

# IR-BASE: An Integrated Framework for the Research and Teaching of Information Retrieval Technologies

Pável Calado   Ana Cardoso-Cachopo   Arlindo Oliveira

Instituto Superior Técnico / INESC-ID

January 10th, 2007



# Outline

- 1 Introduction – what IR-BASE is
- 2 Who/what IR-BASE is for
- 3 Advantages of IR-BASE
- 4 How to use IR-BASE
- 5 Example Applications
- 6 Present State
- 7 Conclusions and Future Work



# Introduction – what IR-BASE is

IR-BASE is an architecture for the development of modular components for building IR systems, including:

- Datasets
- Available components pool
- User-developed components
- Guidelines for component development
- Documentation
- Web-site



# Who/what IR-BASE is for

- Researchers  
Allows a rapid development of prototypes.
- Teachers  
Provides examples for classes and project assignments.
- Students  
Provides a quick startup for their projects.



# Advantages of IR-BASE

- Reusable implementations for the components
- Well tested, modifiable, expandable components
- Ability to construct different combinations
- Easy experimentation of new ideas
- Open source development
- Cooperation between users worldwide



# How to use IR-BASE

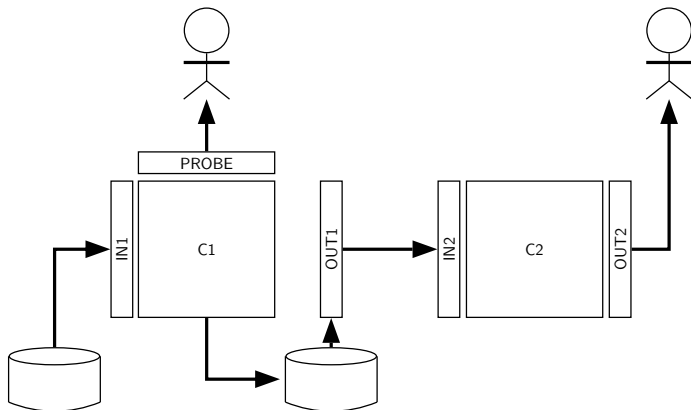


Figure: Example of an IR system built with IR-BASE components.

# Example Applications

- “Webpage Ranker”
  - 1 Web crawler
  - 2 Indexer
  - 3 Document ranking module
- “Paper Finder”
  - 1 Digital paper library
  - 2 Items 2 and 3 from previous application can be reused
- It is possible to change indexing/ranking strategies for any of the applications by replacing components 2/3.



# Present State

- Simple components
  - Developed by different students
- No web service

Major difficulty: students get jobs, dropout of college, new students only in the following year.





# Conclusions and Future Work

- Behind schedule
- Need to boost component development
- Need to increase appeal to students

