Translating Ontologies in Real-World Settings with ESSOT

Mihael Arcan¹, Mauro Dragoni², and Paul Buitelaar¹

¹Insight Centre for Data Analytics, NUI Galway
²Fondazione Bruno Kessler (FBK), Shape and Evolving Living Knowledge Unit (SHELL)

Kobe, Japan, October 19th 2016
Our Goal:

to demonstrate why a domain-aware machine translation system can help domain experts in translating ontologies
Presentation’s main points

Our Goal:

to demonstrate why a domain-aware machine translation system can help domain experts in translating ontologies

ESSOT: a collaborative knowledge management architecture for ontology translation
Our Goal:
to demonstrate why a domain-aware machine translation system can help domain experts in translating ontologies

ESSOT: a collaborative knowledge management architecture for ontology translation

Background & Motivation
Why we did this...
Our Goal:

to demonstrate why a domain-aware machine translation system can help domain experts in translating ontologies

ESSOT: a collaborative knowledge management architecture for ontology translation

Background & Motivation
Why we did this...

Implementation
The machine translation system and the domain-expert facilities
Our Goal:

to demonstrate why a domain-aware machine translation system can help domain experts in translating ontologies

ESSOT: a collaborative knowledge management architecture for ontology translation

Background & Motivation
Why we did this...

Implementation
The machine translation system and the domain-expert facilities

Validation
Real-world scenarios in which ESSOT has been validated
Motivation 1 – Breaking language barriers

- Ontologies created by people do not know other languages than their mother tongue

- Collaborative work and alignment in a multi-lingual scenario.
  - end-user applications
Artefacts are created in specific languages due to policy constraints.
Translation Task: Documents vs Ontologies

Example: “vessels” → Gefäß or Schiff
**Translation Task: Documents vs Ontologies**

Example: “"vessels" → Gefäß or Schiff

**DOCUMENT**

- Context 1 (Gefäß): The **blood vessels** are the part of the circulatory system that transports blood throughout the human body. There are three major types of **blood vessels**

- Context 2 (Schiff): Ships are typically large **ocean-going vessels**; whereas **boats** are smaller, and typically travel most often on inland or coastal waters.
Translation Task: Documents vs Ontologies

Example: “'vessels' → Gefäß or Schiff

- **Context 1 (Gefäß):** The blood vessels are the part of the circulatory system that transports blood throughout the human body. There are three major types of blood vessels.

- **Context 2 (Schiff):** Ships are typically large ocean-going vessels; whereas boats are smaller, and typically travel most often on inland or coastal waters.
ESSOT Architecture

Language Experts

Domain Experts

User Facilities

Domain-Specific Ontology

Domain-Aware Translation Service

Request for missing concept translation

Query Expansion for Disambiguated Sentence Selection

Parallel Corpus For Training Models
Construction of the Translation Model
Construction of the Translation Model
Construction of the Translation Model

Word2Vec Machinery

Multi-dimensional Matrix

<table>
<thead>
<tr>
<th>Word</th>
<th>Vector 1</th>
<th>Vector 2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232</td>
<td></td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
<td></td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
<td></td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
<td></td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
<td></td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
<td></td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
<td></td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
<td></td>
</tr>
</tbody>
</table>
Construction of the Translation Model

Word2Vec Machinery

Parallel Corpora

Multi-dimensional Matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232...</td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540...</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511...</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213...</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415...</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165...</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646...</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544...</td>
</tr>
</tbody>
</table>
Construction of the Translation Model

Multi-dimensional Matrix

- vessels: 0.53956, 0.2232...
- ship: 0.48783, 0.6540...
- radiator: 0.51249, 0.6511...
- cruisers: 0.59540, 0.1213...
- sank: 0.48096, 0.5415...
- Wreckage: 0.49036, 0.5165...
- fleet: 0.48104, 0.5646...
- Shaft: 0.16541, 0.6544...

Parallel Corpora

Word2Vec Machinery
Construction of the Translation Model

Word2Vec Machinery

Parallel Corpora

Multi-dimensional Matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232</td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
</tr>
</tbody>
</table>
Online Domain Adaptation
Online Domain Adaptation

Ontology

vessel

blood  0.539
medical  0.512
body  0.493
vein  0.490
disease  0.487
biomedical  0.481
Online Domain Adaptation

