A study of labour market information needs through employers' seeking behaviour

Sonia Sanchez-Cuadrado; Jorge Morato, and Yorgos Andreadakis
Carlos III University of Madrid. Department of Computer Science, 28911 Leganes, Spain
Jose Antonio Moreiro
Carlos III University of Madrid. Department of Library and Information Science, 28903 Getafe, Spain

Abstract

Introduction. The objective of this study is understand the information needs that businesses have while seeking Library and Information Science professionals and analyse how they formulate those needs.

Method. The analysis is performed by examining the professional skills and capabilities demanded in job offers published. A total of 1,020 job offers collected from a Spanish employment agency Website have been analysed for the period between 2006 and 2008.

Analysis. Knowledge representation techniques using thesauri have been used for the automatic content analysis based on natural language processing. Data extracted from the corpora have been analysed statistically.

Results. Results of this study indicate a demand for skills related to technological advances and the management of electronic resources as well as to technical aspects associated with the Informatics domain. The knowledge of languages and the possession of an academic title represent essential factors in the job offers.

Conclusions. This method permits the identification of changes in the information needs and the contexts inherent to the profession. The advantage of using thesauri permits other research groups to reproduce the results. The re-use of semantic categories, common to other fields, facilitates the reproduction of this method with other occupational groups or social roles.

Introduction

The principal task of information professionals is to satisfy the various information needs of users in different contexts, studying their requirements, their seeking habits and their management and use of information. According to information behaviour researchers, information seeking behaviour analysis is the study of the ‘activities a person may engage in when identifying their own needs for information, searching for such information in any way and using or transferring that information’ (Wilson 1999: 249).

To date, information seeking behaviour research has the creation of models using quantitative or qualitative methods. Data are collected by interviews (structured or unstructured), questionnaires, observations, etc. User groups are usually created starting from different occupational groups, social roles, or demographic groups. One of the groups most studied has been library users, but in recent times new user behaviour research has put under scrutiny user behaviour in the field of informatics. In this field, methodologies were initially centred on studying the relationships with the interfaces of information systems, but recently more attention has been drawn to behavioural studies on the use of the Internet with different methodologies and user groups (Choo et al. 1998; Redden et al. 2007).

Some of the contextual parameters of information behaviour are: information goal, knowledge and skills, roles, engagement or work tasks (Frend et al. 2005). In other words, understanding a profession’s information seeking behaviour involves the study the users’ roles, skills and tasks. Information and communication technologies, and especially internet technologies, have caused changes in the employment market which in turn have directly influenced the requirements of the knowledge and dexterities that graduates in the field of library and information science need to possess (Khurshid 1999, 1996; Yousefi 2007; Rodríguez 2007). As the domain evolves the perception of the roles assumed by professionals, change. According to Fournier (2006: 1) “LIS professionals need to be confident that they can prepare for new challenges, deal with information anxiety and information overload and claim new professional roles.”

In 1993, Sanz (1993) underlined the alarming scarcity of these studies in business. This author points to the work done by Macnab at Queen’s University and a project carried out by Spanish National Research Council, in which the differences in information needs in the business context were confirmed. Since then most of the research may be seen in the corporate and business domain (Bigdeli 2007; Youssefi 2007; Ibañez and Ocholla 2004). These studies were performed through interviews, surveys and questionnaires which could present a certain bias (Fowler 1993; Wiseman 1972; Crawford 1997; Nathan 1992) and impede their reproduction and extension.

Research results dating as far back as 1998, indicated an extension of the employment market towards the information technologies sector (Tabak and Bernhard 1998). Other authors (Albitz 2001; Khurshid 2003; Quarmby et al. 1999) have focused their investigation on the skills and knowledge required in the job advertisements. Results suggested that the most appreciated skills were the management of electronic resources, information search, database design and the creation of HTML documents. All these associated with skills in the informatics domain (Quarmby et al. 1999; Albitz 2002). In Khurshid’s studies (2003), metadata schemes such as Dublin Core, the Text Encoding Initiative (TEI), Extensible Mark-up Language (XML) and the Resource Description Framework (RDF) acquired a great importance in the employment offers, because university libraries with large document collections demanded this type of knowledge.

