NERITS
A Machine Translation mashup system using Wikimeta and LOD

Kamel Nebhi, Luka Nerima, Eric Wehrli

Laboratory of Analysis and Technology of Language
University of Geneva, Switzerland

AI Mashup Challenge 2013
Motivation

• Machine Translation (MT) has become increasingly commonplace in everyday use through BabelFish, Bing Translator or Google Translate.
• MT approaches provide good results but none of them exploit contextual information like named entity to help user comprehension.

Solution

• **Named Entity Recognition for the Interactive Translation System**
• **The goal** is to propose a cross-lingual translation by providing detailed information extracted from DBpedia about persons, locations and organizations in the mother tongue of the user.
• This helps at scaling the traditional multilingual task of MT to cross-lingual applications.
Motivation

- Machine Translation (MT) has become increasingly commonplace in everyday use through BabelFish, Bing Translator or Google Translate.
- MT approaches provide good results but none of them exploit contextual information like named entity to help user comprehension.

Solution

- **Named Entity Recognition for the Interactive Translation System**
- **The goal** is to propose a cross-lingual translation by providing detailed information extracted from DBpedia about persons, locations and organizations in the mother tongue of the user.
- This helps at scaling the traditional multilingual task of MT to cross-lingual applications.
ITS-2: an Interactive Translation System

Machine Translation system based on the multilingual parser Fips

A classical syntactic transfer system:

1. Lexical and syntactic analysis with Fips
2. Lexical and grammatical transfer
3. Syntactic and morphological generation

Paul regardait la voiture → Paul was looking to the car
ITS-2: an Interactive Translation System

Machine Translation system based on the multilingual parser Fips

A classical syntactic transfer system:

1. Lexical and syntactic analysis with Fips
2. Lexical and grammatical transfer
3. Syntactic and morphological generation

Paul regardait la voiture → Paul was looking to the car

\[
\begin{align*}
\text{TP} & \quad \text{DP} \quad \text{regardait}_i \quad \text{VP} \\
\text{DP} & \quad \text{Paul} \\
\text{VP} & \quad \text{e}_i \\
\text{NP} & \quad \text{la} \\
\text{NP} & \quad \text{voiture}
\end{align*}
\]
ITS-2: an Interactive Translation System

Machine Translation system based on the multilingual parser Fips

A classical syntactic transfer system:
1. Lexical and syntactic analysis with Fips
2. Lexical and grammatical transfer
3. Syntactic and morphological generation

Paul regardait la voiture → Paul was looking to the car
ITS-2: an Interactive Translation System

Machine Translation system based on the multilingual parser Fips

**A classical syntactic transfer system:**

1. Lexical and syntactic analysis with Fips
2. Lexical and grammatical transfer
3. Syntactic and morphological generation

Paul regardait la voiture → Paul was looking to the car
Wikimeta: a Semantic Annotation platform

Wikimeta is a multilingual semantic annotation tool which performs:
- Named Entity Recognition (person, location, organization, etc.)
- Named Entity Linking

Approach

- Contextual words resource called *Linked Data Interface* (LDI)
- Disambiguation task using *Semantic Disambiguation Algorithm* (SDA)
Wikimeta: a Semantic Annotation platform

Wikimeta is a multilingual semantic annotation tool which performs:

- Named Entity Recognition (person, location, organization, etc.)
- Named Entity Linking

**Approach**

- Contextual words resource called *Linked Data Interface* (LDI)
- Disambiguation task using *Semantic Disambiguation Algorithm* (SDA)

*Washington D.C. (LOC) - Barack Obama (PERS) confronted a trio of controversy, pledging to work with Congress (ORG) to ensure the IRS (ORG) doesn’t abuse its power.*
Wikimeta: a Semantic Annotation platform

Wikimeta is a multilingual semantic annotation tool which performs:

- Named Entity Recognition (person, location, organization, etc.)
- Named Entity Linking

Approach

- Contextual words resource called Linked Data Interface (LDI)
- Disambiguation task using Semantic Disambiguation Algorithm (SDA)

Washington D.C. (LOC) - Barack Obama (PERS) confronted a trio of controversy, pledging to work with Congress (ORG) to ensure the IRS (ORG) doesn’t abuse its power.

http://dbpedia.org/data/Washington_D.C.rdf
http://dbpedia.org/data/Barack_Obama.rdf
http://dbpedia.org/data/United_States_Congress.rdf
http://dbpedia.org/data/Internal_Revenue_Service.rdf
Wikimeta: a Semantic Annotation platform

Wikimeta is a multilingual semantic annotation tool which performs:

- Named Entity Recognition (person, location, organization, etc.)
- Named Entity Linking

