

The Senso Question Answering approach to Portuguese QA@CLEF 2007

José Saias

jsaias@di.uevora.pt

Paulo Quaresma

pq@di.uevora.pt

Universidade de Évora, Portugal



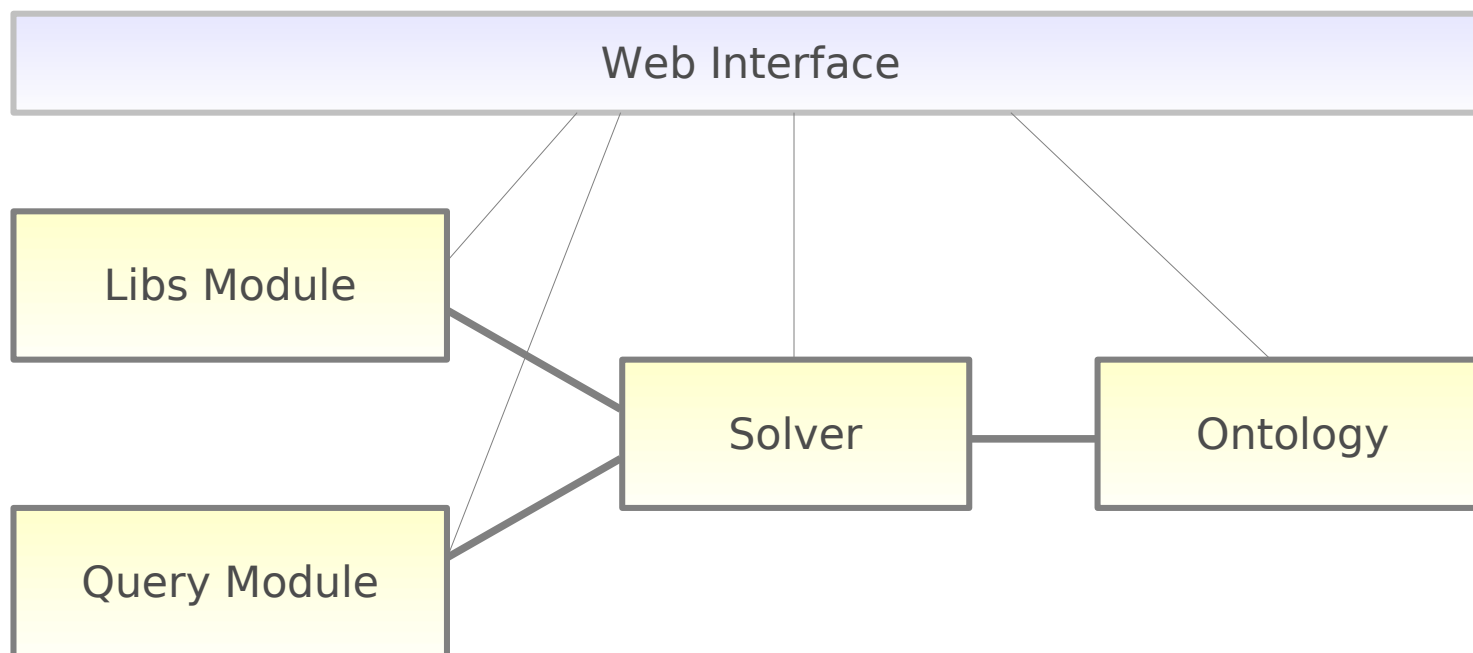
Introduction

- Departamento de Informática, from Universidade de Évora
 - Previous participation in CLEF: 2004, 2005

- This edition: CLEF 2007
 - QA@CLEF Monolingual Portuguese task
 - 200 natural language questions
 - answers could be in
 - newspaper collections from Público and Folha de São Paulo
 - Wikipedia texts

