INFORMATION RETRIEVAL FOR THE MANAGEMENT OF CIVIL ENGINEERING DESIGN CONTENT

Peter Demian, Lecturer in Construction Management, Loughborough University

Keynote Delivered at ICSSE, February 2011, KSR College of Engineering, India

As more and more civil engineering work becomes digital, the amount of information that engineers need to handle is growing exponentially. While this presents opportunities for reusing designs and increasing productivity, there remains the challenging of navigating around huge repositories of digital content. This presentation outlines a research line spanning about one decade in Information Retrieval to manage civil engineering design content, such as drawings and building models. The field of Information Retrieval is concerned with systems that help users to fulfil their information needs. Common examples of Information Retrieval systems include web search engines and library catalogues. The basic aim of the research presented here can therefore be expressed informally as "to develop search engines for civil engineering design content". However, in the case of civil engineering applications of Information Retrieval, *understanding* retrieved information is probably more informant than *finding* the information in the first place. The emergence of Building Information Models in recent years adds some urgency to this line of research.

Problem: Measuring Relevance for Content Reuse

By way of a problem statement, the knowledge life cycle first articulated in the CoMem project (Fruchter and Demian 2002) can be used here. Knowledge is created as designers collaborate on design projects. It is captured, indexed, and stored in an archive. At a later time, it is retrieved from the archive and reused. Finally, as knowledge is reused, it is refined and becomes more valuable, leading to more knowledge creation. The efforts presented here focus on the knowledge reuse phase and build on previous work that addresses knowledge creation, capture, indexing and archiving.

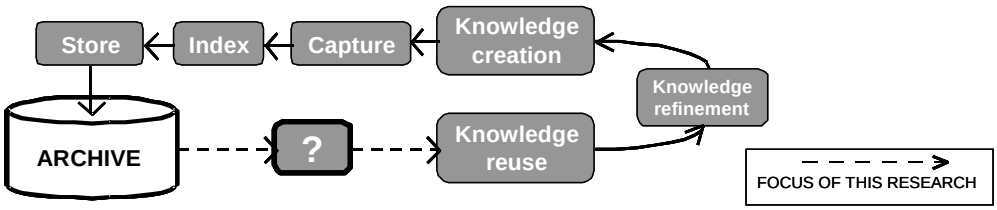

Figure 1: The knowledge life cycle. Knowledge is created, captured, indexed, and stored in an archive. At a later time, it is retrieved from the archive and reused. As it is reused, it becomes refined. This research focuses on the knowledge reuse phase.

There are two possible modes of interaction with a repository of content (Baeza-Yates and Ribeieo-Neto 1999). The first is the *retrieval approach*, in which the user has a precise information need and he/she translates it into a query. The system takes this query as its input and returns a set of (ranked) items as its output. The second mode is the *exploration approach* in which the user explores the repository with only a vague information need. In this mode, the *process* of exploration is often equally as important as the items that are eventually retrieved in satisfying the user's information need. This is because that process of exploration contributes to the *understanding* of the content that is found. For both modes of interaction, it is necessary to quantify the relevance of the items in the repository, either to retrieve the most relevant items, or to guide the user's exploration.

Measuring Relevance for Exploration

In the original CoMem project (Demian and Fruchter 2005), the relevance measure was used to support design reuse from archives of building models created during previous construction projects. The envisaged scenario was that a designer working on a CAD model could (at the press of a button) explore an archive of previous projects to reuse any content that was relevant to the designer's current design task. The relevance of previous content would be indicated by the system, and hence needed to be measured. At the first attempt, standard information retrieval techniques were applied: the text vector model and its refinement latent semantic indexing. Next, an attempt was made to improve the performance of the relevance measure by considering contextual building model elements. When comparing any two building components, it was attempted to improve retrieval performance by considering not only the two components being compared but also components related or linked to them. From this original CoMem study, It was concluded that good retrieval results could be achieved even with the sparse texts in building models. The more complex relevance computations (e.g. latent semantic indexing and tree matching methods) did not perform significantly better than the more simple ones (i.e. the classic vector model).

Retrieval – The Importance of Context and Granularity in Search Result Interfaces

In a more recent project (Demian and Balatsoukas 2011, Balatsoukas and Demian 2010), CoMem was extended to the *retrieval* (rather than *exploration*) mode of interaction. A version of CoMem was developed which allows the user to submit text-based queries. Particular attention was paid to the *search result interface* of such retrieval systems for civil engineers. It was noted that information about the design and construction of buildings can be structured in a particular way. This is especially correct given the increasing complexity of building product models and

Building Information Models. In addition, engineers usually have distinct information needs, particularly the need to *understand* retrieved information. Both these factors (the nature of engineering content and the information needs of engineers) make general retrieval techniques for computing relevance and visualizing search results less applicable in civil engineering Information Retrieval systems. It t transpired from that recent study that *granularity* is a fundamental concept that needs to be considered when measuring relevance and visualizing search results in information retrieval systems for repositories of civil engineering design content. A prototype system, called CoMem-XML, was developed and evaluated in terms of the time needed for users to find relevant information, the accuracy of their relevance judgment and their subjective satisfaction with the prototype. A user study was conducted where test subjects were asked to complete tasks using various forms of the prototype, to complete a satisfaction questionnaire, and to be interviewed. It was found that users performed better and were more satisfied when the search result interface of the CoMem-XML system presented only relevant information in context. On the other hand, interfaces that present the retrieved information out of context (i.e. without highlighting its position in the parts hierarchy) were less effective for participants to judge relevance.

Concluding remarks

Advances in information technology allow us to store and access more and more information. With vast volumes of information becoming available, the problem of "information overload" is increasingly recognized and documented. Repositories of civil engineering content have also grown in line with this general trend. In construction, the sheer quantity and heterogeneity of content in engineering applications has been recognized by many researchers. The need to apply and adapt Information Retrieval in civil engineering will only increase for the foreseeable future.

Acknowledgements

The work in the USA was funded by Stanford University's UPS Endowment and the Center for Integrated Facility Engineering. The work in the UK was funded by the Leverhulme Trust through a Fellowship to the author. I was privileged to work with gifted individuals: Dr Renate Fruchter and Dr Panos Balatsoukas.

References

Baeza-Yates R. and Ribeiro-Neto B., 1999. Modern Information Retrieval, Addison-Wesley, Harlow, UK.

Balatsoukas P and P Demian, 2010. “Effects of granularity of search results on the relevance judgment behavior of engineers: Building systems for retrieval and understanding of context.” Journal of the American Society for Information Science and Technology, 61(3), 453-467.

Demian, P. and R Fruchter, 2005. “Measuring Relevance in Support of Design Reuse from Archives of Building Product Models.” ASCE Journal of Computing in Civil Engineering, 19(2), 119-136.

Demian P and P Balatsoukas, accepted for publication 2011. “Information Retrieval from Civil Engineering Repositories: the Importance of Context and Granularity.” Submitted to ASCE Journal of Computing in Civil Engineering.

Fruchter, R. and P Demian, 2002. “CoMem: Designing an Interaction Experience for Reuse of Rich Contextual Knowledge from a Corporate Memory.” Artificial Intelligence in Engineering Design, Analysis and Manufacturing. AIEDAM Special Issue: Human-computer Interaction in Engineering Contexts, 16(3), 127-147.