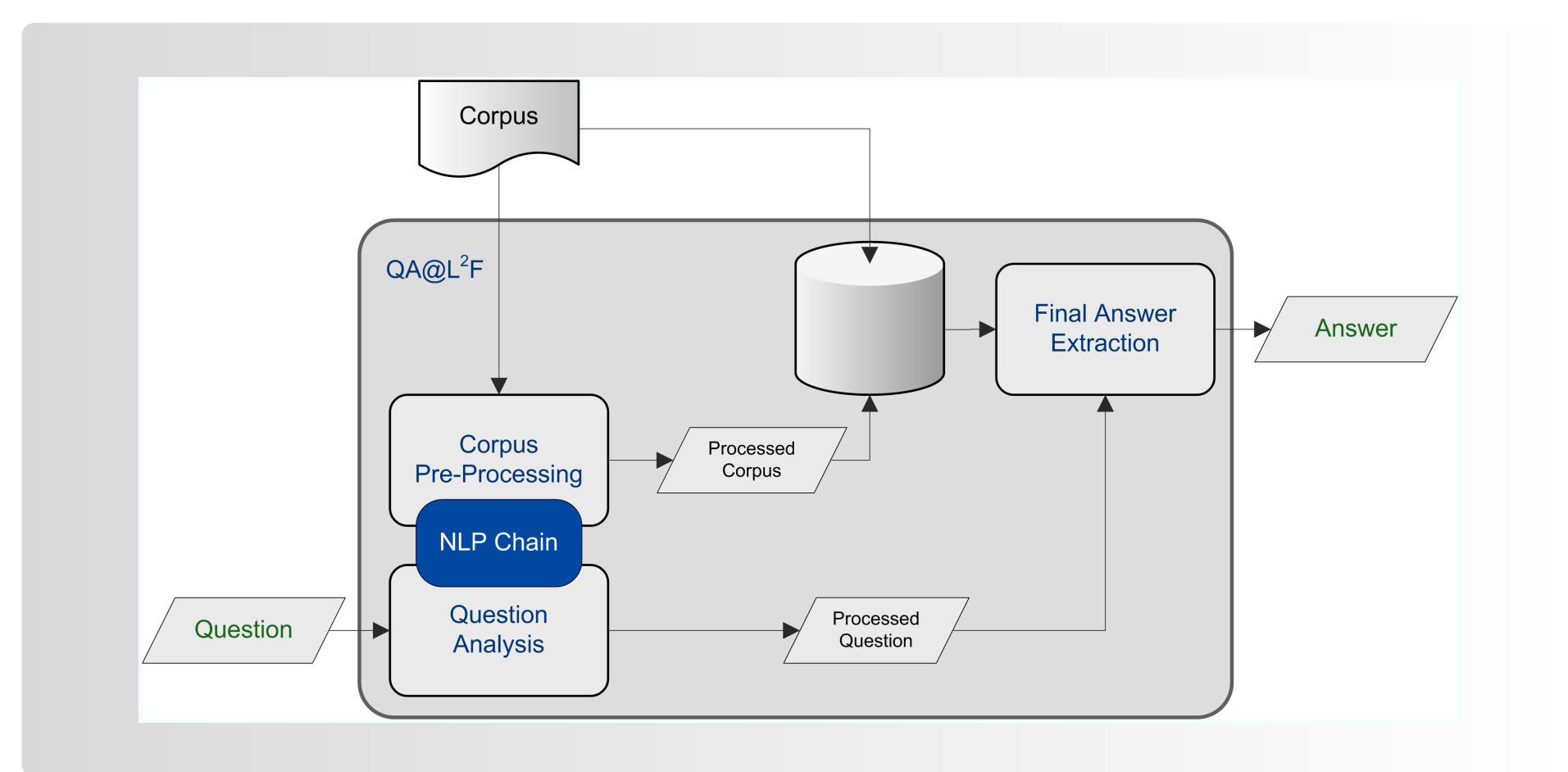


QA@L²F represents the first QA system developed at the INESC-ID's Spoken Language Systems Laboratory participating on the QA@CLEF evaluation track.

