Shifting Contexts: relating the user, search and system in teaching IR

Frances Johnson

Manchester Metropolitan University

Abstract 'Shifting Contexts' explores the benefits that are to be had, with respect to teaching and learning in IR, in delivering a curriculum based on the analysis of the retrieval system in the context of the user perspective of searching for information. Three user contexts are identified, based on students' experience and provide a motivation to relate IR practice and principles in learning how to search, build and/or design a search and retrieval system and its interface. The complementary perspectives enable insight to be gained into our use and interactions with IR systems in characterising search, as well as in the explanations of the techniques and technologies for IR. The goal is to provide an education in the practice of IR and in the concepts and principles that underpin the discipline.

Introduction

If we are to understand the subject of Information Retrieval (IR) we need to be taught in a curriculum that not only covers the core topics of indexing, search and their systems, but that also provides for an analysis of these in the contexts of users, their queries, searches and uses of information. This chapter explores teaching IR in the contexts of the student's experiences with IR systems and, with the goal of facilitating an understanding of IR from the insights gained from the complementary perspectives of the

user and the system. Defining a curriculum for teaching Information Retrieval has been called upon and investigated over the years. Bawden et al (2007), for example, identified a set of 28 topics from an analysis of IR and its related, yet distinct, subjects of Information seeking (IS) and Human Information Behaviour (HIB). It was considered necessary to broaden the topic to cover and integrate the aspects of IR, IS and ISB in order that a coherent and reasoned course curriculum could be constructed from the set of topics, with a particular perspective or target group in mind. Saracevic and Dalbello (2001), in defining a curriculum for a digital library course, not only posed the question *what* to teach, but also *how* and *why* and demonstrated the integrative function of the digital library context for learning about the different aspects of IR. The challenge for the curriculum for IR is not so much that the core techniques or knowledge of the field have developed to the extent that it is hard for the educators to keep abreast, it is the challenges posed by the broadening of IR and its horizons impacting on our experience of the technology. In this chapter, IR refers to its broadest sense whereupon the human or user perspective, the study of information behaviour, and the system perspective, the retrieval system, are complementary in providing the insights for the understanding of IR. In a well organised curriculum this broad scope of the topic can be used as an advantage in teaching the subject. To this end, the approach presented and discussed here is to use an aspect of IS and HIB, specifically user interactions and search processes, as a *context* in which to analyse, and understand, the technology and the design of IR systems. Furthermore the broadening of horizons for IR, in particular in search engine applications is also considered for the impact it has had on our experience with IR systems, and in turn on the user contexts employed in the teaching of IR.

This chapter focuses not so much on the curriculum content for IR, but on the impact the broadening of IR has had on the analysis of the curriculum for its effective delivery. Decisions are to be made regarding what should be taught at undergraduate and at postgraduate level and, given the extent of topics covered in IR, choices must be made to suppress or explain specific detail according to the target audience. However, it is important to note that the fundamental techniques for IR, what we aim to teach, perhaps somewhat surprisingly have not changed (Croft et al., 2010), but have stood the test of time precisely because they are grounded in the discipline – in the study of the properties of text, documents, collections, language, users, information needs and communication. A course in IR is more than a course in engineering a retrieval 'solution', it is a study of the subject area and, as such, the study of the theory, concepts, principles and practice. The broadening of IR for the study of the complementary relations between users and the systems provides insight for the understanding of the discipline, and user contexts may be used to facilitate this analysis in the delivery of the curriculum.

The broadening of the horizons of IR has also had impact on our experience of IR systems and, as such, on the user contexts in which to relate the principles and practices of IR. This chapter explores a range of possible 'contexts' many of which have been employed on the IR units on the undergraduate and postgraduate courses in Librarianship, Information Management and Web Development at the Manchester Metropolitan University, essentially to set up 'need to know' scenarios in learning IR. The shifts in these contexts, in accordance with student experience of IR, will be explored as well as an assessment of their effectiveness in teaching the core practices and principles and, ultimately to gain an understanding of IR.

1.1 What is IR: an overview for the curriculum

Sparck Jones and Willett (1997) define IR in their explanation that

> document retrieval subsumes two related, but different, activities: indexing and searching. Indexing refers to the way documents […] and request [...] are represented for retrieval purposes. Searching refers to the way the file is examined and the items in it are taken as related to a search query.

This succinctly defines the field in terms of its core yet belies the range of topics typically found in an IR curriculum as required to teach the subject – information organisation, indexing, database, file structures, inverted index, Boolean, probabilistic retrieval, ranking, web retrieval, relevance, satisfaction, evaluation of information and human information seeking and searching, to indicate a few. Whilst wide ranging, the extent and depth of coverage can be determined and distinguished, in part, by the aims and perspectives of the host course. Croft et al's (2010) textbook, aimed primarily at undergraduates in Computer Science, indicates what would be an exemplary programme in which the components of a search engine are developed in detail – including crawling to acquire the information, text processing for creating the inverted index, indexing (based largely on statistical properties of text), query processing, ranking algorithms and their retrieval models, evaluation and performance metrics, and techniques beyond index search, including classification and clustering. Indicative of the broadening horizons for IR, the later chapters consider extensions of the core material, suitable for postgraduate study and including advanced techniques to capture document content and search techniques for applications such as social and peer-to-peer search.

