Towards language-agnostic alignment of product titles and descriptions: a neural approach

Daniel Stein
eBay Inc., Aachen, Germany
danstein@ebay.com

Dimitar Shterionov
Dublin City University, Dublihn, Ireland
dimitar.shterionov@adaptcentre.ie

Andy Way
Dublin City University, Dublihn, Ireland
andy.way@adaptcentre.ie

ABSTRACT

The quality of e-Commerce services largely depends on the accessibility of product content as well as its completeness and correctness. Nowadays, many sellers target cross-country and cross-lingual markets via active or passive cross-border trade, fostering the desire for seamless user experiences. While machine translation (MT) is very helpful for crossing language barriers, automatically matching existing items for sale (e.g. the smartphone in front of me) to the same product (all smartphones of the same brand/type/colour/condition) can be challenging, especially because the seller's description can often be erroneous or incomplete. This task we refer to as item alignment in multilingual e-commerce catalogues. To facilitate this task, we develop a pipeline of tools for item classification based on cross-lingual text similarity, exploiting recurrent neural networks (RNNs) with and without pre-trained word-embeddings. Furthermore, we combine our language agnostic RNN classifiers with an in-domain MT system to further reduce the linguistic and stylistic differences between the investigated data, aiming to boost our performance. The quality of the methods as well as their training speed is compared on an in-domain data set for English–German products.

1 INTRODUCTION

In the field of natural language processing (NLP), recent advances in deep learning have led to neural methods surpassing traditional rule-based or statistical ones for various tasks. One important drawback of such methods is the demand for large amounts of high-quality training data. While the situation of freely available (multi-)lingual training material continues to improve, obtaining suitably tailored in-domain content continues to be a bottleneck for training realistic models.

In the e-Commerce domain, whenever large proportions of the content are user-generated, even aligning and evaluating entries (mono- or cross-lingual) becomes extra difficult. These entries often

fall short in terms of linguistic quality, and many are incomplete or contain incorrectly labelled data. Due to cultural preferences, legal restrictions in between countries, or other limitations, cross-country meta data (e.g. an overarching catalogue tree) is also likely to differ.

The work we summarize in this paper aims to align (eBay) cross-language item titles[1] that belong to the same product, via text similarity approaches. With this work we aim to provide a basis for three higher-level tasks that are of interest to eBay: identifying in-domain comparable data, content synchronization and detecting erroneous/misaligned entries (e.g. items that share the same product code but are actually distinct). With real-time application and fast training cycles in mind, we implement several neural methods and compare their performance in terms of accuracy and speed on selected English to German catalogue entries.

The paper is organized as follows: in Section 2 we present an overview of existing methods that tackle similar problems; in Section 3 we provide further motivation to our work by elaborating on use-cases; in Section 4 we describe the data we used; in Section 5 we discuss the methods and methodology we undertook; our experiments and results are summarized in Section 6; we conclude and raise points for future work in Section 7.

2 RELATED WORK

The exponential growth of multilingual content on the web, including commercial content, has created the necessity of higher-level taxonomic organization. Prytkova et al. [18] assess the necessity for alignment of multilingual taxonomies and propose several methods, including a string-similarity method using the Wikipedia taxonomy as a translation or a mapping medium. Among others, Fu et al. [4] and Spohr et al. [23] argue that, in the context of cross-lingual ontology matching, the quality of the Machine Translation (MT) system used is of major importance. The work of Nikoulina et al. [15] investigates cross-lingual search in library catalogues using MT adapted with a corpus of bilingual queries. The work of Guha and Heger [5] and, more recently of Sloto et al. [22], present challenges and solutions for MT for an e-Commerce vendor. In one of our approaches we also exploit MT to translate the German catalogues into English, and then compute string (text) similarity between items from the two catalogues in order to identify which matching entries are correctly aligned and which are not.

