

# *Expansion of Multilingual Queries for IR in Economics*

**Mercè Lorente**

**Institut Universitari de Lingüística Aplicada**

**Universitat Pompeu Fabra**

**[merce.lorente@upf.edu](mailto:merce.lorente@upf.edu)**

# Session Contents

- Working on IR for LSP
- Partners
- Project Antecedents
- Applied Objective
- Secondary Objectives
- Knowledge Bank for Economics
- Expansion of Queries
- RECCI System
- Other Related Research: PhD thesis

# Working on IR for LSP

- Users and Query Types
- Vague Borders between Terminology and General Language
- Conceptual Relations in Specialized Fields
- IR Techniques
- Methods for Expanding Queries
- Improving Precision (Relevance)
- Noise Reduction
- Costs of Specific Resource Building

# Project

- RICOTERM-2
- *Terminological and Discursive Control for Information Retrieval in Specialized Communicative Fields, by means of Specific Linguistic Resources and a Query Re-Elaborator.*
- HUM2004-05658-C02-00
- PR Mercè Lorente

# Partners

- Universitat Pompeu Fabra (coord.) (7)
- Universidade de Santiago de Compostela (5)
- Euskal Herriko Unibertsitatea (3)
- Universiteit Antwerpen (1)

# Project Antecedents

- RICOTERM 2001-03 Project (TIC2000-1191)
  - User needs
  - Evaluation of IR Systems
  - Taxonomies and Ontologies
  - Linguistic Analysis
  - Genome Bank Development
- Redirection of New Projects

[Sugeriments](#) | [Mapa](#) | [Ajuda](#)



## Banc de Coneixement sobre el Genoma Humà

El Banc de coneixement GENOMA ha estat desenvolupat pel grup IULATERM, de l'Institut Universitari de Lingüística Aplicada, en el marc dels projectes TEXTERM (BFF 2000-0841) i RICOTERM (TIC 2000-1191).

Conté informació textual, documental, terminològica i conceptual sobre genòmica humana. És una eina de consulta per a traductors, redactors i especialistes en la matèria.

**BT** banc terminològic/ontològic

**Corpus** corpus textual

**BD** banc documental

**BF** banc factogràfic

# Genoma

© IULATERM 2003

[Mercè Lorente Casafont](#)

[Modificar dades](#) | [Sortir](#)

# Applied Objective

- RECCI System Design and Prototype:  
Query *RE-Elaborator for Internet Search Engines*.
  - User-Search Engine Interface located in a Specialized Site
  - Query Expansion from Simple Queries
  - Multilingual Expanded Queries
  - Interaction with *MKB for Economics*

