Finding the Best Approach for Multi-lingual Text Summarisation: A Comparative Analysis

Elena Lloret
University of Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
elloret@dlsi.ua.es

Manuel Palomar
University of Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
mpalomar@dlsi.ua.es

Abstract

This paper addresses the problem of multi-lingual text summarisation. The goal is to analyse three approaches for generating summaries in four languages (English, Spanish, German and French), in order to determine the best one to adopt when tackling this issue. The proposed approaches rely on: i) language-independent techniques; ii) language-specific resources; and iii) machine translation resources applied to a mono-lingual summariser. The evaluation carried out employing the JRC corpus – a corpus specifically created for multi-lingual summarisation – shows that the approach which uses language-specific resources is the most appropriate in our comparison framework, performing better than state-of-the-art multi-lingual summarisers. Moreover, the readability assessment conducted over the resulting summaries for this approach proves that they are also very competitive with respect to their quality.

1 Introduction

In the current society, information plays a crucial role that brings competitive advantages to users, when it is managed correctly. However, due to the vast amount of available information, users cannot cope with it, and therefore research into new methods and approaches based on Natural Language Processing (NLP) is crucial, thus resulting in considerable benefits for the society. Specifically, one of these NLP research areas is Text Summarisation (TS) which is essential to condense information keeping, at the same time, the most relevant facts or pieces of information. However, to produce a summary automatically is very challenging. Issues such as redundancy, temporal dimension, coreference or sentence ordering, to name a

few, have to be taken into consideration especially when summarising a set of documents (multi-document summarisation), thus making this field even more difficult (Goldstein et al., 2000). Such difficulty also increases when the information is stated in several languages and we want to be capable of producing a summary in those languages, thus not restricting the summariser to a single language (multi-lingual summarisation). The generation of multi-lingual summaries improves considerably the capabilities of TS systems, allowing users to be able to understand the essence of documents in other languages by only reading their corresponding summaries.

Therefore, the aim of this paper is to carry out a comparative analysis of several approaches for generating extractive[1] multi-lingual summaries in four languages (English, French, German and Spanish). These approaches comprise the use of: i) language-independent techniques; ii) language-specific resources; and iii) machine translation resources applied to a mono-lingual summariser. In this way, we can study the advantages and limitations of each approach, as well as to determine which is the most appropriate to adopt for this type of summaries. Although the language-specific resources are limited and perform differently for each language, the results indicate that this approach is the best to adopt, since for each language, more specific information could be obtained, benefiting the final summaries.

The remaining of the paper is organised as follows. Section 2 introduces previous work on multi-lingual TS. Section 3 describes the proposed approaches for generating multi-lingual summaries in detail. Further on, the corpus used, the experiments carried out, the results obtained together with an in-depth discussion is provided

---

[1]Extractive approaches are those ones which only detect important sentences in documents and extract them, without performing any kind of language generation or generalisation.

in Section 4. Finally, the conclusions of the paper together with the future work are outlined in Section 5.

2 Related Work

Generating multi-lingual TS is a challenging task, due to the fact that we have to deal with multiple languages, each of which has its peculiarities. Attempts to produce multi-lingual summaries started with SUMMARIST (Hovy and Lin, 1999), a system which extracted sentences from documents in a variety of languages, by using English, Japanese, Spanish, Indonesian, and Arabic preprocessing modules and lexicons. Another example of multi-lingual TS system is MEAD (Radev et al., 2004), able to produce summaries in English and Chinese, relying on features, such as sentence position, sentence length, or similarity with the first sentence.

More recently, research in multi-lingual TS has been focused on the analysis of language-independent methods. For instance, in (Litvak et al., 2010b) a comparative analysis of 16 methods for language-independent extractive summarisation was performed in order to find the most efficient language-independent sentence scoring method in terms of summarisation accuracy and computational complexity across two different languages (English and Hebrew). Such methods relied on vector-, structure- and graph-based features (e.g. frequency, position, length, title-based features, pagerank, etc.), concluding that vector and graph-based approaches were among the top ranked methods for bilingual applications. From this analysis, MUSE – MUltilingual Sentence Extractor (Litvak et al., 2010a) was developed, where other language-independent features were added and a genetic algorithm was employed to find the optimal weighted linear combination of all the sentence scoring methods proposed. In (Patel et al., 2007) a multi-lingual extractive language-independent TS approach was also suggested. The proposed algorithm was based on structural and statistical factors, such as location or identification of common and proper nouns. However, it also used stemming and stop word lists, which were dependent on the language. This TS approach was evaluated for English, Hindi, Gujarati and Urdu documents, obtaining encouraging results and showing that the proposed method performed equally well regardless of the language. News-

