How to Configure Statistical Machine Translation with Linked Open Data Resources

Ankit Srivastava, Felix Sasaki, Peter Bourgonje, Julian Moreno-Schneider, Jan Nehring, Georg Rehm
DFKI GmbH, Alt-Moabit 91c, 10559 Berlin, Germany
firstName.lastName@dfki.de

Abstract

In this paper we outline easily implementable procedures to leverage multilingual Linked Open Data (LOD) resources such as the DBpedia in open-source Statistical Machine Translation (SMT) systems such as Moses. Using open standards such as RDF (Resource Description Framework) Schema, NIF (Natural language processing Interchange Format), and SPARQL (SPARQL Protocol and RDF Query Language) queries, we demonstrate the efficacy of translating named entities and thereby improving the quality and consistency of SMT outputs. We also give a brief overview of two funded projects that are actively working on this topic. These are the (1) BMBF funded project DKT (Digitale Kuratierungstechnologien) on digital curation technologies, and (2) EU Horizon 2020 funded project FREME (Open Framework of e-services for Multilingual and Semantic Enrichment of Digital Content). This is a step towards designing a Semantic Web-aware Machine Translation (MT) system and keeping SMT algorithms up-to-date with the current stage of web development (Web 3.0).

1 Introduction

In a 2001 article in the Scientific American (Berners-Lee et al., 2001), Berners-Lee and collaborators first publicised the concept of the Semantic Web, sometimes called Web 3.0. The initial aim of the Semantic Web was to provide standards through which people can publish documents and data, allowing computer programs to combine and link data from many datasets in order to perform a task just like a human. In a nutshell, the Semantic Web is about making links so that a person or a machine can explore the web of data.

The World Wide Web Consortium (W3C) provides standards promoting common data formats and protocols that constitute the basic technology for the Semantic Web. These are:

- Resource Description Framework (RDF): A formalism to represent data on the web as a labelled graph of objects and their relations
- Uniform Resource Identifier (URI): A compact sequence of characters used to identify resources on the web
- Ontologies: Hierarchical vocabularies of types and relations, allowing more efficient storage and use of data by encoding generic facts about objects. RDF Schema is one such formalism or knowledge representation language.

According to W3C,[1] Linked Data lies at the heart of what Semantic Web is about. The collection of Semantic Web technologies (mentioned above and detailed further in Section 2) provides an environment where an application such as a Machine Translation (MT) system can query data and draw inferences using vocabularies linked on the web. In this paper, we describe an algorithm (in Section 3) using these open standards and tools in order to automatically identify named entities, retrieve their translations from linked data ontologies and feed them to a Statistical Machine Translation (SMT) system. We summarise our experimental results on this semantic web-aware SMT in Section 4, followed by a discussion on the limitations of this

[1] http://www.w3.org/standards/semanticweb/data

Proceedings of the 38th Conference Translating and the Computer, pages 138–148,
London, UK, November 17-18, 2016.

approach in Section 5. After giving an overview of two funded projects actively working in this area as well as comparing our approach to previous works in Section 6, we conclude our paper in Section 7.

2 Tools of the Trade

The main goal of this paper is to provide a workable technique for integrating linked open data resources into a machine translation system.

The term Linked Data, coined in 2006, refers to the ability of the Web to link related data as opposed to just linking related documents. It refers to a set of best practices[2] for publishing and linking structured data on the web. Linked Open Data (LOD) typically refers to linked data with open sharing licenses. These links enable both humans and machines to explore the web of data.

In recent years, there has been a tremendous growth (Schmachtenberg et al., 2014) in both the quality and quantity of data available as linked data on the web. This data can describe named entities such as people, organisations, locations, etc. in multiple languages. This fact coupled with the increased move towards the publication of multilingual language resources such as WordNets and Wikipedia using linked data principles (Chiarcos et al., 2011) has led to a significant increase in the availability of Semantic Web information relevant to Natural Language Processing applications including machine translation.

Typical examples of LOD resources include DBpedia Knowledge Base (Auer et al., 2007), FreeBase (Bollacker et al., 2008), BabelNet (Navigli and Ponzetto, 2012), JRC-Names.[3] In our experiments, we focus on DBpedia, but any of the aforementioned resources with a SPARQL endpoint can be plugged in our SMT system.

