"Improving QA using NER"

A. Toral, E. Noguera, F. Llopis, R. Muñoz
{atoral,elisa,llopis,rafael}@dlsi.ua.es
DLSI - University of Alicante
Contents

- Introduction
- System description
- Evaluation
- Conclusions & future work
Introduction

• Question Answering (QA)
  – given a query, return exact answer
  – uses Natural Language analysis
  – active field of research (TREC, CLEF)
  – interactive systems, time constraint
Introduction

• Information Retrieval (IR)
  - given a query, returns relevant docs
  - statistical techniques, efficient

• Passage Retrieval (PR)
  - IR improvement
  - returns fragments instead of whole docs
  - ability to locate most relevant fragments
Introduction

● Named Entity Recognition
  – introduced at MUC conferences
  – detect and classify entities in text
  – Types:
    • person, location, organization
    • date, time, numeric expressions
<table>
<thead>
<tr>
<th></th>
<th>QA</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>input</td>
<td>query</td>
</tr>
<tr>
<td>1</td>
<td>output</td>
<td>answer</td>
</tr>
<tr>
<td>2</td>
<td>resources</td>
<td>set docs</td>
</tr>
<tr>
<td>3</td>
<td>load</td>
<td>high (NLP)</td>
</tr>
<tr>
<td>4</td>
<td><strong>IR+QA</strong></td>
<td><strong>query</strong></td>
</tr>
<tr>
<td></td>
<td>QA only applied to relevant docs</td>
<td>rel docs</td>
</tr>
</tbody>
</table>
Introduction

• Approach
  – reduce even more input of QA
  – apply NER to the output of IR
    • QA answer concides entity type
      – Who is the president of Ireland? --> Person
    • Study the query, extract answer type
    • Tag relevant passages with that entities
    • Apply QA only to passages with entities
System description

Diagram:
- Documents
- NER Gazeteers
- NER Rules
- IR-n
- Passages IR-n
- Queries
- Query Rules
- NER
- Passages NER
- Evaluation
System description

- IR-n (passage retrieval system)
  - Indexing module
    - transforms set docs into structures
  - Retrieval module
    - sorts the passages according to a query
      - select relevant passages
      - sort passages
System description

- DRAMNERI (NER system)
System description

• Answer type recognition
  - Interrogative particles (who --> person)
    • **Who** is the president of Ireland? --> person
  - Interrogative particle + Hyponym
    • which + hyponym(location) --> location
      - **Which** european country consumes more alcohol?
System description

- **Named Entity Identification**
  - Groups of capitalised words or numbers (jointed by prepositions)

- **Named Entity Recognition**
  - Triggers, key words
    - i.e. *Mr. Smith* --> person
  - Rules (regular expressions)
    - i.e. rule "PER PREP ART PER --> person" would match "Jorge de la Varga"
Evaluation

• Data collection, EFE1994
  - QA CLEF 2004 (spanish)
  - 215,738 docs
  - 200 queries
Answer entity type detection
- 150 out of 200 correctly classified
- From the remaining 50:
  - 46 not classified (do not fit to any entity type). NER not applied
  - 4 incorrectly classified
- Recall 75%, Precission 97.4%
Evaluation

• Measure used
  – Mean Reciprocal Rank (MRR)
  – Value assigned to each question
    • inverse value of 1st passage with the answer
    • 0 if no passage has the answer
## Evaluation

### IR-n (1) vs. IR-n + DRAMNERI (2)

<table>
<thead>
<tr>
<th>Sentences per Passage</th>
<th>MRR 1</th>
<th>MRR 2</th>
<th>output improvement</th>
<th>output reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.15</td>
<td>0.17</td>
<td>13%</td>
<td>42%</td>
</tr>
<tr>
<td>4</td>
<td>0.19</td>
<td>0.21</td>
<td>11%</td>
<td>32%</td>
</tr>
<tr>
<td>6</td>
<td>0.21</td>
<td>0.23</td>
<td>10%</td>
<td>28%</td>
</tr>
<tr>
<td>8</td>
<td>0.22</td>
<td>0.24</td>
<td>9%</td>
<td>26%</td>
</tr>
<tr>
<td>10</td>
<td>0.22</td>
<td>0.24</td>
<td>4%</td>
<td>24%</td>
</tr>
<tr>
<td>12</td>
<td>0.22</td>
<td>0.24</td>
<td>4%</td>
<td>23%</td>
</tr>
<tr>
<td>14</td>
<td>0.23</td>
<td>0.24</td>
<td>4%</td>
<td>22%</td>
</tr>
<tr>
<td>16</td>
<td>0.23</td>
<td>0.24</td>
<td>4%</td>
<td>21%</td>
</tr>
<tr>
<td>18</td>
<td>0.23</td>
<td>0.24</td>
<td>4%</td>
<td>21%</td>
</tr>
<tr>
<td>20</td>
<td>0.23</td>
<td>0.24</td>
<td>4%</td>
<td>21%</td>
</tr>
<tr>
<td>30</td>
<td>0.24</td>
<td>0.25</td>
<td>4%</td>
<td>21%</td>
</tr>
<tr>
<td>40</td>
<td>0.24</td>
<td>0.25</td>
<td>4%</td>
<td>20%</td>
</tr>
</tbody>
</table>
Conclusions & future research

• NER plays an important role in QA systems
  - 26% reduction, 9% MRR increment
  - Efficient approach:
    • IR-n 94.8% while NER 5.2% of time
Conclusions & future research

● Future lines of research
  – Deeper study about data reduction and its implication about data quality
    • return only sentences with entities, etc.
  – Embed NER into PR indexed data
    • speed up searching, NER is applied once
  – Apply hybrid NER
    • avoid dependency on explicit knowledge
Questions?