DIFFERENT WORD REPRESENTATIONS AND THEIR COMBINATION FOR PROPER NAME RETRIEVAL FROM DIACHRONIC DOCUMENTS

Irina Illina, Dominique Fohr

MultiSpeech team
Université de Lorraine, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France
Inria, Villers-lès-Nancy, F-54600, France
CNRS, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France

ABSTRACT

This paper deals with the problem of high-quality transcription systems for very large vocabulary automatic speech recognition (ASR). We investigate the problem of automatic retrieval of out-of-vocabulary (OOV) proper names (PNs). We want to take into account the temporal, syntactic and semantic context of words. Nowadays, Artificial Neural Networks (NN) are widely used in natural language processing: continuous space representations of words is learned automatically from unstructured text data. To model the latent topics at document level, Latent Dirichlet Allocation (LDA) has been successful.

In this paper, we propose OOV PN retrieval using (1) temporal versus topic context modeling; (2) different word representation spaces for word-level and document-level context modeling; (3) combinations of retrieval results. Experimental evaluation on broadcast news data shows that the proposed method combinations lead to better results. This confirms the complementarity of methods.

Index Terms -- speech recognition, neural networks, LDA, vocabulary extension, out-of-vocabulary words, proper names.

1. INTRODUCTION

In the case of large vocabulary ASR, out-of-vocabulary words can lead to a performance degradation and sometimes the loss of meaning of the sentence to be transcribed. This is especially true for the missing *Proper Names* (PN). For the task of broadcast document indexing, PNs often contain key information for understanding. In this paper, we address the problem of automatic retrieval of PNs for the purpose of adding them to the vocabulary of the speech recognition system.

In our previous work, lexical and temporal context modeling have been used to augment the vocabulary of the ASR system with new PNs. Our hypothesis is that PNs evolve through time, and that for a given date, the same PNs would occur in other documents that belong to the same period [16][27]. Proposed models were based on mutual information and a bag of words model [14]. Temporal contexts have been proposed before by Federico and Bertoldi [7] to cope with language and topic changes, typical to new domains, and in [18] for OOV prediction in recognition outputs. In contrast to these works, our work extends vocabulary using shorter time periods to reduce excessive vocabulary growth. In this paper we continue to exploit temporal context.

Another issue discussed in our paper is how to represent the semantic context of words. Recently, several fine-grained and coarse-grained semantic models have been proposed. Among coarse-grained methods, the LDA model is widely used [3]. This generative model represents each word as a sample from a mixture of global word distributions, where the mixture weight varies between documents. LDA represents the documents as a distribution over topics. To better link the objective function of LDA with the evaluation metric of the retrieval task, different improvements of LDA have been proposed. The *Linear Discriminative Projection* model for query-document matching at the semantic level uses *clickthrough data* (list of queries and their clicked documents) [10]. The *Replicated Softmax Model* [28] proposes binary distributed representation of the documents with latent variables representing topic features. The *Deep Structured Semantic Model* (DSSM) is proposed in [11] for the web search task using clickthrough data.

Among fine-grained representations, *Mikolov*'s model assumes that semantically close words are close in the continuous space *Word Representations* (*WR*) [22][20][21][6]. The neural network is trained using unlabelled data to capture statistical semantic relations of words. The *Convolution Neural Network*, proposed by [15] for sentence classification tasks, uses Mikolov's WR to pre-train the network. The author suggests that Mikolov's WR can be used as universal or/and initial features for several classification tasks. The *Neural Autoregressive Topic Model* [17] is a network topic model that uses hierarchical distribution of words.

The *GloVe* (Global Vectors) model tries to have the best of both worlds: fine-grained and coarse-grained representations [24]. The model is based on a global log-bilinear regression, and tries to keep meaningful structure of the word space, which is important for word similarity tasks.

In our paper, we study these 2 classes of WR spaces for our task of OOV PN retrieval. We want to use the complementarities of different models by combining different systems. In information retrieval, data fusion is used to accurately estimate the relevance of retrieved documents by combining the information from different system outputs. The system outputs should be compatible, accurate in terms of probability relevance and independent of each other [5]. To have compatible outputs, score normalization techniques bring different scores in common range [26][30] and represent a very important factor in optimizing the combination performance [19]. We study different normalization and combination techniques for our model outputs.

