Resource Selection in Distributed Information Retrieval Scenarios

Ilya Markov

University of Lugano

March 14, 2012
Can the discussed techniques be applied to your problem/research area?

Can the techniques that you use be applied to the discussed problems?
Outline

1 Information Retrieval

2 Distributed Retrieval Scenarios

3 Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches

4 Discussion
Outline

1. Information Retrieval

2. Distributed Retrieval Scenarios

3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches

4. Discussion
Information Retrieval in Practice

Google

Search

About 638,000 results (0.37 seconds)

Everything
Images
Maps
Videos
News
Shopping
More

Lugano
Change location

The web
Pages from Switzerland

More search tools

Università di Lugano - Università della Svizzera italiana
www.usi.ch/en

⇒ Graduate studies - the Master's ... - Faculty of Informatics - Job opportunities at USI

USI - Faculty of Informatics
www.inf.usi.ch/

+ Show map of Via Giuseppe Buffi 13, 6900 Lugano

USI - Università della Svizzera italiana
www.usi.ch/ - Translate this page

Università di Lugano. University of Lugano. Scegli la lingua. Choose the language. 1. Italiano. Italian. 2. Inglese. English. Benvenuto sul sito dell'Università ...
+ Show map of Largo Bernasconi 2, 6850 Mendrisio
⇒ Risorse e servizi online - Concorsi e offerte di lavoro ... - Università di Lugano
Given a query and a set of docs

1. Estimate doc’s relevance

2. Rank docs by relevance
Aggregated Search

Google
Search

About 638,000 results (0.37 seconds)

Everything
Images
Maps
Videos
News
Shopping
More

Lugano
Change location

The web
Pages from Switzerland
More search tools

University of Lugano - Università della Svizzera italiana
www.usi.ch/en

→ Graduate studies - the Master's ... - Faculty of Informatics - Job opportunities at USI

USI - Faculty of Informatics
www.inf.usi.ch/
+ Show map of Via Giuseppe Buffi 13, 6900 Lugano

USI - Università della Svizzera italiana
www.usi.ch/ - Translate this page
Università di Lugano. University of Lugano. Scegli la lingua. Choose the language. 1. Italiano. Italian. 2. Inglese. English. Benvenuto sul sito dell'Università ...
+ Show map of Largo Bernasconi 2, 6850 Mendrisio
→ Risorse e servizi online - Concorsi e offerte di lavoro ... - Università di Lugano

University of Lugano - Directions

Maps Google

Details
Bus: Lugano, Ponte Madonnetta

More reviews
google.com (3)

Local info
Images
Search in Digital Libraries

![Diagram showing the interaction between BROKER and different resource containers in a distributed retrieval scenario.](image-url)
Distributed Retrieval Scenario 1

When many document sources can answer a user query (through dedicated search engines).

*Note: the full content of sources may not be available.*
Blog Search

- Find relevant blogs for a user query
- Blog is a collection of posts/documents
Expert Search

- Find people with relevant expertise
- Each person is associated with docs
When a document source is itself an answer to a user query.
Summary

1. Many document sources can answer a query.  
   *Examples: aggregated search, digital libraries.*

2. A document source is an answer to a query.  
   *Examples: blog search, expert search.*
Outline

1. Information Retrieval
2. Distributed Retrieval Scenarios
3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches
4. Discussion
Outline

1. Information Retrieval
2. Distributed Retrieval Scenarios
3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches
4. Discussion
1. **Scenario**: a document source is an answer to a query.  
   **Problem**: select *relevant* sources.

2. **Scenario**: many document sources can answer a query.  
   **Problem**: select sources that *contain relevant documents*. 
Objective of Resource Selection

Given a query and a set of document sources select sources that are likely to contain relevant documents.
Information Retrieval

Introduction

First Generation Approaches

Second Generation Approaches

Resource Selection

Given a query and a set of docs

Estimate doc’s relevance

Rank docs by relevance
Given a query and a set of sources

1. Estimate number of relevant docs
2. Rank sources by this number
Outline

1. Information Retrieval
2. Distributed Retrieval Scenarios
3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches
4. Discussion
First Generation (CORI)

1. Sources as documents
2. Retrieval on documents
3. Documents are sources
Outline

1. Information Retrieval
2. Distributed Retrieval Scenarios
3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches
4. Discussion
Second Generation (ReDDE)

1. All documents from all sources are merged in a centralized index.
2. Documents are ranked with regards to a query.
3. *Top n* documents are supposed to be relevant to a query.
4. The number of relevant documents in a source $R$ is calculated as a proportion of its documents in the *top n*.
Second Generation (ReDDE)

\[ \hat{R}_{red} = \frac{3}{8}, \quad \hat{R}_{green} = \frac{2}{8}, \quad \hat{R}_{yellow} = \frac{3}{8}, \quad \hat{R}_{blue} = 0 \]
Outline

1. Information Retrieval
2. Distributed Retrieval Scenarios
3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches
4. Discussion
Topics Covered

1. Information Retrieval

2. Distributed Retrieval Scenarios

3. Resource Selection
   - Introduction
   - First Generation Approaches
   - Second Generation Approaches

4. Discussion
Discussion Outline

- Questions & Answers

- Feedback

Your ideas on
- Applications outside IR
- Techniques from outside IR