

Teaching online information retrieval to students of journalism

Charlotte Wien

Department of Journalism, University of Southern Denmark, Odense, Campusvej 55, 5230 Odense M, Denmark

Chw@journalism.sdu.dk

In the beginning of the 1990s it was decided to start educating journalists at two Danish universities. This was decided in the hope that such academic journalists would carry greater insight into that work. Therefore for the curricula of these students of journalism some traditional academic disciplines were introduced. One such academic discipline was information retrieval. It was clear, however, that the information retrieval course had to be designed specifically for the students of journalism, as the information needs of journalists differ from the information needs of more traditional academic disciplines. Thus, this article describes the work done in order to develop such a course for students of journalism. Firstly, it analyses the information needs of journalists on a theoretical basis. Secondly, an empirical study analyses which information retrieval resources are available to Danish journalists. Finally, it combines the theoretical and the empirical findings in arguing that it is necessary to provide students of journalism with a theoretical understanding of how online information retrieval works, and also practical experience with several information retrieval resources that they are supposed to use in their daily work.

Introduction

New knowledge is most often produced by combining information from different sources. To produce knowledge, therefore, requires information retrieval skills. An information retrieval skill is defined as the ability to find information in such a way that non-relevant data (noise) are excluded while relevant information is found. Today new information is mostly produced by researchers and journalists. The major difference between these two information providing groups is the amount of time available to them for producing new information. While the researcher might have three years or more to write a doctoral dissertation, journalists, on the other hand, have on average between three hours and one week to draw up an article. The researcher, therefore, might be satisfied with the retrieval of references, while the journalist usually will require full-text documents or facts.

Nicholas and Martin [1, p. 43] state that newspaper journalists:

‘... are very much in the forefront of the information technology revolution. Any changes in their information needs and information seeking behaviour could provide an early warning sign for other information communities,...’

This is one of the main reasons why this article will focus on newspaper journalists and their information needs. In terms of information retrieval journalists in general have two options: searching for information themselves or asking an intermediary (librarian) to do the search. Applying the latter option, they must be able to express their information need in such a way that the intermediary can develop a search profile from the information need statements.

Having retrieved some documents, the journalist or the intermediary has to decide on which of the retrieved documents are relevant to the information need. This is not as simple as it sounds. Some researchers [2, p. 50–53] within Library and Information

Science (LIS) consider the situation in which a user has an information need as an Anomalous State of Knowledge (ASK): the user does not really know what he¹ wants in order to satisfy his information need. The user has an information need *because* he does not feel that he knows all he needs to know about the subject. Many researchers of LIS think that the judgement of relevance necessitates knowledge about the subject of the information need. This has led some researchers to point out that ‘...it may be possible that the user is distinctly *unqualified* to judge relevance on the articles he has retrieved [4, p. 202–204]’. That information retrieval is a complicated process is reflected by the entire discussion of the concept of relevance and the discussion of user information need within the subject of LIS.

Bearing on the information needs of journalists the findings of Nicholas and Martin [1, p. 44] support that the problems discussed by researchers of LIS certainly apply to newspaper journalists. They found that:

‘Journalists ... often provide vague and generalised descriptions of what they want, [to the intermediary, partly because ...] vague subject specification mirrors the journalists’ own uncertainty as to what they want ...’

The only possible way to avoid such a situation seems to be teaching students of journalism the necessary information retrieval skills. Nobody has ever questioned the necessity of training the intermediaries. It takes several years to become a librarian. Only a few educational institutions, however, include student training in information retrieval. Basically, such training can be organised in three ways:

1. By teaching the students how to retrieve information from a series of information retrieval resources.
2. By teaching the students how to express their information need in such a precise manner that an intermediary can transform them into search statements and assist in the process of deciding which of the retrieved documents are relevant.
3. By combining the two approaches. That

is, by teaching the students the basic theoretical problems involved in information retrieval regarding the expression of information need, and by applying the acquired theory to exercises in various information retrieval resources.