The objective of this study is understand the information needs that businesses have while seeking Library and Information Science professionals and analyse how they formulate those needs. Our research was guided by the following questions:

1. Is it feasible to determine the information seeking behaviour of the employers and the information needs of the labour market, by examining job advertisements?
2. Do the roles, skills and tasks required from information professionals in the labour market present, currently, any particularity?
3. What is the interrelation amongst the different parameters and contexts? For example, does increasing the demand for informatics knowledge, augment the need for other factors?

The method applied in this study is herewith explained. The method of data compilation and the creation of the resources needed for the performance of the analysis are also detailed. Then, the results are presented, in which the distinct roles, skills and tasks are analysed. Moreover, a study of the interrelationships amongst the distinct parameters is realized and finally, the conclusions are presented.

Method

The market's information needs have been studied using an unobtrusive observation methodology, based on a mixed analysis (qualitative and quantitative) of job advertisements seeking information professionals. This approach shows employers' and headhunters' information behaviour when they seek such professionals through job advertisements.

Phases of this study include:

1. Information sources: selection and development of a job advertisements corpus, as source of information.
2. Thesaurus development and enrichment: development of a thesaurus, with the domain's terminology about a specific work role, along with the tasks and skills involved in this role. In the case that a domain thesaurus is not available, this phase is realized in two stages:
   - Development: initially, a thesaurus is developed adapting a similar one or creating one from scratch.
   - Enrichment: the thesaurus is evaluated. If it is necessary the thesaurus is adjusted and enriched with the categories and terms required.
3. Content analysis of the job offers by means of automatic indexing using the thesaurus. This step includes a statistical analysis of the main categories of the thesaurus and the correlations amongst them.
4. Evaluation: characteristics of the information needs are evaluated.

Selection of information sources

The documents used to compile the elaborated corpus were 1,020 job advertisements, in Spanish taken from the documentacion.com.es Website (Loeve and Sola 2008) corresponding to the period between 2006 and 2008 (the site ceased operations in 2008). This particular resource represented at that time the best communication channel specializing in the domain of library and information science in Spain. Its main characteristics are: it was dedicated to the Spanish ambit; it was organized by years and months and provided a history of requests, which facilitates temporal investigation.

Thesaurus development

A study of the information resources related to employment is performed to elaborate the ontological model. The thesaurus contains the academic profiles of professionals in library and information science. This thesaurus includes concepts, relationships and instances and is composed of diverse taxonomies. It reflects different standardized and ad-hoc classifications, which are united by relationships and lists of instances. The principal resources that comprise the reference thesaurus were manually established and are listed below:

- Knowledge. Using the Library and Information Science Thesaurus of the Institute of Scientific and Technological Studies (previously known as CINDOC), we have extracted the terminology relative to library and information science and other related domains. We have based our work on 1,480 terms from this thesaurus, which are distributed in seven categories: information science: history, theory and systems; documents and information sources; information science research and methodologies; professionals and users; information representation and retrieval; information systems and; information technologies. Knowledge categories that have been selected were: information science sources; information science research and methodology; information representation and retrieval; information systems and information technology; and academic degrees (Mochon and Sori 2007).
- Competences. transferable competences and specific competences. The transferable and specific graduates' competences which have been selected to be incorporated in the reference thesaurus come from the proposed classification in the library and information science White Paper of the Agencia Nacional de Evaluación de la Calidad y Acreditación (2004).
- The professional profile of library and information science graduates. Analysis of the employment demand managed by the Centre for Information and Orientation for Employment of Complutense University of Madrid (2005).
- Skills elaborated ad-hoc, used by human resources, were also added as a category.