1 Architecture

- *Corpus* Pre-Processing, done before the question's submission
- Question Analysis, gathers the question's relevant information
- Final Answer Extraction, retrieves the question's answer



2 NLP Processing Chain

- Performs all the linguistic analysis used by QA@L²F
- Built upon the following NLP tools
 - Palavroso [3], responsible for morpho-syntactic analysis
 - MARv [5], responsible for morpho-syntactic desambiguation
 - RuDriCo [4], responsible for token splitting and concatenation
 - XIP [1], responsible for returning the input organized in chunks, connected by dependency relations

3 Corpus Pre-Processing

- The *corpus* is processed and stored on separate databases
 - Relation-concepts database
 - ▷ Linguistic patterns for Portuguese were built
 - ▷ Ken Loach is the author of *Land and Freedom*

CULTURE				
id	culture	author	confidence	count
1	Land and Freedom	Ken Loach	99	4

- Named-entities database
- Raw newspaper *corpus* is also stored
- The WikiXML collection and its database structure was used [2]

4 Question Analysis

- Relies on a frame builder responsible for identifying:
 - the answer extraction script to be called
 - the target entity
 - all the named entities
 - auxiliar words, such as verbs

5 Final Answer Extraction

5.1 Linguistic Pattern Matching

- Main target: definition questions
 - **Question:** O que é a TVI?
 - **Snippet:** falando sobre a Televisão Independente (TVI) de inspiração cristã
 - **Pattern:** NounPhrase (Acronym)
 - **Answer:** Televisão Independente

5.2 Linguistic Reordering

- Main targets: definition and list questions
- Uses the Wikipedia
- Builds linguistic patterns with the question's target entity
 - Definition questions: *target entity* + inflected verb to be
 - ▷ **Question:** Quem foi Ésquilo?
 - ▷ **Pattern:** Ésquilo (foi|é) <answer>
 - ▷ **Answer:** poeta trágico grego
 - Questions having a list (of one or more items) as an answer: inflected verb to be + *target entity*
 - ▷ **Question:** Diga uma escritora sarda
 - ▷ **Pattern:** <answer> (foi|é) escritora sarda
 - ▷ **Answer:** Grazia Deledda

5.3 Named Entities Recognition

- Main target: factoid questions
- Merges the named entities in the question with those stored on database

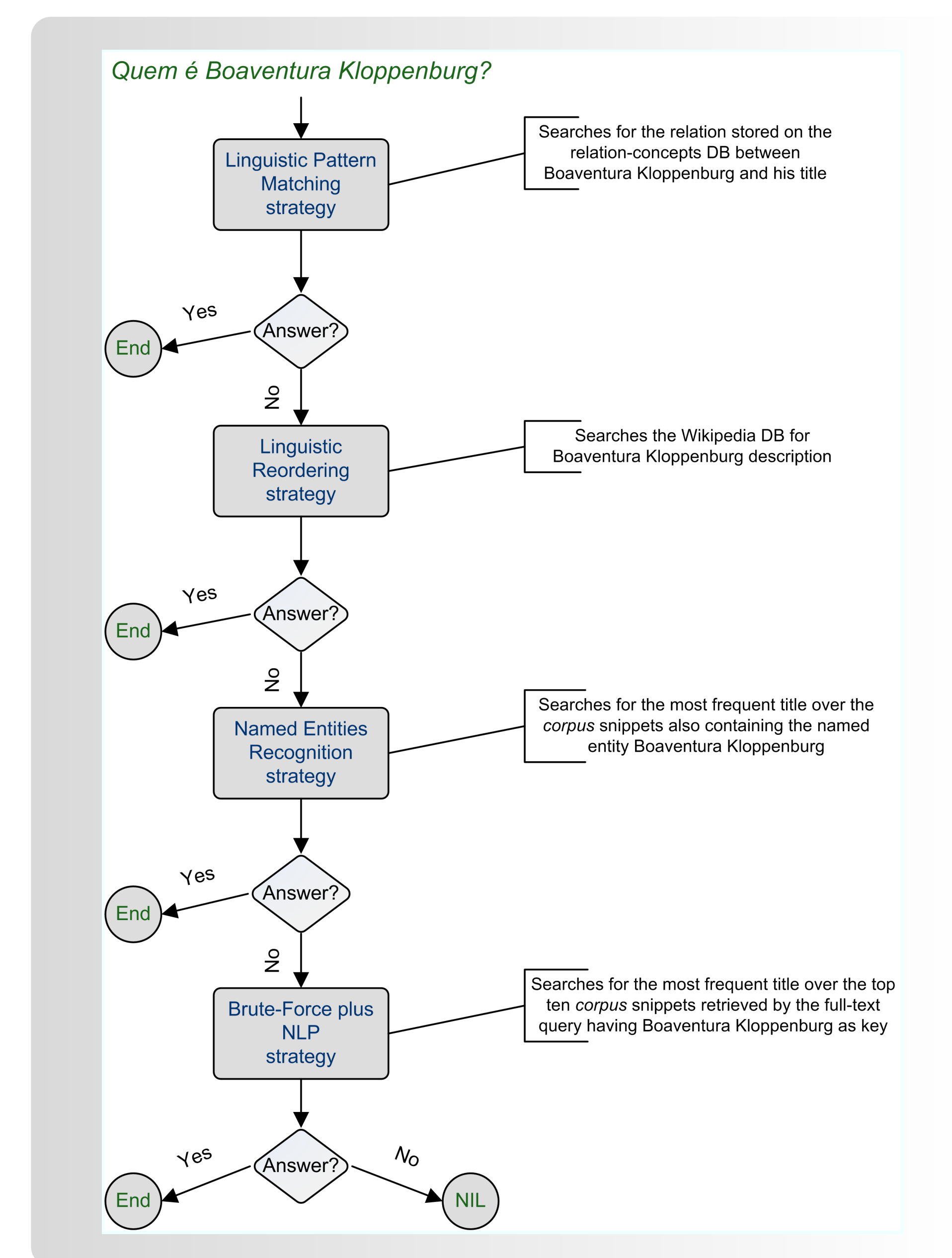
- Gathers the *corpus* snippets where the question's named entities are found
- Returns the most frequent named entity of the expected answer type

