From legal texts to legal ontologies and question-answering systems

Paulo Quaresma
pq@di.uevora.pt
Spoken Language Systems Lab / Dept. of Informatics
INESC-ID, Lisbon / University of Évora
Portugal
Some contextualization... :-) 

Paulo Quaresma  
pq@di.uevora.pt  
Spoken Language Systems Lab / Dept. of Informatics  
INESC-ID, Lisbon / University of Évora  
Portugal
Some contextualization... :-)

From legal texts to legal ontologies and question-answering systems
Overview

– Motivation
– Challenges and objectives
– General architecture
– Examples
– Conclusions and future work
Motivation

– **Access to laws or Court decisions**
  
  • FIRE track “Adhoc retrieval from legal documents” (adapted)
    
    – U: I'm searching for situations where someone marries to an already married person.
    
    – S: In case 123 Mary married Peter, who was already married to Susan.
    
    – OR
    
    – S: Section 5 of the Hindu Marriage Act states: “...”
Motivation

– **Access to laws or Court decisions**
  
  • FIRE track “Adhoc retrieval from legal documents” (adapted)
    
    – U: I'm searching for situations where someone divorced because the husband/wife left home.
    
    – S: In case 999 Ann divorced John after he returned to UK and sent no news for more than 7 years.
    
    – OR
    
    – S: Section 13 of the Hindu Marriage Act states: “...”
Motivation

– Access to laws or Court decisions
  • FIRE track “Adhoc retrieval from legal documents” (adapted)
    – U: I'm searching for situations where someone marries to an already married person.
    –
    – Google-like queries?
      » “married to already married”
      » “bigamy”
    – Not likely to obtain the relevant laws and cases!
Motivation

– Access to laws or Court decisions
  • FIRE track “Adhoc retrieval from legal documents” (adapted)
    – U: I'm searching for situations where someone divorced because the husband/wife left home.
    – Google-like queries?
      » “divorce” “leave”? “abandon”?
    – Not likely to obtain the relevant laws and cases!
Motivation

– Access to laws or Court decisions
  • FIRE track “Adhoc retrieval from legal documents” (adapted)
    – U: I'm searching for situations where someone divorced because the husband/wife left home.
    – Previous classification of laws and cases accordingly with a taxonomy of legal concepts?
      » Requires user knowledge of the taxonomy
    – Not likely to be adequate to all users/citizens!
Motivation

• Information retrieval systems based in
  – Free text searches
  – Catalogues / thesaurus
• Are limited and inadequate for complex queries/situations!
Motivation

• “Google-like” systems:
  – Unable to fully represent user intentions and document semantic content.
Motivation

• Catalogue-based systems:
  – Structured knowledge (thesaurus)
  – Implies that the user needs to “know” the knowledge representation structure → “unfriendly” interface
  – Unable to represent user intentions and document semantic content
Motivation

• (We need) Content-based systems:
  – Ontologies representing concepts and relations
  – Instances of ontologies representing document semantic content
    • Syntactic and semantic document analysis
  – Queries
    • Syntactic and semantic interpretation
    • Answer as an inference process
Motivation

• Example revisited:
  – U: I'm searching for situations where someone marries to an already married person.
  – ...

Motivation

• Example revisited:
  – Ontology
    • Concepts:
      – “person” “marriage”
Motivation

• Example revisited:
  – “someone marries...”
    • An event, which is an instance of a “marriage”
  – “… already married person”
    • Another “marriage” event, which occurred before the previous one
Motivation

• Example revisited:
  – event(E), time(E, TE), action(E, marriage(X, Y))
  – event(P), time(P, TP), action(E, marriage(Y, Z)),
    X != Z, TP < TE.
Challenges and objectives

• To be able:
  – To automatically analyse texts and to extract the conveyed information
  – To represent the extracted information in adequate knowledge representation structures
  – To support inferences over the knowledge
Challenges and objectives

• To be able:
  – To automatically analyse texts and to extract the conveyed information
  • Natural language processing techniques
    – Lexical, syntactical, semantic, and pragmatic interpretation of texts
    – Symbolic-based and/or statistical-based approaches
Challenges and objectives