Enrichment of the thesaurus

After some initial tests it was realized that the thesaurus was not up to date with the necessary skills’ terminology and this impeded the automatic identification of terms in the job advertisements. So, 479 terms (in Spanish) were added to the thesaurus based on the frequency of those terms in the White Paper (ANECA 2004) and the number of job advertisements that contained those terms. Some examples include: natural language processing, ontology, usability, digitization, content managers, CMS, Joomla, Wiki, Moodle, SEO, search engine positioning, semantic web and RSS.

Also, a new category was created that included skills required by human resources and did not exist in the White Paper. For example, charisma, customer care, commitment, efficiency and methodical character. These terms added were relocated within the thesaurus to generic categories that could incorporate them depending on their predominant meaning in the job offers.

Content analysis of the job advertisements

QDA Miner and WordStat software have been used to process the advertisements. By indexing the documents with the thesaurus we obtain the terms that appear in the documents along with their frequency and co-occurrence. In this study, one of the main characteristics of this content analysis and automatic indexing system is the identification of compound terms and the use of the thesaurus during indexing. Grouping the terms based on semantic relationships allows the frequencies and the co-occurrence values to relate to concepts and not to tokens.

Evaluation

The analysis of the employment offers is realized in levels one, two and three of the hierarchy corresponding to knowledge and competences. We have established three types of competences: transferable competences, specific training competences in library and information science and competences used by human resources professionals. Result values are presented as a percentage of documents. A cluster analysis is performed on the job offers and the course descriptions to obtain the relationships amongst the thesaurus’s families. Then, the matrix co-occurrence between knowledge and competences is analysed and finally an analysis of the results is performed without aggregating the various thesaurus categories.

Results and discussion

http://informationr.net/ir/15-4/paper441.html
Reference thesaurus elaboration result

The reference thesaurus is formed of 1,850 terms distributed in twelve categories to represent knowledge and competences. In the hierarchy, knowledge and competences are differentiated as separate concepts to analyse their representation inside the job offers.

Competences

The first results of the automatic process illustrate the most generic competences. In figure 1 we can observe how transversal competences are the ones that acquire greater importance in the job offers, even surpassing specific graduate competences. An analysis of the transversal competences’ subtypes, illustrate that the most required transversal competences are the instrumental competences which greatly exceed systemic and personal competences (figure 2).

It is important to note that the sum of instrumental, systemic and personal competences in Figure 2 exceed transverse competences in Figure 1. This is because various competences may appear in the same employment offer and so overlapping may arise.

![Figure 1: Competence types](image1)

<table>
<thead>
<tr>
<th>Type</th>
<th>Competences</th>
<th>Advertisements (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental</td>
<td>Computer skills in the field</td>
<td>37.80%</td>
</tr>
<tr>
<td>Instrumental</td>
<td>Knowledge of a foreign Language</td>
<td>37.4%</td>
</tr>
<tr>
<td>Instrumental</td>
<td>Organizational skills</td>
<td>8.3%</td>
</tr>
<tr>
<td>Systemic</td>
<td>Initiative and entrepreneurial spirit</td>
<td>5.6%</td>
</tr>
</tbody>
</table>

![Figure 2: Transferable competences](image2)
Within transferable competences, instrumental competences stand out over personal competences and systemic competences. They acquire a particular relevance in the job offers because they include the knowledge of foreign languages and computer skills in the field. Advertisements appear to give more importance to the computer skills in the field and to knowledge of a foreign language (Table 1). The languages in demand are: German, Spanish, Catalan, French and especially English. Moreover, in Table 1 we can notice standing out the transferable competences: organizational skills, initiative and entrepreneurial spirit and capacity to analyse and synthesize. The rest of the values in the table appear scarcely represented. This low level of representation could be caused by the brief description of the employment offers.