 - Senso: QA System

Senso QA System Architecture

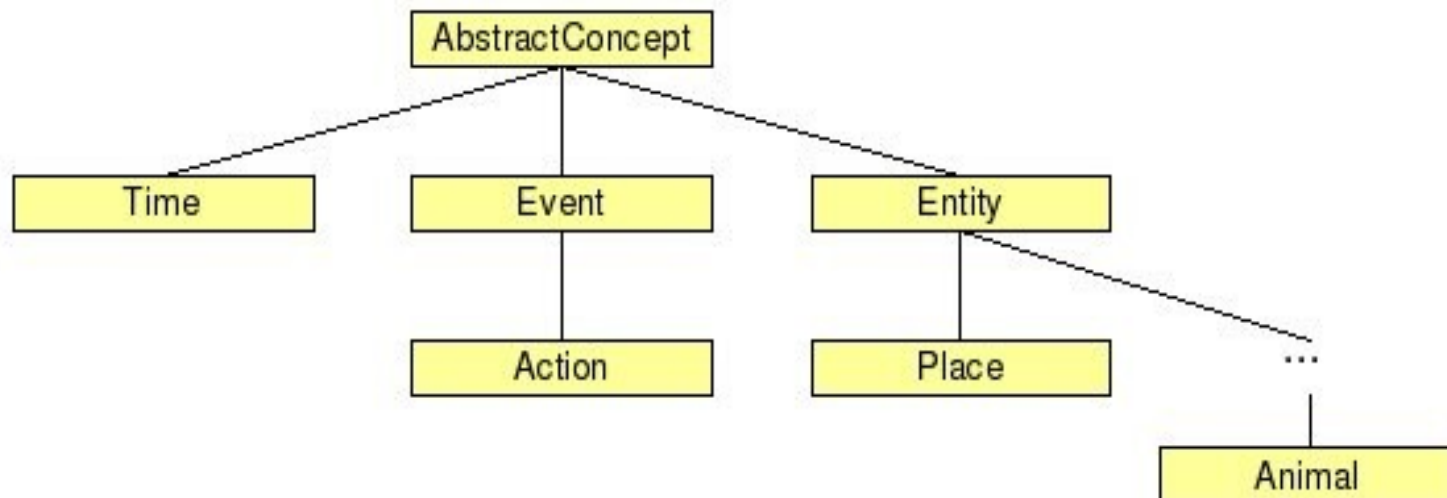


Small starting KB

- sentence understanding
 - requires knowing the used terms or concepts
- starting knowledge base
 - semantic information used in sentence analysis
- *Senso Ontology*
 - has a hierarchy of concepts in Portuguese
 - concept relations are common sense
 -

System Modules

- **Ontology**
 - provides semantic information
 - document retrieval (candidate answer docs)
 - search term expansion
 - concept semantic equivalency or *IsA*/specialization



System Modules

- Libs Module
 - contains collections of text documents
 - indexes all collections
 - text morpho-syntactical analysis

- Query Module
 - manages interrogation
 - query group association, for topic related questions
 - produces a text search query (*Candidate Answer Docs*)
 - based on “question text + ontology”

System Modules

- Solver
 - search possible answers in the identified relevant texts (*Candidate Answer Docs*)
 - logic-programming based tool
 - ad-hoc answer selector
- Web Interface
 - easier and friendly usage
 - analysis of each intermediate step
 - search for documents (or queries) and read them

Web Interface

Senso Interrogação - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://localhost:8080/senso/qa/query?a=:

Google

Senso

Universidade de Évora
Departamento de Informática

UNIVERSIDADE DE EVORA

Interrogação clef0012

introduzida em Mon May 07 19:11:00 WEST 2007

Questão

Quem é Boaventura Kloppeburg?

query visl sintaxe interpretação

clef0012 ver ver ver **run**

[gerar termos de pesquisa para lucene](#)
[ver input lucene](#)
[efectuar a pesquisa](#)
[ver output lucene](#)
[gerar o flenqine](#)

[processar](#)
[ver solução](#)

[voltar](#)

Done

QA System: How it works

- Import text collections
 - newspapers and Wikipedia texts
 - split into individual documents
 - temporal context
 - indexed
 - Lucene (<http://lucene.apache.org/>)
 - morpho-syntactical representation
 - syntactical parser PALAVRAS (*Eckhard Bick*)

QA System: How it works

- Question Analysis

- syntactical parser

- Example:

- *“Quem é Boaventura Kloppenburg ?”*

```
QUE: fcl
=SC:pron-indp('quem' <interr> M/F S)    Quem
=P:v-fin('ser' PR 3S IND)              é
=SUBJ:prop('Boaventura_Kloppenburg' <org> F S) Boaventura_Kloppenburg
=?
```

QA System: How it works

- Question Analysis
 - partial semantic representation
 - First-Order Logic expression
 - Discourse Representation Structures (DRS)

```
query(clef0012,  
      [ name(_199, 'Boaventura_Kloppenburg' , ['F', 'S', 'Boaventura_Kloppenburg' ],  
          [ ] ) ,  
        q(_200, 'quem' , ['M/F', 'S', 'quem' ],  
          [ ] ) ,  
        'ser'(_199,_200,  
              [ modif(verb,'ser', ['PR','3S','IND'] ) ] ) ,  
        [ modif(ordObsu), modif(casoSerNProp) ] ),  
      [ ref(_199), ref(_200) ] ).
```

QA System: How it works

- Get Candidate Answer Documents
 - Lucene search query
 - question text
 - first question on group, it's answer – if topic related
 - ontology driven terms
 - detailed analysis
 - First-Order Logic expression (DRS) for the selected docs
 - Facts List: '*question dedicated kb*'

QA System: How it works

- Solver Engine
 - *logic solver*
 - pragmatic interpretation of the query DRS over the Facts List
 - unify query DRS referents with referents from Facts List
 - *ad-hoc solver*
 - specific cases, where the possible solution can be directly detected in the text
 - answer
 - weight, support text, document identifier
 - CLEF Answer
 - highest weight value from all possible solutions

QA System: How it works

Quem é Boaventura Kloppenburg?