Which of the approaches to choose depends on the job situation of the students subsequent to graduation. Approach (1) is useful if it is known with certainty which information retrieval resources will be available to them when they graduate. Whilst approach (2) seems to be the most appropriate if the students can expect to have information retrieval support available to them for all information retrieval tasks. Finally, approach (3) should be used in the case of uncertainty regarding the conditions, in terms of online information retrieval support and resources, the students will face in their future job.

The subject of this article is the matter of choice between the three basic approaches mentioned in teaching information retrieval to students of journalism.

Background

Nicholas and Martin have carefully studied the information needs of journalists. They [1, p. 44] begin by dividing the user group ‘journalists’ into three subgroups, stating that their information needs differ dependent on the kind of newspaper they are working for. In particular journalists working for daily or Sunday newspapers do have the time necessary to research stories. This group of journalists, consequently, has the greatest need for information retrieval skills. Nicholas and Martin [1, p.44–45] found that the information need of this group can be divided into five broad categories:

1. fact checking;
2. current awareness;
3. researching;
4. obtaining context;
5. stimulus.

Of these categories, fact checking, researching and obtaining contexts are the only ones that a teaching program can include, because only these three presuppose systematic information retrieval behaviour. In

the case of stimulus and current awareness, one might assume that browsing would be fully satisfactory.

Nicholas and Martin [1, p. 45–46] found, that there is no ideal label for the various types of information (conceptual, theoretical, historical, descriptive, statistical, or methodological) that journalists request, and that the journalists are generally capable of penetrating even the most complicated texts. However, the viewpoint or angle of a text may be of great importance to them, because they usually demand information written from the point of view of their own work.

The study of Nicholas and Martin shows that one of the main problems for journalists is the fact that they are bombarded with information. They see it as a major problem to cope with this excess supply of information. This explains why it seems difficult to convince journalists of the value of learning information retrieval. Recapitulating the point made in the beginning of this article that journalists lack the ability to state their information needs in a clear manner, and the findings of Nicholas and Martin, these findings support the view that the journalists are in need of a method for expressing information need in an exact manner. Such a method would simultaneously help the journalists to separate information into relevant and non-relevant categories.

In short, daily and Sunday newspaper journalists need information retrieval skills for fact checking, research and obtaining context. Their needs do not relate to any particular type of information. However, they prefer information that fits in with the point they are trying to make. And finally, one of the major problems for journalists is the lack of ability to express their information needs in a precise manner.

The choice between the three approaches mentioned above is conditioned by knowledge about the online information retrieval resources used by, or available to, journalists. The study of Nicholas and Martin [1] does not focus on this. But it seems obvious that FT Profile as well as wire services like Reuters and Associated Press are of great importance. Another related study by Nicholas and

Williams [5, p. 7] shows that NEXIS, Dialog, Textline and Business Briefing are also widely used for information retrieval by journalists. And finally the importance of the Internet is growing at a tremendous speed.

Most educational institutions for journalists obviously aim at educating journalists in a national context. This very national context must be an important determinant in deciding on the information retrieval tools that the students of journalism are to be trained in using. Thus, for the purpose of teaching information retrieval to students of journalism it is necessary to focus on one particular country. For this article, Denmark was chosen. This was due to several factors. Firstly, because the country is rather small only a few institutions are educating journalists. Secondly, because the number of newspapers is limited. And thirdly, because the author is responsible for developing a course in information retrieval for students of journalism at the University of Southern Denmark.

Journalists in Denmark have for several decades been educated at one particular educational institution (Danmarks Journalisthøjskole) which is dedicated to educating journalists only. In the middle of the 1990s it was decided to break the monopoly of this institution by allowing two Danish universities (University of Southern Denmark and Roskilde University) to educate journalists also. The aim was, amongst others, to introduce some of the traditional academic disciplines into the curricula of students of journalism and hopefully to educate a new type of journalist – one who would be carrying greater insight.

