Ontology-driven Vaccination Information Extraction

Liliana Ferreira, António Teixeira, João Paulo Silva Cunha

Institute of Electronics and Telematics Engineering of Aveiro, Portugal

June 2008
1 Motivation

2 Knowledge and Rules Representation

3 System Overview

4 Results

5 Future Work
The creation of computable interpretable representations allows the development of clinical decision support systems.
The creation of computable interpretable representations allows the development of clinical decision support systems.

Vaccination area - publicly availability of the Portuguese Vaccination Plan (PVP) in portuguese!
The creation of computable interpretable representations allows the development of clinical decision support systems.

Vaccination area - publicly availability of the Portuguese Vaccination Plan (PVP) in portuguese!

Method used to model and populate a vaccination ontology;
The creation of computable interpretable representations allows the development of clinical decision support systems.

- Vaccination area - publicly availability of the Portuguese Vaccination Plan (PVP) in portuguese!
- Method used to model and populate a vaccination ontology;
- System which recognizes vaccination information on medical texts.
At birth, it is recommended the vaccine against tuberculosis (BCG) and the first dosage of the vaccine against hepatitis B (VHB), provided that the weight of the new-born is greater than or equal to 2000g.

Sentence with Linguistic Patterns

AT [age] [vaccine] {against} [disease] ([acronym]) AND [dosage] [vaccine] {against} [disease] ([acronym]), {restriction} [weight].
Partial View of the Vaccination Ontology

- Disorder
- Disease
- Reaction
- Vaccine
- Body Part
- Interaction
- Drug Interaction
- Physical Interaction
- Allergy
- Weight
- Pregnancy
- age
- dosage
- applied_on
- against
- originates
- has an
- inst_of
- Arm
- Leg

Liliana Ferreira, António Teixeira, João Paulo Silva Cunha

Ontology-driven Vaccination Information Extraction
Ontology Population

- Ontology written in OWL;
- Rules developed in SWRL:
  a) prevention of vaccine interaction with a component of the vaccine (e.g. allergy interaction check);
  b) prevention of vaccine interaction with physical condition of the patient (e.g. pregnancy)

Ontology-driven Vaccination Information Extraction
Ontology Population

- Ontology written in OWL;
- Rules developed in SWRL:
  - prevention of vaccine interaction with a component of the vaccine (e.g. allergy interaction check);
  - prevention of vaccine interaction with physical condition of the patient (e.g. pregnancy)
- Information about the entities was automatically extracted and automatically added to the ontology.
- Inter-instances relationships were filled, and an attempt to automatize this step was developed.
Annotated Vaccination Plan

Liliana Ferreira, António Teixeira, João Paulo Silva Cunha

Ontology-driven Vaccination Information Extraction
Ontology-driven Vaccination Information Extraction

Liliana Ferreira, António Teixeira, João Paulo Silva Cunha
First Approach: Inter-instance relationships added manually as OWL properties.

Frequent Pattern Mining, using Association Rule Mining.
Association Rules describe how often entities are mentioned together:

- *disease* $\Rightarrow$ *vaccine* (80%) states that 4/5 times a disease individual (*Tuberculose*) is mentioned it is followed by a vaccine individual (*BCG*).
- The relationship *BCG ‘is_against’ Tuberculose* can be inferred.
Relationships between individuals

- $I$ as the set of all entities identified,
  $$I = \{ \text{age, disease, acronym, bodypart, dosage, interaction, reaction, weight} \}$$

- $D$ a list with the annotations made for each sentence of the PVP.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Annotation vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>${ \text{age, acronym} }$</td>
</tr>
<tr>
<td>101</td>
<td>${ \text{age, age, weight} }$</td>
</tr>
<tr>
<td>102</td>
<td>${ \text{age, dosage, disease, disease, acronym, dosage, acronym} }$</td>
</tr>
<tr>
<td>103</td>
<td>${ \text{acronym, acronym} }$</td>
</tr>
<tr>
<td>104</td>
<td>${ \text{age, acronym, acronym} }$</td>
</tr>
</tbody>
</table>
Results Information Extraction

Results for each annotation type

<table>
<thead>
<tr>
<th>Correct matches</th>
<th>DISEASE ACRONYM</th>
<th>AGE</th>
<th>BODYPART</th>
<th>DOSAGE</th>
<th>INTERACTION</th>
<th>REACTION</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>225</td>
<td>294</td>
<td>181</td>
<td>14</td>
<td>90</td>
<td>10</td>
<td>156</td>
<td>8</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>225</strong></td>
<td><strong>188</strong></td>
<td><strong>14</strong></td>
<td><strong>93</strong></td>
<td><strong>10</strong></td>
<td><strong>156</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td>Recall</td>
<td>1,00</td>
<td>1,00</td>
<td>0,96</td>
<td>1,00</td>
<td>0,97</td>
<td>1,00</td>
<td>1,00</td>
</tr>
<tr>
<td>Precision</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
</tr>
<tr>
<td>F - measure</td>
<td>1,00</td>
<td>1,00</td>
<td>0,98</td>
<td>1,00</td>
<td>0,98</td>
<td>1,00</td>
<td>1,00</td>
</tr>
</tbody>
</table>

Overall Information Extraction Results

<table>
<thead>
<tr>
<th>Correct Matches</th>
<th>Partially Correct matches</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>978</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Recall</td>
<td>0,991</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>1,000</td>
<td></td>
</tr>
<tr>
<td>F - measure</td>
<td>0,995</td>
<td></td>
</tr>
</tbody>
</table>
Relationships between individuals

Results

Number of rules = 3

- DISEASE ⇒ ACRONYM 79.68%
- AGE ⇒ ACRONYM 69.73%
- ACRONYM ⇒ DOSAGE 20.83%
Future Work

- Improve inter-instance relationship extraction: link-grammars;
- *Active* or a *passive* validation of medical reports;
- Use available knowledge sources like wiki dictionaries.
The End...

Thank you...

Liliana Ferreira
5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2008)
12-13 June, 2008 - Barcelona - Spain