

A Survey to Text Summarization Methods for Turkish

Çağdaş Can Birant

Dokuz Eylül University, Computer
Engineering Department, Turkey

Özgün Koşaner

Dokuz Eylül University, Linguistics
Department, Turkey

Özlem Aktaş

Dokuz Eylül University, Computer
Engineering Department, Turkey

ABSTRACT

Nowadays, people deal with a huge amount of data, especially while they are surfing on internet. So, this makes the topic of automatic summarization is very important and in the forefront. In this paper, a review for text summarization methods in Turkish is presented. Brief summary of the methods used for automatic text summarization in the literature, and also brief definitions of summary, abstraction and automatic text summarization are given.

General Terms

Natural Language Processing, Text Summarization.

Keywords

Text Summarization, Natural Language Processing, Turkish.

1. INTRODUCTION

With the commercialization and the public access of internet in the 1990's, there had been a surge in the number and size of the documents created each second. However, as Torres-Moreno [1] indicates, the number of the documents which have been annotated with any mark-up language still remains small when compared to unstructured text documents. Text documents are often analyzed in a perfunctory and very superficial way [1] and the means to categorize and organize these texts are far from being perfect. Also, each text genre requires different approaches to be processed. This renders the implementation of text analysis, text mining and information extraction a difficult task [1]. However, there are many studies try to make these difficult tasks easier.

1.1 Definitions of Summary, Abstract and Automatic Text Summarization

Humans have a limited capacity of processing data and thus are incapable of handling the vast amount of data created each day. In order to get the main theme of any text, one has to read the whole text. This may seem an easy task for short text, however it would be impossible for a human to read and process very long texts and also all texts in the world, even impossible to read texts in one particular genre. Also documents in different language increase this difficulty. And if someone is reading to perform a task, not just for his or her own leisure activity, reading becomes a time and resource consuming task. Therefore, one needs something that would help him or her to process large amounts of text in a more efficient way. At this point, summaries or abstracts step in. As the National Information Standards Organization (NISO) states, “[a] well-prepared abstract enables readers to identify the basic content of a document quickly, to determine its relevance to their interests, and thus to decide whether they need to read the document in its entirety. The abstract may provide an introductory overview of its topic or argument for readers to whom the document is of marginal interest, and make a reading of the full document unnecessary” [2]. Also summaries as short as 17% of full text length sped up

decision-making by almost a factor of 2 with no statistically significant degradation in accuracy [3].

Summaries may be written by the author of the documents, professional summarizers or a third party[1]. Some of the documents, especially scientific articles often come along with summaries written by the author, or some professional summarizes these documents for the scientific journals, etc. Sometimes companies hire staff to summarize large amount of documents to get some job done. However, although some of the texts, especially scientific texts, come along with summaries, this is not possible for all texts. Also summaries provided by professional summarizers or third parties are not sometimes welcome since the cost of production of a summary by a professional is very high and the reliability is controversial, because the preparation of abstracts requires an intellectual effort and a general familiarity with the subject [4]. In addition to this, the summaries are almost always influenced by the summarizer's background, attitude or disposition [4].

These facts put forth several valid reasons in favor of the – automatic – summarization of [1]:

- i) Summaries reduce reading time.
- ii) When researching documents, summaries make the selection process easier.
- iii) Automatic summarization improves the effectiveness of indexing.
- iv) Automatic summarization algorithms are less biased than human summarizers.
- v) Personalized summaries are useful in question-answering systems as they provide personalized information.
- vi) Using automatic or semi-automatic summarization systems increase the number of texts processed.

There is a slight distinction between an abstract and a summary; a summary often being an integral part of a text frequently placed at the end of the document. However, as it has been indicated before abstract and summary are umbrella terms and they may be subcategorized into different kinds of abstracts and summaries. Also, although an abstract appears at the beginning of a document, it is and should be written last since it gives a brief representation of the whole text. Therefore, any distinction between an abstract and a summary cannot be seen in terms of their position in the text, because they are both written at the end of the writing process to give a glimpse of the document. Also, this thesis is about automatic text summarization; although the product of this process may be called as abstract or summary, the process itself is called summarization, so the distinction in definitions would not pose a problem for the automatic text summarization.