In the context of SMT, leveraging translations from Linked Data resources can be likened to plugging external knowledge sources such as terminology banks and translation memories. The major difference is that linked data is stored in a different data format (NIF based on RDF) and is accessed using a dedicated query language (SPARQL).

In this section, we describe in brief the enabling technologies, standards, and software used in our experiments to configure SMT with linked data.

2.1 RDF and RDFS

Resource Description Framework (RDF) is a XML-like syntax providing the foundation for representing and processing machine readable data.

RDF is a graph-based model whose basic building block is an entity-attribute-value triple. There are three fundamental concepts of RDF:

- Resources are objects referenced by an identifier or URI
- Properties describe relations between resources
- Statements assert the properties of resources in the form of entity-attribute-value triple, consisting of a resource, a property, and a value. The value can either be a resource or a literal (atomic values such as language codes)

RDFS (RDF Schema) is a primitive ontology language. It is a vocabulary used to define helpful properties (such as rdfs:label for language name) in Resource Description Framework (RDF). An in-depth exposition is provided in The Semantic Web Primer (Antoniou et al., 2012).

[2] http://www.w3.org/DesignIssues/LinkedData.html
[3] https://data.europa.eu/euodp/en/linked-data

2.2 NIF 2.0

NIF 2.0[4] (Natural Language Processing Interchange Format) is an RDF-based format that aims to achieve interoperability between NLP tools such as SMT engines, parsers and annotated language resources such as DBpedia. Its integration with WC standard ITS 2.0[5] (Internationalization Tag Set) makes it attractive to multilingual applications.

The primary use case of this standard is to serve as an input and output format for web services, that enables seamless pipelining or combination of various language processing web services in sequence (Hellmann et al., 2013).

An important characteristic of this standard relevant to NLP is that the atomic unit is a character rather than a word. Thus, if a sentence has 23 characters (including spaces between words), the resource or sentence spans from 0 to 22.

2.3 SPARQL

SPARQL[6]is recursive acronym for SPARQL Protocol and RDF Query Language. It is a query language (like SQL) primarily for linked data, used to retrieve information from RDF-encoded data including NIF. It is a W3C recommended standard. In simple terms, if the data such as a multilingual lexicon is stored as a linked data (NIF standard), then SPARQL is a tool to retrieve information from the linked data such as translations in the required target language.

2.4 DBpedia

DBpedia[7] is a linked open dataset (extracted from Wikipedia) consisting of 4.58 million entities in up to 125 languages and 29.8 million links to external web pages. DBpedia Spotlight[8] is an open-source tool for automatically annotating mentions of DBpedia resources in text. Note that the translations may be prone to error on account of being user generated.

2.5 Moses

Moses[9] (Koehn et al., 2007) is an open-source SMT system used in our experiments as a test bed for Semantic Web-enabled MT. We have employed Phrase-based Statistical Machine Translation system with standard configurations, as specified in Section 4. The translations from the LOD such as DBpedia are inserted in a forced decoding framework, wherein the translation of selected named entities are chosen from DBpedia instead of the Moses decoder.

3 Methodology

Having touched upon the basic building blocks for configuring a SMT system with LOD resources in Section 2, we now describe the framework to interface a Moses-based SMT system with a DBpedia-based LOD resource. The procedure comprises of 6 steps as enumerated below.

1. Convert the text to be translated into a NIF document

2. For each sentence, API call e-NER (Named Entity Recognition) service

3. For each of the named entities (marked in NIF), API call the e-linking service, that is, annotate named entities in the document using DBpedia Spotlight mentioned in Section 2

[4]http://persistence.uni-leipzig.org/nlp2rdf/
[5]http://www.w3.org/TR/its20/
[6]http://www.w3.org/TR/rdf-sparql-query/
[7]http://wiki.dbpedia.org
[8]https://github.com/dbpedia-spotlight/
[9]http://www.statmt.org/moses/

4. For each link (named entity resource identified in the DBpedia LOD), retrieve the translation in target language using a SPARQL query for attribute *rdfs:label* which contains the language identifier

5. Integrate these translations in the Moses decoder. Encode the named entity and its translation in a format compatible with the Moses decoder (enabled with the xml-input feature)

6. Display translated output in the appropriate format

We will illustrate the mechanism behind each step in the framework with the help of an example sentence. Consider translating an English sentence (from the IT-domain) *"MS Paint is a good option."* into German. Note that all the procedures below are carried out by freely available web service API calls, the source code for which can be found at `https://github.com/freme-project` for FREME web services[10] and at `https://github.com/dkt-projekt` for DKT web services.[11] More information about these projects and their tools can be found in Section 6.

