Acquisition of axioms in ontology learning

Ana B. Rios-Alvarado
Information Technology Laboratory
CINVESTAV - Tamaulipas, Mexico

October, 31th 2012

Supervisor:
Dr. Ivan Lopez-Arevalo
Outline

1. Introduction
2. The research problem
3. Objectives
4. The proposed approach
5. Experiments and Results
6. Conclusions
7. References
Motivation

- Ontologies allow the construction of more “intelligent” applications: semantic search, automated reasoning, among others.
Some definitions...

What is an **ontology**?
- “... an element that defines the **basic terms** and **relations** contained into the **vocabulary** of a topic area as the rules for combining terms and relations to define extensions of a conceptualization” [Neches et al., 1991]

What is **ontology learning**?
- “Ontology learning is the set of methods and techniques used for building an ontology from scratch, enriching, or adapting an existing ontology in a semi-automatic fashion using several knowledge and information sources” [Gomez-Perez, 2003]
Some definitions ...

What is ontology learning from text?

- It is the process of deriving high-level concepts and relations as well as the axioms from unstructured information to form an ontology.
Some definitions ...

**What are the elements of an ontology?**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts (classes)</td>
<td>Ideas to formalize</td>
</tr>
<tr>
<td>Taxonomic relationships</td>
<td>Relation <em>is-a</em> or <em>subClassOf</em></td>
</tr>
<tr>
<td>Non-taxonomic relationships</td>
<td>Interaction between elements</td>
</tr>
<tr>
<td>Instances</td>
<td>“real-world” objects</td>
</tr>
<tr>
<td>Axioms</td>
<td>Theorems on relations to be satisfied by elements</td>
</tr>
</tbody>
</table>
Example: Travel ontology

- Taxonomic relation: Beach is a Destination
- Non-taxonomic relation: Activity isOfferedAt Destination
- Axiom: disjointWith(UrbanArea, RuralArea)
- Instance: “Atlanta” is an instance of City
Ontology learning from text

The aspects and tasks in ontology learning from text are structured as a set of layers [Cimiano, 2006]

∀x (country(x) → ∃y capital_of(y,x) ∧ ∀z (capital_of(z,x) → y=z))

disjoint(river, mountain)
capital_of ≤_r located_in
flow_through( domain:river, range:geopolitical_entity )
capital ≤_c city, city ≤_c geopolitical_entity
c := country := < i(c), ||c||, Ref_c(c) >
{country, nation}
river, country, nation, city, capital ...
Some techniques for obtaining the vocabulary

- Linguistic analysis
- Statistical analysis
  - Term weightning TD-IDF
  - Mutual Information
- Based on WordNet
- Latent Semantic Semantic Analysis
- Text clustering

The use of clustering techniques relies on the assumption that similar terms share similar syntactic contexts
Some techniques for obtaining taxonomic relations

- Lexical databases (for example, WordNet)
- Linguistic approaches
- Co-occurrence analysis
- Lexico-syntactic patterns

The lexical patterns occur frequently in many text genders

- Use of web search and Wikipedia as knowledge source

The web search uses as knowledge base the whole Web
Some techniques for obtaining taxonomic relations

Lexical patterns

- An expression $A$ is a **hyponym** of an expression $B$ if the meaning of $B$ is part of the meaning of $A$ and $A$ is a subordinate of $B$. By contrast, an expression $B$ is a **hypernym** of $A$ if $B$ includes the meaning of $A$ and $B$ is a superior to $A$.

<table>
<thead>
<tr>
<th>Hearst’s Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$, and other $B$</td>
</tr>
<tr>
<td>$A$, or other $B$</td>
</tr>
<tr>
<td>$A$ is a $B$</td>
</tr>
<tr>
<td>$B$, such as $A$</td>
</tr>
<tr>
<td>$B$, including $A$</td>
</tr>
<tr>
<td>$B$, specially $A$</td>
</tr>
</tbody>
</table>
Some techniques for obtaining taxonomic relations

Querying the Web

Given a query, it is possible to do the analysis of the number of hits retrieved by a search engine.
Some the techniques for automatic axiom extraction

**What is an axiom?**

Assertions (including rules) in a logical form that together comprise the overall theory that the ontology describes in its domain of application.