So, the contribution of our paper is threefold. First, we exploit the temporal and topic context modelling. Secondly, we apply fine-grained and coarse-grained models to model semantic context. Finally, we suggest combining the advantages of different model families of WRs.

Section 2 introduces the proposed methodology. Sections 3 and 4 describe the experimental sets and the evaluation results.

2. PROPOSED METHODOLOGY

2.1. General approach

Our methodology is based on the temporal and lexical context proposed in [14]. Furthermore, we propose to use *topic* information.

We use text documents from the diachronic corpus that are contemporaneous with each test document to be transcribed. So, we have a test audio document (to be transcribed) which contains OOV words, and we have a diachronic text corpus, used to retrieve OOV PNs. We assume that a PN from the test corpus will co-occur with other PNs in *diachronic documents* (DD). These documents can be from the same time period (as in [14]) and/or from the same topic. These co-occurring PNs might contain the targeted OOV words. The idea is to exploit the relationship between PNs for a better lexical enrichment. An augmented vocabulary is dynamically built for each test document to avoid an excessive increase in vocabulary size.

Our methodology, similar to [8], contains 5 steps:

A) *In-vocabulary (IV) PN extraction from each test document*: For each test document, we extract IV PNs from the automatic transcription obtained using our standard vocabulary. The goal is to use these PNs as anchors to collect linked new proper names from the diachronic corpus.

B*) Selection of DDs and extraction of new PNs from them*: only DDs that correspond to the *same time period and/or same topic* as the test document are considered. After POS-tagging of these DDs, meaningful words are kept: verbs, adjectives, nouns and PNs. Among these PNs, we create a list of those that do not belong to our standard vocabulary (OOV_PN).

C) *Lexical and semantic context extraction from DDs:* The goal is to extract the most relevant OOV PNs. After extracting the list of the IV PNs from the test document, and the list of the new PNs from DDs, we build their lexical and semantic contexts. For this, a high-dimensionality WR space is used (see Section 2.2). We hope that in this space semantically and lexically related words will be in the same region of the space.

D*) Ranking of new PNs:* the cosine-similarity metric is calculated between the IV PNs found in the test documents and each new OOV PN occurring in diachronic set in the WR space.

E) *Vocabulary augmentation:* to reduce the vocabulary growth, we select the top-N OOV PNs according to the cosine-similarity metric and add them to vocabulary. OOV PN pronunciations are generated using a phonetic dictionary or an automatic phonetic transcription tool [12].

Using this methodology, we expect to extract a reduced list (compared to the baseline, cf. Section 4.1) of all the potentially missing PNs.

2.2. Different word representation spaces

In step B, the selection of DDs can be performed using the time period or using the topic information. To model the topics, we use LDA and for time information we use the date of the test document.

In step C, we propose to use the Cosine-based representations, and those of Mikolov and GloVe. In the following, we will detail the using of each representation.

2.2.1 Topic context modeling

LDA is a generative model [4], where each document is presented as a mixture of topics. Topic distribution is assumed to have a *Dirichlet* prior. Each word is drawn from one of the document's topic.

In our approach, LDA is used to create a list of relevant documents from the diachronic corpus (step B). Our aim is to select the DDs that match the topics of the document to be transcribed. After this, relevant OOV PN words can be retrieved.

2.2.2 Time context modeling

Cosine-based method. Here, the *Bag of Words* (*BOW)* vector space document representation is used: each document is considered as a set of words, disregarding the word order. During step C of our approach, each document (diachronic or test) is represented as a BOW vector of meaningful words. For each PN, a word vector is computed as the sum of all BOW vectors in which this PN occurred. After this, the cosine similarity between the test BOW vector and all the PN vectors is computed (step D).

Mikolov's Neural Networks word representation space. The goal of Mikolov's model [22][20][21] is to capture a large number of semantic and syntactic word relationships using huge amounts of unstructured text data. Linguistic regularities and patterns are learned using continuous distributed context representation of words, maximizing accuracy and minimizing computational complexity. Compared to classical NNs, the non-linear hidden layer is removed and the projection layer is shared for all words. In [8], we proposed to use this model for our OOV PN retrieval task. At step C of our general approach, Mikolov's model can be used: each word in this space is represented by a continuous vector of high-dimensionality.

GloVe word representation space. This is a global log-bilinear regression model using unsupervised training for WRs [24]. Like Mikolov's model, it takes into account semantic relationships between the words but the training is performed on the global co-occurrence counts. GloVe outperforms state-of-the-art models on word analogy and word similarity. As previously, only step C is modified: each word is represented in the GloVe-space.