Convert into NIF: All the web services for various NLP applications including MT hosted by the DKT and FREME are NIF-enabled. The NIF core technology provides classes and properties to describe the relations between substrings, text, documents, and their URI schemes or identifiers (Hellmann et al., 2013).

Listing 1: Representing a sentence in NIF.

```
<http://freme-project.eu/#char=0,26>
    a       nif:Context , nif:RFC5147String , nif:Sentence ;
    nif:anchorOf            "MS paint is a good option." ;
    nif:beginIndex          "0" ;
    nif:endIndex        "26" ;
    nif:firstWord       <http://freme-project.eu/#char=0,2> ;
    nif:isString        "MS paint is a good option." ;
    nif:lastWord        <http://freme-project.eu/#char=25,26> ;
    nif:refContext          <http://freme-project.eu/#char=0,26> ;
    nif:word                <http://freme-project.eu/#char=9,11> ,
                            <http://freme-project.eu/#char=3,8> ,
                            <http://freme-project.eu/#char=12,13> ,
                            <http://freme-project.eu/#char=19,25> .
```

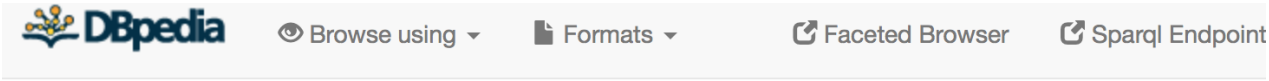
From Listing 1, we observe how the English sentence (source language) *"MS Paint is a good option."* is assigned a URI including the character spans 0 through 26 (first line). There are various attributes or properties such as all the words, firstWord, and lastWord. However the most important line for our purposes is the whole sentence denoted by *nif:isString*.

Tag the Named Entities and link with DBpedia entries: Herein we have combined steps 2 and 3 mentioned above into one process.

Figure 1 shows a screen-shot of a typical lexical entry on DBpedia for the entity *Paint (software)* linked to the phrase *MS Paint* in our sentence. Figure 2 displays the same entry focusing on the concepts *rdfs:label* and *owl:sameAs* which lists links to the same entity

[10] Of particular interest is the web service named e-entity/dbpedia-spotlight.

[11] Of particular interest are the services DKTBrokerStandalone/nifTools, e-NLP/Sparqler, and e-SMT.



About: Paint (software)

An Entity of Type : SkilledWorker110605985, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Paint (formerly Paintbrush for Windows) is a simple computer graphics program that has been included with all versions of Microsoft Windows. It is often referred to as MS Paint or Microsoft Paint.

Property	Value
dbo:abstract	▪ Paint (formerly Paintbrush for Windows) is a simple computer graphics program that has been included with all versions of Microsoft Windows. It is often referred to as MS Paint or Microsoft Paint. The program mainly opens and saves files as Windows bitmap (24-bit, 256 color, 16 color, and monochrome, all with the .bmp extension), JPEG, GIF (without animation or transparency, although the Windows 98 version, a Windows 95 upgrade, and the Windows NT4 version did support the latter), PNG (without alpha channel), and single-page TIFF. The program can be in color mode or two-color black-and-white, but there is no grayscale mode. For its simplicity, it rapidly became one of the most used applications in the early versions of Windows—introducing many to painting on a computer for the first time—and is still widely used for very simple image manipulation tasks. (en)
dbo:wikiPageID	▪ 321796 (xsd:integer)
dbo:wikiPageRevisionID	▪ 683143936 (xsd:integer)
dbp:caption	▪ Paint on Windows 10 featuring its ribbon in user interface (en)