Information Retrieval as taught on courses in Information or Web Technology is likely to cover similar ground of IR architecture and techniques with emphasis on the web, for example, as given in Levene's (2006) textbook, *An introduction to search engines and web navigation*

techniques. Likewise a textbook aimed at the student of Library and Information Science will also cover the tools and techniques for IR but with greater focus on the principles of organising and the representation of information, such as classification and the library catalogue, and in the broader context of providing information and its services. Others, such as *Web Dragons* from Witten et al (2007) provide an understanding of how search engines work whilst focusing on the impact of the technology on our lives and our interaction with the world.

1.1.1 Educational goals: the fundamentals

The breadth and depth of the IR curriculum may well be tailored to the student, as potential computer scientist, web technologist, information scientist or librarian, and the technical detail or particular practice suppressed or explained accordingly. Yet, in every course there must be a common educational goal to teach the fundamental techniques for IR which as Belew (2000) points out are as relevant as they were pioneering. The goal in providing an education in IR, the analysis and appreciation of IR, demands not only the teaching of the practice of IR - the techniques and how the systems work -, but also an analysis of 'why' they are as they are – the theory, concepts and principles that have underpinned its development. Thus in learning about IR we seek to understand

- the properties and patterns of language, text and collections that have led to statistical and probabilistic IR or the 'best match' model;
- the semantic indeterminacy and resultant noise of document representations and searchers' requests which Blair (2006) refers to as "the IR problem";
- the characteristics and patterns of our information seeking behaviour and our ability to interpret the retrieved information; - and,

- how we use and interact with IR systems for effective retrieval as Sparck-Jones and Willett (1997) pointed out, in characterising IR, will increasingly complement and become of importance as the field develops.

To effectively teach IR, consideration must be given to how the learner can gain an appreciation of the contexts for IR for the complementary insights to be gained into the principles that underpin the subject. Teaching IR in the context which is familiar to the student, as a user of IR in searching for information, may help in organising a coherent curriculum and for the insights to be gained from the complementary analysis of the users and their searches and the system. In this broad view of IR, the analysis of the user context and of the core concepts and techniques for IR enhances the understanding of both, and resultantly, our ability to apply, use and reason about IR.

1.2 Teaching in contexts

Teaching in context is a fashionable concept with the purpose of providing a motivation for the learner to have explained the ideas and principles of a subject as dictated by the given context (Light et al., 2009). In this chapter the pedagogy of 'teaching in contexts', and its relations such as 'scaffolding' are not examined in any detail, rather 'contexts' is used simply with the aim to make the subject of IR both interesting, accessible and possibly less abstract to the learner. Consideration of the context can also help determine what needs to be learnt as dictated by the perceived needs of the context, and indeed the resultant 'need to know' set up in the mind of the student. Three teaching contexts in IR can be identified of '*Search*', '*Build*' and '*Design*' where the educational goal remains the same but the shifts in the context are adopted as students' experience of IR systems has

broadened. The remainder of this chapter explores what the student may be motivated to learn about the practice of IR, its techniques and technology, so as to be able to participate in the use and design of IR systems. Consideration is also given to the effectiveness of the contexts in helping the understanding of and the reasoning about the principles of IR. In effect we begin to unravel, so as to purposefully employ, the reciprocal relationship of the user perspective, their searching behaviour and the design and development of the IR system, that is the understanding of IR in principle and in practice. The first of these 'contexts' of learning how to search goes back to the early days of IR and what we might consider to be the 'traditional' IR system.

2. The search context

By the 1980s 'computerized information retrieval' was firmly established with the development of the large bibliographic databases, search system hosts and online searching. Searchers were trained, known professionally as search intermediaries, with knowledge of both the procedural and the conceptual skills and techniques for online search. These skills would be taught and honed as part of a course in Library and Information Science on information access covering cataloguing, classification, indexing and online databases and search systems. Students were thus motivated to learn how to search professionally, on behalf of a client with an information need, and to search effectively with knowledge of the central concepts, principles and techniques of information retrieval systems.

Learning to search, utilised as a 'context', requires the student to find out how information is stored, indexed and searched in the IR system and how its features facilitate search. Knowledge of the record structures,

the parsing rules and the stemming and stopwords applied in the creation of the searchable inverted file structures reveals to the searcher the representation against which their query is matched. Equipped with this knowledge of the logical storage structures and information processing the student learns the requirement to specify precisely the set of query terms that the sought documents will contain. Understanding that the words in the searcher's expression, and not the concepts, are matched in the full text Boolean based system, the student learns to make use of field searching and the Boolean and word proximity operators with a sense of how they affect retrieval. Proximity operators are used to bring together terms separated in the word index, truncation to expand the word forms searched, field searching to add precision to the search and Boolean operators to strategically broaden and narrow a search. Thus the student learns to use the search features to gain control of the search and to build detailed and often sophisticated searches, honing the search expression and manipulating the exchange towards a desired outcome.