• To be able:
  – To represent the extracted information in adequate knowledge representation structures
• Ontologies
  – Creation
  – Mapping
  – Merge
  – Population
Challenges and objectives

• To be able:
  – To support inferences over the knowledge
  • Inference-engine over knowledge represented by ontologies
Architecture (UEvora approach)

• Two main modules
  – Information extraction and representation
    • Extracts information from documents and creates a knowledge base;
  – Information retrieval
    • Processes queries, accesses the knowledge base and generates answers.
Document Collection
(Portuguese sentences)

Syntactical Analyser
Palavras

New document collection
(sentences parsed)

Semantical interpretation
(DRSs generation)

New Collection of Documents
(DRSs)

Ontology

Semantic/Pragmatic interpretation

Knowledge base
(Facts extracted from documents)
Information Extraction

• Syntactical analysis
  – PALAVRAS [Eckhard Bick] (Portuguese)
  – C&C Parser [Clark and Curran] (English)

• (Partial) semantical analysis
  – DRS: entities and conditions [Kamp]
    • Boxer [Bos]

• Semantical/pragmatical analysis
  – Ontology+DRS -> new DRS -> KB
Information Retrieval

- Query syntactic analysis
- Query semantic analysis
- Semantic-pragmatic query analysis
- Answer inference
  - Knowledge base inference
    - Logic programming framework
      - PROLOG + ISCO [Abreu01]
    - OWL inference engines: Pellet, F-OWL, Euler, JENA, JESS, ...

From legal texts to legal ontologies and question-answering systems
Example

- Syntactical analysis (C&C parser)
  - Veronica married Peter who was married to Susan.

Veronica (NNP N) married (VBD (NP\NP)/(S\dcl\NP)/NP) Peter (NNP N) who (WP (NP\NP)/(S\dcl\NP)) was (VBD (S\dcl\NP)/(S\pss\NP)) married (VBN (S\pss\NP)/PP) to (TO PP/NP) Susan. (NNP N)

From legal texts to legal ontologies and question-answering systems
Example

- Semantical analysis (Boxer)
Example

• Semantic-pragmatic analysis

\[
\begin{align*}
\text{x0} & \quad \text{x1} & \quad \text{x2} \\
\text{name}(\text{x0}, \text{veronica}, \text{per}) & \\
\text{name}(\text{x1}, \text{peter}, \text{per}) & \\
\text{name}(\text{x2}, \text{susan}, \text{nam}) & \\
\end{align*}
\]

\[
\begin{align*}
\text{x3} & \quad \text{x4} \\
\text{marry}(\text{x3}) & \\
\text{event}(\text{x3}) & \\
\text{agent}(\text{x3}, \text{x1}) & \\
\text{patient}(\text{x3}, \text{x2}) & \\
\text{marry}(\text{x4}) & \\
\text{event}(\text{x4}) & \\
\text{agent}(\text{x4}, \text{x0}) & \\
\text{patient}(\text{x4}, \text{x1}) & \\
\end{align*}
\]
Example

• Knowledge base - Facts

name(x0, veronica). name(x1, peter). name(x2, susan).

event(x3). event(x4). rel('<', x3, x4).

marry(x4). agent(x4, x0), patient(x4, x1).
marry(x3). agent(x3, x1). patient(x3, x2).
Example

• Knowledge base - Rules

\[\text{marry}(E, X, Y) \leftarrow \text{marry}(E), \text{agent}(E, X), \text{patient}(E, Y).\]
\[\text{marry}(E, X, Y) \leftarrow \text{marry}(E), \text{agent}(E, Y), \text{patient}(E, X).\]
\[\ldots\]
Question-Answering

- Situations where someone married to an already married person?
  - “Who married to a married person?”
- After query analysis:
  - P1:
    - event(E1). event(E2).
    - rel('<', E2, E1).
    - marry(E1, P1, P2).
    - marry(E2, P2, P3).
Question-Answering

- Inference engine (logic programming): gnuProlog
  
  - ?- name(P, N),
    event(E1). event(E2),
    rel('<', E2, E1),
    marry(E1, P, P2), marry(E2, P2, P3).

