InfoMem – A Hierarchical Search Engine for the Enterprise

Dr. Renate Fruchter,
Director, Project Based Learning Lab,
Stanford University.
fruchter@stanford.edu

Siddharth Jonathan J.B.,
Department of Computer Science,
Stanford University.
siddharth.jonathan@cs.stanford.edu

Abstract

With the spiraling accumulation of information in enterprises, it is becoming increasingly difficult to leverage effectively, the information implicit in data that is available from the past to make informed decisions for the present. Most Architecture/Engineering/Construction Corporations for example, have a wealth of information related to past projects which can be effectively put to use for decision support and guiding the design of new projects. InfoMem seeks to solve some of these problems by providing a system that lets users assimilate and reuse the information present in such databases. InfoMem has a couple of features that differentiate it from a standard information retrieval (IR) system. Most notably, it leverages the hierarchical relationships that exist in data stored in databases. Such containment relationships in the form of referential integrity constraints and joins, once identified, can be used to improve precision and recall in information retrieval. In addition, InfoMem also addresses the search results visualization problem. In the context of an IR system for hierarchical information, visualization of the results is a challenging UI problem in itself. Currently, InfoMem uses the Squarified TreeMap algorithm [Bederson, B.B., Shneiderman, B., and Wattenberg, M] for search results visualization. We are also working on a few extensions to the TreeMap algorithm which we mention in the section on future work. For such a system to be useful in a practical setting where there is a constant influx of new data and edits or deletions of existing data, it is also imperative to remain up to date. InfoMem is architected taking into account such considerations. As a test bed, we used data from ThinkTank [http://pbl-ThinkTank.stanford.edu:8090/ThinkTank/Login.jsp], an online discussion forum for a project oriented course at Stanford University. We studied InfoMem along two dimensions. Firstly from an information retrieval perspective, we evaluated precision and recall. Secondly, from a user experience perspective, we investigate the value add that InfoMem provides to users in terms of not only reducing their time spent to satisfy their information need but also in helping them discover new information when they are browsing and not necessarily searching with a specific target in mind. We also develop some theory to understand how familiarity impacts information retrieval. There are two types of familiarity, namely familiarity with the problem domain and familiarity with the data in the database. We study how each of these variables interact with the value added by InfoMem by observing a typical user in action. We end by mentioning a few exciting extensions to InfoMem in terms of user personalization, usage of community feedback and data mining.

