

---

## Computer Science Colloquium

---

### Database and Software Research

Profs. Cindy Chen and Kajal Claypool  
Computer Science Dept, UMass Lowell

Wednesday, 16 February 2005, Olsen 311  
Refreshments at 2:30, Talk from 3:00-4:00

---

#### **Analytical and Experimental Evaluation of Stream Based Joins**

Henry Kostowski

Continuous queries over data streams have gained popularity as the breadth of possible applications, ranging from network monitoring to online pattern discovery, have increased. Joining of streams is a fundamental issue that must be resolved to enable complex queries over multiple streams. However, as streams can represent potentially infinite data, it is infeasible to have full join evaluations as is the case with traditional databases. Joins in a stream environment are thus evaluated not over entire streams, but on specific windows defined on the streams. In this talk, we present windowed implementations of the traditional nested loops and hash join algorithms. The talk will be followed by a brief presentation on ongoing work on sensors by Michael Librodo.

---

#### **Finding a Needle in the Haystack: A Technique for Ranking Matches between Components**

Naiyana Tansalarak

Searching and subsequently selecting reusable components from component repositories has become a key impediment for not only component-based development but also for achieving the overall usability of component development environments and the ultimate re-usability of the components themselves. Component matching, a fundamental aspect of the component search problem, has been a well-studied problem, resulting in many different matching techniques such as keyword, facet, signature and specification matching techniques. However, each matching technique individually applied for component search often yields a large number of (sometimes irrelevant) hits. In this talk, we present a disciplined combination of the different matching techniques to provide a ranked match between the query and library components that reflects the overall “goodness” of the match.

---

#### **Using Caching Techniques to Improve Multiple Sequence Alignment Performance**

Xun Tu

This talk presents a new caching technique to improve the performance of Multiple Sequence Alignment (MSA). A nested two level of B+Tree cache hierarchy is implemented to speed up the lookup of cached pairwise alignment results. Also a new cache data structure is implemented that could decrease the cache size significantly. The first level of a B+Tree is always in the memory and the second level of a B+Tree can be in the memory or in the disk depending on the system configuration. Preliminary experimental results have found that our caching technique can significantly improve the performance of the pairwise alignment step in MSA by using both pairwise alignment algorithms: Needleman-Wunsch algorithm (Global algorithm ) and Smith-Waterman algorithm ( Local alignment ). A brief analysis is presented for deciding when to cache or not cache the pairwise alignments.

---

#### **A Secure XML Querying System**

Wei Li

Current database querying system has to deal with a rapidly growing amount of data and a large number of users, while enforcing secure access control. The performance of the system is often restricted by the authorization process. In our querying system, both cache and user access map are used to check accessibility. Experiments are also carried out to analyze the performance.

---