To bridge the language barrier in the context of cross-lingual information retrieval (IR), Eigen-analysis has also been used. In [9]

---

[1]Within this work an *item* is viewed as a container that encapsulates data such as title, description, identifier, etc., about a specific product that is for sale; multiple items can represent the same product; items are organized in the leaves of the catalogue trees we investigate.

Table 1: Aligned entries from the training material: sample English entry, one of the UPC-aligned German entries, and a human translation of the German title to better read the differences.

EN:	Schleich 13818 Hanoverian Foal Toy Figure, Brown, For Ages 3+
DE:	Schleich Hannoveraner Fohlen 13818
trans.:	*Schleich Hannoverian Foal 13818*
EN:	Siku 4066 Crown Big x 580 Forage Harvester Scale 1:3 2 NEW! °
DE:	SIKU Krone Big X 580 Maishäcksler (Spielware) NEU
trans.:	*SIKU Crown Big X 580 Forage Harvester (Toy) NEW*
EN:	FLOWER SHOW JIGSAW PUZZLE Heye Puzzles 1000 Pieces Bright Lily (NEW FOR 2016)
DE:	Carletto 3329739 Bright Lily Standard 1000 Teile SV
trans.:	*Carletto 3329739 Bright Lily Standard 1000 Pieces SV*

latent semantic analysis (LSA for short, also known as LSI – latent semantic indexing) has been applied on concatenated multi-lingual documents for cross-lingual document IR. Vinokourov et al. [27] uses canonical correlation analysis (CCA) for cross-lingual semantic text representation.

In text classification word representations that map words (or tokens) into vectors in a common vector space are commonly used. The works of Mikolov et al. [11], Pennington et al. [16], Turian et al. [25] and Peters et al. [17] have delivered high-quality word representations – word embeddings – induced through neural networks trained on monolingual data. Word embeddings have proven to be effective in numerous NLP tasks, such as sentiment analysis, textual entailment, and MT. Lai et al. [8] exploit pre-trained Skip-gram word embeddings for their Recurrent Convolutional NN (CNN) models for text classification. Word embeddings have also been successfully used for twitter sentiment classification [21, 24].

Mueller and Thyagarajan [13] present an RNN adaptation of a Siamese architecture [2, 7] with Long Short-Term Memory (LSTM) units [6] for computing text similarity. Another LSTM approach for the task of text entailment is presented in [19]. To improve the performance of their network, the authors exploit a word-to-word attention mechanism. Attention mechanisms have been successfully used for NLP tasks such as machine translation [1, 10], sentence summarization [20] and digit classification [12].

In our work, we are driven by a real-world application scenario. Therefore, we aim at a system that is not only robust, and with high predictive capabilities, but is also optimized towards speed and code sustainability in a large commercial environment. We draw a road map over different LSTM RNN network methods with and without attention as well as with and without pre-trained embeddings. We experiment with original as well as with machine translated data.

3 USE-CASES

Measuring the similarity of entries (or items) in cross-language e-commerce catalogues is essential for aligning products in the catalogue trees. Identifying which items represent the same product(s) across catalogues in different languages is fundamental for three use-cases:

UC1 **find in-domain comparable data.** Identifying the same or comparable catalogue entries is a way to create parallel corpora for training domain-specific MT engines. Given the high volume of data being published on a daily basis on e-Commerce websites, such corpora could encapsulate enough parallel text for high-quality MT. In addition, the organization of products in hierarchical catalogues allows data to be categorized comparatively from domain-specific to more general-domain responding to different MT requirements [28].

UC2 **synchronize content**. Sub-parts of product descriptions across language sites can be used for complementary knowledge exchange (e.g. by using MT to fill or enhance missing parts) thus improving the quality of product descriptions.

UC3 **detect erroneous/misaligned entries**. Automatically detecting erroneous/misaligned entries would help e-Commerce vendors to further improve the cohesion of their catalogues.

4 DATA

We considered two catalogue trees by eBay, one in English (EN) and one in German (DE). The catalogue entries contain the title of the item for sale (or an *item title*) with a maximum length of 80 UTF-8 characters (which can be noisy with characters representing emojis, for example). Table 1 shows examples of item titles together with their human translations. Next to the title for the item itself, we have access to other meta-information provided by the seller such as colour, quantity, the manufacturer, or other product specifics. However, their precision and coverage over the whole data is not complete.