The common points of these definitions are that an automatic text summary should be condensed and shorter version of the original text document or documents, the process being automatic; contain the most important information; be coherent and be sensitive to user's needs. Following these common points, automatic text summarization can be defined as below:

DEFINITION 1. *automatic text summarization is an automated process that produces a coherent and condensed version of a document or a group of documents, retaining the information considered as the most important, according to the needs of the user.*

As DEFINITION 1 suggests automatic text summarization helps people to process the ever-increasing BigData efficiently, which people, as humans, are incapable of handling [1].

2. AUTOMATIC TEXT SUMMARIZATION METHODS

Torres-Moreno asserts that there are three families of approaches for automatic text summarization [1]:

- i. summarization by extraction;
- ii. summarization by abstraction;
- iii. summarization by sentence compression.

According to Radev, Hovy and McKeown, extraction is the process of identifying important segments of a text; abstraction is the process of reformulating and fusing these important segments in novel terms; and compression is the process of trimming unimportant material [5].

Differently from its beginning, automatic text summarization is currently an interdisciplinary field of research benefiting from the expertise of numerous fields other than Natural Language Processing (NLP) such as computer science, artificial intelligence, statistics, cognitive sciences, Natural Language Generation, Machine Learning, linguistics, discourse analysis. In this section, it is focused only on automatic text summarization studies, sometimes emphasizing the contribution of other fields.

The concept of automatic text summarization dates back to Luhn [4], where Luhn describes some exploratory research on automatic methods of obtaining abstracts. At the same time Baxendale [6] will add information from the original resource. In 1963, Vasiliev presented a report to the UNESCO and gave the state of automatic abstracting at that time. This study makes mention of a *statistical approach*, a *descriptor approach* and a *semantico-logical approach* to automatic text summarization [7].

In 1969, Edmundson develops a new approach to automatic summarization [8]. Edmundson, on the contrary to the previous studies, focuses not only on the presence of high-frequency content such as keywords, but also the pragmatic words (cue words); title and heading words; and structural indicators (the location of the words in a sentence). Following this study, Rush, Salvador & Zamora [9] devised an algorithm using contextual inference, the location method, the cue method, the frequency data and coherence considerations to select or eliminate sentences from the document.

Later, Pollock & Zamora [10], customized the Rush-Salvador-Zamora algorithm for chemistry, more especially for pharmacodynamics. The study used a specialized database comprising of pharmacodynamics texts, where the previous

studies used more general databases comprising of novels, textbooks, and magazine or newspaper articles. The study claimed that automatic summarization was more successful with some particular text genres than others.

In 1978, Yale Artificial Intelligence Project announced its new software FRUMP (Fast Reading Understanding and Memory Program) which was devised for skimming and summarizing newspaper articles [11]. The FRUMP system used a structure called *sketchy script*, some kind of pragmatic and semantic knowledge frame in addition to its linguistic knowledge. The pragmatic/semantic knowledge comprised a high-level reasoning system and the low-level (linguistic) text analyzer was sensitive to this high-level reasoning [11].

The 1980's are relatively silent for automatic text summarization studies, but 1990's and 2000's witnessed an explosion in research on automatic text summarization due to the commercialization of the Internet. Spärck-Jones [12] defined summarization as an information management process and devised a process structure for a summarizing task, based on human summarizing, automatic summarizing and discourse interpretation and representation. Spärck-Jones' approach makes use of discourse analysis approaches such as Rhetorical Structure Theory (Mann and Thompson, 1987) and has references to other linguistics resources such as Halliday & Hasan [13], Rumelhart [14] and Kintsch and van Dijk [15]. The system uses linguistic sources, domain sources similar to the *sketchy scripts* used in FRUMP and communicative sources to perform summarization.

In 1995, Kupiec, Pedersen & Chen focused on document extracts as a particular kind of computed document summary and developed a trainable summarization program using a statistical framework. They used a feature set comprising of the *sentence length cut-off feature*, *fixed-phrase feature*, *paragraph feature*, *thematic word feature* and *uppercase word feature* to build the summarizer [16].

The same year, McKeown and Radev [17] presented a multi-document summarizer system (SUMMONS) for articles on the same event. SUMMONS was comprised of a content planner, which selects information from an underlying knowledge base, and a linguistic generator which chooses the words to refer to the concepts contained in the selected information and which combines these words into grammatical sentences.