<table>
<thead>
<tr>
<th>Axiom</th>
<th>DL Syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>subClassOf</td>
<td>$C_1 \sqsubseteq C_2$</td>
<td>Human $\sqsubseteq$ Animal $\sqcap$ Biped</td>
</tr>
<tr>
<td>equivalentClass</td>
<td>$C_1 \equiv C_2$</td>
<td>Man $\equiv$ Human $\sqcap$ Male</td>
</tr>
<tr>
<td>disjointWith</td>
<td>$C_1 \sqsubseteq \neg C_2$</td>
<td>Male $\sqsubseteq \neg$ Female</td>
</tr>
<tr>
<td>sameIndividualAs</td>
<td>${x_1} \equiv {x_2}$</td>
<td>{President Bush} $\equiv$ {G. W. Bush}</td>
</tr>
<tr>
<td>differentFrom</td>
<td>${x_1} \sqsubseteq \neg {x_2}$</td>
<td>{john} $\sqsubseteq \neg$ {peter}</td>
</tr>
</tbody>
</table>
Some the techniques for automatic axiom extraction

<table>
<thead>
<tr>
<th>Kind of axioms</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class expression</td>
<td>Allow relationships to be established between class</td>
<td>subClassOf(Car, CompactCar)</td>
</tr>
<tr>
<td>Object property</td>
<td>Characterise and establish relationships between object property expressions</td>
<td>subPropertyOf(hasMother, hasParent)</td>
</tr>
<tr>
<td>Assertion</td>
<td>Axioms about individuals</td>
<td>sameIndividualAs(President Bush, G. W. Bush)</td>
</tr>
</tbody>
</table>

- Lexical patterns [Del Vasto Terrientes, et al., 2010]
- Transforming rules [Volker, et al., 2007a] and statistical analysis [Volker, et al., 2007b][Volker and Rudolph, 2008]
- Inductive logic programming [Lisi and Straccia, 2011]
Related work

∀x ( country(x) → ∃y capital_of(y,x) ∧ ∀z ( capital_of(z,x) → y=z ) )

disjoint(river, mountain)
capital_of ≤_R located_in
flow_through( domain:river, range:geopolitical_entity )
capital ≤_c city, city ≤_c geopolitical_entity
c := country := < i(c), ||c||, Ref_c(c) >
{country, nation}
river, country, nation, city, capital ...

Völker et al., 2007a
Völker and Rudolph, 2008
Völker et al., 2007b
Lisi and Straccia, 2011
Del Vasto Terrientes et al., 2010
Cimiano et al., 2005
Jiang and Tan, 2010
Ochoa et al., 2011
Related work

Ontology learning from text approaches

<table>
<thead>
<tr>
<th>Phase</th>
<th>Technique</th>
<th>Domain</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Formal concept analysis</td>
<td>Tourism and Finance</td>
<td>Statistical analysis</td>
</tr>
<tr>
<td>Taxonomic relations</td>
<td>Formal concepts are partially ordered</td>
<td>Sport Event and Terrorism</td>
<td>Rule based algorithm</td>
</tr>
<tr>
<td>Non-taxonomic relations</td>
<td>-</td>
<td>Oncology and Finance (Spanish texts)</td>
<td>Semantic roles</td>
</tr>
<tr>
<td>Axioms</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
The problem

- The most of learned ontologies are limited to non-taxonomic relationships
The problem

- The most of learned ontologies are limited to non-taxonomic relationships

- Limitations in previous works:
  - Assignment a label to a group with a semantic context
  - Dependency on corpus or linguistic databases for a specific domain
  - Limited contexts into corpus for the lexical pattern matching
  - Ontology learning process without the axiom layer
  - The axiom learning approaches start from seed concepts and structures well defined
Research questions...

- Is it possible to build an ontology from textual resources with high level of expressiveness?
- Is it possible to extract the vocabulary and their relationships by text clustering and web search?
- Is it possible to extract axioms using natural language processing techniques?
Research questions...