To gather parallel training, test and validation data, we used the 12-digit universal product code (UPC) as well as its superset European Article Number (EAN-13) (again, entered by the sellers) to extract aligned items. The UPC and EAN numbers are unique per product and are shared among catalogues in different languages.

For each item, we also know which category (cars, toys, books, etc.) they are placed in, and we know from preliminary experiments that not all of them offer a fair challenge: for example, movies have a strong localization bias (e.g. "Soylent Green" has the title "... Jahr 2022 ... die überleben wollen" – "... year 2022 ... those that want to survive" when backtranslated from German), whereas music CDs are often verbatim. Thus, we restrict ourselves to the categories *Home and Gardening*, *Toys* and *Cameras & Photos*. We present details about our data in Table 2.

Table 2: Summary of the training/test/validation data.

category	reason for selection	# unique UPCs	# vocabulary entries for		
			EN	DE	EN + DE
Home and Gardening	large variety in the items / brands	800k	105,429	65,124	129,490
Toys	high amount of used items with more user-generated content	800k	99,862	61,586	121,351
Cameras & Photos	lots of technical product descriptions	160k	30,997	23,956	42,335

5 APPROACHES

First, we built three neural models in a language-agnostic way to compute the similarity between item titles and identify whether they are the same item in the different catalogues. Second, we used MT to reduce the cross-lingual problem to a monolingual one and then retrain the models. As stated in previous work, such a language-aware approach depends on the quality of the MT system [4, 23]. This MT system is used in the production environment of eBayfor the English-German direction and has been optimized on title content.

We also experimented with and without pre-trained embeddings. When this is not made explicit in the following of this paper, we assume the embeddings are trained from our parallel data.

5.1 Language-agnostic similarity

In order to identify aligned items without considering the language as a factor, i.e. language-agnostic, we implemented three neural approaches:

ClassifierCat. We concatenate two sequences to form a joint input sequence that is given to a bidirectional LSTM RNN, in which the last hidden state is used in a soft-max layer for classification. The network predicts a probability distribution over n classes; the highest probability indicates to which of the n classes the input belongs. For UC1 and UC3, $n = 2$, same/different, while for UC2, (text synchronisation based on text entailment task), $n = 3$, e.g. positive entailment / negative entailment / contradiction. Our implementation is generic enough to allow both these tasks to be handled.

Siamese. A Siamese neural network combines two (or more) networks that have the same architecture and share the same weights, each of which takes as input one of two (or more) input sequences independently. It has already been successfully applied for text similarity in [13, 14, 26].

At training time the network parameters are optimized to compute a similarity score that would minimize the loss (in our case, mean squared error by default) – more similar input sequences will have a higher score. At test time, the output of the network is simply the similarity score between the input sequences. We use a distance metric, i.e. Euclidean distance, to compute this score. That is, in a multidimensional vector space, the Euclidean distance between the representations of the input sequences expresses their similarity (the smaller the distance, the higher the similarity).

Our Siamese architecture is focused on computing the similarity between two input sequences distance, i.e. at prediction time it will compute a value stating how similar the inputs are. While this is very suitable for UC1 and UC3, it is not suitable for UC2 as it will not handle a third dimension of comparison, as is the case of textual entailment. Accordingly, this approach is used only for handling UC1 and UC3.