With practice the student searcher comes to use the system as a tool and gains a certain control over the search outcome as he concentrates on finding the query terms to represent the information sought; in making assessments of the relevance of the retrieved items; and, in using the search terms strategically to obtain feedback from the system to refine the query or to conclude the search. By understanding how the information is processed and stored in the system and by relating the core principles and techniques of the IR system to the practice of search the student learns how to search, and, ultimately (hopefully) meets Harter's (1986) aim in his textbook *Online Information Retrieval* ,that,

> the reader learn[s] *how to think about* online information retrieval. (p.ix)

Effective retrieval requires knowledge of the domain and the system to formulate a query to capture its intended meaning and to submit for matching to take place on some logical representation of the documents. Learning about the IR system and its tools and techniques thus elucidates the practice of search. In a keystone paper, Swanson (1977) made the comparison of search to the problem solving process of scientific inquiry, thus characterising search as - the formulation of the search expression to test a hypothesis that the expression will retrieve desired items. And, just as with scientific inquiry, the inspection of the retrieved results for relevancy is conducted to test the hypothesis, with some expectation that the testing will be an iterative process. Learning to use the tools and techniques of the traditional (Boolean based) IR system it seems leads to a problem solving approach to the search task towards what is often an unknown final goal.

2.1 The shifting context: from traditional to modern systems

The modern search engine, based on an alternative best match model of retrieval, does not (obviously) present the searcher with command based search and the opportunity to nurture a strategic approach to search dependent on knowledge of the search system. It has been discussed that classic information retrieval techniques are irrelevant in the environment of the web (Savoy and Picard, 2001) and that the availability and regularity of search rarely finds us engaged in a well prepared session to conduct a thorough search on a topic. Teaching IR in the context of learning search may be less convincing to today's students - their experience of regular and instantaneous search would seem to have little call for engagement in a well prepared session to conduct a thorough search on a topic. Whilst the expert searcher may be an effective context to teach IR to students who

may search professionally, on behalf of a client, an alternative context seems justifiably called for to encourage students to find out about the technology that provides their experience of instantaneous search, the modern retrieval system.

2.2 Comparing search in traditional and modern systems

During the 1990s development in search technology, that is to say the emergence out of the labs of the 'modern' statistical based retrieval system, required students from Library and Information Science to take more than a cursory look at ranked retrieval systems. Having been taught the 'traditional' retrieval system in the context of 'search', a comparative study of the functionality of the traditional and modern systems presents a possible basis to learn about the modern IR system. A paper on the historical development of retrieval systems by Hahn (1998) provides just such an opportunity by inviting students to concur with or argue against her conclusion that the lack of search control offered by a search engine is so significant that we should not think of these as successors to the traditional online search system. Entering into the debate, as a set exercise on a course in IR, students are motivated to learn about IR techniques and the models that provide both the control and manipulation afforded by the traditional system and the apparent ease by which we query the search engine which has lead Markey (2007) to observe that,

> for the vast majority of people's information needs, doing one's own searching is convenient, immediate and instantaneous – connect to the internet, launch a Web browser, type a query into a search engine's dialog box, browse ranked retrievals, and link to one or more full-length retrieved documents"

In drawing the comparison, not only are the differences highlighted between the models of Boolean and 'best match', the investigation leads to

contemplation of the notion of what might be termed 'computational off-loading', coined by Navarro-Prieto et al. 1999[1], and used here to refer to the extent to which the user's effort to search is reduced and 'off-loaded' onto the engine's computations to interpret the query and, in effect, judge the relevance of the results. Thus motivated to learn about the inner workings of the modern retrieval system and how search engines work, the student can explore, with a sense of purpose, the effectiveness of the technologies based on the vector space model and/or probability and the further clues used in web retrieval, such as link popularity (as referred to by Brattelle (2005) as pre-programmed relevance calculation) as well as a range of other techniques such as relevance feedback, clustering, user profiling and trends such as 'dice and slice' to bring information to satisfy the user's request and/or to encourage users' interaction with and engagement in search (depending on the stance taken in the debate). Again teaching in context, as we discover and learn more about how the engine processes our queries and returns information, insights are gained into the search activity itself. Have the techniques for pre programmed relevance or 'inferred intent' reduced our search effort for the better, in terms of making search easy requiring less time and thought? Does the answer to this depend on the type of the query? Or are there times when we might find searching less of an effort when we utilise the search features and seek to optimise the precision of the search? Or have our intentions simply been artificially narrowed by the similar results we continually retrieve, leading to a disengagement from the search process opting for the easy fix with an expectation of immediate satisfaction? These are the sorts of questions we might

[1] Navarro-Prieto et al. coined the term to refer to the extent to which a representation reduces the amount of cognitive effort required to understand what is being represented

contemplate informed by our acquired knowledge of IR and, in particular, its application in web search engines.