Gist (Kabadjov et al., 2010) is a multi-lingual summariser that achieves better performance than state-of-the-art approaches. It relies on Singular Value Decomposition, which is also a language-independent method, so it can be applied to a wide range of languages, although at the moment, it has been only tested for English, French and German.

Furthermore, Wikipedia[2] is a multi-lingual resource, which has been used for many natural language applications. It contains more than 18 million articles in more than 270 languages, which have been written collaboratively by volunteers around the world. This valuable resource has also been used for developing multi-lingual TS approaches. For instance, (Filatova, 2009) took advantage of Wikipedia information stated across different languages with the purpose of creating summaries. The approach was based on the Pyramid method (Nenkova et al., 2007) in order to account for relevant information. The underlying idea was that sentences were placed on different levels of the pyramid, depending on the number of languages containing such sentence. Thus, the top levels were populated by the sentences that appeared in the most languages and the bottom level contained sentences appearing in the least number of languages. The summary was then generated by taking a specific number of sentences starting with the top level, until the desired length was reached. Moreover, although the multi-lingual approach proposed in (Yuncong and Fung, 2010) aimed at generating complete articles instead of summaries, it is very interesting and it can be perfectly applied to TS. Basically, this approach took an existing entry of Wikipedia as content guideline. Then, keywords were extracted from it, and translated into the target language. The translation was used to query the Web in the target language, so candidate fragments of information were obtained. Further on, these fragments were ranked and synthesised into a complete article.

Different to the aforementioned approaches, in this paper we carried out a comparison between three approaches: i) a language-independent approach; ii) a language-specific approach; and iii) machine translation resources applied to a mono-lingual TS approach. Our final aim is to analyse them in order to find which is the most suitable for performing multi-lingual TS.

---

3 Multi-lingual Text Summarisation

The objective of this section is to explain the three proposed approaches for generating multi-lingual summaries in four languages (English, French, German and Spanish). We developed an extractive TS approach for each case. In particular, we analysed: i) language-independent techniques (Subsection 3.1); ii) language-specific resources (Subsection 3.2); and iii) machine translation resources applied to a mono-lingual summariser (Subsection 3.3). Next, we describe each approach in detail.

3.1 Language-independent Approach

As a language-independent approach for tackling multi-lingual TS, we computed the relevance of sentences by using the term frequency technique. Term frequency was first proposed in (Luhn, 1958), and, despite being a simple technique, it has been widely used in TS due to the good results it achieves (Gotti et al., 2007), (Orăsan, 2009), (Montiel et al., 2009).

The importance of a term in a document will be given by its frequency. At this point, it is worth mentioning that stop words, such as "the", "a", "you", etc. are not taken into account; otherwise the relevance of sentences could be wrongly calculated. In order to identify them, we need a specific list of stop words, depending on the language used. The language-specific processing in this approach is minimal, so it can be considered language-independent, since given a new language it would be very easy to obtain automatic summaries through this approach.

For determining the relevance of sentences, a matrix is built. In this matrix M, the rows represent the terms of the document without considering the stop words, whereas the columns represent the sentences. Each cell $M[i,j]$ contains the frequency of each term i in the document, provided that such term is included in the sentence; otherwise the cell contains a 0. Then, the importance of sentence S_j is computed by means of Formula 1:

$$Sc_{S_j} = \frac{\sum_{i=1}^{n} M[i,j]}{|Terms|} \quad (1)$$

where

Sc_{S_j} = Score of sentence j
$M[i,j]$ = value of the cell *[i,j]*
$|Terms|$ = total number of terms in the document.