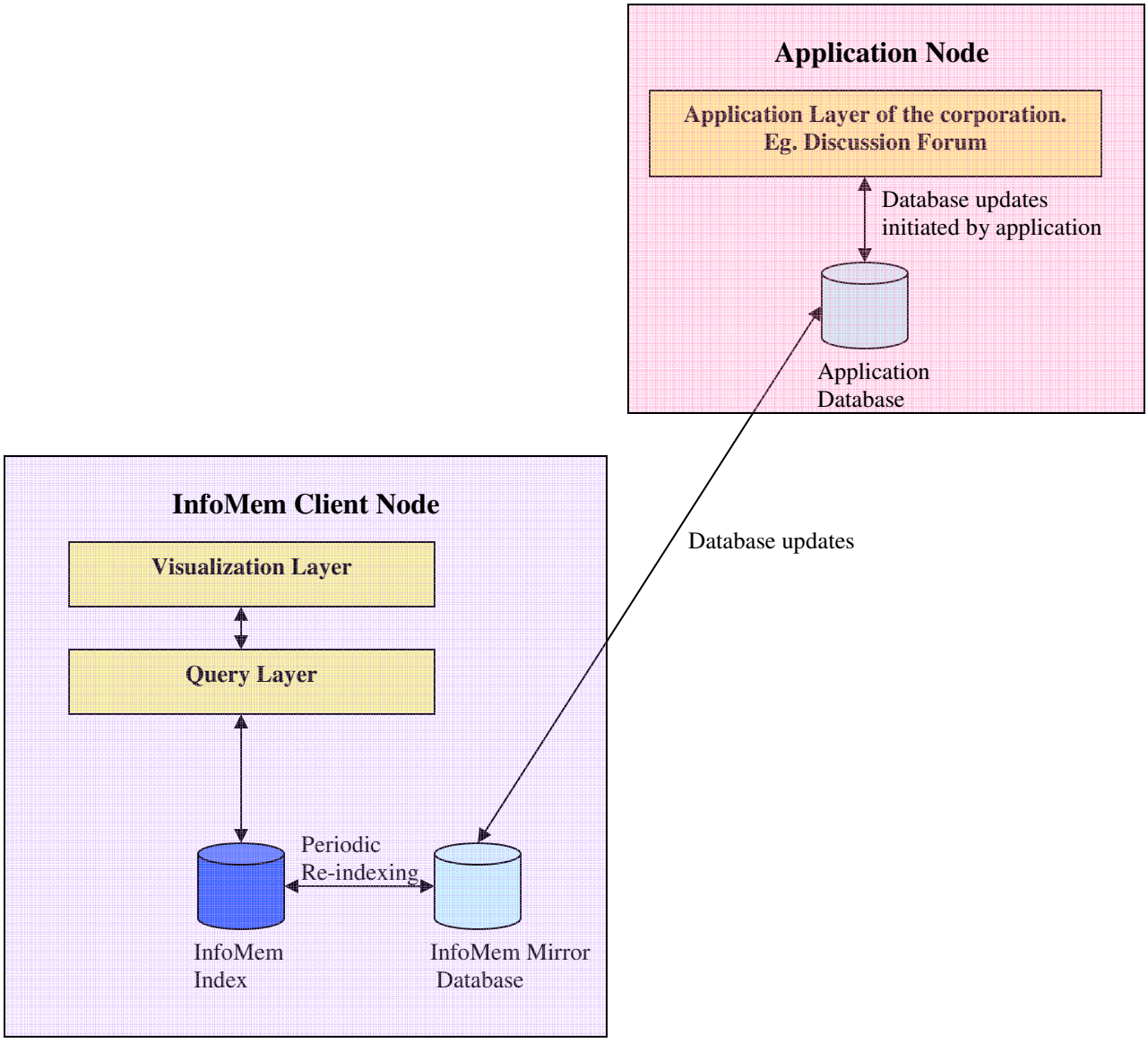

Figure 1: The InfoMem System Architecture

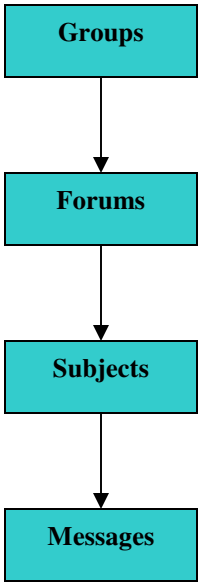

Figure 2: The ThinkTank Hierarchy

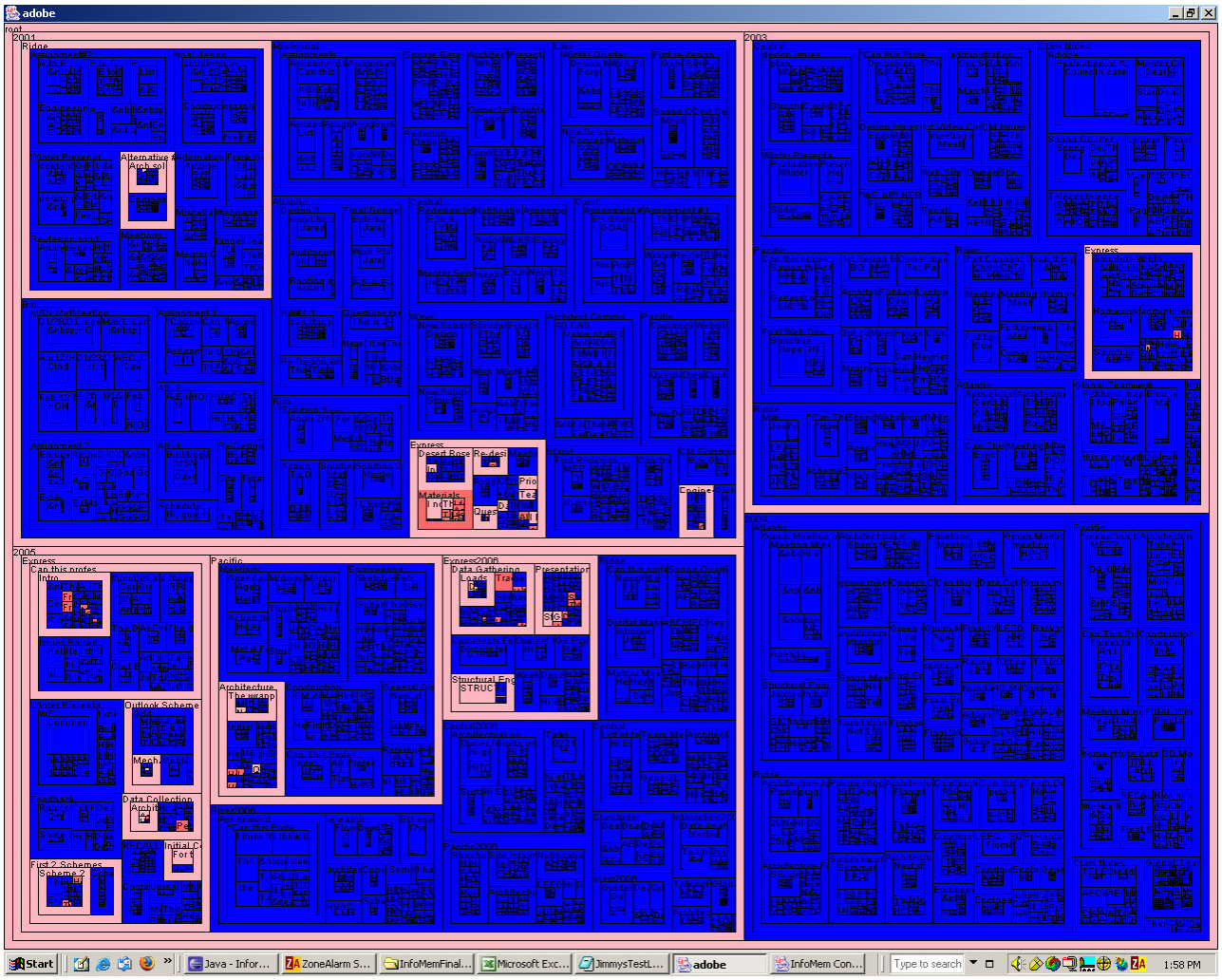