Ontology

vessel

blood 0.539
medical 0.512
body 0.493
vein 0.490
disease 0.487
biomedical 0.481

n-best sentences
Online Domain Adaptation

Ontology

vessel

<table>
<thead>
<tr>
<th>Blood</th>
<th>Medical</th>
<th>Body</th>
<th>Vein</th>
<th>Disease</th>
<th>Biomedical</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.539</td>
<td>0.512</td>
<td>0.493</td>
<td>0.490</td>
<td>0.487</td>
<td>0.481</td>
</tr>
</tbody>
</table>

sentence 1

vessels 0.595
fleet 0.481
shaft 0.480

sentence 2

vessels 0.595
blood 0.493
medical 0.490

sentence 3

vessels 0.595
ship 0.539
radiator 0.512
Online Domain Adaptation

Ontology

Translation Service
Online Domain Adaptation

Ontology

vessel

blood 0.539
medical 0.512
body 0.493
vein 0.490
disease 0.487
biomedical 0.481

n-best sentences

vessels 0.595
fleet 0.481
shaft 0.420

sentence 1

sentence 2

vessels 0.595
blood 0.493
medical 0.490

sentence 3

vessels 0.595
ship 0.539
radiator 0.512

Gefäß
Translation Request Pipeline

User:
1. Select concept to translate
2. Invoke translation suggestion service
3. Send request
4. Receive rank
5. Select translation to adopt

OTTO Service:
1. Receive request
2. Extract semantic information from the request
3. Retrieve the set of candidate translations
4. Send rank
ESSOT: User Interfaces
ESSOT: User Interfaces

Multilingual component

Select language: English

Translation in the language: English

Concept name: agricultural method

Concept description: Practices used to enhance crop and livestock health and prevent weed, pest or disease problems without the use of chemical substances.

[Suggest translation]
Multilingual component

Entity translation

Entity name: agricultural method
Translation: agricultural method

Entity description: Practices used to enhance crop and livestock health and prevent weed, pest or disease problems without the use of chemical substances.

Translation: Practices used to enhance crop and livestock health and prevent weed, pest or disease problems without the use of chemical substances.

Save Cancel
## ESSOT: User Interfaces

**Multilingual component**

**Entity translation**

**Entity name:** agricultural method  
**Translation:** agricola  
**Select language:** English  

**List all Concepts**

Number of concepts in the Domain Model: 62

<table>
<thead>
<tr>
<th>Concept</th>
<th>Description</th>
<th>Concept translation</th>
<th>Description translations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Activity</strong></td>
<td>A type of action performed by an agent in general sense.</td>
<td>attività</td>
<td>Le pratiche vegetali e animali usati per promuovere la salute e la prevenzione delle malattie, parassiti e infestanti problemi senza l'uso di sostanze chimiche.</td>
</tr>
<tr>
<td><strong>agricultural method</strong></td>
<td>Practices used to enhance crop and livestock health and prevent weed, pest or disease problems without the use of chemical substances.</td>
<td>agrario metodo</td>
<td>Le tecniche agricole utilizzate in europa.</td>
</tr>
<tr>
<td><strong>European agricultural method</strong></td>
<td>Agricultural techniques used in Europe.</td>
<td>metodo agricolo europeo</td>
<td></td>
</tr>
<tr>
<td><strong>Animal origin processed product</strong></td>
<td>Any product of animal origin canned, cooked, frozen, concentrated, pickled or otherwise prepared to assure its preservation in transport, distribution and storage, but does not include the final cooking or preparation of a food product for use as a meal or part of a meal such as may be done by restaurants, catering companies or similar establishments.</td>
<td>animale sorgente processed prodotto</td>
<td></td>
</tr>
</tbody>
</table>
ESSOT: The OTTO Translation Service

Upload the ontology to be translated:
Choose File No file chosen

or provide the internet address to it:
Enter ontology address

Submit Data

The OTTO Translation System was developed to enhance English monolingual ontologies in RDF (n-triples) or Owl with multilingual lexical knowledge. Currently, OTTO translates from English into German, Spanish, Italian, Irish, Slovenian and Czech.

Example of an RDF representation in the financial domain (see rdf/n-triples file).

This service is brought to you by http://nlp.insight-centre.org/

RDF file was generated by Saffron.

http://server1.nlp.insight-centre.org/otto
Example

Input

```json
{
  "label2translate": "vessels",
  "concept_context": [
    "blood",
    "medical",
    "disease",
    "biomedical",
  ],
  "translate2": "de"
}
```

Output

```json
{
  "possible_translations": {
    "blutgefäßen": -15.8438,
    "gefäßen": -2.4100,
    "halsgefäße": -2.6682
  },
  "time": "24 wallclock secs",
  "source_label": "vessels",
  "best": "gefäßen"
}
```

http://server1.nlp.insight-centre.org/otto/rest_service.html
Evaluation Procedure

- **Qualitative evaluation:**
  We collected subjective judgments from the language experts involved in the evaluation of the tool on general usability of the components and to provide feedback for future improvements.