Marcu [18], uses Rhetorical Structure Theory [19] and refers, heavily, to discourse markers and lexico-grammatical constructs. The study develops a rhetorical parsing algorithm to hypothesize rhetorical relations between textual units and to map natural language texts onto discourse trees automatically.

Barzilay & Elhadad [20] devised a technique to summarize texts without requiring its full semantic interpretation using topic progression in the text derived from lexical chains. The study merges several robust knowledge bases such as the WordNet, a part-of-speech tagger, a shallow parser for nominal groups and a segmentation algorithm.

Carbonell and Goldstein [21] combines query relevance with information novelty. The Maximal Marginal Relevance (MMR), which emphasizes *relevant novelty*, aims at reducing the redundancy while maintaining query relevance, especially for multi-document summarization.

Witbrock & Mittal [22] proposes a summarization method that uses statistical models of the term selection and term

ordering process to produce even briefer and compacter coherent summaries in a style learned from a training corpus.

Hovy & Lin [23] proposes a summarization system (SUMMARIST) which summarizes a text in three phases:

- i) topic identification which includes position module, cue phrase module, topic signature module, discourse structure module, topic identification integration module;
- ii) topic interpretation which includes concept counting and the wavefront, interpretation using topic signatures;
- iii) summary generation which includes a microplanner and a sentence generator.

Knight & Marcu [24] sets of from the fact that previous studies had focused on only extraction of information; however, simply combining textual segments would not yield coherent outputs. This study uses a decision-based model to reduce sentence size and compress sentences.

Radev, Jing & Budzikowska [25], present a multi-document summarizer (MEAD) which generates summaries using cluster centroids, which consist of words which are central not only to one article in a cluster but to all the articles, produced by a topic detection and tracking system.

Saggion and Lapalme [26] present a text summarization system that takes a raw technical text as input and produces an indicative informative summary. SumUM first the topics of the document, and then elaborates on some of these topics according to the reader's interest.

During 2000's a series of Document Understanding Conference (DUC) yielded very fruitful research on automatic text summarization. Marcu [27] proposes a system of discourse-based summarization algorithms (DUC-2001) for both single documents and collections of documents. The single document summarization employs the following steps:

- i) derive the discourse structure of the text given as input
- ii) determine the important sentences in the input document
- iii) determine all co-reference links in the input document
- iv) increase summary coherence and compactness
- v) generate summary

DUC-2001 uses the following stems for summarizing document collections:

- i) pre-process the collection
- ii) select and order the sentences that summarize the collection
- iii) resolve third person pronouns
- iv) rank headlines
- v) generate summaries

In 2004 Erkan & Radev introduce LexRank which uses a stochastic graph-based method for computing the relative importance of the textual units [28]. Erkan & Radev presented a new approach to define sentence salience based on graph-based centrality scoring of sentences. The authors claim that

constructing the similarity graph of sentences provides a better view of important sentences compared to the *centroid* approach [28].

Barzilay & McKeown [29] discuss the "sentence fusion" method in 2005 within a multi-document summarization system (MultiGen) and argue that sentence fusion involves bottom-up local multisequence alignment to identify phrases conveying similar information and statistical generation to combine common phrases into a sentence; thus producing abstract that contain sentences not found in any of the original documents.

Fernández, SanJuan, & Torres-Moreno present a Neural Network approach in 2007, inspired by statistical physics of magnetic systems, to model documents as neural network whose Textual Energy is studied [30]. The model yielded good results in automatic summarization and Topic Segmentation.

In the same year, Svore, Vanderwende & Burges (2007), presents a novel approach to automatic single-document summarization based on neural networks, called NetSum. The study is the first to use both neural networks for summarization and third-party datasets for features, using Wikipedia and news query logs [31].

Saggion (2008) presents a set of adaptable summarization components together with well-established evaluation tools [32]. The toolkit (SUMMA) includes resources for the computation of summarization features which are combined in order to provide functionalities for single-document, multi-document, querybased, and multi/cross-lingual summarization.

Filippova (2010) devises a method called *multi-sentence compression* a syntax-lean method which requires little more than a tokenizer and a tagger [33]. The method is a graph-based method which produced compressed and grammatical sentences which does require neither a parser nor handcrafted linguistic rules.