One new academic discipline introduced at the University of Southern Denmark was 'Information Retrieval and Content Analysis'. It was clear from the outset that this course could not and should not be an exact copy of the courses given to students of more traditional academic disciplines. These students are mainly taught how to use reference databases, printed or online bibliographies within a specific scientific field. Given the time constraint of journalistic work and the fact that journalists might require information on practically every

possible subject on earth, such course work would be a complete waste of time.

Another important aspect to be considered is the fact that no one knows for certain which information retrieval tools will be available to the students subsequent to graduation. Thus, there is no point in teaching them how to use tools available to journalists in 1999 when these tools might be outdated in 2001. New and much more effective ones might have emerged then. They have to be taught in such a way that their knowledge will not be outdated in two years' time.

A survey of the existing textbooks showed that there are basically two types:

- User manuals or tutorials for various information retrieval systems, or materials describing the content of selected information retrieval resources. Using such materials would only lead to an inductive method of teaching, and it would lead also to the choice of approach (1) mentioned above.
- Theoretical introductions to the scientific field of information retrieval in general. Using such material would lead to a deductive method of teaching and would imply the choice of approach (2) mentioned above. An example of such a textbook is the excellent one by Harter [6].

No available teaching material seemed to combine theoretical problems with practical exercises on selected information retrieval resources that would be relevant to journalists. If approach (3) was selected a new textbook therefore had to be written. However, selecting the basic approach to teaching the students of journalism information retrieval necessitated a study revealing which information retrieval resources and support were in fact available at Danish newspapers.

In short, we were faced with two sets of problems: to choose which basic approach to adopt for the course, and to recognise that if approach (3) was chosen then none of the available material seemed appropriate for the purpose.

Method

Before deciding which approach to adopt, it was necessary to investigate which online

information retrieval support and resources were available for the journalists working at newspapers in Denmark. Therefore, the largest newspapers in Denmark were contacted. It was decided to only include the largest newspapers in the study. The reason for this was: firstly, the assumption that the large newspapers will be the major future employers for the students; secondly, that many of the small newspapers do not themselves produce articles relying on a large amount of research, context or fact checking. The smaller papers are mainly local papers dealing with local news. Finally, the larger newspapers were assumed to be in front in terms of use of information retrieval sources as well as in information retrieval support and resources provided to the journalists. The definition of a newspaper used in this article is that the newspaper should be published at least five days a week. The reason why such a definition is necessary is that several Danish newspapers publish both a Sunday and a daily issue. Most of the large Sunday newspapers are a special version of the daily newspaper. Some of the daily newspapers, however, do not have a Sunday issue. And finally some of the Sunday newspapers are only issued as Sunday newspapers. It also has to be a 'newspaper': that is a paper that promulgates news and not, for example, commercial tenders. In terms of defining what was meant by the 'largest', it was decided to include only the newspapers that had a weekly net circulation larger than the mean value.

According to DRB [7, p. 181–195] there are 42 newspapers in Denmark that meet the definition. Several of the newspapers have a higher net circulation number for their Sunday issue than for their daily issue. Therefore, in order to decide which are the largest ones, it was necessary to calculate the total amount of circulation per newspaper per week. The total amount of newspapers circulated in Denmark per week is 11.547.436 (1998). The largest newspaper has a net circulation of 1.595.312 per week, while the smallest one circulates 10.224 per week only. The mean value of circulation per week is 274.939. Only nine of the 42 newspapers have

a circulation number greater than the mean. These nine newspapers were selected for the study. Table 1 presents the nine newspapers and their net circulation.

Having selected these nine newspapers, the next step was to investigate the availability to the journalists of online information retrieval support and resources. A newspaper was considered to provide online information retrieval support, if at least one of the employees assists with online searching. The employee might be a librarian, a system manager, or a fellow journalist. However, assisting with information retrieval had to be included in the formal job description. Concerning the online information resources, it was decided to define an online information resource as an online service for information retrieval if it had been subscribed for. Thus, for example, 'Dialog' or the 'Internet' was considered *one* online information resource.

