Project Overview

• Electronic infrastructure to triangulate knowledge
  – an individual
  – of collegial communities
  – of the world at large
• Independent, interacting information repositories, each customized to its individual user.
  – Automated data gathering
  – Active observation of user activity
  – Adaptation to individual query needs
  – Inter-haystack collaboration
Project MIT9904-07: Haystack: Per-User Information Environment
David R. Karger and Lynn Andrea Stein

Progress Through December 1999

• Kernel Implementation
  – Haystack is a simple operating system
  – Separate privileged operation (persistence, security, etc.) from data-based services (field extraction, indexing, etc.)
  – Integrate persistent, transaction-safe object store

• Initial Exploration of Adaptation and Clustering
  – Query history should influence future query behavior
  – We integrated query infrastructure and improved UI
  – Ready to begin application of learning techniques
  – Explored use of clustering to organize Haystack data
  – Preliminary theoretical work, prototyping
Research Plan for the Next Six Months

- Complete kernel implementation
- Integrate RDBMS
  - Research issues include separating structured from unstructured parts of query and reintegrating responses from heterogeneous information sources
- Implement and evaluate learning for query adaptation
  - Ranking, relevance feedback, and query expansion
- Begin clustering integration
  - Algorithmic work to account for nonstandard space
  - User interface work for optimal browsing and presentation