QIC: Query In Context for Educational Collections

Students’ demand for rich, multimedia content to be incorporated into their learning process has driven teachers to use online resources. With its accessibility and explosive growth of content, the NSDL repositories are in a prime position to provide the quality material teachers and other knowledge seekers need. Still, teachers are dissatisfied with this resource. They are frustrated with the time-consuming manual effort to retrieve and review each link when searching for the appropriate material. This often leads to settling for “good enough” content. As these collections become larger, the number of hits on a keyword based search will increase. To sustain and increase the utilization of NSDL’s quality resources, it is important that a more sophisticated methodology for query and retrieval be developed.

Query In Context for Educational Collections (QIC) is a research project to revolutionize individual search by shifting the burden of information overload from the user to the computer. This is accomplished through context-sensitive text mining methodologies. The major components of QIC’s portable unified knowledge discovery system are context sensitive retrieval, semantic query analysis, and concept extraction. Augmenting NSDL’s NCore search component with context-sensitive methodologies extends the search engine’s capabilities through a modular interface. As context-sensitive text mining research learns more about the variables that support increases in user satisfaction, QIC can be extended to support online searches by minimizing human intervention and increase the relevance of search results.

This research supports sustainability through increased user satisfaction. By organizing the more relevant information first, an NSDL user’s time spent in the selection phase of the discovery process is reduced, which makes NSDL’s quality repositories and its partners more attractive. Reducing search time should increase content utilization by encouraging repeat use. Finally, higher usage should encourage more new content which in turn will increase visitation frequency.

The NSDL has supported a number of grants analyzing instructor usage of digital libraries. From this research the following key characteristics that influence instructor’s search and selection behavior were identified:

- Teachers focus on domain knowledge over pedagogy in most selections.
- Teachers make use of opinion leadership, selecting content from known colleagues or recommendations by their associates.
- There’s a higher frequency searching for material to augment a single class than to design a new course. This material is often used in course redesigns.
- The data also suggests a long learning curve (12 months or longer). (Manduca, Iverson, Fox, 2005)

Effective searching, however, remains an issue. McMartin et al. indicated instructors search for the ‘perfect’ image when preparing for a class but will often “settle for something that is ‘good enough’.” This trade-off is caused by a “lack of efficient search strategies” and the feeling that “searching for materials can be time consuming” (McMartin, Iverson, Manduca, Wolf, Morgan, 2006). QIC builds on this research to develop its concept
extraction module.

It is expected that as collections become larger, current keyword-based search strategies will exacerbate these frustrations (Recker, M., 2006). Our research utilizes innovative ideas to design efficient information retrieval (IR) and text mining algorithms for large, multimedia libraries.

QIC’s goal is to minimize human intervention in the extraction process and reduce the number of contextually inaccurate results displayed. Its unique approach synthesizes user preferences, their situational context, and the informational needs to provide users with results relevant to what they want, rather than presenting ‘cookie cutter’ answers. This approach will improve user effectiveness and thus satisfaction.

Figure 1: User Context-sensitive Categories

Figure 1 illustrates context information supporting a user’s quest for information. A user’s static context may include their role (student/teacher), areas of interest (science), and level of education (K-12, undergraduate, etc). Variations of this information are found in login profiles that are standard in an NSDL pathway and other digital libraries. This information does not change frequently and is considered static.

Situation context provides information in terms of where and when. If the data request is made in the middle of a school term it can be inferred to be needed for a class redesign rather than a course redesign. Some studies have shown sentiment or opinion information may be extracted from recommendation systems or blog comments a user may have written in reference to material stored in the repository (Pang, Lillian, 2008 and Hu, Liu, 2004).

The third category captures the user’s “information world” – e.g. read documents and visited Web pages – thus reflecting the user’s interests. This is an area where the use of text mining techniques has most often been proposed (Mei, Zhai, 2006 and Raymond, 2003 and Fan, Xu, Friedman, 2007).

Certain category variables, when combined with some basic rules may give insight into a user’s search to improve context understanding. For example, a 5th grade science teacher from Galveston, Texas (static context variables) types ‘wind’ as a search variable on September 10, 2009 (situation context). We can infer this teacher is not developing a new course because the school year has just started. Most likely they are looking for resources for a lecture or assignment (rather than a test). Because of their location, Galveston Texas, we might rank hurricanes high and specifically give a high ranking to Hurricane Ike, which hit Galveston, Texas on September 1, 2008. While work has been done separately in all three context-sensitive categories, this project proposes the
harmonious incorporation of all three into text mining processes and will augment them with a rules based engine. Figure 2 shows what the output might look like.

**Wind Types**

**EXPLORE! Curriculum Project**

- [Image](#)
- [Science](#)

**Wind Effects**

- [Image](#)

**Wind Dynamics and Forests**

In this activity, students will set up a model forest using plastic bottles to observe changes caused by differences in wind speed and forest density. An extension to the activity will allow students to explore the concept of evaporation.

**References**


By utilizing data about a user's preferences, search behavior, and information retrieved within the current session, user context-sensitive text mining should provide a more personalized ranked and grouped set of relevant information, thus reducing a user's manual effort in the discovery process. Outliers or results which may lead to accidental discovery or learning will be ranked lower but will not be removed.

Organizing and managing the continuous expansion of digital data is a challenge. QIC helps by integrating digital libraries via a portable platform that supports and improves the discovery process. Techniques to order search results better serves the needs of the users which improves digital library utilization, and, ultimately, encourages the seeking of knowledge and exploration of ideas.

QIC is a starting point to develop a platform portable knowledge discovery system framework that can be tailored to different types of users, content, and digital formats. Evaluations of our results will add to the current research to better understand how educators use digital libraries. This in turn will be a feedback loop to improving extractions results.


