, Challenges and Opportunities for Information Retrieval: Report from SWIRL 2012." *SIGIR Forum* 46:2-32.

Association for Information and Image Management. 2014. "Search and Discovery." http://www.aiim.org/Research-and-Publications/Research/Industry-Watch/Search-and-Discovery-2014

Athukorala, Kumaripaba, Dorata Głowacka, Giulio Jacucci, Antti Oulasvirta and Jilles Vreeken. 2015. "Is Exploratory Search Different? A Comparison of Information Search Behavior for Exploratory and Lookup Tasks." *Journal of the Association for Information Science and Technology*. Wiley Online Library. doi:10.1002/asi.23617

Bourne, Charles P. and Trudi Bellardo Hahn. 2003. *A History of Online Information Services 1963-1976*. Cambridge, MA: MIT Press.

Büttcher, Stefan, Charles L.A. Clarke and Gordon V. Cormack. 2010. *Information Retrieval: Implementing and Evaluating Search Engines*. Cambridge, MA: MIT Press.

Cleverley, Paul H., Simon Burdett and Laura Muir. 2015. "Exploratory Information Searching in the Enterprise: A Study of User Satisfaction and Task Performance." *Journal of the Association for Information Science and Technology* early view doi:10.1002/asi.23595

Findwise 2015 Findability Survey 2015 http://www2.findwise.com/findabilitysurvey2015

Ford, Nigel. 2015. *Information Behaviours*. London: Facet Publishing.

Furnas, George W., Landauer, Thomas K., Gomez, Louis M. and Dumais, Susan T. 2007. "The vocabulary problem in human-system communication." *Communications of the ACM* 30:964-971

Guy, Ido, Tal Steier, Maya Barnea, Inbal Ronen and Tal Daniel. 2012. "Swimming Against the Streamz: Search and Analytics over the Enterprise Activity Stream." In *CIKM'12. Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. New York, NY: ACM, pp. 1587-91. doi:10.1145/2396761.2398478

Hamilton, Alex. 2014. "Gartner Predicts Third of Major Firms will Suffer Information Crisis by 2017." *Techradar-pro* http://www.techradar.com/news/world-of-tech/gartner-predicts-third-of-major-firms-will-suffer-information-crisis-by-2017-1229785

Hawking, David. 2011. "Enterprise Search." In *Modern Information Retrieval: The Concepts and Technologies Behind Search*, ed. Ricardo Baeza-Yates and Berthier Ribeiro-Neto. New York: Addison Wesley, pp. 1-37.

Hass, Robert W. 1977. "SIC System and Related Data for More Effective Market Research" *Industrial Marketing Management* 6:429-35.

Hunink, Ivo, Rene van Erk, Slinger Jansen and Sjaak Brinkkemper. 2010. "Industry Taxonomy Engineering: the Case of the European Software Ecosystem." In *ECSA '10 Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*. New York, NY, ACM, pp. 111-18. doi:10.1145/1842752.1842779

Lemming, Reynold. 2015. "Why is Information the Elephant Asset? An Answer to this Question and a Strategy for Information Asset Management." *Business Information Review* 32:212-19.

Lewis, Ronald. 1995. "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use." *International Journal of Human-Computer Interaction* 7: 57-78.

MacConell, Jane. 2015. "The Organization in the Digital Age: Trends, Challenges, Practices." NetStrategyJMC. http://www.digital-workplace-trends.com

Morville, Peter and Jeffrey Callender. 2010 *Search Patterns*. Sebastopol, CA: O'Reilly.

Nagy, Helmut, Tassillo Pellegrini and Christian Mader. 2011. "Exploring Structural Differences in Thesauri for SKOS-based Applications." In *I-SEMANTICS 2011, Proceedings of the 7th International Conference on Semantic Systems*. New York, NY: ACM, pp. 187-90. doi: 10.1145/2063518.2063546

Poynter, Kieran. 2008. *Review of Information Security at HM Revenue and Customs: Final report*. London: Her Majesty's Stationery Office.

Rosenfeld, Louis, Peter Morville and Jorge Arango. 2015. *Information Architecture for the Web and Beyond*. Sebastopol, CA: O'Reilly Media.

Russell-Rose, Tony and Tyler Tate. 2013. *Designing the Search Experience*. London: Morgan Kaufmann.

Sacco, Giovanni Maria and Yannis Tzitzikas, eds. 2009. *Dynamic Taxonomies and Faceted Search: Theory, Practice, and Experience*. Berlin: Springer.

Shiri, Ali. 2012. *Powering Search. The Role of Thesauri in New Information Environments*. Medford N.J.: Information Today.

Tunkelang, Daniel. 2009. *Faceted Search*. San Rafael: Morgan & Claypool.

White, Martin and Stavri G. Nikolov. 2013. *Enterprise Search in the European Union: A Techno-Economic Analysis. Institute for Prospective Technological Studies*. Luxembourg: Publications Office of the European Union. EUR Number 26000EN. doi:10.2791/17809

White, Martin. 2015. *Enterprise Search,* 2nd ed. Sebastopol, CA: O'Reilly.

White, Ryen.W. and Resa R. Roth. 2009. *Exploratory Search. Towards New Search Horizons*. San Rafael: Morgan & Claypool.