- Is it possible to build an ontology from textual resources with high level of expressiveness?
- Is it possible to extract the vocabulary and their relationships by text clustering and web search?
- Is it possible to extract axioms using natural language processing techniques?

Hypothesis

Considering that text clustering can get the vocabulary of a corpus, web search can obtain taxonomic relationships, and natural language processing techniques allows discover axioms, it is possible adapt and integrate these techniques for ontology learning.
Objectives

General objective

Obtain an approach for ontology learning getting the vocabulary, taxonomic relationships, and key axioms from textual resources
Objectives

General objective
Obtain an approach for ontology learning getting the vocabulary, taxonomic relationships, and key axioms from textual resources

Particular objectives
- Obtain a method to get the vocabulary from text using a clustering algorithm
- Obtain a method to extract taxonomic relationships between concepts
- Obtain an unsupervised method for axiom learning
The proposed approach

Acquisition of axioms in ontology learning

The proposed approach
The proposed approach: main tasks

1. **Pre-processing**
   - collect corpus, POS tagging, lemmatization, delete stopwords

2. **Extract topics**
   - obtain the representation model of input corpus’s content, feature selection (nouns/verbs)

3. **Discovering taxonomic relations**
   - build queries, execute web search

4. **Axiom learning**
   - linguistic analysis, named entity recognition

5. **Evaluation of the obtained ontology**
   - improvement the obtained ontology, evaluation and comparison with “gold standard” ontologies
Topic extraction

- Axiom schema
- Instances
- Taxonomic Relationships
- Vocabulary

Topic extraction
Acquisition of axioms in ontology learning

The proposed approach

**Topic extraction**

This phase involves the construction of a representation model from corpus and the implementation of clustering algorithm.
Topic extraction: the representation model

For obtaining the vocabulary:

- Using linguistic analysis the pairs $\langle$verb, subject$\rangle$ and $\langle$verb, object$\rangle$ are considered for building a pair-term matrix

\[
\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\text{Verb} & \text{chance} & \text{village} & \text{turtle} & \text{effort} & \text{Hanoi} & \text{performance} & \text{season} & \text{Vietnam} & \text{dune} \\
\hline
\text{offer} & \checkmark & \checkmark & & & & & & & \\
\text{see} & & & \checkmark & \checkmark & \checkmark & \checkmark & & & \\
\text{make} & & & & & \checkmark & & & & \\
\text{visit} & & & & & & \checkmark & \checkmark & & \\
\text{climb} & & & & & & & & \checkmark & \checkmark \\
\hline
\end{array}
\]
Topic extraction: the representation model

For obtaining the vocabulary:

- Using linguistic analysis the pairs \(<\text{verb}, \text{subject}>\) and \(<\text{verb}, \text{object}>\) are considered for building a pair-term matrix

<table>
<thead>
<tr>
<th>Verb</th>
<th>chance</th>
<th>village</th>
<th>turtle</th>
<th>effort</th>
<th>Hanoi</th>
<th>performance</th>
<th>season</th>
<th>Vietnam</th>
<th>dune</th>
</tr>
</thead>
<tbody>
<tr>
<td>offer</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>see</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>make</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>climb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- The *Mutual Information* is the measure used for the association strength between two words, a verb \((v_i)\) and a noun \((n_j)\)

\[
mi(v_i, n_j) = \log \frac{p(v_i, n_j)}{p(v_i)p(n_j)}
\]
Topic extraction

For topic extraction:

- *Clustering by Committee* can assign words to different clusters using sets of representative elements *committees* for discovering unambiguous centroids that describe the members of a possible class.
Topic extraction

For topic extraction:

- *Clustering by Committee* can assign words to different clusters using sets of representative elements *committees*) for discovering unambiguous centroids that describe the members of a possible class
- This method only creates clusters of terms, but it does not create a hierarchical structure
Discovering taxonomic relations

- Axiom schema
- Instances
- Taxonomic Relationships
- Vocabulary

Discovering taxonomic relations
Acquisition of axioms in ontology learning

Discovering taxonomic relations

This phase comprises the identification of hypernymy/hyponymy relationships applying a method that combines lexical patterns and web search
Discovering taxonomic relations