3. The build context

The 'search' context provides a structure of sorts for student learning in IR, and its effective relating of IR techniques and tools to the practice of search helps the learner to understand and be able to reason about IR. However, for many students, especially those in Computing courses the context may be unconvincing. The 'learning to search' context may even be considered too specialised for today's LIS student who comes with considerable experience in searching on the Internet and whose expectations in studying IR increasingly leans towards the developer perspective to learn how to build, implement or provide access to a collection. Thus as the educational goals of the host course develops from the perspective of the user to developer, the context for teaching IR shifts. The information retrieval system itself, specifically the search engine, is something that all students will have used and, in itself, provides an interesting, and familiar context to learn about the subject. The search engine, and the lists of pages in response to their searches, is the most likely student response if asked to name a most frequently used piece of software. The ubiquity of the search engine at once provides a strong motivational context to find out more about IR.

Chau et al (2003) at the University of Arizona make a compelling case for using the familiar (domain specific) web search engine to interest and motivate students in learning fundamentals of Computer Science (CS). They point out the variety and diversity of core skills CS students need to acquire including databases, data structures, algorithms, web servers and

web based interfaces. Furthermore, in a web based environment, students will need to be taught application development using and integrating different systems and tools (with script languages and Internet protocols) and project management. Perhaps they allude to a motivational context in their decision to base the project on the building of a domain specific Web search engine in stating that:

> [w]e believe the project is useful for helping students understand some key computer science and information system concepts, acquire sufficient background in Web computing technologies, and obtain experience with various types of real-life challenges in system development projects.

Courses intending to use the 'build' context to learn about computing basics as well as specific IR techniques may use existing toolkits that provide the components needed to create a search engine. Tools, such as those listed on the web site SearchTools.com, can be directed to take a list of Web sites from a user as seed URLs, collect Web pages based on these, index the pages, and set-up a user interface for querying and browsing. Whilst these are interesting for students to look at and to learn about the basic architecture of a search engine with its components of spider, indexer, query processor and search engine, they may not provide sufficient technical detail on the information processing involved for the educational purposes of understanding data structures and algorithms in a 'build' project. For this reason, course tools (e.g., the programming exercises in Croft et al's (2010) textbook make use of Galago, open source designed for the teaching of search engines) can provide a transparency of the results from each component to favour later processing in a way that is not possible in the integrated toolkits. Chau et al (2003) provide students with two simple tools, called the *AI Spider* and the *AI Indexer1*, to collect and store and index web pages for import into databases whereupon the query engine can be implemented and customized, for example, to improve the ranking of the results or to provide additional functionality such as applica-

tion of linguistic analysis and clustering to the search results (Chen et al 2002 Competitve Intelligence Spider). These individual spidering and indexing tools are widely available to build a searchable index between terms and document for use in a project. *SMART*, developed by Salton, is probably the earliest indexing tool (Salton and McGill 1983) and is widely used in traditional and web based engines. The java based Lucene toolkit is a more recent tool and software library that has become increasingly popular for document indexing due to its ease-of-use and fast indexing speed.

The learning opportunities in the 'build' are considerable as students are involved in working with the programs of search software, including parsing, indexing and computing similarity between query and documents; and the ability to affect the processing in amending the software helps develop an understanding of the theory and core concepts in IR, such as term distribution and the discriminating power of a good index term. Belew (2000) takes this further in intimating that the impact of search engines is such that IR techniques and theory may be perceived as central to the discipline of computer science itself (and thus its curriculum), possibly on par with database but providing a distinct technology that is described in probabilistic rather than in absolute terms.

Thus, it would appear that there is a strong rationale in terms of learning opportunities in IR for the build project to motivate the CS students. Furthermore on obtaining a solid grasp of the fundamentals of IR at undergraduate level there are opportunities to develop student experience in IR at postgraduate level. On understanding the concepts and principles of IR, and the technology of IR, students may be exposed to the related technology and/or applications to provide further functionality to the IR project. Hearst (2005) writes about her experience to introduce NLP in to

what she calls an 'applied class' to extend the usefulness of projects built using basic IR. Her experience may be taken to be promising in that students involved in an IR build project quickly learn the limitations of the existing tools with regards to the functionality they want to build in, but also as a caveat in trying to extend the project and pack too much into the course:

> In fact, I find that this is a key aspect of teaching an applied class: learning what is possible with existing tools, what is feasible but requires more expertise than can be engineered in a semester with existing tools, and what is beyond the scope of current techniques.

Hearst provides the example of a project that provides search on a collection of blogs, enhanced with NLP technology for text categorisation and the content analysis of the retrieved blogs for, say, the analysis for the emergence of new terminology within the blogs. The broadening horizons for IR certainly presents a range of such possibilities to interest students in the development of, and, application of IR technologies to enhance their projects and resultantly their learning in IR. However, cognisant of the extent of required learning for such enhanced projects, and the danger of losing sight of what is essential to learn in the course, Jones (2007) describes a successful problem solving approach to deliver a course that focuses on the interest generated by emerging digital environments, such as multilingual IR or agent based discovery, for students to research and develop solutions using advanced techniques and applications.