Figure 1: Screen-shot of the DBpedia resource

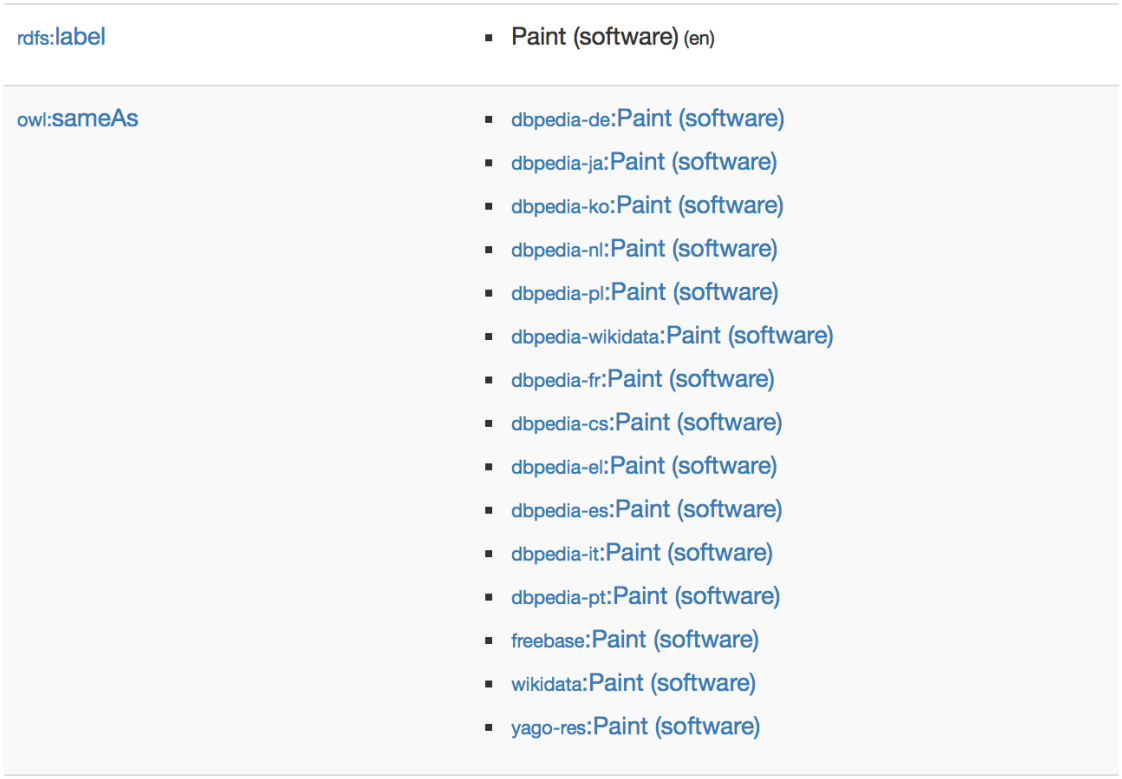


Figure 2: Screen-shot of the DBpedia multilingual entries

(Microsoft Paint) in different languages identified by a 2-digit language code. For example, *de* denotes German language.

The sentence is parsed by the FREME DBpedia-Spotlight web service and all entities or terms which occur in our LOD resource (DBpedia) are annotated with the property *itsrdf:taIdentRef*. A fragment of a NIF document with the disambiguated term and link to DBpedia entry is shown in Listing 2:

Listing 2: Output from FREME NER in the NIF format.

```
<http://freme-project.eu/#char=0,8>
    a                       nif:RFC5147String , nif:Word ;
    nif:anchorOf            "MS-Paint" ;
    nif:beginIndex          "0" ;
    nif:endIndex            "8" ;
    nif:nextWord            <http://freme-project.eu/#char=9,11> ;
    nif:referenceContext    <http://freme-project.eu/#char=0,26> ;
    nif:sentence            <http://freme-project.eu/#char=0,26> ;
    itsrdf:taIdentRef
                <http://dbpedia.org/resource/Paint_(software)> .
```

Query for Target Entity Translation: As stated in Step 4 of the procedure, we use DBpedia SPARQL endpoint available at `https://dbpedia.org/sparql`. This can be directly invoked from inside Java code in the DKT and FREME e-services. Essentially, the DBpedia database is stored as a triple store or a Graph store alluding to the entity-attribute-value triple structure of the RDF data.

The SPARQL query snippet shown in Listing 3 helps us retrieve German *(de)* translations for each annotated named entity or resource (denoted by `http://dbpedia.org/resource/`

`Paint_(software)` in our example), using the attribute *rdfs:label*.

