











Construction of posting lists

- Overview
 - "document" now means preprocessed document
 - One pass through collection of documents
 - Gather postings for each document
 - Reorganize for final set of lists: one for each term
- Look at algorithms when can't fit everything in memory

7

- Main cost disk page reads and writes
 - Terminology: disk block = disk page





















