

# JaTeDigo - A Natural Language Interface for a Movies Database



Ana Raquel Guimarães  
L2F, INESC-ID/IST

Luísa Coheur  
L2F, INESC-ID/IST

Nuno J. Mamede  
L2F, INESC-ID/IST

Ana Mendes  
L2F, INESC-ID/IST

<http://www.l2f.inesc-id.pt>

## JATEDIGO

### What is it?

- A Natural Language Interface for a Movies Database

### JaTeDigo goals

- Easy to use, fast and credible
- Easy to extend
- Able to answer “traditionally difficult questions”, such as:
  - *Who plays in MOVIE1 and MOVIE2?*

1

## DATABASE

### Sources

- Internet Movie DataBase (IMDB)
- Oscars Academy website
- Cinema PTGate website

### Data

- movies cast
- biographies
- best movie/director
- best (supporting) actor/(supporting) actress

2

## ARCHITECTURE

### Named Entity Recognition

- Full-text query  
*SELECT DISTINCT films.title  
FROM films  
WHERE match(title)  
AGAINST (“Who directed King Kong?”)*

- Disambiguation
  - ambiguity situations:
    - \* people with the same name;
    - \* movies with the same name;
    - \* movies with people names;
  - disambiguation types:
    - \* user disambiguation
    - \* context disambiguation

### Local grammar creation

#### Example of user disambiguation

**Resultado**

Existe mais que um “king kong” na base de dados de filmes. Escolha um dos seguintes

- King Kong (2005)  
Elenco Principal
  - Naomi Watts
  - Jack Black
  - Adrien Brody
- King Kong (1976)  
Elenco Principal
  - Jeff Bridges
  - Charles Grodin
  - Jessica Lange

3

## ARCHITECTURE (CONT.)

### Question Interpretation

*In which movie did Meg Ryan and Tom Cruise performed together?*

- Syntactic and semantic pattern matching are used to...
- ... create dependencies (XML)

#### Example of dependency

```
target_which_films_main_act_two("meg ryan", "tom cruise")  
<DEPENDENCY name="TARGET_WHICH_FILMS_MAIN_ACT_TWO">  
<PARAMETER ind="0" num="23" word="meg ryan"/>  
<PARAMETER ind="1" num="24" word="tom cruise"/>  
</DEPENDENCY>
```

- Dependencies are used to create scripts...  
*script-which-films-main-act-two.pl ACTOR 'meg ryan' ACTOR 'tom cruise'*
- ... that generate SQL queries

4

## EVALUATION

### Experiments

- First experiment with examples and second experiment without
- 5 users for each experiment/10 questions each user

#### Web-page for the first experiment

- First experiment results:
  - Right answer: 29
  - Wrong answer: 4
  - No answer: 17 (mainly due to misspelling)
- Second experiment results:
  - Right answer: 18
  - Wrong answer: 2
  - No answer: 30

5

## REMARKS

- People are influenced by examples
- Urgent: speller integration
- Available at:  
<http://www.l2f.inesc-id.pt/~arog/index.pl>

6