ClassifierAttn. On long sequences LSTMs do not perform well as they need to compress all the information of a sequence in one context vector, i.e. the last state of the network. To solve this problem, attention mechanisms have been introduced [10] which allow the network to focus on parts of the sequence(s) that have the greatest importance. We implemented two attention mechanisms: (i) word-by-word attention inspired by [19] which we refer to as AttentionRTE[2], and (ii) Soft Dot Attention which we refer to as AttentionDot. We use soft attention instead of hard attention as we aim to provide a smooth representation of the encoded sequence where the important points are weighted accordingly, rather than select only a single point of interest and ignore the rest. Furthermore, we select dot attention as it is very fast (e.g. compared to additive attention mechanisms) and has shown to be very effective. The implementation of the ClassifierAttn model is similar to the ClassifierCat when it comes to the underlying LSTM network(s). However, there are two networks instead of one; two separate input sequences are provided: one from the L1 catalogue and the other from the L2 catalogue; also while for the ClassifierCat and the Siamese models a joint vocabulary is used, for the ClassifierAttn we use two different vocabularies: one for each language.

During the preprocessing step (prior to training the NN models), each sentence is tokenised and a beginning-of-sentence and end-of-sentence tokens (<bos> and <eos> respectively) were added to identify these positions. In addition, when joining two sentences for the ClassifierCat a <break> token was used inbetween to identify the joining point.

5.2 Language-aware similarity

The ClassifierCat approach and the Siamese approach use a shared embedding across both languages. We can thus compare their language-agnostic performance against a data set where the English titles are translated by a production facing English–German title translation service provided by eBay.

For the attention approach, the architecture contains separate embeddings for source and target. For this, we conducted early experiments using word2vec [11] embeddings trained on the individual languages, using MT on the tokens, and applying CCA to transform the English embedding space into the German. Then, we initialised the ClassifierAttn embedding layer with these pre-trained embeddings. This approach is similar to [3].

---

[2]We refer to this model as RTE, as the attention we use follows the work on Reasoning about Textual Entailment (RTE)

Table 3: Performance on the test set for all methods for all data (first four rows: combined; below: per category).

Category	Method	Time [min]	Epochs	Time/Epoch [min]	Precision	Recall	F1
all	ClassifierCat	440	20	22.0	0.955	0.974	0.964
	Siamese	243	13	18.7	0.837	0.961	0.896
	AttentionDot	455	17	26.8	0.962	0.944	0.953
	AttentionRTE	470	17	27.6	**0.962**	**0.972**	**0.967**
Toys	ClassifierCat	167	21	7.9	0.922	0.979	0.950
	Siamese	233	34	6.85	0.741	0.881	0.805
	AttentionDot	300	33	9.0	0.966	0.990	0.978
	AttentionRTE	314	34	9.2	0.955	0.981	0.968
'Cameras & Photos'	ClassifierCat	93	24	3.8	0.963	0.978	0.970
	Siamese	40	10	4.0	0.500	0.992	0.665
	AttentionDot	60	13	4.6	0.945	0.971	0.958
	AttentionRTE	73	16	4.5	0.958	0.980	0.969
'Home & Gardening'	ClassifierCat	139	28	4.9	0.887	0.952	0.918
	Siamese	155	33	4.7	0.882	0.92	0.903
	AttentionDot	170	29	5.9	0.958	0.964	0.960
	AttentionRTE	132	22	6.0	0.957	0.978	0.967

Table 4: Results on the combined three categories using additional language knowledge via MT and multi-lingual embeddings

Method	Precision	Recall	F1
ClassifierCat	0.955	0.974	0.964
+ machine translation	0.958	0.972	0.965
Siamese	0.837	0.961	0.896
+ machine translation	0.925	0.976	0.950
AttentionDot	0.962	0.944	0.953
+ multi-lingual embedding	0.912	0.970	0.940
AttentionRTE	**0.962**	**0.972**	**0.967**
+ multi-lingual embedding	0.927	0.949	0.938

5.3 Implementation details

Our implementation uses PyTorch[3] and gensim[4] as packages for neural network support and embeddings, respectively. The toolkit we developed within the scope of this work consists of a pipeline with the following component classes:

(1) data handling: to handle the large volume of data we implemented a set of scripts that (i) ingest the hadoop output; (ii) extract information per field from the hadoop output; (iii) align data based on defined field, e.g. UPC or EAN number; (iv) filter non-unique tuples; and (v) convert the data into a suitable format for each of the aforementioned methods.
(2) operational components: these include scripts for invoking training models with the aforementioned methods and testing with these models.
(3) pipelines: we developed several pipelines that would wrap different scripts together in order to provide an end-to-end functionality (i) data alignment (based on UPCs etc.); (ii)

---

[3]https://pytorch.org/
[4]https://radimrehurek.com/gensim/

training word2vec, CCA, ClassifierCat, ClassifierAttn, Siamese; (iii) machine translation; (iv) training data preparation (including splitting into training, test and development sets) and (v) visualisation of results.