Once the score for each sentence is calculated, sentences will be ranked in descending order, and the top ones up to a desired length will be chosen to become part of the summary.

Apart from its simplicity, the advantage of this techniques is that it can be used in any language. However, its main limitation is that the relevance of the sentences is only determined through lexical surface analysis, and therefore, semantics aspects are not taken into account.

3.2 Language-specific Approach

Our second proposed approach is very similar to the first one, but instead of term frequency, it employs language-specific resources for each of the target languages. For determining the relevance of sentences, this approach analyses the use of Named Entity Recognisers (NER) and the identification of concepts, by means of their synsets in WordNet (Fellbaum, 1998) or EuroWordNet (Ellman, 2003). On the one hand, named entities can indicate important content, since they refer to specific people, organisations, places, etc. that may be related to the topic of the document. On the other hand, the identification of concepts involves semantic analysis, and therefore, we can identify synonyms or other types of semantic relationships.

These types of resources (NERs and resources like Wordnet) have been commonly employed for generating specific types of summaries (Hassel, 2003), (Bellare et al., 2004), (Chaves, 2001). Moreover, in (Filatova and Hatzivassiloglou, 2004) it was proven that approaches that took into consideration named entities as well as frequent words were appropriate for TS. In light of this, we decided to develop a similar approach, but relying on named entities and concepts.

In particular, we focus on four languages (English, French, German and Spanish). The named entities are identified using different NERs, depending on the language. In this way, we use LingPipe[3] for English, the Illinois Named Entity Tagger[4] (Ratinov and Roth, 2009) for French, the NER for German[5] proposed in (Faruqui and Padó, 2010), and Freeling[6] for Spanish. For detecting concepts, we rely on WordNet for English and EuroWordNet for the remaining languages. Thanks

---

[3]http://alias-i.com/lingpipe/
[4]http://cogcomp.cs.illinois.edu/page/software_view/4
[5]http://www.nlpado.de/ sebastian/ner_german.html
[6]http://nlp.lsi.upc.edu/freeling/

to these types of resources, this approach uses semantic knowledge, instead of only lexical, as in the case of the term frequency in the language-independent approach.

For computing the relevance of the sentences, a matrix (M) is also built, where the rows represent the entities or concepts of the document and the columns, the sentences. Each cell $M[i, j]$ contains the frequency of appearance of either each entity or concept. As in the previous approach, stop words are not taken into consideration, and in those cases where neither the entity nor the concept is included in the sentence, a 0 is assigned to the cell. Once the matrix has been filled in, Formula 2 is then used to compute the relevance of sentences:

$$Sc_{S_j} = \frac{\sum_{i=1}^{n} M[i, j]}{|NE + Concepts|} \qquad (2)$$

where

Sc_{S_j} = Score of sentence j
$M[i, j]$ = value of the cell *[i,j]*
$|NE + Concepts|$ = total number of named entities and concepts in the document.

The highest scored sentences, up to a specific length, will be extracted to build the final summary.

The advantages of this approach with respect to the previous one (i.e. the language-independent) is that semantic analysis is applied by using resources such as WordNet or EuroWordNet. This allows us to group synonyms under the same concept. For instance, the words *harassment* and *molestation* represent the same concepts (since they both belong to the same synset in WordNet), so they are grouped together in this approach, whereas in the previous one, where only the frequency of terms is taken into consideration, they are considered two distinct words. In contrast, the drawback of this approach is that such kind of resources may not be available for all languages, and therefore we might have problems in applying this approach. Moreover, the error these resources introduce (e.g. NERs) may negatively affect the performance of the summariser.

3.3 Machine Translation Resources applied to a Mono-lingual Approach

The idea behind this approach is to use an existing mono-lingual summariser for a specific language and then employ a machine translation system for obtaining the summaries in the different languages. In particular, we employ the TS approach proposed in (Lloret and Palomar, 2009) that generates extractive summaries for English. The reason for employing such summariser is its competitive results achieved compared to the state of the art. Briefly, the main features of this approach are: i) redundant information is detected and removed by means of textual entailment; and ii) the Code Quantity Principle (Givón, 1990) is used for accounting relevant information from a cognitive perspective. Therefore, important sentences are identified by computing the number of words included in noun-phrases, taking also into consideration the relative frequency each word has in the document. Once the summaries have been generated, Google Translate[7] is used to translate the summaries into the different target languages (i.e., French, German and Spanish), since it is a free online language translation service that can translate text in more than 50 languages.