Querying the Web

For finding hypernym relations between elements in each topic:

- A set of queries is built considering the following:
  - Lexical patterns + contextual information
  - Lexical patterns + related information
- Each query set is executed on a web search
- Candidate hypernyms are extracted from the retrieved pages
- A score is computed using:

  \[
  \text{ScoreCandHypernym} = \frac{\text{hits}(\text{LexicalPattern}(\text{term}, \text{CandHypernym}))}{\text{hits}(\text{CandHypernym})}
  \]

- The hypernym with greater score is selected
Discovering taxonomic relations

Querying the Web

- Using lexical patterns + contextual information

  term + **lexical pattern** + terms with the higher frequencies in the input corpus

- Example:
  
  museum + **and other** + cash + travel + product
Discovering taxonomic relations

Querying the Web

• Using lexical patterns + contextual information

term + lexical pattern + terms with the higher frequencies in the input corpus

• Example:
  museum + and other + cash + travel + product

• Using lexical patterns + related information

term + lexical pattern + the most representative terms in the WordNet synset

• Example:
  museum + and other + collection + object + display
Axiom learning approach

- Axiom schema
- Instances
- Taxonomic Relationships
- Vocabulary
Axiom learning approach

Considerations:

1. High evidence of named entities
Axiom learning approach

Considerations:

1. High evidence of named entities
2. An *instance level* and *class level* into hierarchical structure
Acquisition of axioms in ontology learning

The proposed approach

Axiom learning approach

Considerations:

1. High evidence of named entities
2. An *instance level* and *class level* into hierarchical structure
3. Contextual information

1. “In Wexford the **November Opera Festival is an international event**”

2. “The Elephanta **Festival is a classical dance and music event** on Elephanta Island usually held in February”

3. “The **Grenada National Museum** in the center of town incorporates an old French barracks dating from 1704”

Some obtained relations are:

- `instanceOf(November Opera Festival, festival)`
- `instanceOf(The Elephanta Festival, festival)`
- `subClassOf(festival, event)`
- `disjointWith(festival, museum)`
Axiom learning approach

Definition 1
A class is a set of individuals which share similar characteristics or properties. It is denoted by $C$.

Definition 2
An instance $x$ is an object given for a specific class.

Definition 3
A $instanceOf$ relation associates an instance as a member of a specific class. It is denoted by $instanceOf(x, C)$.

Example:
- $C = \text{“river”}$ and $x = \text{“Nile”}$
- $instanceOf(\text{“Nile”}, \text{“river”})$
Axiom learning approach

Definition 4
Let $A$ and $B$ be classes, $B$ is a subclass of $A$ if all the instances of $B$ are a subset of $A$.

Definition 5
Let $A$ and $B$ be classes, $A$ and $B$ are disjoint classes if the set of instances of $A$ is different to set of instances of class $B$.

Definition 6
Let $A$ and $B$ be classes, $A$ and $B$ are equivalent classes if the set of instances of $A$ is the same to set of instances of class $B$. 
Axiom learning approach
Example of disjoint classes
Axiom learning approach

Example of equivalent classes

Case 1:

Alchemy:Country  OpenCalais:Country  dbpedia-owl:country

equivalentClass(Alchemy:Country, OpenCalais:Country)
equivalentClass(Alchemy:Country, dbpedia-owl:country)
equivalentClass(OpenCalais:Country, dbpedia-owl:country)
Axiom learning approach

Example of equivalent classes

Case 1:

\begin{align*}
\text{Australia} & \quad \text{Scotland} \\
\text{Honduras} & \quad \text{Nicaragua} \\
\text{El Salvador,} & \quad \text{El Salvador,} \\
& \quad \text{...} \\
\text{...} & \quad \text{...} \\
\text{Alchemy:Country} & \quad \text{OpenCalais:Country} \\
& \quad \text{dbpedia-owl:country} \\
equivalentClass(\text{Alchemy:Country, OpenCalais:Country}) \\
equivalentClass(\text{Alchemy:Country, dbpedia-owl:country}) \\
equivalentClass(\text{OpenCalais:Country, dbpedia-owl:country})
\end{align*}

