MacFarlane, A. (2007). Pedagogic challenges in Information Retrieval – teaching mathematics to Postgraduate Information Science students. Paper presented at the 1st International Workshop on Teaching and Learning in Information Retrieval (TLIR 2007), 10-01-2007 - 10-01-2007, London, UK.

**CITY UNIVERSITY LONDON**
EST 1894

[City Research Online](City Research Online)

**Original citation**: MacFarlane, A. (2007). Pedagogic challenges in Information Retrieval – teaching mathematics to Postgraduate Information Science students. Paper presented at the 1st International Workshop on Teaching and Learning in Information Retrieval (TLIR 2007), 10-01-2007 - 10-01-2007, London, UK.

**Permanent City Research Online URL**: http://openaccess.city.ac.uk/4439/

**Copyright & reuse**

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

**Versions of research**

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

**Enquiries**

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

# Pedagogic challenges in Information Retrieval – teaching mathematics to Postgraduate Information Science students

Andrew MacFarlane

Department of information Science, Northampton Square, LONDON EC1V 0HB

*andym@soi.city.ac.uk*

**Understanding of mathematics is needed to underpin the process of search, either explicitly with Exact Match (Boolean logic, adjacency) or implicitly with Best match natural language search. In this paper I outline some pedagogical challenges in teaching mathematics for information retrieval to postgraduate information science students. The aim is to take these challenges either found by experience or in the literature, to identify both theoretical and practical ideas in order to improve the delivery of the material and positively affect the learning of the target audience. Some ideas are put forward to resolve these issues and to promote discussion.**

*Keywords: Information Retrieval, information science, teaching, Boolean logic.*

## 1. INTRODUCTION

The author teaches an Information Retrieval module which is delivered to a variety of postgraduate MSc students at City University, London. These courses include both Library and Information Science and Information Systems and Technology courses. The purpose of the module is to teach formal ideas and practical search methods to information scientists/managers who will act as *search intermediaries* between information users (such as lawyers, doctors, etc) and a given resource (in the case of the module, this means resources held on IT systems). Search intermediaries are needed, because many information users do not have the requisite search skills in order to specify a query that will obtain documents they require to fulfil their information need e.g. a lawyer who needs documents on case law for a particular client. Various mathematical skills are needed for this role such as knowledge of Boolean logic used for Exact Match search. In this paper I outline some pedagogical challenges in teaching mathematics to information science students, and propose some methods to resolve them. The paper is structured as follows. The characteristics of the student body are described in section 2. The problem of teaching mathematics in higher education as applied to information science students are described in section 3, outlining some ideas to resolve these pedagogical challenges in section 4. I outline a way forward in the conclusion.

## 2. STUDENT CHARACTERISTICS

Given the categories of students' characteristics described in D'Andrea (1999); the attributes of the students who take the authors course are outlined:

- Many (but not all) have one years experience in the information profession (either as a search intermediary or as a librarian). [Knowledge on entry/Personal characteristics]
- A first degree in a subject other than information science or information studies/management. [Demographic information]
- A wide variety of learning styles: some are *deep learners* who expect to work independently, while others are only prepared to do the minimum possible in order to pass the course and are therefore *surface learners*. [Learning style]

There are therefore a wide variety of students who have vastly different levels of experience and expectations from the course. Some students may become search intermediaries when they leave the University and find work, others may become librarians and the skills gained on the course may only be intermittently used. More importantly, the student body has a variety of mathematical skills on entry to the course.

## 3.  THE ISSUE OF MATHEMATICS IN HIGHER EDUCATION

In this section the issue of teaching mathematics in higher education is examined and the impact of various issues on our information retrieval students, given their particular characteristics. In order to do this the following themes are addressed: the effect of mathematics teaching at school, the role of the institution in teaching mathematics, the issue of who should teach mathematics to information retrieval students, the attitude of the students and the effect this has on their knowledge, useful approaches to teaching mathematics and finally specific issues in teaching mathematics to information retrieval students.