We also implemented a docker version of our toolkit; the quality of our implementation was continuously controlled through a series of regression tests.

6 EXPERIMENTS

From the data entries, we randomly assigned English entries with their German counterparts whenever they share the same UPC. For the experiments in this paper, we make the assumption that the UPCs are already known to the system. This means that development set and the test set contain entries from a withheld 2% of the parallel data (30K positive matches out of 145K for all categories) i.e. this pair match has not been encountered in the training. We created negative samples of double the size by randomly assigning titles from the same category but with a different product code.

For a fair comparison, we limited the embedding size of all methods to 100, and kept the number of hidden dimensions consistent to 50. Training was conducted with a batch size of 64, a patience of 5 and a maximum number of epochs of 100 (which was never exhausted in any training setting).

We summarize our experimental results for the language-agnostic case in Table 3; our results for the language-aware case are presented in Table 4.

Quality. First, our results for the language-agnostic case show that the attention-based models and especially the AttentionRTE outperform both the ClassifierCat as well as the Siamese network for the '**Toys**' and '**Home & Gardening**' categories as well as for **all** the data. The differences on the precision, recall and F1 metrics for the ClassifierCat, the AttentionDot and AttentionRTE models are quite small – for precision between 0.007 and 0.071; recall: 0.110 and 0.300 – indicating that these systems perform in a similar way. However, the observed difference between the AttentionDot,

Table 5: Attention for "3 packs 30 photos black frame fujifilm fuji instax mini film polaroid 7s sp 1"

attention	0.00	0.01	0.03	0.49	0.12	0.04	0.03	0.04	0.00	0.11	0.02	0.12
English token	\<sos>	3x	fuji	instax	mini	black	frame	schwarzer	rahmen	sofortbildfilm	sofortbild	\<eos>

Table 6: Excerpt attention for UPC-aligned "12 en 1 professional kit accessories bundle for gopro hd hero 4 3 2 1 sj4000"

attention	[..]	0.01	0.16	0.04	0.00	0.29	0.14	0.00	0.03	0.01	0.08	[..]
English token	[..]	kit	von	zubehör	für	gopro	hero	4.00	silber	schwarz	hero	[..]

AttentionRTE, the ClassifierCat and the Siamese networks range between precision: 0.076 and 0.463; recall: 0.021 to 0.109; and F1: 0.064 to 0.305, with the Siamese networks always underperforming. Furthermore, the Siamese network did not converge as well, when trained on the **'Cameras & Photo'** data. One reason for this might be that the similarity between items in a given category is too high for proper discrimination. Especially in the **Camera & Photo** category, where items have a distinct product code even for subtle changes in the title (obviously, "New Camera BRAND-XYZ Black model 5" and "New Camera BRAND-XYZ Black model 6" are always distinct product codes), the Siamese network fails to converge at all.

For the MT-supported experiments, while the impact on the ClassifierCat was negligible, our results show that MT aids the Siamese network. In particular the difference in scores for the Siamese network in comparison to the other methods is reduced for precision from 0.125 to 0.037, for F1 from 0.071 to 0.017 and in terms of recall, it performed best.

For the attention-based methods, we pre-computed word2vec skip-gram embeddings on the training parts, with the same size as the methods would use, i.e. 100. Then, we trained a linear CCA transformation via token-level MT system on the proportion where the MT could be linked to a target token. After applying the transformation on the whole embedding space, we then seeded the built-in embedding layer of the AttentionDot and the AttentionRTE with this shared embedding space. In our setting, this did not improve the performance (cf. Table 4), at least not in a setting where the pre-trained embedding is drawn from the same material as the actual classification.