Firstly, the newspapers were asked if they had a library, and whether online information retrieval resources were available there (national and international). Secondly, they were asked if a librarian or other professional among the staff was in charge of carrying out online information retrieval. The next questions related to whether or not the journalists themselves had online access and to which of the resources and finally, whether the journal-

ists themselves in fact did carry out searches either in the library or from their desktop.

Findings

All of the newspapers, except one (*Århus Stiftstidende*), claimed to have a library in accordance with the definition of this article, and all of them, except one (*JydskeVestkysten /Kolding Folkeblad*), had at least one librarian among the staff. For one of the newspapers the librarian was not carrying out searches for the journalists (*Fyens Stiftstidende*). In terms of available online information retrieval resources Table 2 presents how many different resources could be accessed from the library.

Table 2 confirms the assumption that the larger the newspaper, the more information retrieval resources are subscribed to. There seems to be one exception only from that pattern, as *Politiken* subscribes to the same amount of information retrieval resources as does *Ekstra Bladet*. The reason why *Ekstra Bladet* and *Politiken* subscribe to so many information retrieval resources is due to the fact that these two newspapers share their online information retrieval resources, and that they furthermore serve as agents for several of the information retrieval resources in Denmark.

Table 3 presents the results of the study on which online information retrieval

Table 1. The nine largest Danish daily and Sunday newspapers, net circulation and total weekly circulation (1998)

Name of newspaper	Net circulation		Total weekly
	Daily	Sunday	
(a) Ekstra Bladet	222,697	259,130	1,595,312
(b) Morgenavisen Jyllands-Posten	175,687	270,349	1,324,471
(c) Berlingske Tidende	154,687	201,475	1,129,597
(d) Politiken	147,153	197,205	1,080,123
(e) B.T.	138,263	138,263	967,841
(f) JydskeVestkysten/Kolding Folkenblad	95,663	111,103	685,081
(g) Aalborg Stiftstidenende	70,333	85,522	507,520
(h) Fyens Stiftstidende	66,147	94,276	491,158
(i) Århus Stiftstidende	62,780	79,183	455,863

Teaching online information retrieval to students of journalism

Table 2. Online information retrieval resources available at the library of the newspapers

Name of newspaper	Subscription to online information resources		Total
	National	International	
(a) Ekstra Bladet	9	16	25
(b) Morgenavisen Jyllands-Posten	10	4	14
(c) Berlingske Tidende	4	1	5
(d) Politiken	9	16	25
(e) B.T.	4	1	5
(f) JydskeVestkysten/Kolding Folkenblad	3	1	4
(g) Aalborg Stiftstidenende	3	4	7
(h) Fyens Stiftstidende	3	4	7
(i) Aarhus Stiftstidende	2	1	3

Table 3. Online information retrieval resources available at the desktop of the journalists

Name of newspaper	Subscription to online information resources		Total
	National	International*	
(a) Ekstra Bladet	8	1	9
(b) Morgenavisen Jyllands-Posten	5	1	6
(c) Berlingske Tidende	2	1	3
(d) Politiken	8	1	9
(e) B.T.	2	1	3
(f) JydskeVestkysten/Kolding Folkenblad	3	1	4
(g) Aalborg Stiftstidenende	4	1	5
(h) Fyens Stiftstidende	4	1	5
(i) Aarhus Stiftstidende	2	1	3

*The Internet

resources are available to the journalists themselves.

As can be seen the newspapers are restrictive in terms of allowing journalists themselves to search the large international databases except for the Internet. Several of the newspapers had a company policy stating that the journalists themselves could only have access to the free or national services. Such policy was made for economic reasons.

