



Models have seen			
Model	Document	Query	Satisfy
Boolean	set of terms	boolean expression over terms	evaluate boolean expression
Vector dictionary of	<i>t</i> -dimensional vector	<i>t</i> -dimensional vector	vector measure of similarity
t terms			by score
Extend	more information ?	more information ?	?

Start to enhance model

Properties of terms within documents

- Frequency of term in doc
- Where in doc?
- Relative position of multiple query terms
- Special use? (e.g. in title, font, ...)
- Occurs in anchor text of another doc. pointing to this doc.



Properties of terms within documents

Vector model gave us

- Frequency of term in doc
- Property of each occurrence of term in doc. - Where in doc?

 - Relative position of multiple query terms
 - Special use? (e.g. in title, font, ...)
- Found when evaluate another document - Occurs in anchor text of another doc. pointing to
 - this doc.

Model • Document: bag of terms + attributes - Includes sequence of terms. Why? Query: sequence of terms

- Can make more complicated
- · Satisfying: in current search engines, documents "containing" all terms - AND model
 - "containing" includes anchor text of pointers to this doc from other docs
- Sometimes relax if no documents contain all • Ranking: wide open function
 - info beyond documents and query ?



keep summary information for documents

Data Structure for Collection: Invert

- for each term, keep list of:

 documents in which it appears
 positions at which it occurs in each doc.
 attributes for each occurrence
- keep summary information for documents
- · keep summary information for terms

8

Inverted Index for Collection

7

• for each term, keep POSTINGS LIST of:

- each document in which it appears

 each position at which it occurs in doc.

- attributes for each occurrence

• Core structure used by query evaluation and document ranking algorithms