Performance. On an Nvidia M40 GPU, our simplified models are quite quick to train as is clear from the measured time in Table 3. Even when using all the combined data, the maximum total training time is 470 minutes; maximum time per epoch is 27.6 minutes. It is also obvious that the attention-based methods are slower, but more robust. These times are very promising for a real-world application.

Note on attention. The attention-based methods have the benefit of yielding extra information on a word-level. While a detailed analysis of their performance is beyond the scope of this paper, we anecdotally found that attention puts more weight on product names than on brands; this behaviour is expected, as these are the trigger words that most often make the difference. We present two examples in Table 5 and Table 6. It is interesting to see that some German prepositions like "von" (of/from) or "mit" (with) gain high

attention as well; we believe that the attention mechanism learns to identify these trigger words when associated with a product title. This seems beneficial to our case study since accessories to products have their own UPC and could be identified with these trigger words making it easier to differentiation between "a case *for* Samsung" and "a case *for* iPhone".

7 CONCLUSIONS AND FUTURE WORK

In this paper, we described and evaluated three neural approaches for cross-language item title alignment. The initial experiments are encouraging. We showed that the methods work fast and reasonably well on this data set.

As future work, apart from embarking on the typical journey of feature engineering, we intend to increase the data challenge by limiting the development and the test set to entirely unseen UPCs (currently, we only ensure that no title pair was encountered in training).[5] In addition, we aim to use informed negative sampling (measured by, e.g. catalogue tree approximity).

For the language pair itself, additional focus to the language specifics could be applied, such as compound splitting for German and normalization of abbreviations/units.

While our preliminary experiments with pre-trained cross-language embeddings did not yield overall better result than the strongest attention-based systems we plan to further investigate this topic by experimenting with different embeddings. This approach is not limited to bilingual data and could be applied to a much larger monolingual data collection before training a shared embedding space. Among others we consider investigating the applicability of MUSE[6].

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of the 6th International Conference on Learning Representations (ICLR 2015)*. San Diego, CA, USA. 15pp.

[2] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1993. Signature Verification Using a "Siamese" Time Delay Neural Network. In *Proceedings of the 6th International Conference on Neural Information Processing Systems (NIPS 1993)*. Denver, Colorado, USA, 737–744.

[3] Manaal Faruqui and Chris Dyer. 2014. Improving Vector Space Word Representations Using Multilingual Correlation. In *Proceedings of the 14th Conference of the European Chapter of theAssociation for Computational Linguistics, EACL 2014*. Gothenburg, Sweden, 462–471.

[5] Given the high scores we have seen on our test sets, UPC-distinction seems to be the most dire next step. Since the constraint on the UPCs leads to different data splits, the results are not comparable; however, as a first impression, early experiments indicate accuracy scores around 0.91 (ClassifierCat) and 0.93 for the Attention-based methods.
[6] https://github.com/facebookresearch/MUSE

[4] Bo Fu, Rob Brennan, and Declan O'Sullivan. 2009. Cross-Lingual Ontology Mapping – An Investigation of the Impact of Machine Translation. In *The Semantic Web*, Asunción Gómez-Pérez, Yong Yu, and Ying Ding (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–15.

[5] Jyoti Guha and Carmen Heger. 2014. Machine translation for global e-commerce on eBay. In *Proceedings of the Eleventh Conference of the Association for Machine Translation in the Americas (AMTA 2014)*, Vol. 2. Vancouver, Canada, 31–37.

[6] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (Nov. 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[7] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese Neural Networks for One-shot Image Recognition. In *ICML Deep Learning Workshop*.

[8] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. [n. d.]. Recurrent Convolutional Neural Networks for Text Classification. In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence (AAAI 2015)*. Austin, Texas, USA, 2267–2273.

