Teaching information retrieval (IR) as a philosophical problem

Clare Thornley

School of Information and Library Studies, University College Dublin, Belfield, Dublin 4, Ireland.
clare.thornley@ucd.ie

This paper discusses a method of teaching IR within the context of the philosophical problem of meaning to a mixed discipline undergraduate class. It introduces the distinction between content and context in meaning and the respective roles they play in the effective communication between an IR system and an IR user. It shows how teaching IR as a problem of meaning and interpretation can provide a conceptual framework in which students can effectively learn about both the social context in which IR operates and technical approaches to meaning representation.

Teaching methods, Philosophy, Information Retrieval, Curricula

1. THESIS

In this paper I argue that the fundamental concepts of IR concern some of the most profound philosophical problems about the nature of meaning. IR should be taught as a problem of meaning and technical solutions should be introduced within this context. Why? This enables students to gain skills in tackling complex intractable problems and to understand them from both a philosophical and an information science perspective. Thus this method of teaching develops problem analysis skills and deepens knowledge of the subject area.

2. MEANING AND IR

In my view the core topic in IR is the nature of the relationship between the meanings of documents (to include non-textual documents such as images and music) stored in an IR system and the meaning of the query and the broader information need of the user. Understanding the fundamentals of IR is about understanding how this relationship can result in the effective transfer of information to the user. This understanding of IR as primarily a problem of meaning and of communication between the user and the IR system, particularly in relationship to the work of the philosopher Wittgenstein (1953), has been discussed at length by Blair (1990, 2006). In terms of theories of meaning and teaching IR Belew's (2000) textbook combines instruction on the nature of meaning and interpretation with a technical grounding in IR.

The importance of meaning has been disputed by some, notably by Salton and McGill (1983) and, more recently, by Ingwersen (1992) and Ingwersen and Jarvelin (2005), who argue that the effective retrieval of information is not the equivalent to a match in meaning between query and document. I argue that IR requires both some communication of meaning and some transfer of information (Thornley and Gibb, forthcoming) and that these are related but not identical processes. Thus meaning is central to IR; however, to understand how meaning operates in IR one must also acknowledge its primary role as one of information transfer.

3. WHAT IS A PHILOSOPHICAL PROBLEM?

A philosophical problem can normally be recognised by the fact that it is concerned with fundamentals and not really possible to solve in a definite manner. Harre (2001), in his analysis of Wittgenstein, argues that philosophical problems are both perennial and intractable. This is normally because they are concerned with understanding complex and often irreconcilably contradictory aspects of human experience. Meaning is a philosophical problem in so far as the question of how we actually use words to communicate mental events and describe physical events/objects in the world is a complex issue which concerns how humans think and interact with each other and the environment.

The particular philosophical problem of meaning is concerned with what meaning is and how words can, in fact, be used to communicate. Within philosophy the debate on this matter has two main perspectives: the content approach and the context approach. The content perspective, as proposed by Frege (1892), argues that meaning derives from the relationship between words, objects in the world and the mental experience of understanding those words. The philosophical problem then becomes an exploration of how this relationship between physical word, physical object and mental event can work. The context, also known as the social practice approach to meaning, proposed by Wittgenstein (1953) and Putnam (1973), argues that words carry meaning primarily due to their relationship to the shared social and physical context or environment which language users share. Thus we can communicate effectively because we don't just have words (content) but we also have context which helps to sort out problems of ambiguity and different interpretations. The philosophical problem is then changed from an

investigation into the mental event/physical object relationship to an investigation into how the environment and context help us to 'make sense' of meaning.

4. WHY IS THIS IMPORTANT FOR IR?

Thus in philosophy there is a debate between the relative role of content and context in the effective communication of meaning. If IR is understood as a 'communication process' between an IR system and a user then this philosophical debate is pertinent to IR as it concerns itself with a very similar problem. What factors facilitate the user and the system 'communicating' effectively or using the same words to describe the same thing, and what factors inhibit this, causing a mismatch between user and system and leading to ineffective searching and irrelevant results?