### 3.1 Effect of mathematics teaching a school

There is clear evidence that there is a decline in mathematical skills in students entering University: Croft (2002: p151) states that the performance in university entry tests of students with grade N A-level maths taken in 1991 is equivalent to that of students who obtain grade C today. This is a worrying trend, and while it does not effect the author directly (he does not teach undergraduates) it will have a considerable knock on effect, as students who have less skills in maths, gradually filter through the higher education system to postgraduate level. A further problem is that many of the students will not have studied maths since doing their GCSE and many subjects such as calculus are no longer taught at that level (Appleby and Cox, 2002: p6). Because of the gap between the mathematical skills of students and the requirements placed upon them is growing wider, there is an increasing need to take steps to address the issue. Another important aspect is one of attitude – this gap alluded to above causes real fear the student body and they may develop avoidance strategies (Appleby and Cox, 2002). Many of the students have completed Arts or Humanities degrees and are not comfortable with mathematics – Croft (2002: p145) calls this type of students the 'maths anxious'. The effect of the transition of the teaching of mathematics at school is that reliance cannot be placed on certain subjects having been taught to the students, and/or at the level required.

### 3.2 Institutional factors

The increasing lack of mathematical skills is not just a problem for the author and their department; it is a problem for the whole institution. How much support can a teacher depend on from the University? There is conflicting evidence as to the usefulness of mathematical support centres. Croft (2002: p155) points out that there is a danger that departments and schools will rely on these support centres and not develop their course material. This may fail to address the problems students have as resources for such support centres are limited – the wrong strategy chosen by the institution could lead to these support centres being overwhelmed. However it has been shown that such centres are useful (Lawson, 2003): students at Coventry University are very happy with a centre providing drop in support for maths problems and use it heavily. The author believes that such centres are useful, but should not be overused. The question to answer here is when should the department offer the support needed and when should the services of a University mathematical support centre be called upon – this question will be dealt with in the next section.

### 3.3 Who should teach mathematics to IR students?

Croft (2002: p147) poses the question of who should teach mathematics to students e.g. when it is appropriate for either a mathematics support centre or mathematics department to teach maths and when is it appropriate to be done in house? Croft (2002: p148) outlines the problems with both strategies. If a mathematics department teaches information science students, they will not have the same background in IR as the author, and will therefore not be able to give the students' context. Mathematics lecturers may not understand the often negative feelings the students have for mathematics and that they are not mathematics students. These lecturers may feel that they have been dumped in a support role and are taken away from the advanced mathematics teaching they would like to do. However as the author does not teach mathematics full time, he is unaware of the precise details of mathematics teaching at school. Because of the lag between students leaving school and taking the authors courses, it may be difficult for us to develop strategies to deal with students problems over the course of time. There is a danger of a turf war developing between my department and the mathematics department over who should do this kind of teaching. Croft (2002: p149) argues that mathematics as a discipline is unique within each subject and for the most part the author agrees with this assessment. The teaching of maths in IR is very context driven; for example the author teaches the student body set theory within the context of searching, how to form search sets and manipulate them with various strategies. The author does not feel the need to call upon the services of the mathematics department to support his teaching; however a mathematics support centre provided by the university could be useful in some circumstances (see below).

## 3.4 Students attitude and knowledge

It is important to consider what the effect the students attitude and characteristics (specified in section 2) has on the knowledge they bring with them and what they are required to do on the information retrieval course. One particular problem is that many of the students may adopt strategies that try to avoid genuine engagement with the mathematical material provided to them. There is a real tension here between supporting students and encouraging independent learning in them (Appleby and Cox, 2002: p15). The student body who are all at postgraduate level, are particularly encouraged to be independent. One way of tackling this problem is to encourage the use of mentors in the student body. The mentor is someone who is comfortable with mathematics – this kind of student is admitted to our courses. This strategy is useful in a couple of ways. Firstly it gets students communicating with each other, and provides peer support. Secondly it allows students to discuss their problems in a non-threatening environment (providing the mentor is sympathetic of course) with fellow students who are not involved in assessing them. It is hoped that this would encourage independent learning on both the part of the mentor and the mentored. In order to match the knowledge of the student body with what they will need for an IR course and their future career it is important to consider the learning in mathematics required or my students. Table 1 shows the 'Mathematical Assessment Task Hierarchy' or MATH taxonomy, defined by Smith et al (1996) and based on Bloom's Taxonomy.

| Group A | Group B | Group C |
|---|---|---|
| Factual Knowledge<br>Comprehension<br>Routine use of procedures | Information transfer<br>Application in new<br>situations | Justifying & interpreting<br>Implications, conjectures<br>and comparisons<br>Evaluation |

**TABLE 1 – THE MATHEMATICAL ASSESSMENT TASK HIERARCHY  (MATH) TAXONOMY**

Each of the groups in table one is a building block (Croft, 2002: p144) e.g. Group B depends on knowledge gained in Group A, which in turn depends on Group C. The student body will need at least Group B knowledge and will certainly need most of Group C (advanced understanding of conjectures and theorems are not necessary). The problem is that students will often not have the requisite skills in Group A. This issue will be tackled in section 4.

