JOINT BOARD OF MODERATORS

GOOD PRACTICE GUIDE – INFORMATION MINING

Introduction

There is increasing concerns amongst a number of academics of the ability of students to source quality information especially when they are undertaking research into, what is to them, new areas of knowledge. This paper, prepared in consultation with the JBM, provides advice on good practice.

Background

Civil engineering knowledge is based partly on the ability to search out and interpret information, often from multiple sources. Most professionals utilise a hierarchy of written information, which may be loosely ordered in terms of reliability as codes of practice, refereed journal papers, technical reports, conference papers, un-refereed papers, newspaper articles, and the unrefereed internet. Engineers need to know how to extract high quality information properly from the hierarchy, and then how to assess the information extracted. Universities have an important part to play in training students in information handling. The rapid development of e-resources at universities, the use of the internet from childhood, and proliferation of means of sharing information through web 2.0 and other developments in IT have rendered past assumptions about good library provision, and user training redundant. The following questions are relevant.

(1) How to find information? Students need to be aware of the hierarchy of information sources, which is technology independent. An understanding is required of the best methods for data mining (e.g. the choice between search engines and bibliographic databases, and how to make best use of both)?

(2) Is the extracted information correct? What is its provenance and accuracy? This may involve finding out whether the information is received wisdom, is in common practice, and whether it has been independently confirmed. Students should develop a proactive approach towards checking the reliability of information. Concepts should be understood, and then critically appraised. Equations may have to be checked (even by simple dimensional considerations), and re-derived if necessary. Published data should be examined with regard to inherent error (e.g. measurement, round-off, etc.), uncertainty, typographic error, incorrect English, and unethical procedures (e.g. falsification of observations). Information may be incorrect even in documents at the topmost layer.

(3) Why is this important? On graduation, engineers have a professional responsibility to ensure they understand up-to-date information relevant to their work. All students should be made aware of the societal, legal, financial, and reputation consequences of engineering failure. There is no legal defence for ignorance.

(4) How are individual academic institutions addressing the problems associated with universal online access to information? Reports published by RIN/JISC e.g. JISC The Digital information seeker, (2010); Proctor, R. and others Adoption and use of web 2.0 in scholarly communications Phil Trans Royal Society, A (2010), 4039-4056; RIN:

Researchers use of academic libraries and their services (2007); highlight the challenges libraries face in developing good practice in information seeking. Material not immediately available via Google is likely to be ignored; traditional catalogues and databases are by-passed, and engineers are often slower than other disciplines in adopting more sophisticated tools. While users may discriminate between academic papers and general web chit chat, their starting place may not be apposite for retrieval of the full hierarchy of material. It is important that there is evidence of some understanding between the academic department and the information professionals in the library/institution knowledge store on how to educate the post Google generation

(5) How could this be taught? Case studies are useful. While attempts are made to identify plagiarism, what is done to highlight good practice in terms of using the hierarchy of knowledge? One idea would be to set students a design exercise, during which they would undertake calculations using information from a mock publication (containing a few embedded errors – typographic, conceptual, etc.), where the result leads to catastrophic failure. The lecturer could then hold a class to consider the lessons learnt. The exercise could be set nationally by an experienced practitioner through ICE in conjunction one or two representatives from universities. Such an exercise could be very attractive to students, and be a means of engaging them directly with ICE.

(6) Examining? Marks awarded for essays and project reports should include a component related to how well information sources have been mined, with the expectation that the majority of sources come from the upper levels of the information hierarchy. In other words, reference lists should not be dominated by links to websites but primarily cite journal papers, codes of practice, and textbooks, although these may have been delivered electronically.

By ensuring buy-in at an early stage and getting students used to retrieving information from reputable sources, this culture would carry on once they become members of one of the JBM member Institutions.

Universities are encouraged to include ‘information mining, interpretation, and critical understanding’ in their courses.

Alistair Borthwick
Mike Chrimes
Barry Clarke
Leon Heward-Mills
[December 2010]