Figure 3. InfoMem Screen Shot for a typical use case of hierarchical search navigation

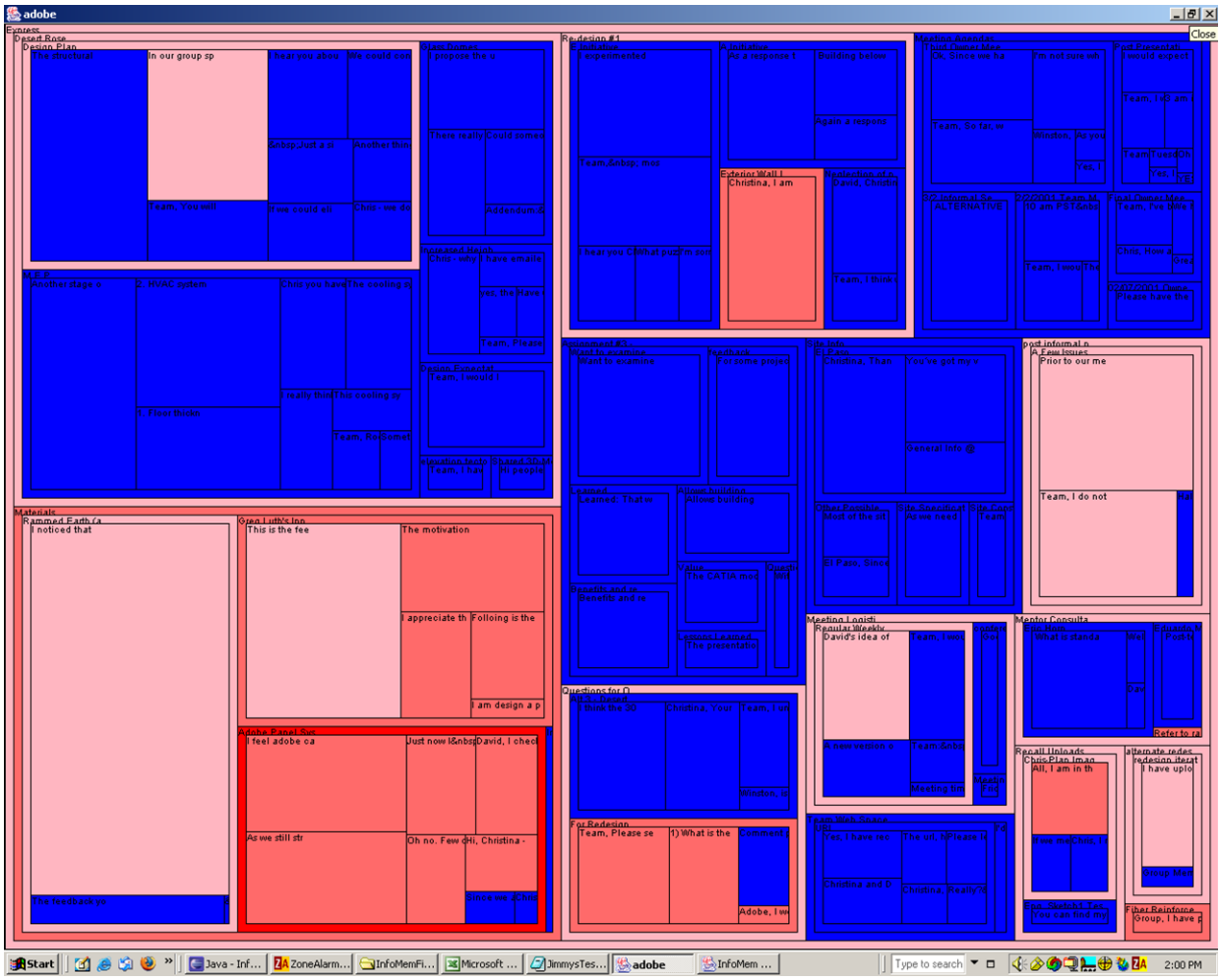

Figure 4. InfoMem Screen Shot for a typical use case of hierarchical search navigation after zooming in.