The presented approaches give 6 systems: using time information or topic information to choose DDs, using Cosine, Mikolov's or GloVe's representations.

2.3. List combination strategies

Different OOV PN retrieval systems can retrieve different OOV PN lists while achieving similar performance. This can be explained by the fact that each system uses some kind of context modeling (temporal, lexical, semantic, topic). So, the WR spaces are different and can lead to different retrieved OOV PN lists.

To combine several systems and to make the system outputs comparable, normalization techniques are required to scale the output scores. State-of-the-art normalization techniques are heuristic and their performance seems to depend on the used dataset [23]. Concerning the combination techniques, several studies have shown that scores are more informative than rankings for retrieval tasks [30][1][29].

In this paper, we propose to combine different system and we study different combination strategies. If we assume that each OOV PN retrieval system gives a list of OOV PNs (*list1, list2, etc.*), the goal of these combination methods is to keep only the *N-best* OOV PNs. The lists are sorted by word scores and this corresponds to rank. We present our combination methods only for combination of 2 lists. Extension for more lists is straightforward. As different system outputs are not compatible in terms of scores, score normalization techniques have been used before the system combination. We use score normalisation from [23] because they show a good performance in data retrieval task:

- ***Standard:*** shift min to 0 and scale max to 1;
- ***Sum:*** shift min to 0 and scale the sum to 1;
- ***ZMUV:*** shift mean to 0 and scale variance to 1.

After the score normalisation, we propose to use the following combination methods, based on ranking or scoring.

Let *list1* and *list2* be two lists of OOV PNs to combine, sorted by normalized word scores.

Method 1 (rank-based): take the top *N/2* OOV PNs from *list1* and the top *N/2* OOV PNs from *list2* to obtain a resulting list of N OOV PNs.

Method 2 (score-based): concatenate *list1* and *list2*, sort the resulting list according to scores and keep only the *N-best* OOV PNs.

Method 3 (rank-based): find OOV PNs common to *list1* and to *list2*, take the top *N/3* words of this list. Complete resulting list by taking *(2/3 * N)* of OOV PNs according to Method 1. In this method, the words common to both lists are favoured.

3. EXPERIMENTS

In this paper, s***elected PNs*** correspond to the new PNs that we were able to extract from DDs using our methods. ***Retrieved OOV PNs*** correspond to the *selected PNs* that are present in the test documents. Using these *selected PNs*, a specific augmented lexicon for each test document was built. Results are presented in terms of Recall (%): the number of retrieved OOV PNs versus the number of OOV PNs.

3.1. Development and test corpora

Development corpus: seven audio documents of development part of ESTER2 (between 07/07/2007 and 07/23/2007) [9]. The aim of the ESTER2 evaluation campaign (2007 to 2009) was to evaluate automatic radio broadcasts rich transcription systems for the French language.

Test corpus: 13 audio documents from RFI (*Radio France International*) and *France-Inter* (test part of ESTER2) (between 12/18/2007 and 01/28/2208).

Automatic transcription of the development and test corpus is performed using the ANT system, trained on 200-hour broadcast news audio files [13].

The average number of occurrences of all PNs (IV and OOV) in development and test documents with respect to 122k-word ASR vocabulary is presented in Table 1. To artificially increase the OOV rate, we randomly removed 223 PNs occurring in the development and test set from our 122k ASR vocabulary. Finally, the OOV PN rate is about 1.2%.

File	Word occ	IV PNs	IV PN occ	OOV PNs	OOV PN occ
Dev	4525.9	99.1	164.0	30.7	57.3
Test	4024.7	89.6	179.7	26	46.6

Table 1. Statistics per file of the development and test corpora.

3.2. Diachronic corpus

The French *GigaWord* corpus is used as the diachronic corpus: newswire text data from *Agence France Presse* (AFP) and *Associated Press Worldstream* (APW) from 1994 to 2008. The choice of GigaWord and ESTER corpora was driven by the fact that one is contemporary with the other, their temporal granularity is the day and they have the same textual genre (journalistic) and domain (politics, sports, etc.).

4. EXPERIMENTAL RESULTS

We used the development corpus to set the parameters and we used the test corpus for the evaluation. The best results per period are highlighted in bold in Tables.