Case 2:

\begin{align*}
\text{Fever} & \quad \text{Malaria} \\
\text{Hepatitis} & \quad \text{dehydration,} \\
& \quad \text{...} \\
\text{Alchemy:HealthCondition} & \quad \text{OpenCalais:MedicalCondition} \\
equivalentClass(\text{Alchemy:HealthCondition, OpenCalais:MedicalCondition})
\end{align*}
Acquisition of axioms in ontology learning

The proposed approach

Axiom learning approach

disjointClass(C1, C2)
equivalentClass(C1,C2)
subClassOf(C1,C2)....

Class Level

Extract context

POS-Tagger
Syntactic Parser

Axioms

Instance Level

Named Entity Recognition Tool

C1
C2
C3
C4
C5

NE1
NE2
NE3
NE4
NE5
NE6
Test Corpus

- Tourism: “Lonely Planet”\(^a\)
  - manually constructed
  - 1801 HTML files
  - 96 concepts
  - 103 taxonomic relations
  - 278 named entities

\(^a\)http://olc.ijs.si/lpReadme.html

- Sport Events: “SmartWeb Project”\(^b\)
  - manually constructed but independent of the corpus
  - 400 files
  - 235 concepts
  - 185 taxonomic relations

\(^b\)http://www.dfki.de/sw-lt/olp2_dataset
Evaluation measures

• Precision (P)

\[ P = \frac{CorrectlySelectedEntities}{TotalSelectedEntities} \]  

• Recall (R)

\[ R = \frac{CorrectlySelectedEntities}{TotalDomainEntities} \]  

• F-Measure (F)

\[ F = \frac{2 \times P \times R}{P + R} \]
Experiments

Phases evaluated in the experiments:

- Axiom schema
- Instances
- Taxonomic Relationships
- Vocabulary

- Axiom learning
- Discovering taxonomic relations
- Topic extraction
Phase: Topic extraction

Objective: Selection of syntactic-dependency parser

Test corpus: one file sample of “Lonely Planet”

<table>
<thead>
<tr>
<th>Tool</th>
<th>Number of words</th>
<th>Multi-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minipar</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>StanfordParser</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>LinkGrammar</td>
<td>52</td>
<td>0</td>
</tr>
</tbody>
</table>

Minipar parser can identify multi-words as:

- Salt Lake City
- Majestic Taj Mahal hotel
- St. Mary Magdeline Church
- geographical region
- olympic games
Phase: Topic extraction

Objective: Selection of similarity measure

Test corpus: “Lonely Planet” - 10856 words

<table>
<thead>
<tr>
<th>Measure</th>
<th>Number of groups</th>
<th>F Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hindle</td>
<td>10</td>
<td>0.1302</td>
</tr>
<tr>
<td>Jaccard</td>
<td>11</td>
<td>0.1176</td>
</tr>
<tr>
<td>PMI$^a$</td>
<td>25</td>
<td>0.1136</td>
</tr>
<tr>
<td>Cosine</td>
<td>13</td>
<td>0.1052</td>
</tr>
<tr>
<td>NGD$^b$</td>
<td>1</td>
<td>0.0555</td>
</tr>
</tbody>
</table>

$^a$Pointwise Mutual Information

$^b$Normalized Google Distance

Examples of groups using Hindle measure in CBC algorithm:

- {accomodation, business, traveler, hotel, food}
- {July, March, August, week, day, time, year, month}
Phase: Discovering taxonomic relations

Objective: Evaluation the taxonomy relation extraction

Example:

- Query with contextual information:
  - region + **lexical pattern** + cash + travel + product

- Query with related information:
  - region + **lexical pattern** + extended + spatial + location
Phase: Discovering taxonomic relations

Objective: Evaluation the taxonomy relation extraction

Example:

"gold standard"

The obtained precision is 53.3%
Phase: Axiom learning

Objective: Evaluation of “instanceOf” relation by classes

Evaluation of \textit{instanceOf} relation for the classes: \textit{City}, \textit{Country} and \textit{Holiday}