Listing 3: Code Snippet for a SPARQL Query.

```
PREFIX dbpedia: <http://dbpedia.org/resource/>

SELECT distinct *
WHERE {
    <http://dbpedia.org/resource/Paint_(software)>
    rdfs:label ?label
    filter langMatches( lang(?label), "de" )
}
```

Moses Decoder Integration: Once we have identified the entities and obtained their translations from a LOD resource such as DBpedia, the next step is to plug it in a machine translation system such as Moses. Essentially we treat the SMT system as a black box thus making it theoretically possible to substitute any MT system for Moses as per the user requirements.

The Moses decoder is "forced" to use translations for the named entities tagged by the linked data instead of relying on its own translation models and phrase tables. We achieve this by invoking the Moses decoder with the *xml-input*[12] feature turned on, demonstrated by a command-line code snippet in Listing 4. The phrase *MS Paint* has its translation *Microsoft Paint* (retrieved by the SPARQL query) hardcoded before the Moses decoder is initiated.

Listing 4: Code Snippet for a command-line call of Moses.

```
% echo '<np translation="Microsoft Paint">MS Paint</np>
is a good option ." | moses -xml-input exclusive -f moses.ini
```

Display Translated Output: Once we have translated the whole sentence, the procedure is complete, and the translation can either be simply displayed as a *plaintext* string (*"Microsoft Paint ist eine gute wahl."*) or encoded in *NIF* format as shown in Listing 5. The property *itsrfd:target* is associated with linking the translated string along with the target language code (*de*) to the remainder of the NIF document. This format or any other RDF-style (linked data) format is just so that the output can be further chained as input in subsequent NLP applications in a seamless manner.

Listing 5: NIF representation of a sentence and its translation.

```
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .
@prefix nif:   <http://persistence.uni-leipzig.org/
                 nlp2rdf/ontologies/nif-core#> .


<http://dkt.dfki.de/documents/#char=0,26>
    a          nif:RFC5147String , nif:Context , nif:String ;
    nif:beginIndex "0"^^xsd:nonNegativeInteger ;
    nif:endIndex   "26"^^xsd:nonNegativeInteger ;
    nif:isString   "MS paint is a good option." ;
    itsrdf:target  "Microsoft Paint ist eine gute wahl.\n"@de .
```

[12] See Section 4.8.2 in `http://www.statmt.org/moses/manual/manual.pdf`.

4 Experimental Results

Section 3 outlined the core ingredient of this paper, that is, a recipe on how to source translations from linked data ontologies (e.g. DBpedia) into a statistical MT system (e.g. Moses). In this section, we examine the overall benefits, if any, of interfacing a SMT system with Semantic Web resources.

We trained a Moses-based SMT system (Koehn et al., 2007) to translate from English (source language) into German (target language). The set of parallel sentences for training, and the development and test sets for tuning and testing respectively were sourced from the data provided for the WMT 2016 shared task on machine translation of IT domain (Bojar et al., 2016) available at `http://www.statmt.org/wmt16/it-translation-task.html`.

For the purposes of this experiment, we chose this corpora setting, i.e. training a SMT system with large amounts of out-of-domain data (typically European parliamentary proceedings and newswire corpus) augmented with small amounts of domain-specific data (IT-domain corpora such as Libreoffice, Ubuntu, Chromium) in order to translate 1000 IT-domain answers from Batch 3 (the same test set as that employed in the shared task). Batch 1 was used for tuning the translation, language and reordering models (development set). Table 1 outlines the size of the training data.

corpus	entries	words
Chromium browser	63K	551K
Drupal	4.7K	57.4K
Libreoffice help	46.8K	1.1M
Libreoffice UI	35.6K	143.7K
Ubuntu Saucy	182.9K	1.6M
Europarl (mono)	2.2M	54.0M
News (mono)	89M	1.7B
Commoncrawl (parallel)	2.4M	53.6M
Europarl (parallel)	1.9M	50.1M
MultiUN (parallel)	167.6K	5.8M
News Crawl (parallel)	201.3K	5.1M

Table 1: Size of corpora used for SMT.