3.1 The shifting context: from build to design

IR taught in non computing departments such as Library, Information Science and Web Technology may also find the 'build' context motivational. In these types of courses less emphasis will be given to the acquired and

applied knowledge and skills in programming and computing. However, the engineering approach to build a search and retrieval system, again, provides a familiar and potentially interesting context on which to enhance the required learning of the concepts and principles of information organisation and retrieval. Students on these courses would be expected to be involved in the organisation and representation of information as well as the implementation of search technologies for the retrieval of information. The availability of integrated software development toolkits and commercial systems provides the opportunity for these students to be involved in the 'design' of search and retrieval systems suitably informed by an acquired understanding of IR.

Ruthven et al (2008) describe a project developed at the University of Strathclyde which requires students to draw on course material to engineer a design for the storage and retrieval of a given collection of documents or information objects. This course used the Lucene toolkit, thus focusing less on the computing skills required in the implementation and more on the design, as they explain:

> The use of Lucene meant that the group did not have to invest time in implementing low-level retrieval and indexing code but could concentrate on appropriate design decisions for their documents, e.g. whether to use stemming, to use index fields or whole texts.

The driving force, as explained was the constructivist principles of learning, which goes beyond the motivational notion of 'teaching in context' based on the idea that active engagement with material provides learners with the basis and the means to construct knowledge.

The context has thus shifted from development to emphasise the requirements for the design in providing search and retrieval. Decisions regarding the provision of Boolean, free text searching, fielded or faceted search, vocabulary browse, clustering and ranked retrieval with or without relevance feedback are based on knowledge of the index and search meth-

ods of the system, again motivating the student to have explained these techniques. The potential learning opportunities set up by the design context, however, goes further. The design decisions are not only informed by knowledge of the back end information processing but also by knowledge of IR as a system involving the user, the reasoning about search behaviour and assumptions made regarding the user's interactions with the system. The final section thus concentrates on 'design' as a teaching context and again considers the learning opportunities presented in finding out about the retrieval technology when related to the user's activity of search which the system is designed to support.

4. The design context

Students may be involved in a 'design project' via a series of exercises or in a complete user centered design project with finished product. In either scenario the context sets up the need to learn about search and information behaviour as varied as it is complex. There are times when we want to be able to specify our query, to gain control of the search and go for an exact match on the stored information. Boolean may be off-putting to the untrained user, but there are alternative models that can be used to design for the specific search. Faceted search, for example, is gaining popularity for its functionality, especially on e-commerce websites. On the other hand there are times when we do not come to the engine with our honed search formulation, but with the immediate feedback of retrieved results we can quickly learn how best to ask for the information. It has been noted, usually with a degree of concern, that users typically enter only 2 or 3 keywords to represent their query. However at the same time it has been noted that the general public have learnt how to search and are more likely

to enter a promising set of keyword or phrase, such as 'body mass index' rather than a question 'Am I fat?' (in this instance, to retrieve documents to find out if you are overweight). And, there are other times when our research has not a finite answer and we seek a more tempered approach to searching, along the lines of the berry picking model (Bates, 1989) wherein thought is involved in associating prior knowledge and in interpreting and recognising the information retrieved. Whilst factors will end the query, with this sort of inquiry it is difficult to judge the outcome of the search - the line between "satisfiction" and satisfaction may be impossible to decipher. It is, however, an approach to search which involves the elements of exploration and learning that make it quite distinguishable from the immediateness of the 'popular' query, stopping where we have found something - anything – and from the approach where we have a good idea of what we are looking for and seek to optimise the precision of the retrieved results.

A class exercise can be used to encourage students to think about these three main search activities beyond the immediately apparent surface procedural activity of querying a search engine. In observing the search behaviour of oneself and of others insights are gained into the activity of search and the processes deciphered by relating thoughts and actions, if at all, to established search models, such as Bates's (1989) berrying picking or Kuhlthau's (1988) stages of [re]search. Such practical exercises have always had a place in the IR course, for example to demonstrate inter-inconsistency in indexing documents, and which aim to reveal the issues and challenges for IR (rather than as a piece of valid and tested research). Observing search behaviour is, of course, non trivial and for the purpose of a class task it is facilitated by setting up the (non validated, but functional) categories of the '*anything will do dialogue'*, the '*precision dialogue'* and

the '*concept dialogue'* as are outlined above and to identify differences in the user information queries and system requirements.