## 3.5 Approaches to teaching mathematics

In section 3.4 the building block approach advocated by Croft (2002) was mentioned. Rather than expecting the students to have Group C knowledge in the MATH taxonomy, it must be accepted that some remedial material needs to be delivered at the Group A and B levels. Table 2 shows the building blocks for mathematics required for information retrieval:

| Group A | Group B | Group C |
|---|---|---|
| Numeracy.<br>Set theory.<br>Transformation rules e.g.<br>commutativity, associativity.<br>Statistics and probability. | Forming Boolean queries<br>analysed from a users<br>information need. | Search strategies  (different<br>uses of  Boolean and<br>Adjacency  operators and<br> terms). Evaluation of<br>results. |

**TABLE 2 – BUILDING BLOCKS FOR MATHEMATICS REQUIRED FOR INFORMATION RETRIEVAL**

The mathematics required for information retrieval can be divided into three broad areas: numeracy, discrete mathematics and probability & statistics. Numeracy is helpful in building the other two areas. It is important for the sake of those who require it, that confidence is built up on numeracy material first. Even then, the student body may resist for reasons given above. But it is important that the underlying theories and axioms are delivered (Ahmed et al, 2002), rather than just the procedures – these skills are required as a search intermediary. The use of additional and unassessed modules to help students could be considered (Appleby and Cox, 2002), but is this a realistic option in an already tight curriculum? And will students attend these extra courses? This is very unlikely – any mathematical training required should therefore be done within the framework of the information retrieval course.

What is the best way to deliver mathematical material to the student body? Ahmed et al (2002: p38) outline a choice of two models for delivering maths to learners. The first of these is a *transmitting model*: facts and ideas are transmitted to the student, who is just a passive recipient. Ahmed at al (2002) assert that this type of maths teaching is inadequate and has damaged learning in higher education. As the students are postgraduates, this

model is completely inappropriate – they must be actively engaged with the material presented to them. The traditional didactic approach of teaching is not relevant for postgraduate information retrieval courses. The other model Ahmed et al (2002) describes is the *transaction model*. In this model the student is involved in solving real problems and encourages active learning with the material. This material may sometimes be basic, for example with numeracy problems, but it is still possible to encourage active learning: the mentor system described above could be helpful.

## 3.6 Specific issues in teaching mathematics for information retrieval

In the previous section the three broad areas of mathematics that are of concern were outlined. The first, and most fundamental, is numeracy. As stated above many of the student body have completed arts or humanities degrees and have been avoiding mathematics since leaving school. Duffin (2002: p132) describes the shock for many students who are confronted with their numeracy problems on finishing their undergraduate degree. The author has had experience of students who have been very upset over mathematical material delivered to them in the lecture room. Some institutional support may be useful here (see above), but it is still possible to the students some material that is within the context of information retrieval. Duffin (2002: p133) points out that numeracy teaching, even at a basic level, must be targeted at a University trained mind.

Things are worse when the issue of discrete mathematics is considered. Burn (2002: p32-33) points out that students who have had no prior experience of thinking abstractly will find problems with this type of mathematics. Just delivering set theory without some context will not work with the student body. It is best to take a number of specific examples, show how these examples work in practice, and them move to the general case. With this understanding they will be able to think more clearly. The author provides a number of case studies useful for this purpose.

Discrete mathematics is not the most problematic of all the mathematical ideas applied to information retrieval. Many of the theories in information retrieval and evaluation methods require the knowledge of probability and statistics. Some of these ideas can be very difficult to master. So while there is a requirement to deliver this material to postgraduate information retrieval students, there is a real worry that students will not want to actively engage with the material. The author does deliver this material, but only scratches the surface and teaches it at a very simplified level. Davies (2002) suggests that the use of real data in order to deliver statistics is a good strategy: the author does this with some simple example of how term weighting works.

