1. The Vector Space Model in a Nutshell
- Documents and queries: feature vectors,
- Similarity score: cosine of enclosed angle,
- Search: compute similarity and sort results,
- Corpus matrix C: contains all document vectors (normalized),
- Compute similarities: $s = Cq$ for a given query $q$,
- Very sparse: more than 99% zero entries!

2. Dimensionality Reduction
- Multi-scalar CPUs prefer dense vector computations,
- Text indexing issues: Synonyms, etc.,
- Our new method: Rare Term Vector Replacement,
- Improves the precision of the search results.

Retrieval Precision (Reuters Corpus)

3. Parallel Dimensionality Reduction
- Task / Data / Combined parallelization possible.

Execution Time (Reuters Corpus)

4. Parallel Query Processing
- Hybrid partitioning: split features and documents into equal parts,
- Implemented using MPI: supercomputer-grade middleware,
- Single-precision floating point: avoids the memory bottleneck.

Performance Baseline  Hybrid Partitioning

5. Improved Response Time
- Hybrid partitioning exploits the memory hierarchy,
- Delivers super-linear speed-up over serial, in-memory system,
- Disk-based systems are not considered here.

Speed-up  Efficiency

6. Improved Throughput
- The standard for parallel search engines is index replication.
- Can a parallel program outperform multiple serial programs? Yes!
- Parallel queries/serial programs vs. serial queries/parallel program:

Time to Execute Multiple Queries

7. Concurrent Index Maintenance
- Concurrency required for adding / removing / changing documents,
- MPI thread support: one communicator per thread.

Planes of Communication  Cooperating Threads

8. Architecture & Deployment
- Fault tolerance - MPI applications terminate upon node failure,
- Dynamic process & machine management limited with MPI,
- Distributed ensemble of smaller parallel computers,
- Front-End, Query Processing, Document Analysis and Back-End,
- Parallel HPC Search Engine Architecture

9. Summary & Conclusions
- We have parallelized the basic operations for a search engine,
- Measurements indicate a reasonable performance,
- The middleware was not built for this - but it can be bent into shape.

10. Future Extensions
- Clustering - conduct the parallel search within clusters,
- Text processing pipeline - from text to vectors,
- Server-side result set caching, server-side cursors,
- Hybrid computing - support accelerators.