The aim of learning about user search behaviour is to reveal the challenges for the design of the search interface. Searching is not a homogeneous activity. Korfhage (1991) argued that designing for the simple search can inhibit the searcher and suggests what is needed is a range of search tools from which users could pick the most suitable for their current query. The user may have a preference for simple search and immediate results (and for computational off-loading), but the observed brief query and subsequent scanning of the retrieved results may signify the searcher engaged in a more complex cognitive or conceptual process, making mental connections to concept build and learn about the search topic. Effectively the 'design' context, in encouraging the student to think about search behaviour and the resultant challenges for the IR system guides the student towards learning about IR principles and practices in the realm of HCIR (Human Computer Information Retrieval). HCIR, as coined by Marchionini (2006), is to design IR systems in a way that reflects the needs and behaviour of its users during the interactive communicative process of IR. If the goal of HCIR is to bring the user back into the system by interacting with the information sought and taking responsibility and control for successfully retrieving information, we need to design not only for known item searching but also for exploration and the high level learning and investigative processes in which searchers engage to understand the concepts about which they seek information. The concept dialogue in the class exercise thus sets the greatest challenge in the design task, and optimises the potential for learning about IR, in aiming for what might be considered to be 'low computational off-loading' and high user engagement with the information and the search features at the interface.

The ensuing shift in the context, but not necessarily of the course content, is perhaps realised with reference to the original 'search' context which encouraged the searcher to *think* about search by relating the principles and practices of IR to the activity of search. The design context focuses on IR at the interface, the presentation of the tools and techniques for IR to facilitate the searcher in searching and in interacting with the information, concentrating the mind on the query and encouraging the searcher to think about search.

4.1 Search Design

The task to design a search and retrieval system given an appreciation of the user's search activity and requirements, aims to provide a motivation for the student to learn about IR. To achieve this, the design task may involve the actual design and build of the system, although the educational goal of the teaching context may also be achieved in paper prototyping and/or comparative reviews of existing search systems and specifically their interfaces. Depending on the duration of the course, the latter may be the only feasible choice and may be preferable in deterring the student from getting too caught up in practicalities of implementation which can distract from the aim of the exercise. Designing the system requires an understanding of IR techniques and technologies to decide on the best search model for the proposed system. Understanding the user search activity, especially the requirements for exploratory searching, further informs design for user interaction. For example important design decisions need to be taken to show to the user how or why their results were found and to enable the searcher's interaction with the results in focusing on finding what they want. Further inspiration into how the study of the

search interface is informed by the theory and practice of IR (and HCI) can be drawn from recent textbooks focusing on the search interface, such as Hearst (2009). Of course there is the danger that design and the interface further broadens the study of IR requiring the student to delve into the topics of Interface Design, Human Computer Interaction (HCI), usability and user testing. Whilst it is possible that an interesting and highly motivating course could be built around the intersection of IR and Interface Design, we need to keep reminded of the intention of 'teaching in context'. For this reason, the analysis of information visualisations used in the design of the search interface may be used in effect as a teaching resource to meet the goal of learning IR.

Visualisation is about exploration and understanding (presenting the information in a way that facilitates its interpretation and its use). Clustering technologies, for example, can be used to group results or generate categories or facets to guide search. Thematic visualisations of the clusters of information aim to present to the user an overview of a subject domain, or a fuller picture of the document set. Early examples of search results visualization are discussed in Korfhage (1991) and Hearst (1999) which cluster, based the statistical similarity of documents within a set, and then map the resulting groups into a galaxy type visualisation (e.g ,Korfhage's 2D (VIBE) or 3D (GUIDO)) These visualisations may be particularly useful to a searcher new to a topic as it provides a snap-shot of the significant areas in the subject. It may also be of use to more established researchers as an aid to spotting gaps in the knowledge of the subject and where new connections might be established. Throughout the 1990s and into the 2000s a wide range of experimental visualisation systems have been developed, such as InfoCrystal (Hearst, 1999) and TileBars (Hearst, 1999), Kartoo (online) and Grokker (online), which also pre-

sent potentially useful search aides in the visualisations of concepts and the view of their relationships within and across documents Conceptual visualisation is a major development in the promotion of searching for information as a holistic process, as compared to the atomised bytes of the linear text based results. It offers a way of exploring and comprehending the complexity and connectedness of information, rather than the relatively straightforward task of retrieval. It aims to assist in the learning and investigating involved in searching for information firstly, to allow access to subject areas the searcher did not know existed and did not know to ask questions about; and, secondly, to expand the searcher's broader understanding of the domain of interest. Perhaps this is akin to what Marchionini and Shneiderman (1988) allude to in their writing on hypertext retrieval systems, claiming that when an information system presents results it engages in a process of structuring knowledge and resultantly, "*the systems themselves affect how users think when using them*" The evaluation of these visualisations may not demonstrate their value and a desire from users to stay with the status quo may present significant barriers to the development of the visual concept based search interface. However, and the important point made with regards to teaching IR is that their analysis serves the goal of bringing together and developing the student's knowledge of user search behaviour, the system retrieval techniques and their relations in developing an understanding and ability to reason about modern principles and practices of IR.