  - P = x0, N = veronica
Evaluation - CLEF

- 200 queries – CLEF05 (newspaper domain)
  - 25% correct and well-supported answers
  - 52% without answer!
  - 1.5% correct answers but not well-supported
  - 11% not correct answers—too many (or too few) words
  - 10.5% incorrect answers
Evaluation - CLEF

• Main problem
  – 52% questions without answer!

• Cause?
  – Mainly because the ontology was not “good enough” – it didn't represent the referred concepts and their relations; it didn't allow to represent correctly the document semantic content!
Ontologies

• The proposed approach requires:
  – Ontologies to represent concepts and relations
  – The inference of instances \(\rightarrow\) ontology population
Ontologies

• Problems:
  - How to choose/create the ontology?
    • Use existent ontologies
    • Automatic merge of existent ontologies
    • Automatic creation of domain specific ontologies from the analysis of the documents
Ontologies

- Ontology mapping and mapping:
  - Cássia Santos PhD [Santos2009]
- Multi-agent system based on argumentation and/or negotiation
  - Lexical agent
  - Structural agent
  - Semantical agent
Ontologies

• Ontologies mapping [Santos2009]

Matcher A  Matcher B  ...  Matcher N

Composite Framework

Consensus
Ontologies

• Ontologies mapping
  • OAEI: Ontology Alignment Evaluation Initiative
  • Dataset: “real” ontologies
    • BibTex MIT, BibTex UMBC, BibTex Karlsruhe, INRIA
  • Matchers: OAEI
    • ASMOV, DSSim, Falcon, Lily, Ola, OntoDNA, PriorPlus, RiMON, Sambo, SEMA, TaxoMap, XSOM, ...
Ontologies

• Automatic creation of domain specific ontologies – [Saias2010]
  – Lexical, syntactical and semantical analysis of documents
  – Identification of:
    • NER – Named Entities Recognition
    • Actions
    • Semantic Role Labeling
      – Relations – triples subject-verb-object
Ontologies

• Automatic creation of domain specific ontologies – [Saias2010]
  – Use of:
    • Top-level/upper ontologies (SUMO, DOLCE, ...)
    • Wordnet
    • Wikipedia
  – To map new concepts and to extend the existent ontology
Ontologies

• Automatic creation of domain specific ontologies – [Saias2010]
  – Example:
    • Top-level ontology has the concept “animal”
    • Document refers “cat”
    • Wordnet states that a cat is a mammal and mammals are animals
    • So, the to-level ontology can be extended to have the two new concepts!
Ontologies

- How to populate ontologies?
  - DRS represent document content
  - Entities and conditions can be used to populate ontologies
    - person(a), name(a, 'Veronica')
    - Class “person”; “name” is a property of class “pessoa”
    - “a” is an instance of class person
Ontologies

• Another possible approach:
  – How to populate ontologies?
    • Machine learning approach
      – Information Extraction techniques
        • Support-Vector Machines
        • Conditional Random Fields
      – Requires a previously tagged set of documents (learning set), which should be representative of the document collection
Evaluation - CLEF

- 200 queries – CLEF08 (newspaper domain)
  - 25% → 46.5% correct and well-supported answers
  - 52% → 32% without answer!
  - 1.5% → 1% correct answers but not well-supported
  - 11% → 5.5% not correct answers– too many (or too few) words
  - 10.5% → 15% incorrect answers
Conclusions

- Proposed an ontology/knowledge-based (legal) question-answering system
  - Natural language processing
  - Knowledge bases / Ontologies
  - Logic Programming framework
Future Work

• Improve “all” modules
• Improve ontologies!
  – Use of adequate top-level ontologies
  – Ontology merge
  – Extract more instances – ontology population
    • Symbolic + machine learning approaches
• Apply the system to the legal domain and evaluate the results.