  - **Research Questions 1:** Does the proposed system provide an effective support, in terms of the quality of suggested translations, to the management of multilingual ontologies?
  
  - **Research Questions 2:** Do the MoKi functionalities provide an effective support to the collaborative management of a multilingual ontology?

- **Quantitative Evaluation:**
  We collected objective measures concerning the effectiveness of the translations suggested by the embedded machine translation service.

  - 3 metrics: BLEU, METEOR, TER
  
  - 3 language pairs: EN→DE, EN→IT, EN→ES
  
  - 6 ontologies:
    - Organic.Lingua, Presto
    - TheSoz, Geoskill, DOAP, STW
Qualitative Evaluation Results

- Language Experts view
  - Pros: Easy to use for managing translations (9)
  - Cons: Usable interface for showing concept translations (3)

- Approval and Discussion
  - Pros: Pending approvals give a clear situation about concept status (4)
  - Cons: Discussion masks are not very useful and the approval process might be improved (8)

- Quick Translation feature
  - Pros: Best facility for translating concepts (8)
  - Cons: Improvable interface design (3)
Quantitative Evaluation Results

- General robustness of the approach
- EN$\leftrightarrow$DE: wrong translations needed many editing operations
- Computational time to improve
  - avg. 4.1 seconds per request
Conclusions

- ESSOT integrates the OTTO domain-adaptable semantic translation service and the MoKi collaborative knowledge management tool in an effective pipeline for multilingual ontology management.

- ESSOT provides helpful suggestions for performing ontology translation (RQ1) and the provided interfaces are usable and useful for supporting the language experts in the translation activity (RQ2)

- The translation quantitative evaluation shows significant improvements over the Microsoft Translation system

- Future work:
  - to extend this strategy to other problems (e.g. ontology enrichment);
  - improve the approval management task;
  - improve the service efficiency.

SEE YOU LATER AT THE DEMO SESSION !!!!!!
It’s time for questions...

Mauro Dragoni
Fondazione Bruno Kessler

https://shell.fbk.eu/index.php/Mauro_Dragoni
dragoni@fbk.eu
<table>
<thead>
<tr>
<th>Ontology Label</th>
<th>Microsoft</th>
<th>OTTO</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>driver</td>
<td>Treiber</td>
<td>Fahrer</td>
<td>Fahrer</td>
</tr>
<tr>
<td>stroke</td>
<td>Strich</td>
<td>Schlaganfall</td>
<td>Schlaganfall</td>
</tr>
<tr>
<td>demonstration</td>
<td>Vorführung</td>
<td>Demonstration</td>
<td>Demonstration</td>
</tr>
<tr>
<td>race</td>
<td>Rennen</td>
<td>Rasse</td>
<td>Rasse</td>
</tr>
<tr>
<td>fine</td>
<td>feine</td>
<td>Geldbuße</td>
<td>Geldbuße</td>
</tr>
<tr>
<td>bonapartism</td>
<td>Bonapartismus</td>
<td>bonapartism</td>
<td>Bonapartismus</td>
</tr>
<tr>
<td>shamanism</td>
<td>Schamanismus</td>
<td>shamanism</td>
<td>Schamanismus</td>
</tr>
</tbody>
</table>
## ESSOT Training Data

<table>
<thead>
<tr>
<th>Corpus Name</th>
<th>Domain</th>
<th>→ German</th>
<th>→ Italian</th>
<th>→ Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>JRC-Acquis</td>
<td>legislative</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>DGT</td>
<td>legislative</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ECB*</td>
<td>financial</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>EMEA*</td>
<td>medical</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Europarl</td>
<td>proceedings of the EP</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Gnome*</td>
<td>information technology</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>KDE*</td>
<td>information technology</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>OpenOffice*</td>
<td>information technology</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>PHP*</td>
<td>information technology</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>OpenSubtitles13*</td>
<td>translated movie subtitles</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>TED</td>
<td>talk transcriptions</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MultiUn</td>
<td>United Nation Documents</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wikipedia Titles</td>
<td>terminology expressions</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Number of parallel sentences: 10,222,259 | 23,441,275 | 37,583,385
The Organic.Lingua and Presto Scenarios

The **Organic.Lingua** project: multilingual services for discovery, retrieval, exploitation and extension of content related to Organic Agriculture and AgroEcology, particularly:

- find resources in languages different from the ones in which the query has been formulated;
- manage meta-data information for resources in different languages;
- contribute to evolving content, e.g. providing services for content generation.