5.4 Brute-Force plus NLP

- Performs a full-text query
- Uses the information collected during the question analysis phase as key
- Gathers the best classified snippets
- Returns the most frequent named entity of the expected answer type

6 Relaxing Constraints Mechanism

- When a question is not answered using the appropriate strategy
- QA@L²F relaxes its constraints and applies a set of answer extraction strategies
- NIL is returned only if none of the strategies finds an answer



7 QA@L²F Results at QA@CLEF

Right	Wrong	ineXact	Unsupported	Total	Accuracy
28	166	4	2	200	28/200 = 14%

- Anaphora and ellipsis were not handled
- NIL was returned 152 time, being correct in 11 of those
- IneXact answers were all incomplete answers
 - **Question:** Quem é George Vassiliou?
 - **Snippet:** George Vassiliou, presidente de Chipre entre 88 e 93, lançou, por sua vez,
 - **QA@L²F's answer:** presidente de Chipre
 - **Correct answer:** presidente de Chipre entre 88 e 93
- In the Unsupported answers, the Wikipedia pages title were returned to support the answer, instead of a sentence
 - **Question:** Diga uma escritora sarda.
 - **QA@L²F's answer:** Grazia Deledda
 - **Support:** Grazia Deledda
 - **Correct support:** Grazia Deledda (Nuoro, 27 de setembro de 1871 - Roma, 15 de agosto de 1936) foi uma escritora e poeta sarda, vencedora do Prémio Nobel de Literatura de 1926.

[1] Salah Aït-Mokhtar, Jean-Pierre Chanod, and Claude Roux. A multi-input dependency parser. In *Proceedings of the Seventh IWPT (International Workshop on Parsing Technologies)*, Beijing, China, October 2001.

[2] Information and Language Processing Systems group at the Informatics Institute at the University of Amsterdam.

[3] José Carlos Medeiros. Análise morfológica e correção ortográfica do português. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal, 1995.

[4] Joana Paulo Pardal and Nuno J. Mamede. Terms Spotting with Linguistics and Statistics. pages 298–304, November 2004.

[5] Ricardo Ribeiro, Nuno J. Mamede, and Isabel Trancoso. Using Morphosyntactic Information in TTS Systems: comparing strategies for European Portuguese. In *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003, Faro, Portugal, June 26-27, 2003. Proceedings*, volume 2721 of *Lecture Notes in Computer Science*. Springer, 2003.