In a similar way to philosophy, this is problem in IR which can never be totally resolved or solved. Despite recent developments in implicit feedback in IR (White, Jose and Ruthven, 2006), even if an IR system could somehow 'read the mind of a user' to help design searches the user often only has a impartial or partly developed idea of their exact information need. Thus it is fundamentally difficult to totally remove the ambiguous and unknown from the IR process.

5. WHY IS THIS EDUCATIONALLY IMPORTANT?

What are we trying to teach in IR, or how do we want our students to change during the course, and why? If we wish to engage our students we need to teach them just how important and fundamental a problem IR actually is. They need to learn the technical approaches to the problem of meaning representation and also the fundamental reasons why this problem matters and why it is hard to solve. This provides, not only background knowledge of the philosophy of meaning, but also skills in problem analysis and assessment which can then be transferred to many other areas. This enriches the teaching of IR and also makes it relevant to a wider range of students. The trend towards modularisation, in which more students taking IR courses will not necessarily have computer or information science backgrounds, presents a new challenge to teaching and requires imaginative solutions to make IR accessible and interesting to a range of students.

6. WHY BOTHER WITH PHILOSOPHICAL THEORY IN IR?

This question of the relevance of theory in IR research has been raised, as it appears to be the case that effective improvements in IR systems can be made without any particular reference to complex theories of meaning either in a philosophical or linguistic sense (Sparck-Jones, 2000, 2004). In education, however, we are not only trying to make improvements, incremental or otherwise, to particular IR systems, but trying to train minds to appreciate and understand the nature of the problems they are grappling with. It is not clear to us what technology will be available to our students by the time they are at the peak of their research careers. It is therefore not wise to limit their knowledge to the confines of current solutions. Also, even if our students do not go onto research careers in IR, it is in our interests as a discipline to ensure that our students have a proper understanding of the fundamentals and importance of IR which they may use to inform their future work in a wide variety of ways. Finally IR is an excellent way to introduce students to conceptual puzzles which have practical manifestations, e.g. meaning representation, relevance judgements. The skills gained in identifying and tackling these issues can be used in many contexts.

7. HOW THIS WORKS IN PRACTICE

I teach a 2nd year undergraduate course using this approach to IR. University College Dublin has a modularised system which means that the class are from the whole range of University subjects, some are information studies and computer science students, many, however, have no previous background in IR at all. I need to make IR accessible and interesting to students who, in many cases, have no plans to use the subject directly for their future research or career. I split the course into two main sections. The first six weeks covers the importance of IR in society and an introduction to the problem of meaning whilst the second six weeks covers different approaches in IR to addressing this problem in terms of providing meaningful access to recorded information.

The purpose of the first section of the course is to show how IR is a problem which has multiple levels or aspects, often with conflicting properties, which have to work together to provide information. Firstly I introduce a theoretical overview of the problem by covering a simple model of the different levels of meaning (lexical, syntax, semantic, pragmatic) and also work by Buckland (1991) on different aspects of information (information as thing, information as process, information as knowledge). Secondly I use examples and case studies (which are contributed by the students in many cases) to show how different participants at different times have different levels of access to these levels of meaning and aspects of information and the implications of this for successful 'communication' between the IR system and the IR user. I also use specific examples of how different factors, such as previous knowledge, prejudice, lack of context, could mean that different people develop different interpretations of the meaning of the same event or document. It proved useful to use a wide range of examples as I found some

students responded very well to visual examples and others could relate better to musical or textual documents. This discussion is then developed to show how storing and recording meaning in IR systems has particular implications for the potential of different interpretations due to the loss of context and the passage of time. During this part of the course I use philosophy when appropriate and always in way which explicitly relates to IR problems. In particular I avoid lengthy background information about the detailed views of philosophers.