The **PRESTO** project: the creation of a system for the customization of serious games scenarios based on virtual reality. Ontologies are used for:

- manage interoperability between different 3D libraries;
- support developers in developing behaviors of virtual agents.
Construction of the Translation Model

- Word2Vec for building semantic vectors of words
  - Wikipedia pages used as a single huge textual file for building word vectors.
- Parallel corpora used for building a generic SMT model
  - Moses Toolkit has been used.
- Word vectors generated by W2V are used for filtering parallel sentences by detecting the ones that are more relevant.
- When a translation is requested:
  - the context of the concept is extracted from the ontology, packed and sent to the service;
  - vectors associated with each label are aggregated in order to generate a unique vector describing the full concept context; the same is done for the sentences containing the label to translate;
  - The aggregated vector is compared with the vectors of the relevant sentences;
  - Closest sentence is chosen for extracting the concept translation.
Statistical Machine Translation for Ontology Translation

Parallel Corpora

Probabilistic Model

\[ \text{arg max } p(f|e)p(e) \]

Translation Model

Translation Model

n-best sentences
## Problem Space of Ontology Translation within SMT

<table>
<thead>
<tr>
<th>Source text:</th>
<th>Out-of-Vocabulary (OOV) Problem</th>
<th>Out-of-Domain Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>acroscopic</td>
<td></td>
<td>vessels</td>
</tr>
</tbody>
</table>

| Reference translation: | akroskopisch | Gefäße |
| (gold standard)        |              |        |

| Target transl. (generic): | acroscopic | Schiffe$^1$ |
| Target transl. (specific): | acroscopic | Gefäße |

$^1$ Google Translate
Domain-specific data selection from available resources

Ontology to be translated
labels:
- Administrative Act
- Insurance
- Equity

Parallel Resources
source sentences:
- sentence1
- sentence2
- sentence3

Word2Vec
- two-layer neural network, trained on large amount of text (Wikipedia)
- produces a distributed representation of words

Combining vectors of labels and its direct parent

parent_of Equity

Equity
Domain-specific data selection from available resources

Intestinal infectious diseases (*ICD label*)
- infection (0.882381)
- gastroenteritis (0.879408)
- disease (0.874399)
- ...

Ataxia and anemia could be infectious (*k*-sentence)
- infection (0.743936)
- encephalitis (0.721071)
- syndrome (0.711319)

COSINE SIMILARITY between related words from sentences and ontology labels
Online Domain Adaptation
Online Domain Adaptation
Online Domain Adaptation

Word2Vec Machinery
Online Domain Adaptation

Word2Vec Machinery

Multi-dimensional Matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232</td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
</tr>
</tbody>
</table>
Online Domain Adaptation

Word2Vec Machinery

Multi-dimensional Matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232</td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
</tr>
</tbody>
</table>

Parallel Corpora
Online Domain Adaptation

Multi-dimensional Matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232</td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
</tr>
</tbody>
</table>

Word2Vec Machinery

Parallel Corpora

Generic Translation Model
Online Domain Adaptation

**Parallel Corpora**

**Word2Vec Machinery**

**Generic Translation Model**

**Multi-dimensional Matrix**

<table>
<thead>
<tr>
<th></th>
<th>0.53956</th>
<th>0.2232</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
<td>...</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
<td>...</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
<td>...</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
<td>...</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
<td>...</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
<td>...</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
<td>...</td>
</tr>
</tbody>
</table>

**n-best sentences**
Online Domain Adaptation

Multi-dimensional Matrix

<table>
<thead>
<tr>
<th>Word</th>
<th>Multiplicative Coefficient</th>
<th>Cumulative Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>vessels</td>
<td>0.53956</td>
<td>0.2232</td>
</tr>
<tr>
<td>ship</td>
<td>0.48783</td>
<td>0.6540</td>
</tr>
<tr>
<td>radiator</td>
<td>0.51249</td>
<td>0.6511</td>
</tr>
<tr>
<td>cruisers</td>
<td>0.59540</td>
<td>0.1213</td>
</tr>
<tr>
<td>sank</td>
<td>0.48096</td>
<td>0.5415</td>
</tr>
<tr>
<td>Wreckage</td>
<td>0.49036</td>
<td>0.5165</td>
</tr>
<tr>
<td>fleet</td>
<td>0.48104</td>
<td>0.5646</td>
</tr>
<tr>
<td>Shaft</td>
<td>0.16541</td>
<td>0.6544</td>
</tr>
</tbody>
</table>
Domain-specific data selection from available resources

... serious inflammation of the **vessels** ...
... **vessels** inflammation which may result in very rare ...
... arteries or of the blood **vessels** supplying the brain ...
... puncture of non-compressible **vessels** within the last month ...

**Statistical Machine Translation System**

- **Decoder**
- **Translation model**
- **Language model**