Lenkova & McKeown (2011) provide a comprehensive overview the 50 years of research in summarization [34]. They also discuss the challenges which are still open in the field of summarization such as language generation and deeper semantic understanding of language. They start by categorizing summaries and explaining how summarization systems work. Later they elaborate on the steps and methods used in summarization process.

Torres-Moreno (2012) presents another algorithm (ARTEX) for Automatic Text Summarization which calculates a simple inner product, comprising of a document vector (text topic) and a lexical vector (vocabulary in the sentence), between each sentence [35]. The algorithm generates summaries by assembling the highest ranked sentences, retains the salient information of each sentence of the document and it does not require any linguistic knowledge.

Litvak & Vanetik (2014) present a new model for extractive summarization and try to obtain a summary that preserves the information coverage as much as possible [36]. They use a new tensor-based representation that describes the given document set in terms of its topics. Later these topics are ranked via Tensor Decomposition, and a summary from the sentences of the highest ranked topics.

In 2014, Torres-Moreno give a more comprehensive account of automatic text summarization beginning with the

foundations of the topic and discussing the most important concepts, methods, systems and evaluation systems [1].

2.1 Automatic Text Summarization Studies in Turkish

Although automatic text summarization studies, and natural language processing studies in general, can be traced back to Köksal in the early 1980's, this section will only make mention of the ones which are related to the subject of this study, leaving aside many NLP studies, which are important in their own rights [37].

It can be claimed that the first systematic study of NLP and automatic text summarization started with Oflazer & Kuruöz, in which they developed a POS tagger for Turkish based on a full-scale-two-level specification of Turkish morphology [38]. The tagging tool integrated morphological analysis, multi-word and idiomatic construct recognition, morphological disambiguation, and root and lexical form statistic compilation; where the second and the third functionalities were implemented by a rule based subsystem. The tagger uses a multi-word construct processor to detect and tag fixed expression such as duplications and other forms of semantic coalescences such as proper nouns which may generate spurious or incorrect results. The authors claimed that the tagger runs with 98-99% accuracy with minimal user intervention.

Later in 2003, Tür, Hakkani-Tür & Oflazer presented the results of their study on information extraction from unrestricted Turkish text using statistical language processing methods [39]. The system developed in the study used statistical models. The authors of the study built a model which used surface forms of the words and their underlying forms, i.e. their morphological structure. This was done by using a preprocessing module which tokenized the data, analyzed these tokens and gave the most probable morphological analyses. The system used a topic segmentation module to define the topic clusters in the data using the word-based model and the stem-based model in combination. Later the system used a module to extract named entities, i.e. names of persons, locations, organizations, monetary values, percentage, dates and times. The authors claimed that their system reached an F-measure of 91.56%.

One of the most important studies for automatic text summarization in Turkish is Bilgin, Çetinoğlu & Oflazer's initiative for developing a WordNet for Turkish in 2004 [40]. The authors started with a "first set of concepts" comprising of 1310 base concepts of the EuroWordNet project. After translating the first set of 1310 base concepts, the authors attempted to automatically extract synonyms, antonyms and hyponyms for these base concepts. Then they expanded the WordNet to 5000 base concepts during a second phase and then to 8000 base concepts in a third phase. The system run at 68% accuracy.

In 2004 Karakaya and Güvenir conducted a study to integrate text classification and text summarization to compile and extract information from large bodies of texts [41]. The system, named ARG, was a two-phase algorithm in which the paragraphs were classified according to given topics and then each topic was summarized into an automatically generated report. ARG, differently from other systems in the literature, used paragraphs as the unit of analysis, instead of the whole text. Also they used a user supervision for text classification to enable the users with a better option for expressing their

information needs. As Karakaya and Güvenir explained, in ARG;

- i. the user determines the subject topics
- ii. the user splits one or more article into their paragraphs and distributes them into topics
- iii. Each topic is indexed using paragraphs in step ii
- iv. Other documents are split into paragraphs
- v. Paragraphs in step iv are classified according to given subject topics in step i
- vi. Each topic is summarized
- vii. Summaries are compiled and outputted as a report

However this algorithm needs too much intervention from the user and classification and summarization of large bodies of text could be a burden for the user.

Another text classification study was conducted by Amasyalı & Diri in 2006, to determine the authors of documents and the gender of the author, and to classify the genre of documents [42]. They used an n-gram model and Naïve Bayes, Support Vector Machine, C4.5 and Random Forest methods to classify documents.