- The journalists are not in any way trained in information retrieval and, therefore, giving them access to, for example, Dialog or FT Profile was assumed not to be cost efficient.
- The librarians were considered to be able to perform the searches much more efficiently.
- Furthermore, three of the newspapers said that some of their departments have their own independent budget and that they subscribe to other services than the ones men-

Teaching online information retrieval to students of journalism

Table 4. Frequencies of subscriptions for various information retrieval resources by the newspapers: national and international resources

National information retrieval of resources	Newspapers subscribing	International information retrieval resources	Newspapers subscribing
Polinfo (H)	8	Internet	9
Berlingske Avisdata (T)	8	FT Profile	8
Borsen (T)	5	Dialog (H)	3
Aktuelt	5	Reuter Business Briefing (F)	3
Public-com (F)	5	Affairs Data (T)	3
Greens Directory (F)	3		
Artikelbasen (R)	3		
Danbib (R)	3		

H: Database host, T: Full-text database, F: Fact-database and R: Reference database

tioned here. These services are mainly trade directories and encyclopaedias. These are not included in Table 3. It is, however, clear that the journalists are expected to carry out some searches themselves (the cheap ones in the local databases) and to ask for assistance from the librarian for other searches (in expensive international databases).

Table 4 presents the frequencies of subscription for the various databases. The table includes information retrieval resources subscribed to by more than two newspapers only. The reason for this was again, that two of the newspapers share resources.

As can be seen, the newspapers subscribe to all three types of information retrieval resources: full-text, fact-databases and reference databases. With regard to the international resources, the Internet and FT Profile are very important. Among the national ones Polinfo and Berlingske Avisdata are the most significant. Polinfo can best be explained as the 'Danish FT Profile': It is a database host that includes several Danish newspapers in full-text but also some fact-databases, while Berlingske Avisdata is the text archives of *Berlingske Tidende*. Berlingske Avisdata, Børsen and Aktuelt are full-text newspaper databases. Publi-com is a fact-database that includes company budgets and account balances from all major Danish companies. Greens Directory is a local com-

pany directory. Artikelbasen is a reference database, which contains references to articles from 750 Danish newspapers and journals and finally Danbib is a reference database that includes references to all documents in Danish libraries.

In summary, Polinfo and Berlingske Avisdata are the national information retrieval resources that are subscribed to by most newspapers, while FT Profile and the Internet are the most important international resources. It is important to note, however, that the newspapers also subscribe to national fact-databases and national reference databases.

Discussion

The study has shown that most of the newspapers have a library and that they have a librarian, who is expected to carry out or assist with searches for the journalists. This finding supports the view that the students should be taught to express their information needs rather than perform searches themselves. On the other hand, the journalists are also to a certain extent expected to carry out searches themselves: they have access to some, mainly local, online information retrieval resources. This finding supports the view that students need practical hands-on training in using some information retrieval resources.

The findings of this study, therefore, suggest that the students should be taught the basic theoretical problems involved in information retrieval, but also that they should be trained in using certain information retrieval resources. This means that approach (3) should be adopted as the basic concept for teaching information retrieval to students of journalism.

Since the students of journalism can expect to have online information retrieval support, the most important topic to teach them was considered to be how to express an information need. Such skills will be useful to them when searching with or without support. The first aim of the course is, therefore, to teach the students how to express information needs as a question. This helps the student to choose between relevant and non-relevant information retrieved: the relevant information is the information that provides part of or the complete answer to the question. The question, however, is not in itself a search statement. The next step, therefore, was to introduce the basics of Boolean logic, Common Command Language, truncation, organising knowledge (controlled and uncontrolled vocabularies), the essentials of how indexing in databases works, and basics about the differences between reference, full-text and fact-databases and database hosts. These topics were selected because they were considered to be general to all information retrieval resources today and in the future.

