First CAMELEON SHARED TASK

Carlos Ramisch
2013-05-15

Table of Contents

Multilingualization of ontologies

Ontologies are language-independent abstract representations of a portion of human knowledge. However, in practice, the labels and descriptions of the concepts and relations represented in an ontology are often expressed using natural language, to allow humans to more easily interpret the information encoded. Having such ontologies enables the pursuit of the ultimate goal of the semantic web, which is to use the information represented in these ontologies to allow automatic reasoning on large amounts of data.

Therefore, towards the multilingual semantic web, there is a need for efficient automatic techniques that allow the translation of ontologies from one language to another. On the one hand, general-purpose machine translation systems, like Google translate, present lexical choice and ambiguity resolution problems, yielding bad-quality translations due to the limited available context and the specialized domain modelled by the ontology. On the other hand, the use of specific hand-crafted lexicons is onerous and such resources are only rarely available for a given language pair and domain. Furthermore, most of the available resources are targeted to English or use English as a pivot language.

Task description

The goal of this first edition of the CAMELEON shared task is to explore the problem of automatic ontology translation. The teams will be free to use whatever resources, tools and techniques they consider relevant to solve the problem. We will provide a set of ontologies in English, French and Portuguese as test set. The goal of the systems is to generate as output the labels for an ontology in a language L2 having as input the ontology in language L1.

Therefore, each registered team will have the possibility to implement a system that translates between the three language pairs in each direction. The systems will be evaluated in the following translation directions:

- *Input*: English ontology, *Output*: French ontology
- *Input*: English ontology, *Output*: Portuguese ontology
- *Input*: French ontology, *Output*: English ontology
- *Input*: French ontology, *Output*: Portuguese ontology
- *Input*: Portuguese ontology, *Output*: English ontology
- *Input*: Portuguese ontology, *Output*: French ontology

If a team chooses which translation directions (language pair + direction) they want to target, from one to all. Each translation direction will be evaluated through a separate set of evaluation scores. To measure the performance of a system, we will compare each automatic translation obtained with the reference manual translation, using standard ontology matching measures like exact match and edit string distance.

We would like to remember that this shared task is not at all a competition. It is an opportunity to focus on a given task which is of common interest within CAMELEON.

Furthermore, we invite the teams to provide manual translations to the labels of each provided ontology. We will then use the manual translations to enrich the ontologies and to apply metrics to evaluate the precision of multiple translations.

Team registration

Each team willing to participate in the shared task should send, before December 21, 2012, an email to both task organizers carlosramisch at gmail.com and cassia.ts at gmail.com, containing:

- Name of the team members and a name identifying the team;
- The translation directions targeted (please use the numbers 1 to 6 above, for instance, directions 3 and 5)
- A brief description (max. 1 paragraph) of the intended system.

In this first edition, each participant team should include at least one member of the CAMELEON project.

The data set

The MultiFarm dataset is composed of a set of ontologies describing the domain of organising conferences (the domain being well understandable for different researchers). The ontologies were developed within the OntoFarm project . These OntoFarm ontologies were manually translated into eight different languages (Chinese, Czech, Dutch, French, German, Portuguese, Russian, and Spanish) and the corresponding alignments between these ontologies were also manually set. For this shared task, we will use the MultiFarm translations in English, French and Portuguese.

%The details about the input and output formats will be posted on the shared task website soon.

Open test set

Attention: A new version of the dataset was made available on May 01, 2013. Please use this last version as the previous one contained a spurious "null" class.

We are making available a set of three ontologies for testing. This data is made available using the standard format OWL . We expect the participant systems to receive as an input one ontology in OWL, and the output will be another OWL ontology. The output ontology will be identical to the input one, except for the rdfs:label elements. For example, if the input ontology has labels in French, the output ontology will have the same labels translated into Portuguese, like in the example below:

```
<owl:Class rdf:about="http://cmt_fr#c-6425593-9645104">
    <rdfs:label xml:lang="fr">membre du comité de programme</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://cmt_fr#c-2724520-3587324"/>
    <rdfs:subClassOf rdf:resource="http://cmt_fr#c-4497134-7374494"/>
    <owl:disjointWith rdf:resource="http://cmt_fr#c-9412558-6009078"/>
</owl:Class>
```

```
<owl:Class rdf:about="http://cmt_fr#c-6425593-9645104">
    <rdfs:label xml:lang="pt">membro do comitê de programa</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://cmt_fr#c-2724520-3587324"/>
    <rdfs:subClassOf rdf:resource="http://cmt_fr#c-4497134-7374494"/>
    <owl:disjointWith rdf:resource="http://cmt_fr#c-9412558-6009078"/>
</owl:Class>
```

- Download the open test set

The evaluation will be performed both on the open test set and on a blind test set. The participants should send the output of the systems on the open test set before ~~April 30~~ May 31, 2013. The blind test set will be made available soon after receiving the outputs of the systems on the open test set, and participants will be requested to return the outputs in 24h. The blind test set consists of a set of different ontologies in the same domain of conference organization. The evaluation will be performed using standard measures of ontology alignment calculated with the OAEI API .