Table 1: Transferable competences subtypes found in job advertisements

<table>
<thead>
<tr>
<th>Type</th>
<th>Competence</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental</td>
<td>Capacity to analyse and synthesize</td>
<td>2.3%</td>
</tr>
<tr>
<td>Systemic</td>
<td>Creativity</td>
<td>1.7%</td>
</tr>
<tr>
<td>Systemic</td>
<td>Self-learning</td>
<td>1.5%</td>
</tr>
<tr>
<td>Instrumental</td>
<td>Oral and written communication in native language</td>
<td>1.3%</td>
</tr>
<tr>
<td>Personal</td>
<td>Teamwork</td>
<td>1.0%</td>
</tr>
<tr>
<td>Personal</td>
<td>Interpersonal communication skills</td>
<td>1.0%</td>
</tr>
<tr>
<td>Personal</td>
<td>Information management skills</td>
<td>0.4%</td>
</tr>
<tr>
<td>Instrumental</td>
<td>Decision making</td>
<td>0.4%</td>
</tr>
<tr>
<td>Systemic</td>
<td>Leadership</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

Figure 3: Library and information science competences in job offers

The specific library and information science skills that appear in the advertisements are illustrated in Figure 3. The skills standing out are: Knowledge of the professional environment in libraries and information science, information technology - telecommunications, information analysis and representation and identification and evaluation of information sources and resources.

Knowledge

Table 2 contains the principal families related to thesaurus knowledge, along with examples. These categories group together hundreds of specific terms.

<table>
<thead>
<tr>
<th>Generic Descriptor</th>
<th>Specific Descriptor Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Technology (&quot;Tecnologías de la Información&quot;)</td>
<td>Library computing, hardware, software, Internet</td>
</tr>
<tr>
<td>Information Systems (&quot;Sistemas de Información&quot;)</td>
<td>Libraries and archives (types, services and techniques), documentation centre</td>
</tr>
</tbody>
</table>
Figure 4 illustrates the relevance of the skills and knowledge according to their presence in the advertisements. This figure shows some generic categories of skills and the number of documents that reference these skills. We have observed that the jobs are not always described using a technical vocabulary for the respective domain. On the other hand, in the advertisements the importance of information technology and information systems is often over-emphasized. The theoretical content which is often related to research is rarely present.

Table 3 illustrates some examples of skills at a more specific level. The objective of this table is to show how the thesaurus has been structured and what kinds of terms have been added to it. Terms that did not exist in the original resources but were found in the advertisements were added to the thesaurus during its elaboration phase. The first column holds the name of the specific descriptor prepended with the initials of the family (or generic descriptor) in which it belongs added to the beginning (note Figure 4). Column 2 contains specific descriptors of the generic descriptor in the respective row of column 1.
Figure 5 illustrates the skills most demanded for the jobs. The importance of skills associated with software, Internet and library computing is clearly represented at this level. These technical skills are prerequisites in seeking a job. However, the fact that the job offers are published in a Website could imply a certain slant towards those variables.

To dig into concrete skills in demand, we have extracted the most frequent terms as a percentage of the number of cases. Without grouping into generic categories or synonyms, the following figures stand out in the list. In Figure 6, the required language skills are presented. In Figure 7, skills in management and editing of digital and Web content are presented, along with other desired knowledge. In Figure 8, we illustrate knowledge and skills related to information retrieval and optimization. Finally, in Figure 9, knowledge centred on information technologies and programming languages is presented. In all these cases, the great importance of technologies, especially those related to the Internet, office automation, data bases and content management, is made apparent.

<table>
<thead>
<tr>
<th>Table 3: Skill subtypes (terms added appear in italics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT Software</td>
</tr>
<tr>
<td>IT Internet</td>
</tr>
<tr>
<td>S&amp;T Science disciplines</td>
</tr>
<tr>
<td>IR: Information representation</td>
</tr>
<tr>
<td>IT: Library computing</td>
</tr>
<tr>
<td>ISys: Archives</td>
</tr>
<tr>
<td>So: Documents</td>
</tr>
<tr>
<td>ISys: Library techniques</td>
</tr>
<tr>
<td>ISR: Research methods</td>
</tr>
<tr>
<td>ISys: Libraries</td>
</tr>
<tr>
<td>ISR: Statistical methods</td>
</tr>
<tr>
<td>ISys: Archiving techniques</td>
</tr>
</tbody>
</table>

Figure 6: Skills in languages

Figure 5: Level 2 demanded skills

English
Catalan
French
German
Italian
Basque
Portuguese

http://informationr.net/ir/15-4/paper441.html
Figure 7: Skills in management and edition of digital and Web content

Figure 8: Skills in retrieval and optimization
To test the relationships amongst knowledge and skills we analyse the similarity matrix to identify clusters. We select the phi coefficient to measure similarity because: 1) it takes into consideration the documents that do not have any occurrence of any of the two elements; and 2) it is not sensitive to the direction of coding.