4.1 Baseline results

The baseline method consists in extracting a list of all the new PNs occurring in a diachronic corpus using a time period corresponding to the test document. As time period, we choose to use a day, a week and a month. Our vocabulary is augmented with the list of extracted OOV PNs.

Using TreeTagger [25], we extracted 160k PNs from 1 year of the diachronic corpus. Of these 160k PNs, 119k are not in our lexicon. Of these 119k, only 151 PNs are present in the development corpus (193 in the test corpus). So, it is necessary to filter this list of PNs to have a better tradeoff between the PN lexical coverage and the increase in lexicon size.

Period to select DDs	Av. of sel.PNs per dev file	Average of retrieved OOV PNs per dev file	Recall (%)
1 day	532.9	10.0	32.6
1 week	2928.4	11.4	37.2
1 month	13131.0	17.6	57.2
1 year	118797.0	24.0	**78.1**

Table 2. Baseline results for development corpus according to time periods.

Table 2 shows that using the DDs of 1 year, we retrieve, on average, 118797.0 PNs and 24.0 OOV PNs per development file (compared to 30.7 in Table 1), which corresponds to a recall of 78.1%.

4.2 Results for different word representation spaces

Mikolov and GloVe softwares, available on the web, are used. Mikolov's NN (*called NN below*), GloVe and LDA are trained on the diachronic corpus (cf. Section 3.2) but only meaningful words are kept.

After several experiments, the best parameter set for Mikolov's NN is 400 for the size of the hidden layer, 20 for the context size and 5 training epochs. We observed that the *Skip-gram* works better than Mikolov's *continuous bag-of-word model*.

For GloVe, window size is 10 and the space dimension is 200. For the LDA experiments, we used 200 topics.

For the month period, the OOV PNs occurring fewer than 6 times in the selected DDs are excluded.

For temporal selection of DDs, to obtain a good recall with a reasonable number of selected PNs, we fixed the number of selected PNs for each time period: 80 for a day, 440 for a week and 2000 for a month. These numbers correspond to the operating point 15% of the average number of selected PNs per development file (cf. Table 2: 15% of 532, of 2928 and of 13131).

Period to select DDs	Method	Selected PNs	Retrieved OOV PNs	Recall (%)
1 day	*temp*Cos	80	7.1	23.3
	*temp*NN	80	7.4	**24.2**
	*temp*GloVe	80	7.3	23.7
1 week	*temp*Cos	440	9.3	30.2
	*temp*NN	440	9.9	**32.1**
	*temp*GloVe	440	9.7	31.6
1 month	*temp*Cos	2000	12.1	39.5
	*temp*NN	2000	14.4	47.0
	*temp*GloVe	2000	14.7	**47.9**

Table 3. Recall results using the temporal information to select relevant DDs. Development corpus.

For topic-based DD selection, DDs are selected from 1 year of the diachronic corpus and ranked according to their relevance using LDA. We keep the same number of selected DDs as with time period selection: 800 documents for a day period, 5300 for one week and 22000 for one month. To be able to compare the results with results obtained by the temporal selection of DDs, we keep the same number of selected PNs, 80, 440 and 2000.

Table 3 shows the OOV PN recall averaged over the development files. For DD selection (step B), temporal information is used (denoted by prefix *temp*). From Table 3, for all studied time periods, the NN and the GloVe systems achieve better performance than cosine-based systems: the high-dimensionality continuous space WR given by NN and Glove is more efficient. NN perform slightly better than GloVe for a day and a week period, while GloVe obtains better results for a month period.

Nbr of selected DDs	Method	Sel. PNs	Retrieved OOV PNs	Recall (%)
800	*topic*NN	80	3.3	10.7
	*topic*GloVe	80	4.0	**13.0**
5300	*topic*NN	440	7.9	25.6
	*topic*GloVe	440	8.9	**28.8**
22000	*topic*NN	2000	13.6	44.2
	*topic*GloVe	2000	15.1	**49.3**

Table 4. Recall results using LDA to select relevant DDs according to topics. Development corpus.

Table 4 shows the recall results using LDA topic information for DD selection (denoted by prefix *topic*). Table 4 results are poorer compared to Table 3 (excluding GloVe for 22000 selected DDs). So, selecting the DDs according to the topics, leads to performance degradation. GloVe manages better recovery of the PNs for the month period.