Evaluation and baseline NEW

We are making available a package containing a script for calculating the evaluation metrics and a baseline translation for each direction. The tool can be downloaded here:

- Download the evaluation and baseline package

Evaluation script

Once you have downloaded and unzipped the *eval-baseline.zip* package, you must place your results in the corresponding folder. In order to evaluate the systems in a systematic way, you have to put the file containing your translated ontology in the folder *translations/direction/ontology* and name it *mysystem.owl*. A *direction* should be named *language1-language2*, where *language1* is the source language and *language2* is a different target language. An *ontology* is one of the three ontologies in the open test set: *cmt*, *confOf* or *conference*.

You MUST use the same filename across the different folders (for example, *mysystem.owl*). Please modify the variables in the script *run-eval.sh* according to the ontologies, directions and system you want to evaluate.

Evaluation metrics

The script *run-eval.sh* evaluates all the systems, directions and ontologies at once. You should edit the script variable *systems* to provide as many system names as translated ontologies you generated. If all the files are in their right place in the folders, the evaluation results will be a series of *direction-system.html* files in the *results* folder. For example, the baseline system *bing* for the direction *en-fr* generated the following results:

algo		**BLEU**					**bing-equal**				**bing-edit**	
test	Average	1-grams	2-grams	3-grams		Prec.	Rec.	FMeas.		Prec.	Rec.	FMeas.
cmt	0.2503	0.4400	0.2380	0.1900		1.00	0.25	0.40		0.56	0.56	0.56
confOf	0.2533	0.5410	0.2300	0.1650		1.00	0.24	0.39		0.72	0.72	0.72
conference	0.2697	0.5690	0.2810	0.1720		0.96	0.21	0.35		0.63	0.63	0.63
H-mean	**0.2574**	**0.5103**	**0.2477**	**0.1750**		**0.99**	**0.23**	**0.38**		**0.63**	**0.63**	**0.63**

All these automatic measures compare somehow the translation obtained by your system with the reference translation, generated by a human translator. In the future, we would like to generate multiple reference translations in order to yield a fairer evaluation.

- **BLEU**: This is a standard measure in the evaluation of machine translation. See this paper for more details. It computes the proportion of common n-grams between the generated translation and the reference translation. In the next columns, values are broken down according to 1-, 2- and 3-grams. It is expected that there are fewer 2- and 3-grams in common than the number of 1-grams. Since the translations are short, we calculate BLEU only for up to 3-

grams (instead of the typical value of up to 4-grams). When including 4-grams, some of the values for the baseline system were 0.0 because the system didn't share any 4-gram with the reference translation.

- **bing-equal** and **bing-edit**: These two measures use an ontology alignment approach to evaluate the systems. Each correspondence between the system and reference labels are considered as a reference alignment. Therefore, we run two simple ontology alignment algorithms: equal and edit. The first one only aligns labels when they are 100% equal. The second aligns similar labels using a Levenstein string distance measure. Then, these alignments are compared with the reference alignments that should have been found if the translations were perfect. This gives an idea of the proportion of 100% correct translations and the proportion of translations where the system translation is reasonably similar to the expected reference translation.

Baseline system

The baseline system uses the Bing standard translation engine, which is freely available, for translating the labels. No domain adaptation or tweaking was performed. The ontologies translated by Bing are placed in the folder translations/direction/ontology/bing.owl.

The baseline evaluation measures are available in the results folder. The baseline translations were generated by Roger Granada.

Presentation of the systems

The strategy and results of each team will be presented in a short paper (max. 2 pages) in English. The format of the paper will be announced later. This paper will not be published publicly, but only internally on the CAMELEON wiki. Each team will have the opportunity to present their results during the Fourth CAMELEON workshop, around July 2013 in Porto Alegre (either in person or by videoconference).

In the future, we would like to transform the resulting set of papers into a larger paper to be submitted to a conference or journal, and the short papers will be used as a starting point. Everybody willing to contribute to this submission will be welcome as a co-author.

The CAMELEON reading group

From the moment when the teams are registered, we intend to create a reading group on multilingualization of ontologies. This reading group will read one article per week, suggested by one of the participants, and then we will have a weekly skype meeting to discuss the paper. The skype meeting will be coordinated by the person who suggested the paper. The goal of this reading group is to familiarize the participants with the domains of ontology multilingualization, ontology matching, machine translation and multilingual lexical resources.

The participation in the reading group is non-mandatory for the registered teams and open to other participants.

Important dates

- Registration open - November 30, 2012
- Test data available - ~~December 14, 2012~~ December 17, 2012
- Team registration deadline - ~~December 21, 2012~~ February 28, 2013
- System outputs due - ~~April 30, 2013~~ ~~May 31, 2013~~ July 26, 2013
- Description paper due (max. 2 pages) - ~~May 31, 2013~~ ~~June 30, 2013~~ August 30, 2013
- CAMELEON workshop - to be announced (December 2013).

Reading suggestions

- A Machine Learning Approach to Multilingual and Cross-lingual Ontology Matching
- Cross-lingual entity matching and infobox alignment in Wikipedia
- Using Domain-specific and Collaborative Resources for Term Translation
- Simple methods for dealing with term variation and term alignment
- Babouk: Focused Web Crawling for Corpus Compilation and Automatic Terminology Extraction