Recuperação de Informação

Cap. 01: Introdução

21 de Fevereiro de 1999

Berthier Ribeiro-Neto

Motivation

- IR: representation, storage, organization of, and access to information items
- Focus is on the *user information need*
- User information need:
  - Find all docs containing information on college tennis teams which: (1) are maintained by a USA university and (2) participate in the NCAA tournament.
- Emphasis is on the retrieval of information (not data)

Motivation

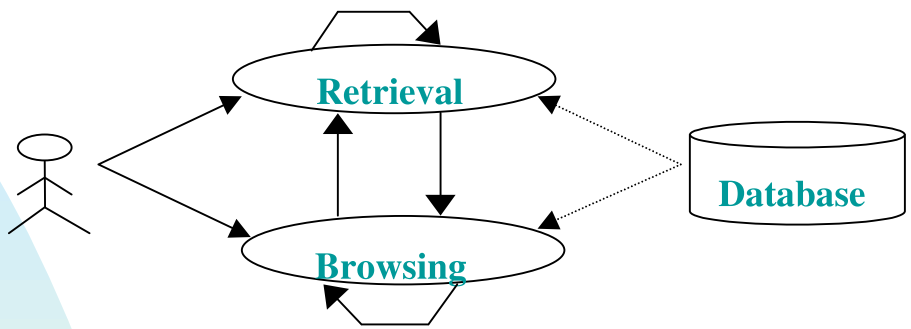
- Data retrieval
  - which docs contain a set of keywords?
  - Well defined semantics
  - a single erroneous object implies failure!
- Information retrieval
  - information about a subject or topic
  - semantics is frequently loose
  - small errors are tolerated
- IR system:
  - interpret contents of information items
  - generate a *ranking* which reflects relevance
  - *notion of relevance* is most important

Motivation

- IR at the center of the stage
  - IR in the last 20 years:
    - classification and categorization
    - systems and languages
    - user interfaces and visualization
  - Still, area was seen as of narrow interest
  - Advent of the Web changed this perception once and for all
    - universal repository of knowledge
    - free (low cost) universal access
    - no central editorial board
    - many problems though: IR seen as key to finding the solutions!

Basic Concepts

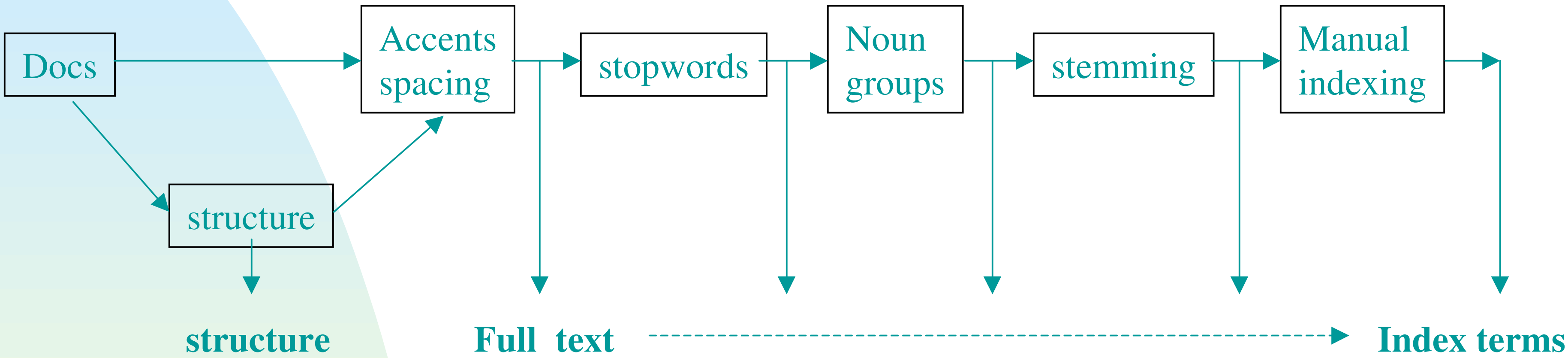
- The User Task



  - Retrieval
    - information or data
    - purposeful
  - Browsing
    - glancing around
    - F1; cars, Le Mans, France, tourism

Basic Concepts

- Logical view of the documents



- Document representation viewed as a continuum: logical view of docs might shift

The Retrieval Process