

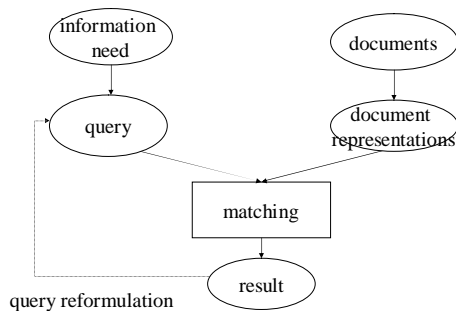
## Processing of large document collections

Part 12 (Closing of the course)  
Helena Ahonen-Myka  
Spring 2005

## QA vs IE

- open domain, closed domain?
  - questions, task definitions
- task definition
  - IE: static
  - QA: question defines the task dynamically
- answer
  - IE: structured template
  - QA: text snippet and/or exact answer (~one slot value of a template?)
- similar components can be used
  - language analysis, semantics (WordNet, word lists,...), pattern matching

## Mapping to the information retrieval process



## TC vs TS vs IE vs QA?

- text categorization (TC)
- text summarization (TS)
- information extraction (IE)
- question answering (QA)

## General issues

- performance requirements
  - building vs. use
    - offline vs. online processing
  - effectiveness vs. efficiency
  - observed performance in some user task
  - evaluation
    - simplified research settings vs. real-life environments

## General issues

- portability, scalability
  - amount and type of manual processing
  - domain/language dependency
  - are some components available off-the-shelf?
    - but do not use heavy processing for simple tasks
    - e.g. linguistic analysis vs. pattern matching
  - static vs. dynamic system/component?

## General issues

- What is our goal?
  - automatic text understanding
    - and automatic processing based on that?
  - probably tools for specific tasks for specific users are more reasonable than very generic, open-domain tools

## Learning goals

- learn to recognize components of applications/processes
- learn to recognize which (kind of) methods could be used in each component
- learn to implement some methods
- (meta)learn to control learning processes (What do I know? What should I know to solve this problem?)

## Exam & exercise points

- exam: Thu 12.5. at 16-20, A111
- points: exam 50 pts, exercises 10 pts
  - required: ~30 pts (= 1-)
- exercise points:
  - 1 exercise -> 1 point
  - 2 -> 2
  - 3 -> 3
  - 5 -> 4
  - 7 -> 5
  - 9 -> 6
  - 11 -> 7
  - 13 -> 8
  - 15 -> 9
  - 16 -> 10

## Kiitos!!!

- ... ja hyvää kesää!!