

Overview of Web Indexing

- Standard IR techniques are the basis, but insufficient
 - Scalability
 - Efficiency
- Google's contributions:
 - Google File System (GFS): distributed file system
 - MapReduce: Software framework for parallel computation
 - Hadoop: Open source implementation of MapReduce