5. Assessment and Conclusions

The topics that may be covered in a course in IR are many, and their effective delivery requires the curriculum content to be analysed and organised

in some way to help the student come to an understanding of IR. In essence the educational goal in teaching the fundamentals of IR is to develop the students' ability to reason about IR, that is to have knowledge not only of the core techniques and technologies for search and retrieval but also to have gained insight into concepts and principles of IR that have lead to the generation of these systems. The analysis of the complementary aspects of IR, the user perspective of searching for information and the system perspective of search and retrieval, suggests that the latter when taught in the context of the former can help make the subject more comprehensible to students of IR. The user context of searching for information and using search technology is something with which the student is familiar and, as such, provides a motivational context in which to have explained the practices and principles of IR Over time and with the developments that have taken place in IR, specifically in search engine technology, the teaching of IR in contexts (as experienced in Manchester) has shifted in accordance with the developing student experience with and expectations of the technology. The context of 'learning to search (expertly)' guides the student in learning about the core concepts and techniques of the 'traditional' online retrieval system and, in the process, gives rise to considerable insight into the processes of search as a skilful and intellectual activity. It also provides a sense of the problem of 'semantic indeterminacy' to which Blair (2006) refers and which begins to explain the problem with the notion that information retrieval is some sort of matching process as might be suggested in the bibliographic description and vocabulary control employed in the traditional retrieval system. The modern retrieval system is however based on the model of retrieval of finding a degree of similarity between the query and document, and again it is valuable exercise to explore and analyse the relationship of the user's search and the system's processing of

information with the aim of finding out how the search engine works. This can effectively take place in a comparison of the Boolean and best match models for search technology and their provision for what might be referred to as the computational off-loading on behalf of the searcher. Students working on this as a piece of assessed coursework evidently found the debate based on Hahn's (1998) assertion a useful basis on which to explore the activity of search and to have elucidated the workings of the search engine. In addition when taught the controlled approach to searching, as in the search intermediary, students have declared that they feel more confident in their ability to search for information in transferring some of the conceptual processes of search and in taking a more controlled and strategic approach to search in the traditional and the modern environments. It may be interesting to attempt to evaluate this in some way, but in the meantime it is assumed that such feedback is as a result of the student learning more about the process of search, or how search and retrieval engines work, or is a piece of insight from learning a bit about both.

Students on courses in Computer Science may find there is more motivation to be had in learning 'how to build' a search engine and the learning opportunities are considerable as students work on the use and possible modifications of the components of search engine software. In using and developing programs for IR and to support real and familiar applications students can learn a lot about the fundamentals of computer science as well as the techniques and theoretical underpinnings of IR. Working with IR programs helps the student to develop a key sense of the statistical and linguistic properties of text which underpins so much of IR and leads to insight that, for example, clustering can take place based on lexical cooccurrence or that a purely algorithmic approach to stemming has its limits and why stemming might be useful to regularise terms in IR.

A final shift in the contexts is from ‘search’ and ‘build’ to ‘design’ to motivate learning in IR for students who may not have a Computer Science background but who may still expect to be involved in the design and development of these systems. The decisions involved in the design of search and browse features and functionality are non trivial and draw on extensive knowledge of user’s information searching behaviour and on working knowledge of search and retrieval techniques and technologies. Thus to inform the design students are encouraged to get involved in exercises to observe and attempt to identify search behaviours in order that they have a solid knowledge of the varied and often complex search activities which the system is designed to support. Whether this exercise is carried out in its own right or, as part of a preliminary investigation in a user centered approach to design, it provides the student with insight into searching tasks and behaviours that helps in making decisions regarding the search and browse features that are possible from the underlying retrieval model and information processing. Particular emphasis is given to the open ended exploratory searching behaviour in which the user of the search system interacts with the information in a far more holistic and iterative way than might otherwise be assumed in the common procedural perception of search. Focusing on what we might refer to as the concept dialogue provides the student of IR, working in the design context, with a basis on which to investigate and begin to appraise the design of the search interface and the potential for information visualisations in reducing the user effort in interpreting and understanding the information and thus facilitate their search. However, in recommending the ‘design context’ and specifically the analysis of the relationship of exploratory search and information visualisations, it is important to be reminded of the goal, which is not so much as to teach how to design a ‘search solution’, but rather to

motivate the student of IR to have explained the concepts and the techniques and to make IR comprehensible in investigating how IR works - and why. Contexts and the exercises that may be set to relate the principles and practices of IR to the activity of search brings together the broadening of IR necessary for the delivery of a coherent curriculum. For some time the aspects of IR, the user perspective of IS and HIB and the system perspectives of retrieval technology have gone about their business in harmony but never quite connecting. But things are changing and, as might be expected, research and development in IR is evolving, as are our teaching contexts, possibly towards shared perspectives and, at the least, towards a complementary understanding of the aspects of IR.

References

Bates, M. J. (1989). The design of browsing and berry-picking techniques for online search interface. *Online Review,* 13, 407-424.