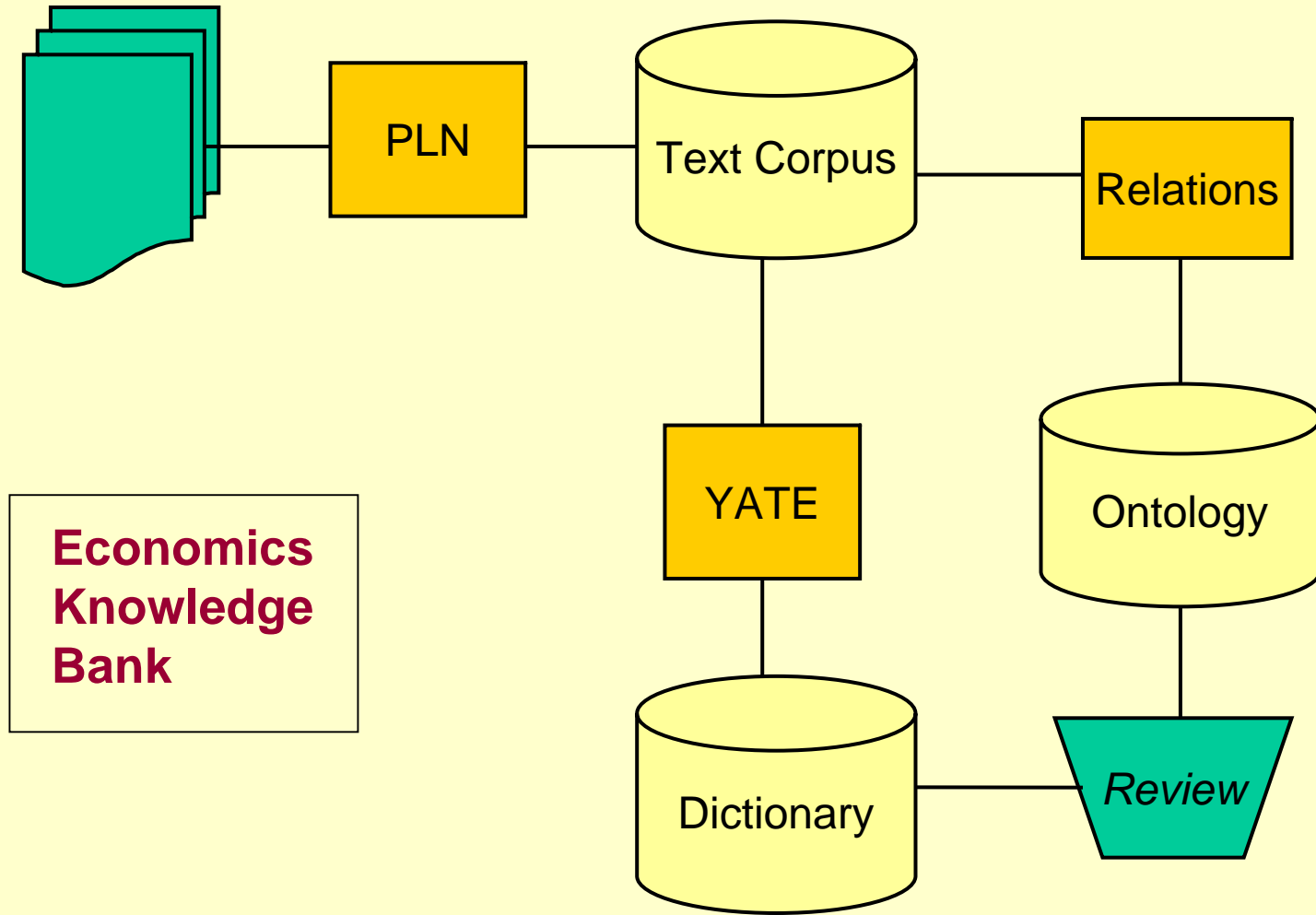


# Secondary Objectives

- MKB for Economics Building
- Economy Lexicon and Discourse Description for Linguistic Strategies Design
- Enrichment of Computational Dictionary for NLP with Semantic Information

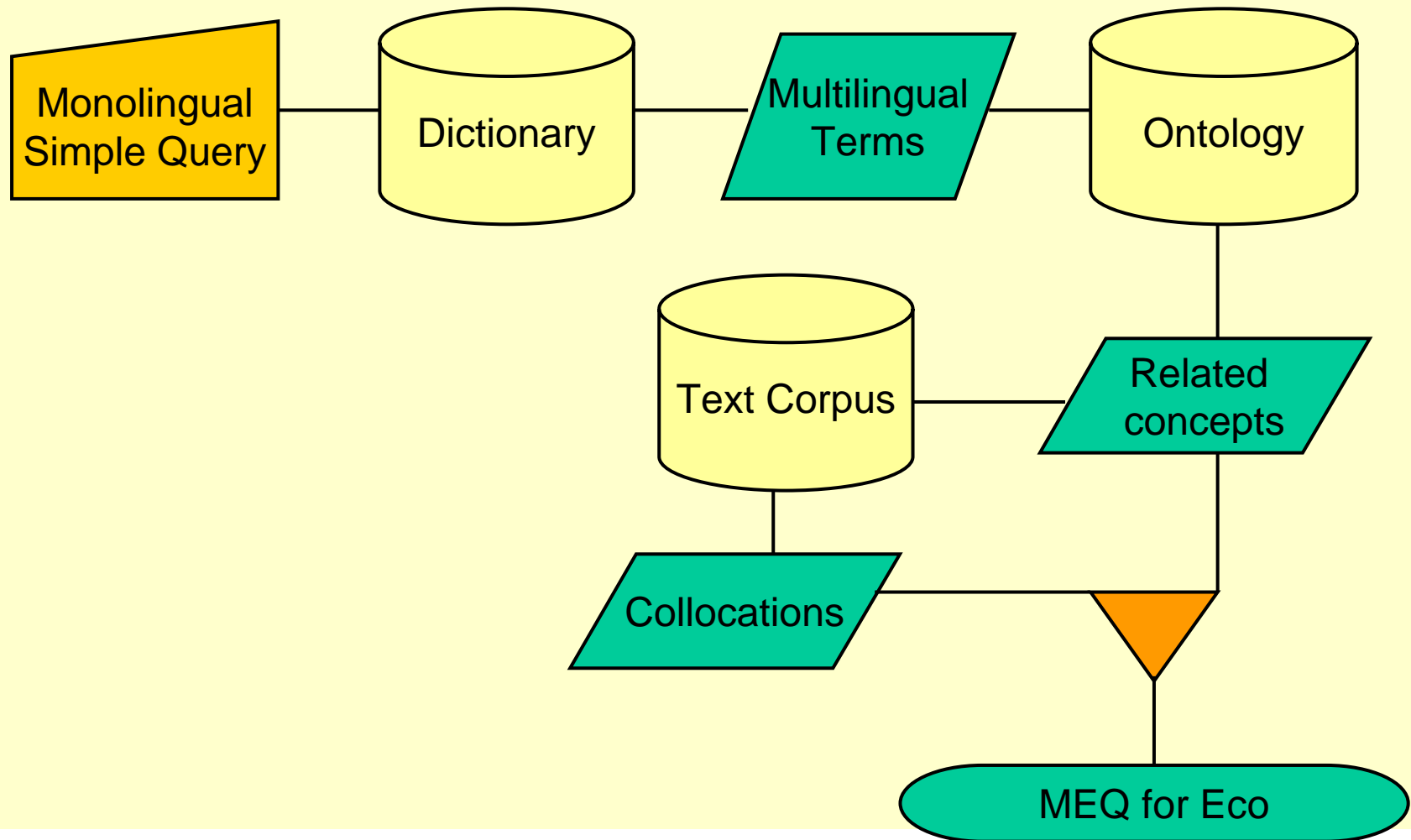
# Knowledge Bank for Economics

- Modular
  - TEXTUAL CORPUS
  - ONTOLOGY
  - DICTIONARY / TERMINOLOGICAL DB
- Non Exhaustive: Experimentation
- Multilingual: Basque, Spanish, Catalan, Galician and English
- Re-use of Existing Sources



# Expansion of Queries

- Tested Methods
  - From an Ontology or Lexical Hierarchy
  - From a Structured Textual Corpus
  - From an Intranet
  - From Keywords or Controlled Vocabulary
- Results
- Method Combinations

**Multilingual Expanded Queries: RECCI**

# Other Related Research

- Bidirectional Re-use of Thesauri and Ontologies
- Lexical Semantics of Specialized Fields
- Multilexical Words and Collocations
- Adaptation of Terminology Automatic Extractor for Economy

Moltes gràcies

[merce.lorente@upf.edu](mailto:merce.lorente@upf.edu)

<http://www.iula.upf.edu>