In the dendrogram in Figure 10, a cluster is observed that separates the possession of an academic title from the rest. Within the skills we notice that skills in information technology, the transferable competences (e.g., English language abilities), specific training competences in library and information science and the competences used by human resources (e.g., customer care or orientation to results) are united and have a higher frequency. On the other hand theoretical skills and skills specific to the discipline remain separated.
Table 4 presents the concurrence of knowledge and competences most demanded in the job offers. Term frequencies with more significant values in respect to the other categories have been formatted as bold. Categories with smaller frequencies have been eliminated for brevity. The transferable competencies at the knowledge of a foreign language, the possession of a relative academic title and the knowledge of information technologies (Internet, software and library computing) do stand out. It can be deducted from the table that knowledge in archive and library science is frequently associated with the possession of an academic title. The knowledge of languages and the possession of an academic title represent the two essential factors in the job offers. The importance of Internet is again decisive especially along with knowledge of languages, software and marketing techniques. It is important to point out, that important concepts such as search engine optimization tools and search engines positioning are not included into marketing techniques even though they are related.

Conclusions

The information needs of businesses requiring library and information science professionals to search for information have been approached, through an unobtrusive observing methodology, using the analysis of employment offers published in Websites addressed to this sector.

Skills and knowledge that library and information science professionals should have, as those are shaped by employers' demands, seem to have their own characteristics. These characteristics do not seem to reflect the typical role of a librarian or an archives expert.

Transferable competencies are the ones that acquire a greater relevance in the job offers (43%). Within the transferable competencies, instrumental competences stand out because they contain competences relative to the knowledge of a foreign language, skills relative to the discipline studied and to organizational skills. This underlines the importance of the knowledge of languages to access the job market in the library and information science domain.

The concurrency matrix shows the value that the employers grant to an academic title (mainly in library and information science). An interesting datum is given between marketing techniques and Internet which when analysed points to what seems to be new tasks for information professionals: Web positioning and search engine optimisation. Information technologies as can be seen in the concurrency matrix are strongly related to the knowledge of Internet and software, or Internet and library computing.

The realignment of a thesaurus for the analysis facilitates the reproduction of results. Also, it permits the identification of changes in the information needs and the contexts inherent to the profession. Moreover, the advantage of using thesauri permits other research groups to reproduce the results and improve the thesaurus until a shared conceptualization is achieved. The extension of the vocabulary and the reuse of semantic categories common to other fields, permits in its turn the reproduction of this method with other occupational groups, social roles, or demographic groups.

The authors consider that this unobtrusive approach is complementary to questionnaires, surveys and interviews which are typical to information behaviour studies.

Acknowledgments

This study has been financed by the Carolina Foundation training programme and the Hispano-Brazilian programme of University Cooperation of the Ministry of Science and Innovation and the mobility programme of the Ministry of Education (PHB2007-0099-PC).

We would like to thank documentacion.com.es webmasters, Javier Leiva and María José Sola, for supplying the corpus of job offers and authorizing its publication for research purposes.

About the authors

Sonia Sánchez-Cuadrado holds a PhD in Library Science and Digital Environment, designing a methodology for the automatic construction of knowledge organization systems. She can be contacted at s.sanchez@ie.inf.uc3m.es.

Jose Antonio Moreiro joined Carlos III University of Madrid in 1991. He is Professor at the Department of library and information science of Carlos III University and leader of the Information Engineering Group researching on knowledge organization systems. He can be contacted at jmore@bib.uc3m.es.