The advantage of this approach is that we do not have to develop a particular approach for each language, because we can rely on existing mono-lingual summarisers. Although machine translation has been made great progress in the recent years, and they can translate text into a wide range of languages, the disadvantage associated to using such tools concerns their performance, since wrong translations can negatively affect the quality of the resulting summary.

4 Experimental Framework

The goal of this section is to setup an experimental framework, thus allowing us to analyse the aforementioned approaches in a specific context. Therefore, the corpus employed and the languages used are described in Subsection 4.1. Then, the evaluation methodology proposed and the results obtained together with a discussion is provided in Subsection 4.2.

4.1 Corpus

We used the JRC multi-lingual summary evaluation data[8] for carrying out the experiments, in order to determine which approach should be more appropriate for the task of multi-lingual summarisation. The corpus consists of 20 docu-

---

[7]http://translate.google.com/
[8]http://langtech.jrc.ec.europa.eu/JRC_Resources.html

	English	French	German	Spanish
No. of words	16,398	18,329	16,837	18,547
Avg. words/document	819.9	916.45	841.45	928.7
Max. words/document	973	1,157	1,025	1,144
Min. words/document	617	698	645	708
No. of NE	511	254	345	326
Avg. NE/document	25.6	12.7	17.25	16.3
Max. NE/document	44	22	37	32
Min. NE/document	3	6	1	1
No. of concepts	3,405	2,376	2,115	3,580
Avg. concepts/document	170.25	118.8	105.75	179
Max. concepts/document	1,353	159	136	231
Min. concepts/document	222	90	78	138

Table 1: Statistical properties of the JRC corpus.

ments grouped into four topics (genetics, Israel-and-Palestine-conflict, malaria and science-and-society). Each document is available in seven languages (Arabic, Czech, English, French, German, Russian and Spanish), and the corpus also contains the manual annotation of important sentences, so it is possible to have four model summaries for each of the documents. Four our purposes, four languages were selected (English, French, German and Spanish), thus dealing with 80 documents.

The type of documents contained in the JRC corpus pertained to the news domain. Table 1 shows some properties of the corpus.

As it can be seen from the table, all the documents have a similar length, the shortest ones having more than 600 words, whereas the longest ones around 1,000 words. Regarding the statistics about the words, it is worth noting that the documents in Romance languages (Spanish and French) have similar characteristics. Analogously, the same happens for the Germanic languages (English and German). However, the highest differences between languages can be found in the number of NE and concepts detected. Whereas for English, the average number of NE is 25, for the remaining languages is at most 17. This depends on the NER employed. The language-specific resources used for detecting concepts (WordNet and EuroWordNet) also influence the number of concepts identified. In this way, Spanish and English are the languages with more concepts.

4.2 Results and Discussion

The JRC corpus was used to generate extractive summaries in four languages (English, French, German, and Spanish), following our three proposed approaches. We generated 20 summaries for each approach and language, thus evaluating 240 different summaries in the end. Two types of evaluation were conducted. On the one hand, the content of the summaries was evaluated in an automatic manner (Subsubsection 4.2.1), whereas on the other hand, their readability was manually assessed (Subsubsection 4.2.2). In addition, a comparison with current multi-lingual TS systems was also carried out (Subsubsection 4.2.3).

4.2.1 Content Evaluation

The automatic summaries were compared to the model ones, using ROUGE (Lin, 2004), a widespread tool for evaluating TS. In this way, the content of the summaries was assessed, since this tool allows to compute recall, precision and F-measure with respect to different metrics, all of them based on how much vocabulary overlap there is between an automatic and model summary. Table 2 shows the F-measure value for ROUGE-1 (R-1), ROUGE-2 (R-2), and ROUGE-SU4 (R-SU4) for each of the proposed multi-lingual TS approaches. R-1 computes the number of common unigram between the automatic and model summary; R-2 computes the number of bi-grams, whereas R-SU4 accounts for the number of bi-grams with a maximum distance of four words in-between.