Following the theoretical introduction, the theory has to be applied to the information retrieval resources that were found to be the most widely used by the newspapers in Denmark. The study has shown, that both the Internet, full-text, fact-, and reference databases are used. And also that both national and international resources are available to the journalists. Therefore firstly, Polinfo, the most frequently used national online information retrieval resource, was selected; secondly, FT Profile was selected for the course. These two databases were considered representative for teaching the students to work with full-text retrieval on a national and international level. Secondly, the students will

work with fact-databases. Here company directories, statistical databases and encyclopaedias will be presented. Thirdly, concerning reference databases the students will learn how to use the local Danish 'Artikelbasen' which contains references to journal and newspaper articles. Finally, the students will be introduced to some of the different types of search engines found on the Internet. Manually indexed: for example Yahoo! And automatically indexed: for example Lycos.

Regarding the introduction of different online information retrieval resources, the students will be given exercises that simulate real life situations. On the basis of a news telegram they will be asked to formulate a question, to develop the appropriate search statements on the basis of the question, and finally, to retrieve the information that is relevant in order to answer the question. This requires a targeted textbook written for journalists. This textbook should consist of two parts. Firstly, a theoretical introduction which should be kept at an absolute minimum. Secondly, it should include a user manual for the information retrieval resources selected. For this purpose a new concept had to be invented: the critical user manual. Most user manuals for information retrieval resources do not inform the user of how to manoeuvre between controlled and uncontrolled vocabularies of the retrieval tool in order to optimise a search in terms of recall and precision. Neither do they inform about lack of index consistency and errors nor how to detect inconsistencies and errors in the indexing.

Conclusion

In terms of selecting the basic approach to teaching information retrieval to students of journalism it was decided to teach the students the basic theoretical problems involved in information retrieval and to apply this theory to exercises on various information retrieval resources. It was decided to include a theoretical introduction to the course for at least three reasons. Firstly, the study of Nicholas and Martin [1] has shown that journalists need to learn how to express their information needs. Secondly, there is a great

difference in the availability of information retrieval resources at Danish newspapers, which means that it would not be possible to teach them how to use all of them. Finally, that such a basic theoretical introduction, it is hoped, can help the students to cope with future systems.

When students have learnt the basics of information retrieval, they will then be taught how to work with a selection of full-text, fact- and reference databases. These will be selected from the information retrieval resources most frequently used by the largest Danish daily and Sunday newspapers. They will also be introduced to the Internet. For that introduction, however, the focus will be placed on the information retrieval resources available on the Net. In this way, we hope to have provided the students with a general knowledge about information retrieval and the practical skills that enable them to find the information they need – also in the year 2001 when they graduate. By then this investigation will be followed by a comparative study comparing firstly, the information retrieval skills of the students of journalism that have graduated from this University to the students of journalism having graduated from the two other educational institutions (Danmarks Journalisthøjskole and Roskilde University) and comparing secondly, the information retrieval skills of the students of journalism to students of other disciplines here at the University of Southern Denmark.

References

1. Nicholas, D. and Martin, H. Assessing information needs: a case study of journalists. *Aslib Proceedings*, 49(2), 1997, 43–52.
2. Belkin, N.J. Ineffable concepts in information retrieval. In: Sparck Jones, K. ed., *Information retrieval experiments*. London: Butterworth, 1981, 44–58.
3. Harter, S.P. Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 1992, 602–615.
4. Hersh, W. Relevance and retrieval evaluation: perspectives from medicine. *Journal of the American Society for Information Science*, 45(3), 1994, 201–206.
5. Nicholas, D. and Williams, P. *Journalism and the Internet – the changing information environment – pre-fieldwork literature review*. 1998. http://www.soi.city.ac.uk/~pw/ji_lit.htm (visited 12 January 1999).
6. Harter, S.P. *Online information retrieval – concepts, principles and techniques*. Bloomington: Academic Press, 1986.
7. DRB. *Media Scandinavia 1998*. Copenhagen: Danish Association of Advertising Agencies, 1998.

Notes

1. To avoid clumsy constructions resulting from trying to refer to two sexes simultaneously, I have used male pronouns exclusively. However, my information seekers are as likely to be female as male and I mean no disrespect to *men* by my usage. See also Harter [3, 603].