The motivation was to increase the potential for occurrence of named entities such as technical terms (e.g. Microsoft Paint) in the test data such that we could demonstrate our linked data-aware SMT system. The phrase-based SMT system was trained with standard Moses configuration settings for language model, word alignments, reordering model, explicitly specified in our system description paper for the WMT 2016 IT-domain Shared Task (Avramidis et al., 2016).[13]

Nearly each of the 1000 segments in the test set had at least 1 named entity tagged and annotated. When comparing, a baseline system (translating entirely from the Moses models) with a system whose named entities were translated by linked data resources, a BLEU (Papineni et al., 2002) score improvement of 0.8 (accuracy improved from 34.0 to 34.8) and TER (Snover et al., 2006) score improvement of 2.5 (error reduced from 56.1 to 53.6) was observed. The linked data-aware system identified and correctly translated 12% more terms (named entities) than the baseline SMT system.

One such example of how a SMT system configured with LOD resources benefited and outperformed a baseline SMT system is seen as follows.

[13] http://www.aclweb.org/anthology/W/W16/W16-2329.pdf

SRC (en): MS Paint is a good option.
MT 1 (de): Frau Farbe ist eine gute wahl.
MT 2 (de): Microsoft Paint ist eine gute wahl.

Consider translating the English sentence (the one used to demonstrate our framework in Section 3) into German. MT1 displays the baseline translation where the SMT decoder is unable to disambiguate the term **MS Paint** as a software and not a person. MT2 configured with linked open data gives us the correct translation.

5 Limitations

There are shortcomings and potential pitfalls such as accuracy of user-generated translations in DBpedia and mismatched entity linking which make the case of optimally exploiting linked data resources in SMT system non-trivial.

- Translations are not always accurate because these are user-generated (from Wikipedia entries) and therefore prone to error
- Mismatched Entity Linking. For instance, *MS Paint* only links *MS* to *Microsoft Paint* and leaves the *Paint* unlinked. The result is that *MS* translates to *Microsoft Paint*, while *Paint* is translated separately thereby generating a double translation in the target language. A viable solution is to combine MS and Paint as one entity (pre-processing)
- There is also the issue of how to handle multiple links or translations for a frequently occurring term (entity disambiguation). A possible solution is to pick the top item, or use domain filters (IT-domain versus general in the case of the entity *Paint*).

6 Project Overview and Related Work

6.1 FREME

The project FREME (`http://www.freme-project.eu`) is a two-year European Union Horizon 2020 funded project (started February 2015) aimed towards Open Framework of e-services for Multilingual and Semantic Enrichment of Digital Content. The project involves 8 partners:[14] the DFKI Language Technology Lab, AgroKnow, iMinds, Institute for Applied Informatics, Istituto Superiore Mario Boella, Tilde, VistaTEC, and Wripl.

It essentially hosts a chain of e-services performing diverse NLP applications with the help of interoperability standards such as NIF. The partners lead four business cases around digital content and linked data. With the help of reusable NLP workflows and pipelines, the FREME project provides access to a set NLP and data services demonstrating monetisation of the multilingual data value chain.

More information is available at `https://github.com/freme-project/e-services` and `https://freme-project.github.io/`.

6.2 DKT

The project Digitale Kuratierungstechnologien (DKT: Digital Curation Technologies (`http://digitale-kuratierung.de`)) is a two-year project (started September 2015) funded by the Bundesministeriums fur Bildung und Forschung (German Ministry of Education and Research).

[14] http://www.freme-project.eu/partners/consortium/

The project involves four Berlin-based partner companies (ART+COM AG, Condat AG, 3pc GmbH, and Kreuzwerker GmbH) and the DFKI Language Technology Lab. The project supports digital curation processes carried out by knowledge workers in multiple sectors (museums, television and media, exhibitions, publishers) through robust, precise, and modular language and knowledge technologies. The main goal is to semi-automate the different curation processes (research, annotation, timelining) to make the knowledge workers more time and cost efficient.

More information on the linked data-aware web services is available at `https://github.com/dkt-projekt`.

6.3 Related Work

There have been several approaches in the past that leveraged linked data in SMT systems. Most approaches either use it as an additional knowledge source and training the models on the dictionaries extracted from such resources, or use it in a post-training framework, either forced decoding named entities like our approach or translating unknown words.[15]

McCrae and Cimiano (2013) primarily integrated the dictionary of translations extracted from LOD resources during decoding and creating a new feature for linked data. They essentially let the Moses decoder decide when to chose translations from LOD and when to translate from its phrase tables. In contrast to our approach on forcing translations of all named entities identified by DBpedia, they employ another ontology called Lemon (Lexicon Model for Ontologies[16]) to translate primarily unknown words, that is translations not found by the decoder.