- 12 relations of \textit{City} class
- 35 relations of \textit{Country} class
- 10 relations of \textit{Holiday} class

<table>
<thead>
<tr>
<th>Class</th>
<th>Tool</th>
<th>Precision</th>
<th>Recall</th>
<th>F Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>AlchemyAPI</td>
<td>0.4000</td>
<td>0.5000</td>
<td>0.4444</td>
</tr>
<tr>
<td></td>
<td>OpenCalais</td>
<td>0.3529</td>
<td>0.5000</td>
<td>0.4137</td>
</tr>
<tr>
<td>Country</td>
<td>AlchemyAPI</td>
<td>0.7631</td>
<td>0.8285</td>
<td>0.7945</td>
</tr>
<tr>
<td></td>
<td>OpenCalais</td>
<td>0.7000</td>
<td>0.8000</td>
<td>0.7486</td>
</tr>
<tr>
<td>Holiday</td>
<td>AlchemyAPI</td>
<td>0.4285</td>
<td>0.3000</td>
<td>0.3529</td>
</tr>
<tr>
<td></td>
<td>OpenCalais</td>
<td>0.4000</td>
<td>0.4000</td>
<td>0.4000</td>
</tr>
</tbody>
</table>
Phase: Axiom learning

Objective: Evaluation of “subClassOf” relation

Test corpus: 30 files of “Lonely Planet” manually annotated
NER tool: AlchemyAPI

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
<th>Total</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeographyFeature</td>
<td>Location, Island, CityTown, ...</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Country</td>
<td>Location, Government Jurisdiction, Kingdom, ...</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>HumanLanguage, MeteorologicalService</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>AdministrativeDivision, ...</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Person</td>
<td>MilitaryPerson, Monarch, Politician, ...</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Region</td>
<td>Location</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Facility</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The obtained precision was 70.37%
Phase: Axiom learning

Objective: Evaluation of “disjointWith” relation

Test corpus: 30 files of “Lonely Planet” manually annotated

NER tool: AlchemyAPI

The obtained precision was 83.80%
Phase: Axiom learning

Objective: Evaluation of “equivalentClass” relation

Test corpus: 30 files of “Lonely Planet” manually annotated
NER tool: AlchemyAPI and OpenCalais

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>EquivalentClass</th>
<th>Other relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlchemyAPI:Organization</td>
<td>OpenCalais:Organization</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>AlchemyAPI:Country</td>
<td>OpenCalais:Country</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>AlchemyAPI:Sports</td>
<td>OpenCalais:SportsGame</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>AlchemyAPI:HealthCondition</td>
<td>OpenCalais:MedicalCondition</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>AlchemyAPI:Organization</td>
<td>OpenCalais:Company</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>AlchemyAPI:Holiday</td>
<td>OpenCalais:Person</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

The obtained precision was 67.50%
Results

Previous works:

- Cimiano *et al.*, 2005
  - Learning concept hierarchies from text corpora using formal concept analysis
  - Tourism and Finance domain
- Jiang and Tan, 2010
  - A semantic-based domain ontology learning system from text
  - Terrorism and Sport event domain
## Results

Comparison with Cimiano et al.’s work (“Lonely Planet dataset”):

<table>
<thead>
<tr>
<th>Element</th>
<th>Author(s)</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept extraction</td>
<td>Cimiano et al., 2005</td>
<td>29.33</td>
</tr>
<tr>
<td></td>
<td>Rios-Alvarado et al., 2012</td>
<td>67.00</td>
</tr>
<tr>
<td>Taxonomic relation</td>
<td>Cimiano et al., 2005</td>
<td>50.00</td>
</tr>
<tr>
<td></td>
<td>Rios-Alvarado et al., 2012</td>
<td>55.00</td>
</tr>
<tr>
<td>Axioms:</td>
<td></td>
<td></td>
</tr>
<tr>
<td><code>instanceOf</code></td>
<td>Cimiano et al., 2005</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Rios-Alvarado et al., 2012</td>
<td>46.67</td>
</tr>
<tr>
<td><code>subClassOf</code></td>
<td>Cimiano et al., 2005</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Rios-Alvarado et al., 2012</td>
<td>70.37</td>
</tr>
<tr>
<td><code>disjointWith</code></td>
<td>Cimiano et al., 2005</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Rios-Alvarado et al., 2012</td>
<td>83.80</td>
</tr>
<tr>
<td><code>equivalentClass</code></td>
<td>Cimiano et al., 2005</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Rios-Alvarado et al., 2012</td>
<td>92.00</td>
</tr>
</tbody>
</table>
Results