Moreover, a t-test was performed in order to account for the significance of the results at a 95% level of confidence. Results statistically significant are marked with a star. As it can be seen from the table, the results for the language-independent (LI) and language-specific (LS) approaches are statistically significant compared to the mono-lingual approach combined with machine translation (TS+MT) in all the cases, except for English. Furthermore, from the results obtained, it is worth noting that the LS approach

Language	Approach	R-1	R-2	R-SU4
English	LI	0.53097	0.31777	0.34873
	LS	**0.56530**	**0.37568**	**0.39828**
	TS	0.52823	0.33011	0.35832
French	LI	**0.55758***	0.33777*	**0.36116***
	LS	0.55638*	**0.35119***	**0.37316***
	TS+MT	0.50054	0.20505	0.24204
German	LI	0.47886*	0.29219*	0.30646*
	LS	**0.52614***	**0.36849***	**0.38002***
	TS+MT	0.41716	0.15985	0.18180
Spanish	LI	0.57920*	0.36234*	0.39296*
	LS	**0.62351***	**0.42975***	**0.45653***
	TS+MT	0.52886	0.24362	0.28623

Table 2: F-measure results for the content evaluation using ROUGE (LI=language-independent; LS=language-specific; TS= mono-lingual; TS+MT=mono-lingual and machine translation).

obtains better results than the LI approach, in all ROUGE metrics, except R-1 for French, where LI and LS obtain very similar results. In addition, the differences between them are statistically significant for German and Spanish. As it can also be seen, the LS obtains the best results for English and Spanish. This may happens because these languages have a lot of specific resources for dealing with them. In contrast, the performance for French and German linguistic resources may not be as accurate as for the other languages, thus affecting the results. Moreover, it is also worth noting that the performance of the LI approach for German is quite low with respect to the other languages. This is due to the fact that the way of writing in German differs from the others in that it is more agglutinative (e.g. arbeitstag[9]); consequently, the frequency for some of the words in the documents will be computed separately (in the previous example *tag* and *arbeitstag* will have different frequencies). This occurs because in the LI approach we do not rely on any specific resources, such as tokenisers or stemmers; we only use the corresponding stop word list for each language.

4.2.2 Readability Evaluation

From Table 2 we can conclude that the LS approach is the most appropriate to tackle multi-lingual TS. However, we are interested in carrying out a readability assessment, so that the summaries generated by our best approach (LS) can be also assessed with respect to their quality. For conducting this type of assessment, we followed

---

[9]day at work

the DUC guidelines[10], and we asked four people (two natives of Spanish and German and two with very advanced knowledge of English and French) to manually evaluate each summary, assigning values from 1 to 5 (1=very poor... 5=very good) with respect to five quality criteria: grammaticality, redundancy, clarity, focus and coherence. Results are shown in Table 3.

	English	French	German	Spanish
Grammaticality	3.4	4.3	4.6	3.1
Redundancy	3.8	5.0	4.3	4.8
Clarity	3.6	3.9	4.6	3.8
Focus	4.4	3.9	4.6	4.6
Coherence	4.0	3.5	4.0	3.5

Table 3: Readability Assessment of the language-specific (LS) multi-lingual TS approach.

In general terms, the results obtained in the readability assessment are very good. This means that using the language-specific approach, the resulting summaries are also good with respect to their quality. Concerning this issue, German summaries obtains the best results, all of them above 4 out of 5. The summaries in the remaining languages perform also very good in the coherence and redundancy criteria. It is worth noting that we generated single-document summaries (i.e., the summaries were produced taking only a document as input), so the chances of redundant information decrease. However, in this criteria we also measured the repetition of named entities, so in this sense, despite relying on named entities and concepts, there was not much repeated information in the summaries.