4.3 List combination results

List combination can probably benefit OOV PN retrieval. OOV PN output lists are combined according to 3 methods presented in Section 2.3. For the score-based combination (method 2), Table 5 shows the recall for different score normalizations. The goal is to choose the best score normalization method. This Table presents the results only for combination method 2, but the two other combination methods give similar comportment and the results are not presented here.

In Table 5, 6 systems are evaluated: three first lines correspond to combination using only the time period for diachronic document selection. In the three last lines, we combine the OOV PN lists obtained using temporal (week or month) and topic (5300 or 22000) document selection. For example, the *topic*NN+*temps*NN line and 440 column correspond to: for the *topic*NN, the selection of 5300 DDs (440 best PNs are selected); for *temps*NN, the selection of the DDs from the same week that the document to transcribe (440 best PNs are selected).

On average, ***Sum*** normalization gives the best results and will be used for further evaluation.

To evaluate the best combination of word space representations, we compute the recall for different DD selection methods (temporal and topic), and for the different combination methods, method 1 (M1), method 2 (M2) and method3 (M3).

Table 6 gives the combination results using *Sum* score normalization. The combination of 3 systems has not improved the recall and is not presented in this paper. For Method 3, time period DD selection gives the same OOV PN lists and, thus, is not presented here.

Methods	Standard		SUM		ZMUV	
Period to select DDs for *temp*XX	**week**	**month**	**week**	**month**	**week**	**month**
Number of selected DDs for *topic*YY	**5300**	**22000**	**5300**	**22000**	**5300**	**22000**
Number of selected PNs	**440**	**2000**	**440**	**2000**	**440**	**2000**
*temp*NN+ *temp*GloVe	32.6	47.9	32.6	46.5	32.6	42.8
*temp*NN+ *temp*Cos	33.5	46.1	33.5	46.1	33.5	43.3
*temp*GloVe+ *temp*Cos	32.6	43.3	32.6	46.5	29.8	41.9
*topic*NN+ *temp*NN	**34.4**	46.5	**34.4**	**52.1**	**34.4**	48.4
*topic*NN+ *temp*GloVe	31.6	47.0	33.0	**52.1**	32.1	49.8
*topic*NN+ *temp*Cos	31.6	45.1	**34.4**	49.3	32.1	47.0

Table 5. Different score normalization results in terms of Recall (%) for combination method 2. Development corpus.

We observe that, on one hand, using only the time period DD selection and different word space representations (first part of Table 6) decreases the recall for month period compared to best individual system *temp*GloVe (Table 3): 46.5% versus 47.9%. For a day and a week period we observe the opposite. On the other hand, combining the systems with time period and topic-based DD selection provides substantial increase in recall: from 32.1% to 36.7% for a week, from 49.3% to 54% for a month (between 9% and 14% relative, for *topic*GloVe+*temp*NN). So, it is better to combine the systems obtained using different DD selection methods (topic and temporal) and different word space representations (GloVe and Mikolov's NN).

On average, rank-based method 3 (M3) gives the best results. So, it is a good idea to use the words common to both lists.

Period to select DDs for *temp*XX		**day**	**week**	**month**
Number of selected DDs for *topic*YY		**800**	**5300**	**22000**
Number of selected PNs		**80**	**440**	**2000**
*temp*NN+ *temp*GloVe	M1	24.7	31.6	46.5
	M2	26.1	32.6	46.5
*temp*NN+ *temp*Cos	M1	24.2	32.6	46.5
	M2	26.1	33.5	46.1
*temp*Cos+ *temp*GloVe	M1	22.8	32.1	46.5
	M2	27.0	32.6	46.5
*topic*NN+ *temp*NN	M1	20.0	36.3	51.6
	M2	24.2	34.4	52.1
	M3	20.5	**37.7**	52.6
*topic*NN+ *temp*GloVe	M1	20.5	34.4	52.6
	M2	23.3	33.0	52.1
	M3	20.9	36.3	52.1
*topic*NN+ *temp*Cos	M1	19.5	33.5	49.3
	M2	**25.1**	34.4	49.3
	M3	20.0	36.3	49.3
*topic*GloVe+ *temp*NN	M1	21.4	36.7	53.5
	M2	24.2	33.0	53.0
	M3	21.4	37.2	**54.0**
*topic*GloVe+ *temp*GloVe	M1	19.5	35.4	53.0
	M2	23.3	30.2	52.6
	M3	21.9	35.8	53.5
*topic*GloVe+ *temp*Cos	M1	20.5	33.5	51.2
	M2	24.2	32.1	49.3
	M3	20.5	35.8	51.6

Table 6. Combination results in terms of Recall (%) according to combination methods (**Sum** score normalization). Development corpus.