## 4.   SOME IDEAS FOR RESOLVING PEDAGOGICAL CHALLENGES

In section 3.6 three separate areas of concern for teaching mathematics in information retrieval were identified namely, numeracy, discrete mathematics and probability/statistics. In section 3.4 the problem of the tension between support and encouraging independent learning was outlined, particularly given that the student body (all postgraduate) are encouraged to be independent (Appleby and Cox, 2002). In this section these particular problems are tackled by looking at the use of diagnostic tests, delivery of material in mathematics and the role of assessment.

## 4.1 Diagnostic tests

There is a significant role for the use of diagnostic tests on entry to gather information on the current knowledge of the student (Croft, 2002: p150). Other methods such as entry-level qualifications and pre-university syllabuses will not provide the up to date information needed (Appleby and Cox, 2002: p11). It should be noted that diagnostic tests are not full proof e.g. the student might under perform on the day. There is also the danger that poor results in these tests may undermine the students' confidence further. However with the correct strategy, diagnostic tests are a useful tool. The plan would be to devise questions for numeracy; discrete maths and statistics at each of the levels described in table 1 and 2 above. It would be important to encourage the students to be honest about their shortcomings, pointing out the tests are not formally assessed and no marks will be assigned. The aim is to identify students' weaknesses and therefore to identify a further set of tests for them to complete in their own time (with a caveat on non-assessed tests mentioned above). These tests could be on-line tests (Beevers and Patterson, 2002: p53) the student would be allowed to complete each test as many times as they wish. The test should contain concrete examples in information retrieval, so that the student can more easily see the purpose of them. This information could be used to set up preliminary tutorials for those who need it or even point the student to university support services in severe cases. A further refinement of this method is to do another test at the end of the course (this could be formally assessed or not) to see if anything had been learnt (Duffin, 2002: p133). This would be a useful reflective mechanism for both the student and the author: the information could be used to improve the delivery of diagnostic tests in future years.

## 4.2 Delivery of material

The author wishes to develop active learners using the transaction model briefly described in section 3.5. The delivery of mathematics for information retrieval is crucial for developing these skills. Ahmed et al (2002: p40) describe three key aspects of developing active learners when delivering such material. The first of these is the mechanisms for mathematics e.g. manipulation of sets using operators such as AND. The next key aspect is the communication between the student and the lecturer. The last key aspect is the student working on his or her own without interference from a lecturer. Each of these aspects is built on each other in the order stated. The diagnostic tests could be used to develop the first key aspect. Lectures, seminars and tutorials could be used to develop the second and hence encourage the third (Ahmed at al, 2002: p43-44). With lectures the author utilises a number of strategies. The use of group work in conjunction with lectures to work on material delivered has proven to be very valuable – this forces the student to engage with the material. For example, each group is given an information need: tutorial tasks require them complete facet analyses of this need and to write a Boolean query from this analysis. A tutorial task is associated with all the lectures. The results of these tasks are then discussed in seminars, with material posted on an e-learning system beforehand together with oral presentations in the seminars. The author tries to ensure that there is a mentor in each group e.g. a student with prior experience as a search intermediary. Another method used to give the lecture material in advance and encourage the student to pursue some topics of research. Setting a series of exercises at the end of the lecture notes, which encourages them to go to the library or search the web, is a useful mechanism for doing this. Another useful strategy is to encourage the students to create a mind map, which they can build on during the series of lectures.

## 4.3 Assessment

Beevers and Patterson (2002: p50) outline the types of assessment available to teachers of mathematics in higher education. These include diagnostic tests, self-tests, formative, summative and continuous. Most of these types of assessment have been discussed in section 4.1 apart from summative assessments – the paper therefore concentrates on this method. It should be noted that one way to make sure that students engage with the material and do not avoid it, is to set an assessment – in fact it is the only sure way to do it! Beevers and Patterson (2002: p55-57) outline some methods of assessments which can be used: examinations (open and closed book), multiple choice questions (MCQ's), reports, project work, presentation and peer assessment. Many of these can be used in conjunction with each other.