4.2.3 Comparison with Current Multi-lingual Summarisers

With the purpose of widening the analysis and verifying our results, we compared our LS approach to several current multi-lingual TS systems, that also produce extractive summaries as a result. In particular, we selected:

- **Open Text Summarizer**[11] (OTS). This is a multi-lingual summariser able to generate summaries in more than 25 languages, such as English, German, Spanish, Russian or Hebrew. In this approach, keywords are identified by means of word occurrence, and sen-

---

[10]http://duc.nist.gov/duc2007/quality-questions.txt
[11]http://libots.sourceforge.net/

tences are given a score based on the the keywords they contain. Some language-specific resources, such as stemmers and stop word lists are employed. It has been shown that this system obtains better performance than other multi-lingual TS systems (Yatsko and Vishnyakov, 2007).

- **MS Word 2007 Summarizer**[12] (MS Word). This summariser is integrated into Microsoft Word 2007 and it also generates summaries in several languages. Since it is a commercial system, the implementation details are not revealed.

- **Essential Summarizer**[13] (Essential). This TS system is a commercial version of the one presented in (Lehmam, 2010). It relies on linguistic techniques to perform semantic analysis of written text, taking into account discursive elements of the text. It is able to produce summaries in twenty languages.

For conducting such comparison, summaries were generated using the aforementioned TS systems in the four languages we dealt with. Then, they were evaluated using ROUGE. Table 4 shows the F-measure results for the ROUGE-1 metric. As before, we performed a t-test in order to analyse the significance of the results for a 95% confidence level (significant results are marked with a star). In most of the cases, our LS approach performs better than the other multi-lingual TS systems, except the OTS which performs slightly better for French and German. Our approach (LS) and OTS performed statistically better than the Essential summariser for German, increasing the results by 20% compared to it. Moreover, for Spanish, LS improves the results of MS Word and Essential summarisers by 9% and 16%, respectively, and this improvement is also statistically significant.

	English	French	German	Spanish
LS	**0.56530**	0.55638	0.52614*	**0.62351***
OTS	0.55732	**0.57745**	**0.53451***	0.60591*
MS Word	0.53591	0.54046	0.48427	0.57396
Essential	0.52622	0.51819	0.43727	0.53978

Table 4: Comparison with current multi-lingual TS systems (F-measure results for ROUGE-1).

---

[12]http://www.microsoft.com/education/autosummarize.aspx
[13]https://essential-mining.com/es/index.jsp?ui.lang=en

5 Conclusion and Future Work

This paper presented a comparative analysis of three widespread multi-lingual summarisation approaches in order to determine which one would be more suitable to adopt when tackling this task. In particular, we studied: i) a language-independent approach using the term frequency technique; ii) a language-specific approach, relying on specific linguistic resources for each of the target language (named entities recognisers and semantic resources); and finally, iii) a mono-lingual text summariser for English, whose output was then inputted to a machine translation system in order to generate summaries in the remaining languages. The experiments carried out in English, French, German and Spanish showed that by employing language-specific resources, the resulting summaries performed better than most of the state-of-the-art multi-lingual summarisers.

In the future, we plan to extend our analysis to other languages as well as to investigate other ways of generating multi-lingual summaries, for instance, employing Wikipedia, as in (Filatova, 2009). This would be the starting point to address cross-lingual summarisation, task that we would like to tackle in the long-term.

Acknowledgments

References

Kedar Bellare, Anish Das Sarma, Atish Das Sarma, Navneet Loiwal, Vaibhav Mehta, Ganesh Ramakrishnan, and Pushpak Bhattacharyya. 2004. Generic text summarization using wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*.

Rui Pedro Chaves. 2001. Wordnet and automated text summarization. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium*, pages 109–116.

Jeremy Ellman. 2003. Eurowordnet: A multilingual

database with lexical semantic networks. *Natural Language Engineering*, 9:427–430.

Manaal Faruqui and Sebastian Padó. 2010. Training and evaluating a german named entity recognizer with semantic generalization. In *Proceedings of KONVENS 2010*, Saarbrücken, Germany.

Christiane Fellbaum. 1998. *WordNet: An Electronical Lexical Database*. The MIT Press, Cambridge, MA.

Elena Filatova and Vasileios Hatzivassiloglou. 2004. Event-Based Extractive Summarization. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 104–111.

Elena Filatova. 2009. Multilingual wikipedia, summarization, and information trustworthiness. In *Proceedings of the IGIR Workshop on Information Access in a Multilingual World*.