Approach

- Contextual words resource called Linked Data Interface (LDI)
- Disambiguation task using Semantic Disambiguation Algorithm (SDA)

Washington D.C. (LOC) - Barack Obama (PERS) confronted a trio of controversy, pledging to work with Congress (ORG) to ensure the IRS (ORG) doesn’t abuse its power.

http://dbpedia.org/data/Washington_D.C.rdf
http://dbpedia.org/data/Barack_Obama.rdf
http://dbpedia.org/data/United_States_Congress.rdf
http://dbpedia.org/data/Internal_Revenue_Service.rdf
The NERITS Approach

An example through NERITS Architecture

Jean-Marc Ayrault vient à Montpellier pour une visite officielle...
The NERITS Approach
Data Retrieval: Translation

http://latlapps.unige.ch/Translation

<Translation>
<input src="fr">
Jean-Marc Ayrault vient à Montpellier pour une visite officielle.
</input>
<output tgt="en">
Jean-Marc Ayrault is coming to Montpellier for an official visit.
</output>
</Translation>
The NERITS Approach

Data Retrieval: Semantic Annotation

http://www.wikimeta.com/

<extraction>
  <NE>Jean-Marc Ayrault</NE>
  <type>PERS.HUM</type>
  <LINKEDDATA>
    http://www.dbpedia.org/resource/Jean-Marc_Ayrault
  </LINKEDDATA>
</extraction>

<extraction>
  <NE>Montpellier</NE>
  <type>LOC.ADMI</type>
  <LINKEDDATA>
    http://www.dbpedia.org/resource/Montpellier
  </LINKEDDATA>
</extraction>
The NERITS Approach

Data Integration

http://dbpedia.org/data/Jean-Marc_Ayrault.rdf

```xml
<rdf>
  <rdfs:label xml:lang="fr">Jean-Marc Ayrault</rdfs:label>
  <dbpedia-owl:orderInOffice>Prime Minister of France</dbpedia-owl:orderInOffice>
  <dbpprop:birthDate>1950-01-25</dbpprop:birthDate>
  <dbpprop:placeOfBirth>Maulévrier, Maine-et-Loire, France</dbpprop:placeOfBirth>
  <dbpprop:spouse>Brigitte Terrien</dbpprop:spouse>
</rdf>
```
The NERITS Approach

Response

Jean-Marc Ayrault is coming to Montpellier for an official visit.

Jean-Marc Ayrault

Occupation: Prime Minister of France
Date of Birth: 1950-01-25
Place of Birth: Mâcon, Saône-et-Loire, France
Spouse: Brigitte Titen

Jean-Marc Ayrault (born 25 January 1950) is the Prime Minister of France, appointed on 16 May 2012. He was the Mayor of Nantes from 1989 to 2012, and led the French Socialist Party group in the National Assembly from 1987 to 2012.
The NERITS Approach

Other Examples

United Nations
Organisation des Nations unies

Established Date: 1945-06-26

Admin Center: in New York City, New York

The United Nations (abbreviated UN in English, and ONU in its other official languages), is an international organization whose stated aims are facilitating cooperation in international law, international security, economic development, social progress, human rights, and achievement of world peace. The UN was founded in 1945 after World War II to replace the League of Nations, to stop wars between countries, and to provide a platform for dialogue.

Montpellier

Population: 252,968
Postal Code: 34000
Mayor: Hélène Mandroux-Coles

Montpellier is a city in southern France. It is the capital of the Languedoc-Roussillon region, as well as the Hérault department. Montpellier is the 8th largest city of the country, and is also the fastest growing city in France over the past 25 years.

Barack Obama

Occupation: President of the United States, 44th
Date of Birth: 1961-08-04
Place of Birth: Honolulu, Hawai, United States
Spouse: Michelle Robinson

Barack Hussein Obama II is the 44th and current President of the United States. He is the first African American to hold the office. In January 2005, Obama was sworn in as a U.S. Senator in the state of Illinois. He would hold this office until November 2008, when he resigned following his victory in the 2008 presidential election. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he was the president of the Harvard Law Review.
The NERITS Approach

User Interface

http://cms.unige.ch/lettres/linguistique/nebhi/nerits/
Conclusion - Further Work

• We have presented a MT mashup system using Semantic Annotation provided by Wikimeta and LOD.
• We have given a first glance on how Semantic Annotation and information from DBpedia can be combined to help user comprehension.

Further Work

• Recognition of Events (September 11 attacks, Fall of the Berlin Wall, etc.)
• Incorporate more knowledge from LOD (points of interest of a Location, people around a Person)
• Add other languages (French-German, German-French)
Questions?

kamel.nebhi@unige.ch
Thank you!
kamel.nebhi@unige.ch