Firstly the use of examinations can be considered, both open and closed book. Beevers and Patterson (2002: p55) assert that closed book examinations are easy to set up and police but because of the time restriction only test lower level cognitive skills. With open book examinations however it is possible to test higher level cognitive skills as the student has all the mathematical material available to them – the difficulty here is assessing what the student knows, what are they allowed to take to the examination? A two-fold strategy could be used. The closed book examination can be used to tackle problems such as numeracy, the fairly low level cognitive skills (Group A in tables 1 and 2), while higher level cognitive skills in discrete mathematics and statistics can be assessed by open book examinations (Group B, C in tables 1 and 2). Students would be allowed to take a sheet of formulas and rules into these closed book examinations – the use of lecture notes would not be allowed, the author wants to force them to engage with the material. The sheet should be constructed such that the student must engage with as much of the material as possible, otherwise they will only study the material necessary for the exam.

The author would not use MCQ's to assess mathematical skills as they do not believe that this type of assessment is useful for their purposes (particularly at levels B and C – see table 2). The students need to manipulate the mathematics to solve real problems, and the author believes that assessment should contain examples that force them to use the ideas. Consider the following scenario: the student is given a real world information need and asked multiple-choice questions. A 'correct' answer cannot be given, as there are many approaches that could be successful. A counter argument is that you could use a single correct answer and derive some poor or irrelevant strategies to choose from. This would still not help in formally assessing students' knowledge – it is desirable for the students to derive their own answer and show the thinking behind it. However, the author does use MCQ's on another module which introduces IR concepts, using fact based answers on issues such as the structure of inverted files, correctly formed queries etc.

Finally I consider the use of reports, presentations and peer assessment. A useful way to use this type of assessment is to set a real problem and ask the students to prepare a report, which will be presented to the class. This task can be done either individually or in groups: the author uses the group method. Peer assessment can be used in a variety of ways. With presentations each student can be given a mark sheet with which to assess the performance of their peers; in group work each student could comment on the performance of their peers on the task given in such areas as gathering data, report writing, chairing, editing and even presentation. One needs to be

careful that students do not either cosy up to each other or launch a vendetta against one of their colleagues. Reports from peers must be vetted very carefully. An example of what could be done here is given them some data to analyse statistically (results from search engines for example). They would then produce a report on this analysis (do an evaluation of the search engines) and present the results to the rest of the group.

## 5. CONCLUSION

Mathematics for information retrieval is an important area of teaching to be addressed. The author has been developing a tutorial task method to assist in the teaching of mathematics to both Library and Information science and Information Systems and Technology students for a number of years now. A number of ideas have been identified that I would like to discuss in the workshop, as well as sharing experiences with colleagues from other Universities. It is becoming increasingly important that the issues addressed in this paper are tackled as changes in all levels of education are going to impact on the students we, as academics, will be teaching. We as a community will need to continue to develop both materials and strategies by looking at the evidence as well as keeping up to date on the literature of mathematics in higher education.

## REFERENCES.

Ahmed, A., Holton, D. and Williams. H. (2002). "Developing active learners". In: Kahn, P and Kyle, J., (eds) Effective *learning and teaching in mathematics and its applications*, Kogan Page, 34-48.

Appleby, J. and Cox, W. (2002). "The transition to higher education". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 3-19.

Beevers, C. and Paterson, J. (2002). "Assessment in mathematics". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 49-61.

Burn, R. (2002). "The genesis of mathematical structures". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 20-33.

Croft, T. (2002). "Mathematics: the teaching , learning and support of non-specialists". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 144-157.

D'Andrea, V. (1999). "Organising teaching and learning: outcomes-based planning". In: Fry, H., Ketteridge, S., and Marshall, S., (eds), *A handbook for teaching and learning in higher education*, Kogan Page, 41-57.

Davies, N. (2002). "Ideas for improving learning and teaching of statistics". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 175-193.

Duffin, J. (2002). "Numeracy in higher education". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 131-143.

Kahn, P and Kyle, J. (2002). "Making learning and teaching more effective". In: Kahn, P and Kyle, J., (eds) *Effective learning and teaching in mathematics and its applications*, Kogan Page, 204-211.

Lawson, D., (2003). "Need a little help figuring it out?", The Times Higher Education Supplement, 6th June, No. 1,592. 23.

Smith G, Wood, L, Crawford, K., Coupland, M., Ball, G., and Stephenson, B. (1996). Constructing mathematical examinations to assess a range of knowledge and skills, International Journal of Mathematical Education in Science and Technology, 30, 47-63.

## Web References

Coventry mathematics support centre – http://www.mis.coventry.ac.uk/maths_centre [visited: 30th September 2006].