Talmy Givón, 1990. *Syntax: A functional-typological introduction, II*. John Benjamins.

Jade Goldstein, Vibhu Mittal, Jaime Carbonell, and Mark Kantrowitz. 2000. Multi-document Summarization by Sentence Extraction. In *NAACL-ANLP Workshop on Automatic Summarization*, pages 40–48.

Fabrizio Gotti, Guy Lapalme, Luka Nerima, and Eric Wehrli. 2007. Gofaisum: A symbolic summarizer for duc. In *Proceedings of the Document Understanding Workshop*.

Martin Hassel. 2003. Exploitation of named entities in automatic text summarization for swedish. In *Proceedings of the 14th Mnordic Conference on Computational Linguistics*.

Eduard Hovy and Chin-Yew Lin. 1999. Automated text summarization in summarist. In *Inderjeet Mani and Mark Maybury, editors,* Advances in Automatic Text Summarization, pages 81–94. MIT Press.

Mijail Kabadjov, Martin Atkinson, Josef Steinberger, Ralf Steinberger, and Erik Van Der Goot. 2010. NewsGist: a multilingual statistical news summarizer. In *Proceedings of the European conference on Machine learning and knowledge discovery in databases: Part III*, pages 591–594.

Abderrafih Lehmam. 2010. Essential summarizer: innovative automatic text summarization software in twenty languages. In *Adaptivity, Personalization and Fusion of Heterogeneous Information*, pages 216–217.

Chin-Yew Lin. 2004. ROUGE: a Package for Automatic Evaluation of Summaries. In *Proceedings of Association of Computational Linguistics Text Summarization Workshop*, pages 74–81.

Marina Litvak, Mark Last, and Menahem Friedman. 2010a. A new approach to improving multilingual summarization using a genetic algorithm. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 927–936.

Marina Litvak, Mark Last, Slava Kisilevich, Daniel Keim, Hagay Lipman, and Assaf Ben Gur. 2010b. Towards multi-lingual summarization: A comparative analysis of sentence extraction methods on english and hebrew corpora. In *Proceedings of the 4th Workshop on Cross Lingual Information Access*, pages 61–69.

Elena Lloret and Manuel Palomar. 2009. A gradual combination of features for building automatic summarisation systems. In *Proceedings of the 12th International Conference on Text, Speech and Dialogue*, pages 16–23.

Hans Peter Luhn. 1958. The automatic creation of literature abstracts. In *Inderjeet Mani and Mark Maybury, editors,* Advances in Automatic Text Summarization, pages 15–22. MIT Press.

Romyna Montiel, René García, Yulia Ledeneva, and Rafael Cruz Reyes. 2009. Comparación de tres modelos de texto para la generación automática de resúmenes. *Sociedad Española para el Procesamiento del Lenguaje Natural*, 43:303–311.

Ani Nenkova, Rebecca Passonneau, and Kathleen McKeown. 2007. The pyramid method: Incorporating human content selection variation in summarization evaluation. *ACM Transactions on Speech and Language Processing*, 4(2):4.

Constantin Orăsan. 2009. Comparative Evaluation of Term-Weighting Methods for Automatic Summarization. *Journal of Quantitative Linguistics*, 16(1):67–95.

Alkesh Patel, Tanveer Siddiqui, and U. S. Tiwary. 2007. A language independent approach to multilingual text summarization. In *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, RIAO '07, pages 123–132.

Dragomir Radev, Tim Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Elliott Drabek, Wai Lam, Danyu Liu, Jahna Otterbacher, Hong Qi, Horacio Saggion, Simone Teufel, Michael Topper, Adam Winkel, and Zhu Zhang. 2004. MEAD - A Platform for Multidocument Multilingual Text Summarization. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*.

Lev Ratinov and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In *Proceedings of the 13th Conference on Computational Natural Language Learning*, pages 147–155.

Viatcheslav Yatsko and Timur Vishnyakov. 2007. A method for evaluating modern systems of automatic text summarization. *Automatic Documentation and Mathematical Linguistics*, 41:93–103.

Chen Yuncong and Pascale Fung. 2010. Unsupervised synthesis of multilingual wikipedia articles. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 197–205.