In the second half of the course I introduce a number of different IR approaches to meaning representation within the framework of how they approach the problem of meaning. Due to the lack of technical background of the students this is at a fairly basic level but my objective is to give them some insight into how the IR systems they use so frequently actually work. I use the content/context debate from the philosophy of meaning to discuss the differences and relationship between more traditional IR approaches and more recent developments incorporating context, to at least some extent, such as Web 2.0. Finally, I spend one session explaining IR testing and used this to introduce the relationship, and also the tension, between reliability and validity in research.

8. COURSE EVALUATION

In general I find this approach to teaching IR is fairly effective in engaging a very diverse range of students. The course evaluation suggests that the use of case studies and specific examples, particularly when students have been asked to provide them from their own experience or from recent news stories, helps them to really understand why IR matters and how different factors can influence the interpretation of meaning. There was a mixed reaction to the issue of how one tackles problems that can't be totally resolved, perhaps because depending on their subject background, students will have different levels of comfort about this. I am considering re-designing the course as problem based learning course, (see for example <http://physics.dit.ie/programmes/pbl.html>) to increase the emphasis on problem investigation skills and would be very interested to hear from any readers who have experience of this.

Students were interested in the basics of how IR systems work and surprised by the historical background to much of current web technology. This may suggest that we should provide a least some history to our teaching of IR and show how previous work in information and computer science has contributed to current systems.

9. CONCLUSIONS

This paper suggests one way of teaching IR which introduces the subject within the context of the philosophical problem of meaning. The focus is on teaching IR but it is also used as tool to introduce some central problems in philosophy. The aim is for both subjects to inform each other and thus to both provide some conceptual framework for IR and some tangible examples to illustrate the problem of meaning. Teaching IR firstly as a problem of meaning and secondly as range of different technical approaches to address this problem is an effective approach for teaching IR to a mixed subject class

10. REFERENCES

- Belew, R.K. (2000) *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*. Cambridge University Press, Cambridge.
- Blair, D.C. (1990) *Language and Representation in Information Retrieval*. Elsevier Science Publishers, Amsterdam.
- Blair, D.C. (2006) *Wittgenstein, Language and Information: 'Back to the Rough Ground!'*. Springer, Dordrecht.
- Buckland, M.K. (1991) *Information and Information Systems*. Greenwood, New York.
- Frege, G. (1892) On sense and reference. *Zeitschrift für Philosophie und Philosophische Kritik*. **100**, 25-50. In Moore, A. W. ed. (1993) *Meaning and Reference*. Oxford, Oxford University Press.
- Harre, R. (2001) Wittgenstein: science and religion. *Philosophy*, **76** (296), 211-237.
- Ingwersen, P. (1992) *Information Retrieval Interaction*. Taylor Graham, London.
- Ingwersen, P. and Jarvelin, K. (2005) *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, Dordrecht.
- Salton, G. and McGill, M.J. (1983) *Introduction to Modern Information Retrieval*. McGraw-Hill International, Tokyo.
- Putnam, H. (1973) Meaning and reference. *The Journal of Philosophy*. **70** (19). In Moore, A.W. ed. (1993) *Meaning and Reference*. Oxford: Oxford University Press.

- Sparck Jones, K. (2000) Further reflections on TREC. *Information Processing and Management*, **36** (1), 37-85.
- Sparck Jones, K. (2004) IDF term weighting and IR research lessons. *Journal of Documentation*, **60** (5), 521-523.
- Thornley, C. and Gibb, F. forthcoming (accepted March 2008) Meaning in philosophy and meaning in information retrieval (IR). *Journal of Documentation*.
- White, R.W., Jose, J. M., and Ruthven, I. (2006), An implicit feedback approach for interactive information retrieval. *Information Processing and Management*. **42** (1), 166-190.
- Wittgenstein, L. (1953) *Philosophical Investigations*. Blackwells, Oxford.