Du et al., (2016) on the other hand leveraged translations from BabelNet dictionaries using both McCrae and Cimiano (2013)'s methods as well as the forced decoding employed by our paper to demonstrate modest improvements in translating English-Polish and English-Chinese data.

It must be noted that the main goal of this paper was to provide a Semantic Web aware method to interface a SMT system with LOD knowledge base via seamless NIF-aware web service API calls. Testing the benefits to a translation system was a secondary outcome. We leave for future work, methods to optimally leverage knowledge from linked data on the Semantic Web and improve SMT system performance, especially in sense disambiguation and translating unknown words.

7 Conclusion

In this paper, we have successfully outlined a procedure to equip an off-the-shelf statistical machine translation system with linked data available on the Semantic Web. With the help of an example, we illustrated a novel machine translation adaptation with the potential for seamless integration into translation and localisation workflows. This is a step towards making MT semantic web-aware.

Acknowledgments

The authors wish to thank the anonymous reviewers for their helpful feedback. The project "Digitale Kuratierungstechnologien (DKT)" is supported by the German Federal Ministry of Education and Research (BMBF), "Unternehmen Region", instrument

[15] While we focus on 2 papers in our related work, a helpful survey of works on Machine Translation using Semantic Web technologies is currently under review and can be found at `http://www.semantic-web-journal.net/content/machine-translation-using-semantic-web-technologies-survey`.

[16] http://lemon-model.net

”Wachstumskern-Potenzial” (No. 03WKP45). More information on the project can be found online at http://www.digitale-kuratierung.de.

References

Antoniou, Grigoris, Paul Groth, Frank van Harmelen, and Rinke Hoekstra. 2012. A Semantic Web Primer. MIT Press, 3rd edition.

Auer, Sren, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference*, pages 722–735.

Avramidis, Eleftherios, Aljoscha Burchardt, Vivien Macketanz, and Ankit Srivastava. 2016. DFKI's system for WMT16 IT-domain task, including analysis of systematic errors. In *Proceedings of the First Conference on Machine Translation*, Berlin, Germany pages 415–422.

Berners-Lee, T., J. Hendler, and O. Lassila. 2001. The Semantic Web. In *Scientific American*, Vol. 284 number 5, pages 34–43, http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

Bojar, Ondrej, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Aurelie Neveol, Mariana Neves, Martin Popel, Matt Post, Raphael Rubino, Carolina Scarton, Lucia Specia, Marco Turchi, Karin Verspoor, and Marcos Zampieri. 2016. Findings of the 2016 Conference on Machine Translation. In *Proceedings of the First Conference on Machine Translation at ACL 2016*, Berlin, Germany, pages 131–198.

Bollacker, Kurt, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pages 1247–1250.

Chiarcos, Christian, Sebastian Hellmann, and Sebastian Nordhoff. 2011. Towards a Linguistic Linked Open Data cloud: The Open Linguistics Working Group. In *TAL*, Vol. 52 number 3, pages 245–275.

Du, Jinhua, Andy Way, and Andrzej Zydron. 2016. Using BabelNet to Improve OOV Coverage in SMT. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, isbn 978-2-9517408-9-1.

Hellmann, Sebastian, Jens Lehmann, Sren Auer, and Martin Brmmer. 2013. Integrating NLP using Linked Data. In *International Semantic Web Conference*, Sydney, Australia.

Koehn, Philipp, Hieu Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, Wade Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *demonstration session of ACL '07*, Prague, Czech Republic. 177–180.

McCrae, John Philip, and Philipp Cimiano. 2013. Mining Translations from the Web of Open Linked Data. In *Proceedings of the Joint Workshop on NLP, LOD and SWAIE*, Hissar, Bulgaria, pages 8–11.

Navigli, Roberto, and Simone Paolo Ponzetto. 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. In *Artificial Intelligence*, Vol. 193, pages 217–250.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jung Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. *ACL 2002, 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA. 311–318.

Schmachtenberg, Max, Anja Jentzsch, and Richard Cyganiak. 2014. Linking Open Data Cloud Diagram. http://lod-cloud.net/.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciula, and John Makhoul. 2006. A Study of Translation Edit Rate with targeted Human Annotation. *AMTA 2006, 7th Conference of the Association for Machine Translation in the Americas*, Cambrdge, MA. 223–231.