In another study, Ercan presented an extractive summarization system focusing on important sentence and key phrase identification [43]. The system used lexical cohesion and cohesion, and identified the topics and the segments of the texts via lexical chains. The study focused on the specificity of the words in the WordNet and using this specificity score identified the key phrases. The study also used co-occurrence analysis to capture the links between actors, places and other objects, i.e. the participants of a state-of-affairs, which could not be captured via lexical chains approach. The author claimed that the system obtained promising results in overall.

Ercan & Çiçekli described a keyword extraction method using lexical chain features which improved the precision significantly in 2007 [44]. The authors tried to devise a way to form lexical chains out of phrases which were not represented in the WordNet already.

Kutlu, Çığır & Çiçekli proposed a generic text summarization method via sentence ranking in 2010 [45]. The system calculated the sentence scores with regard to their surface level features and created the summaries by extracting the highest ranked sentences from the original documents. The system used features such as term frequency, key phrase centrality, title similarity and sentence position. The study was a first in many aspects that it showed the effectiveness of the centrality feature and introduced the use of key phrases in text summarization in Turkish.

In the same year, Özsoy, Çiçekli & Alpaslan proposed two new LSA based summarization algorithms [46]. They presented a generic extractive Turkish text summarization system based on LSA. The authors claimed that the Cross method devised in the study was better than any of other LSA methods.

Uzun-Per proposed a concept extraction system for Turkish in 2011 [47]. The system first pre-processed the characters of Turkish alphabet and then it only used nouns, which were stripped from their inflectional morphemes. Then these nouns were clustered via k-means algorithm and concepts were assigned to this cluster of words using a user interface. Then the corpus of documents was tested against the list of concepts

determined in the previous phase of the algorithm. The author claimed that the system devised in the study achieved 41% accuracy, which was low but higher than other studies in the literature.

In a different study Demir, El-Kahlout, Ünal & Kaya presented their efforts to build the first Turkish paraphrase corpus, which was an important step for creating summaries, in 2012 [48]. They drew parallel sentences from multiple translations of literary texts, two different subtitles of a movie, multiple reference translation of a parallel corpus and human-written paraphrases of news sentences. Their system contained 1270 paraphrastic sentences.

3. CONCLUSION

In this paper various approaches of automatic text summarization has been described. Many important works for Turkish in the literature have also been told and the results of them have been discussed by comparing all approaches.

4. ACKNOWLEDGMENTS

This research paper was generated from the project supported under Contract Number: 2011.KB.FEN.030 by Dokuz Eylül University Scientific Research Projects Coordination Unit.

The authors thank Dokuz Eylül University Scientific Research Projects Coordination Unit for their support.