4.4 Recognition results on the test corpus

In the previous section, we have selected the parameters and the combination system that give the best recall on the development corpus: *topic*GloVe+*temp*NN and M3. This configuration will be used on the test corpus (see section 3.1). For *temp*NN, the DDs are selected using the time period and for *topic*GloVe, the DDs are selected using topic (cf. Table 5 or 6).

We begin by verifying our previous conclusion: the recall using the *topic*GloVe+*temp*NN system increases compared to the best unique system: *topic*GloVe.

The results on the test corpus confirm our conclusions: recall performance increases from 48.8% to 59.2% for a week, from 56.8% to 62.4% for a month for *topic*GloVe+*temp*NN (between 9% and 14% relative) compared to *topic*GloVe system.

The final goal of this word is to add the retrieved OOV PNs in our speech recognition system to decrease the *Word Error Rate* (WER).

In order to incorporate the new PNs in the language model, we re-estimated it for each augmented vocabulary using the large text corpus described in Section 3.2. The best way to incorporate the new PNs in the language model is beyond the scope of this paper.

For generating the pronunciations of the added PNs, we used the G2P CRF approach [12], trained on phonetic lexicon containing about 12000 PNs.

Table 7 shows the results in terms of WER and PNER. *PN Error Rate* (PNER) is also given and calculated like WER but taking into account only proper names. We observe that the augmented lexicon systems slightly decrease the WER, but this improvement is not significant. Regarding the PNER, the performance is significantly improved when an augmented lexicon is used. The combination system is more powerful when 440 or 2000 words are added.

Stand. Lexicon	Method	Augmented lexicon Number of added PNs		
		80	440	2000
WER 31.8	**WER** *topic*GloVe+ *temp*NN	31.7	**31.4**	**31.3**
	WER *temp*Cos	**31.6**	31.5	31.4
PNER 42.7	**PNER** *topic*GloVe+ *temp*NN	40.2	**37.6**	**36.3**
	PNER *temp*Cos	**39.8**	38.1	37.0

Table 7. Combination results in terms of WER (%) and PNER (%) according to combination methods (Sum score normalization, M3). **Test corpus.**

5. CONCLUSION

In this paper, we proposed different continuous word representations for OOV PN retrieval from diachronic documents. Temporal, lexical and semantic contexts of words are modeled at word-level and at document-level. Proposed combination strategies, using temporal and topic information, substantially improved the retrieved PN recall: between 9 and 14% relative. By adding the selected OOVC PNs in the lexicon of our automatic speech recognition system, a significant improvement is achieved in terms of PNER. It is worth noting than adding only 80 OOV PNs, the PNER decreases greatly (6% relative).

This confirms that combining complementarity methods improves performance of ASR system: (1) selecting documents based on temporal versus topic information; (2) retrieval OOV PNs using Neural Network based (Mikolov's NNs) versus co-occurrence based method (GloVe).

ACKNOWLEDGEMENTS

This work is funded by the *ContNomina* project supported by the French national Research Agency (ANR) under the contract ANR-12-BS02-0009.

6. REFERENCES

[1] Aslam, J.A., Montague, M. « Models for Metasearch », *Proceedings of the 24th Annual International ACM SIGIR Conference, 2003.*

[2] Bigot, B., Senay, G., Linares, G., Fredouille, C., and Dufour, R. "Person name recognition in ASR outputs using continuous context models", *Proceedings of ICASSP*, 2013.

[3] Blei, D., Ng, A.-Y, Jordan, M.I. "Latent Dirichlet Allocation*", Journal of Machine Learning Research*, 3; 9953-1022,2003.

