**GoGetIM: Instant Message Retrieval**

Michael Seltzer  
Faculty Advisor: Zachary Ives

**Problem**

- Instant message conversations are fundamentally different from other searchable information
- Current info retrieval systems are not adequate for search and traversal of logged conversations

**Abstract**

- GoGetIM is an instant message retrieval system that allows users to index sets of instant message logs
- GoGetIM’s indexes can be searched with user-specified constraints
- GoGetIM provides the most relevant conversations and allows for robust results navigation

**Indexing**

- Use preexisting AIM logs to build an index
- Parse HTML log files to strip message usernames, times, and message texts
- Find applicable conversation boundaries based on time differences between messages
- Extend Apache Lucene to write customized index of all conversations

**Searching**

- Program user provides subset of usernames, time range, and query in order to search
- Extend Apache Lucene for efficient search of loaded index
- Results ranked based on combination of tf-idf, relative conversation length, and relative average message word density
- Allow easy navigation of results, as program user can restrict returned conversations by username and date without a new search