[9] Michael L Littman, Susan T Dumais, and Thomas K Landauer. 1998. Automatic cross-language information retrieval using latent semantic indexing. In *Cross-language information retrieval*. Springer, 51–62.

[10] Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, (EMNLP 2015)*. Lisbon, Portugal, 1412–1421.

[11] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[12] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. 2014. Recurrent Models of Visual Attention. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS 2014)*. Montreal, Canada, 2204–2212.

[13] Jonas Mueller and Aditya Thyagarajan. 2016. Siamese Recurrent Architectures for Learning Sentence Similarity. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI 2016)*. Phoenix, Arizona, 2786–2792.

[14] Paul Neculoiu, Maarten Versteegh, and Mihai Rotaru. 2016. Learning Text Similarity with Siamese Recurrent Networks. In *Proceedings of the 1st Workshop on Representation Learning for NLP, (Rep4NLP@ACL 2016)*. Berlin, Germany, 148–157.

[15] Vassilina Nikoulina, Bogomil Kovachev, Nikolaos Lagos, and Christof Monz. 2012. Adaptation of Statistical Machine Translation Model for Cross-lingual Information Retrieval in a Service Context. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*. Avignon, France, 109–119.

[16] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, (EMNLP 2014), A meeting of SIGDAT, a Special Interest Group of the ACL*. Doha, Qatar, 1532–1543.

[17] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (NAACL-HLT 2018)*, Vol. 1. New Orleans, Louisiana, USA, 2227–2237.

[18] Natalia Prytkova, Gerhard Weikum, and Marc Spaniol. 2015. Aligning Multi-Cultural Knowledge Taxonomies by Combinatorial Optimization. In *Proceedings of the 24th International Conference on World Wide Web Companion, (WWW 2015) - Companion Volume*. Florence, Italy, 93–94.

[19] Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomás Kociský, and Phil Blunsom. 2015. Reasoning about Entailment with Neural Attention. *CoRR* abs/1509.06664 (2015). arXiv:1509.06664 http://arxiv.org/abs/1509.06664

[20] Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A Neural Attention Model for Abstractive Sentence Summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, (EMNLP 2015)*. Lisbon, Portugal, 379–389.

[21] Aliaksei Severyn and Alessandro Moschitti. 2015. UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado, 464–469.

[22] Steve Sloto, Ann Clifton, Greg Hanneman, Patrick Porter, Donna Gates, Almut Silja Hildebrand, and Anish Kumar. 2018. Leveraging Data Resources for Cross-Linguistic Information Retrieval Using Statistical Machine Translation. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas, (AMTA 2018)*. Boston, MA, USA, 223–233.

[23] Dennis Spohr, Laura Hollink, and Philipp Cimiano. 2011. A Machine Learning Approach to Multilingual and Cross-Lingual Ontology Matching. In *The Semantic Web – ISWC 2011*, Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Noy, and Eva Blomqvist (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 665–680.

[24] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*. Baltimore, Maryland, USA, 1555–1565.

[25] Joseph Turian, Lev Ratinov, and Yoshua Bengio. 2010. Word Representations: A Simple and General Method for Semi-supervised Learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*. Uppsala, Sweden, 384–394.

[26] Nicola Ueffing, José GC de Souza, and Gregor Leusch. 2018. Quality Estimation for Automatically Generated Titles of eCommerce Browse Pages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (NAACL-HLT 2018)*. New Orleans, Louisiana, USA, 52–59.

[27] Alexei Vinokourov, John Shawe-Taylor, and Nello Cristianini. 2002. Inferring a Semantic Representation of Text via Cross-language Correlation Analysis. In *Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS 2002)*. Vancouver, British Columbia, Canada, 1497–1504.

[28] Jian Zhang, Xiaofeng Wu, Andy Way, and Qun Liu. 2016. Fast Gated Neural Domain Adaptation: Language Model as a Case Study. In *Proceedings of the 26th International Conference on Computational Linguistics, (COLING 2016)*. Osaka, Japan, 1386–1397.