[4] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I, Lafferty, John, "Latent Dirichlet allocation" *Journal of Machine Learning Research 3 (4–5): pp. 993–1022, 2003.*

[5] Corft, W. "Combining Approaches to Information Retrieval". *In Croft editor, Advances in Information Retrieval, Kluwer Academic Publishers, 2000.*

[6] Deng, L., Li, J., Huang, J.-T., Yao, K., Yu, D., Seide, F., Seltzer, M., Zweig, G., He, X., Williams, J., Gong, Y. and Acero A. "Recent Advances in Deep Learning for Speech Research at Microsoft", *Proceedings of ICASSP*, 2013

[7] Federico, M. and Bertoldi, N. "Broadcast news LM adaptation using contemporary texts", *Proceedings. of Interspeech*, pp. 239-242, 2001

[8] Fohr, D., Illina, I. "Continuous Word Representation using Neural Networks for Proper Name Retrieval from Diachronic Documents*", Proceedings of Interspeech 2015.*

[9] Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.-F., Gravier, G. "The ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News", *Proceedings of Interspeech,* 2005.

[10] Gao, J., Toutanova, K., Yih, W. « Clickthrough-Based Latent Semantic Models for Wen Search », *ACM SIGIR, 2011.*

[11] Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., Heck, L. "Learning Deep Structured Semantic Models for Web Search using Clickthough Data", *Proceedings of CIKM*, 2013.

[12] Illina I., Fohr D., Jouvet D. "Grapheme-to-Phoneme Conversion using Conditional Random Fields", *Proceedings of Interspeech,* 2011.

[13] Illina I., Fohr D., Mella O., Cerisara C. "The Automatic News Transcription System: ANTS, some Real Time experiments ", *Proceedings of. ICSLP*, 2004.

[14] Illina, I., Fohr, D. and Linares, G. "Proper Name Retrieval from Diachronic Documents for Automatic Transcription using Lexical and Temporal Context", *Proceedings of SLAM,* 2014.

[15] Kim, Y. "Convolutional Neural Network for Sentence Classification", *Proceedings of EMNLP*, 2014.

[16] Kobayashi A., Onoe K., Imai T., Ando A. "Time dependent language model for broadcast news transcription and its post-correction", *Proceedings of ICSPL*, 1998

[17] Larochelle, H., Lauly, S. "A Neural Autoregressive Topic Model*", Proceedings of NIPS,* 2012.

[18] Lee, A. and Kawahara, T. "Recent Development of Open-Source Speech Recognition Engine Julius", *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC),* 2009.

[19] Mamou, J. Cui, J., Cui, X., Gales, M., Kingsbury, B., Knil, K., Mangu, L., Nolden, D., Picheny, M., Ramabhadran, B., Schluter, R., Sethy, A., Woodland, P. "System Combination and Score Normalization for Spoken Term Detection", *Proceedings of ICASSP, 2013*

[20] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. "Distributed Representations of Words and Phrases and their Compositionality", *Proceedings of NIPS*, 2013.

[21] Mikolov, T., Yih, W. and Zweig, G. "Linguistic Regularities in Continuous Space Word Representations", *Proceedings of NAACL HLT,* 2013.

[22] Mikolov, T., Chen, K., Corrado, G. and Dean, J. "Efficient Estimation of Word Representations in Vector Space", *Proceedings of Workshop at ICLR*, 2013.

[23] Montague, M., Aslam, J. "Relevance Score Normalisation for Metasearch", *Proceedings of the ACM 10th Annual International Conference on Information and Knowledge Management (CIKM), 2001*.

[24] Pennington, J., Socher, R., Manning, C. "GloVe:Global Vectors for Word Representation", *Proceedings of EMNLP,* 2014

[25] Schmid, H. "Probabilistic part-of-speech tagging using decision trees", *Proceedings of ICNMLP*, 1994.

[26] Seven, H., Tolun, M. "Comparison of Normalization Techniques for Metasearch*", Lecture Notes in Computer Science, vol. 2457, 2002.*

[27] Sheikh, I., Illina, I., Fohr, D. and Linares, G. "OOV Proper Name Retrieval using Topic and Lexical Context Models", *Proceedings of ICASSP*, 2015.

[28] Srivastava, N., Salakhutdinov, R., Hinton, G."Modeling Documents with a Deep Boltzmann Machine", *Proceeding of Conference on Uncertainty in Artificial Intelligence CUAI*, 2013.

[29] Wu, S. and Crestani, F. "Data Fusion with Estimated Weights" *Proceedings of the 2002 ACM CIKM International Conference on Information and Knowledge Management, 2002.*

[30] Wu, S., Crestani, F., Bi, Y. « Evaluating Score Normalization Methods in Data Fusion *», Lecture Notes in Computer Science, vol. 4182, 2006.*