Comparison with Jiang and Tan’s work (“Sport Event dataset”):

<table>
<thead>
<tr>
<th>Element</th>
<th>Author(s)</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jiang and Tan, 2010</td>
<td>Rios-Alvarado et al., 2012</td>
</tr>
<tr>
<td>Concept extraction</td>
<td>45.00</td>
<td>34.01</td>
</tr>
<tr>
<td>Taxonomic relation</td>
<td>74.00</td>
<td>77.50</td>
</tr>
<tr>
<td>Non-taxonomic relation</td>
<td>69.40</td>
<td>-</td>
</tr>
<tr>
<td>Axioms:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>instanceOf</td>
<td>-</td>
<td>23.52</td>
</tr>
<tr>
<td>subClassOf</td>
<td>-</td>
<td>64.44</td>
</tr>
<tr>
<td>disjointWith</td>
<td>-</td>
<td>85.00</td>
</tr>
<tr>
<td>equivalentClass</td>
<td>-</td>
<td>93.33</td>
</tr>
</tbody>
</table>
Publications


- Ana Rios-Alvarado, Ivan Lopez-Arevalo, and Victor Sosa-Sosa. “A taxonomy construction approach supported by web content”. *Journal of Information Science and Engineering*
Schedule of activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Period (4 months)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2009</td>
</tr>
<tr>
<td>Literature review</td>
<td></td>
</tr>
<tr>
<td>Curses: Data mining/Algorithms</td>
<td></td>
</tr>
<tr>
<td>Collect textual resources</td>
<td></td>
</tr>
<tr>
<td>To obtain the model representation</td>
<td></td>
</tr>
<tr>
<td>To design and implementation of methods for:</td>
<td></td>
</tr>
<tr>
<td>Topic extraction</td>
<td></td>
</tr>
<tr>
<td>Discovering taxonomic relations</td>
<td></td>
</tr>
<tr>
<td>Learning axioms</td>
<td></td>
</tr>
<tr>
<td>To adapt and test of models</td>
<td></td>
</tr>
<tr>
<td>To evaluate the proposed approach</td>
<td></td>
</tr>
<tr>
<td>PhD stay at DERI-NUIG Ireland</td>
<td></td>
</tr>
<tr>
<td>Pre-doctoral examination</td>
<td></td>
</tr>
<tr>
<td>To write an article</td>
<td></td>
</tr>
<tr>
<td>To write thesis</td>
<td></td>
</tr>
<tr>
<td>To submit dissertation</td>
<td></td>
</tr>
</tbody>
</table>

Current period

Finished activities

Next activities
Conclusions I

- An approach for ontology learning from scratch has been presented, the source of knowledge is a set of textual resources on a specific domain.

- Our proposal is inspired by *layer cake model* and it has been worked three main layers:
  - topic extraction
  - discovering taxonomic relations
  - learning axioms

- In the topic extraction layer, the text clustering combined with linguistic analysis is a good technique to obtain the representative vocabulary in a specific domain.
Conclusions II

• For the discovering taxonomic relations, the use of additional information and lexical-patterns into query sets executed on web search showed a good evidence to identify hypernymy/hyponymy relations

• The method for axiom learning is based on identifying named entities as class instances and comparing their context for building the taxonomic structure, and so establish axiomatic relations such as:
  • instanceOf
  • subClassOf
  • disjointWith
  • equivalentClass

• In two corpus, our approach has shown promising results
References


http://dx.doi.org/10.1002/asi.v61:1


References II


Thanks a lot for your kind attention!

Ana B. Rios-Alvarado
arios@tamps.cinvestav.mx