Bawden, D., Vilar,P., Bates,J.,. Steinerov, J., and Vakkari, P. (2007). Information retrieval curricula; contexts and perspectives. In *Proceedings of the First International Workshop on Teaching and Learning of Information Retrieval* (TLIR 2007). London, 10 January, 2007 eWiC proceedings online Accessed 01 March 2010.

Belew, R.K. (2000). *Finding out about: A cognitive perspective on search engine technology and the WWW*. Cambridge: Cambridge University Press.

Blair, D. (2006). *Wittgenstien, language and information 'Back to the rough ground*!, Netherlands: Springer

Brattelle, J. (2005). *Search: How Google and its rivals rewrote the rules of business and transformed our culture.* London: Portfolio, p.15

Chau, M., Huang, Z., and Chen, H. (2003). Teaching key topics in computer science and information systems through a web search engine project. *ACM Journal of Educational Resources in Computing*, 3(3), 1-14

Chen, Hsinchun., Chau, Michael, Zeng, Daniel. (2002). CI Spider: a tool for competitive intelligence on the Web. *Decision Support Systems* 34(1), 1-17

Croft, W.B., Metzler, D. and Strohman, T. (2010). *Search Engines: Information Retrieval in Practice*. New Jersey: Pearsons

Griffiths, J., Johnson, F., and Hartley, R.(2007). User satisfaction as a measure of system performance, *Journal of Librarianship and Science* ,39(3), 142-152.

Jones, G. (2007). Using Research Questions to Encourage Creativity and Assess Understanding in Teaching of Information Retrieval In *Proceedings of the First International Workshop on Teaching and Learning of Information Retrieval* (TLIR 2007). London, 10 January, eWiC proceedings online Accessed 01 March 2010

Hahn, Trudi Bellardo (1998). Text retrieval online: Historical perspectives on Web search engines. *Bulletin of the American Society for Information Science,* 24*(4),* 7-10.

Harter, S.P. (1986). *Online information retrieval: Concepts, principles and techniques*. Academic Press, 1986. p ix)

Hearst, M.(1999). User interfaces and visualization. *In*: R. Baeza-Yates and B.Ribeiro-Neto, eds., *Modern Information Retrieval*. Harlow : Addison-Wesley, 257-323

Hearst, M.(2005). Teaching Applied Natural Language Processing: Triumphs and Tribulations, *ACL 2005 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, Ann Arbor, MI.

Hearst,,M.A. (2009) *Search User Interfaces* Cambridge: CUP

Kartoo [online] http://www.kartoo.com Accessed 20 December 2009

Grokker [online] http://www.grokker.com Accessed 20 December 2009

Kuhlthau, C.C (1988). Developing a Model of the Library Search Process: Cognitive and Affective Aspects', *Research Quarterly* 28(2), 232-242.

Korfhage, R.R. (1991). To see, or not to see - is That the query?. *Proceedings of the 14th Annual international ACM SIGIR Conference on Research and Development in information Retrieval*, Chicago, 13-16 October , 134-141

Levene. M.(2006). *An introduction to search engines and web navigation*. London: Addison Wesley

Light, G., Calkins, S., and Cox, R. (2009). *Learning and teaching in higher education: the reflective professional.* 2nd ed. London: Sage Publications Ltd

Markey, K. (2007). Twenty-five years of end-user searching, Part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, 58(8), 1123-1130

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM,* 49 (4), 41-46.

Marchionini, G. and Sheiderman, B.(1988). Finding facts vs. browsing knowledge in hypertext systems. *Computer*, 21 (1),70-80.

Navarro-Prieto, R., Scaife, M. and Rogers, Y. (1999). Cognitive Strategies in Web Searching. *Proceedings of the 5th Conference on Human Factors & the Web*, http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html. Accessed 01 March 2010

Ruthven, I., Elsweiler, D., and Nicol, E.(2008), Designing for users: an holistic approach to teaching information retrieval. In *Proceedings of the Second International Workshop on Teaching and Learning of Information Retrieval* (TLIR 2008). London, 18 October, eWiC proceedings online Accessed 01 March 2010

Saracevic, T. and Dalbello, M. (2001), A survey of digital library education: libraries in the digital age, *Proceedings of LIDA, Dubrovnik* 23-26 May

Salton. G. and McGill, M.J. (1983). *Introduction to modern information retrieval,* McGraw-Hill Education.

Savoy, J. and Picard, J.(2001). Retrieval effectiveness on the Web. *Information Processing and Management,* 37(4), 543--569, : p. 545)

Sparck Jones, K. and Willett, P. (Eds.). (1997). *Readings in Information Retrieval* . San Franciso, CA: Morgan Kaufmann,

Swanson, D. R. (1977). Information Retrieval as a Trial-and-Error Process. *Library Quarterly, 47*(2). (Reprinted in: *Key Papers in Information Science*. Ed. by Belver C. Griffith. New York: Knowledge Industry Publ., 1980, 328-348)

Witten, I.A.,, Gori, M. and Numerico, T *(2007). Web Dragons*. San Franciso:Morgan Kaufmann,