Valor	Peso	Documentos que suportaram a resposta
bispo de Novo Hamburgo (RS)	92	<p>doc FSP940317-013 respSupport [92] O aborto é crime sob qualquer ótica", disse dom Boaventura Kloppenburg, bispo de Novo Hamburgo (RS).[h01]</p> <p>doc FSP940424-031 respSupport [92] É o caso de d. Boaventura Kloppenburg, bispo de Novo Hamburgo (RS).[h01]</p> <p>(2)</p>

Quando é que foi assinada a Lei Áurea?

Valor	Peso	Documentos que suportaram a resposta
13 de maio de 1888	93	<p>doc w8051 respSupport [95] A Lei Áurea foi assinada em 13 de maio de 1888, extinguindo a escravidão no Brasil.[h11_1].[h11_1].[h11_1.]</p> <p>doc w77212 respSupport [95] A escravidão só foi oficialmente abolida no Brasil com a assinatura da Lei Áurea , em 13 de maio de 1888 .[h11_1.]</p> <p>(2)</p>
13 de maio	91	<p>doc w8051 respSupport [97] Embora muitos discordem, no Brasil, tem-se que o ato da assinatura da ""Lei Áurea"" assinado pela filha do Imperador Pedro II do Brasil Dom Pedro II , Princesa Isabel (1846 - 1921) adquiriu esse nome por ser o dia 13 de maio , escolhido pela princesa, uma das principais datas que acompanham a história brasileira e refere-se ao aniversário de seu falecido bisavô, Dom João VI e não o instrumento (uma pena de ouro) usado pela princesa no ato da sanção e como os historiadores tem repassado ao povo desse país.[h11_1.]</p> <p>(1)</p>

Note on Time Expressions

- “Quando” questions (When, in English)
 - temporal expressions (*this month, next year, on Friday..*)
 - ontology terms related to “*conceito temporal*”
 - document date of publication
 - sentence verb time (future/present/past)

- Example

Interrogação q1157045981031

introduzida em Thu Aug 31 18:39:41 WEST 2006

Questão

Quando é a Feira da Luz?

[processar](#)
[ver solução](#)

[voltar](#)

Note on Time Expressions

interrogação recebida em Thu Aug 31 18:39:41 WEST 2006

Questão

Quando é a Feira da Luz?

Resposta(s)

Valor	Documentos que suportaram a resposta
2006-09	ut1159807260026 ut1159396642316 ut1159395455716 ut1159289776222 ut1157118194242 ut1157118180468 (6)
2006-09-28	ut1159820453835 ut1159807241747 (2)
2009	ut1159806593027 (1)
2007-01	ut1160262332397 (1)
2006-11-29	ut1159806646578 (1)
2006-11-14	ut1160046857790 (1)
2006-10-5	ut1159806575855 (1)
2006-10-3	ut1159789348852 (1)
2006-10-21	ut1159806629514 (1)
2006-10-1	ut1159789337232 (1)
2006-02-11	ut1160348079009 (1)
2005-12	ut1160262363905 (1)
2005	ut1157118215403 (1)
1968-09-27	ut1159806607358 (1)
1958-09	ut1158510355750 (1)

(15 respostas em 21 documentos relacionados)

on Oct 02: "... foi no mês passado."

on Sep 26: "... é este mês."

on Sep 01: "... realiza-se este mês."

on Sep 01: "... é em Setembro."

on Sunday, Oct 08, 2006: "... será em Janeiro."

on Monday, Oct 02, 2006: "... é na quinta-feira."

on Oct 08, 2006: "... foi em Dezembro."

Results

Question Type	#	Right	Wrong	Unsupported	Inexact	Accuracy
Nil	111 returned	12	99	0	0	10.81%
Temporally Restricted	19	4	15	0	0	21.05%
Definition	31	19	5	0	0	61.29%
Lists	10	2	8	0	0	20.00%
Factoids	159	63	90	1	1	39.62%
All Questions	200	84	103	1	1	42.00%

correct answers: 84 of 200

- improvement since our last participation (50 correct answers)

Results by Type

- Most errors
 - could not find an answer
 - 90 wrong NIL returned values
- List and Temporally Restricted questions
 - accuracy around 20%
- Factoids
 - accuracy close to 40%
- Definition questions
 - best relative accuracy: 61.29%.

Discussion

- this year
 - improved results since our previous participation

- some reasons for incorrect answers
 - no candidate documents (Lucene term search), resulting in NILs
 - morpho-syntactical analyser mistakes
 - semantic analyser problems (DRS generation)
 - inference problems and other
 - rules for limited number of scenarios
 - ontology limitations

Discussion

- Future Work
 - improve the way the Lucene text search query is built
 - improve the Senso Ontology
 - other Source Languages besides Portuguese