5. REFERENCES

- [1] Torres-Moreno, J.-M. 2014. Automatic Text Summarization, New Jersey: Wiley-ISTE.
- [2] ANSI/NISO. (1997). Guideline for Abstracts, Z39.14-1997, Maryland: Niso Press.
- [3] Mani, I., Klein, G., House, D., Hirschman, L., Firmin, T. & Sundheim, B. (2002). "SUMMAC: a text summarization evaluation" in Natural Language Engineering, vol. 8 no. 1, pp. 43–68.
- [4] Luhn, H. P. (1958). "The automatic creation of literature abstracts", IBM Journal of Research and Development, vol. 2, no. 2, pp. 159–165.
- [5] Radev, D.R, Hovy, E., McKeown, K. (2002). "Introduction to the Special Issue on Summarization", Computational Linguistics, Volume 28, Number 4, 399-408.
- [6] Baxendale, P. B. (1958). "Machine-made Index for Technical Literature: An Experiment", IBM Journal of Research Development, vol. 2, no. 4, pp. 354–361, 1958.
- [7] Vasiliev, A. (1963), "Automatic Abstracting and Indexing", Automatic Documentation – Storage and Retrieval, UNESCO/NS/WS 1163.112, pp. 1–12, Paris.
- [8] Edmundson, H. P. (1969). "New methods in automatic extraction", Journal of the Association for Computing Machinery, vol. 16, no. 2, pp. 264–285.
- [9] Rush, J. E., Salvador, R. & Zamora, A. (1971). "Automatic abstracting and indexing. II. Production of indicative abstracts by application of contextual inference and syntactic coherence criteria", Journal of the American Society for Information Science, vol. 22, no. 4, pp. 260–274.
- [10] Pollock, J.J., Zamora, A. (1975). Automatic Abstracting Research at Chemical Abstracts Service, Journal of Chemical Information and Computer Science, 15,4, pp. 226-232.
- [11] DeJong, G. F. (1982). "An overview of the FRUMP system", in W. G. Lehnert and M. H. Ringle (eds.), Strategies for Natural Language Processing, New Jersey: Lawrence Erlbaum Associates, pp. 149–176.
- [12] Spärck-Jones K. (1993) "What might be in a summary?" in G. Knorz, J. Krause & C. Womser-Hacker, (eds.), Information Retrieval'93: Von der Modellierung zur Anwendung, Universitätsverlag Konstanz, Constance, Germany, pp. 9–26.
- [13] Halliday, M. A. K., & Hasan, R. (1976) Cohesion in English. London: Longman.
- [14] Rumelbart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, N.J.: Erlbaum, 1977.
- [15] Kintsch, W. & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. Psychological Review, 85 (5), 363-394.
- [16] Kupiec, J., Pedersen, J. & Chen, F. (1995). "A trainable document summarizer", 18th Conference ACM Special Interest Group on Information Retrieval (SIGIR'95), Seattle, WA, ACM Press, New York, pp. 68–73.
- [17] McKeown, K. & Radev, D. R. (1995). "Generating summaries of multiple news articles", 18th Conference ACM Special Interest Group on Information Retrieval (SIGIR '95), Seattle, WA, ACM Press, New York, pp. 74–82.
- [18] Marcu, D. (1998). The rhetorical parsing, summarization, and generation of natural language texts, PhD Thesis, Computer Science, University of Toronto, Toronto, Canada.
- [19] Mann, W.C. and Thompson, S.A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text 8(3): 243–81.
- [20] Barzilay, R. & Elhadad, M. (1997). "Using lexical chains for text summarization", Workshop on Intelligent Scalable Text Summarization (ACL '97/EACL '97), Madrid, Spain, pp. 10–17, 11 July 1997.
- [21] Carbonelli, J. G. & Goldstein, J. (1998). "The use of MMR, diversity based reranking for reordering documents and producing summaries", in A. Moffat & J. Zobel (eds.), ACM Special Interest Group on Information Retrieval (SIGIR '98), Melbourne, Australia, pp. 335–336.
- [22] Witbrock, M. J. & Mittal, V. O. (1999). "Ultra-Summarization: A Statistical Approach to Generating Highly Condensed Non-Extractive Summaries", 22nd Conference SIGIR'99, Berkeley, CA, USA, ACM, pp. 315–316, 15–19 August 1999.
- [23] Hovy, E. & Lin, C. Y. (1999). "Automated text summarization in SUMMARIST", in I. Mani & M. T. Maybury (eds.), Advances in Automatic Text Summarization, Cambridge: MIT Press, pp. 81–94.
- [24] Knight, K. and Marcu, D. 2000. Statistics-based summarization – step one: sentence compression, 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence, Austin, TX, pp. 703–10.

- [25] Radev, D. R., Jing, H. & Budzikowska, M. (2000). "Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies", Workshop on Automatic Summarization (NAACL-ANLP-AutoSum'00), Seattle, WA, ACL, pp. 21–30.
- [26] Saggion, H. & Lapalme, G. (2002). "Generating indicative-informative summaries with SumUM", Computational Linguistics, vol. 28, no. 4, pp. 497–526.
- [27] Marcu, D. 2000. The Rhetorical Parsing of Unrestricted Texts: A Surface-Based Approach. Computational Linguistics, 26 (3), pp. 395-448.
- [28] Erkan, G. & Radev, D. R. (2004). "LexRank: graph-based lexical centrality as salience in text summarization", Journal of Artificial Intelligence Research, vol. 22, no. 1, pp. 457–479.
- [29] Barzilay, R. & McKeown, K. R. (2005). "Sentence fusion for multidocument news summarization", Computational Linguistics, vol. 31, no. 3, pp. 297–328.
- [30] Fernández, S., SanJuan, E. & Torres-Moreno, J.-M. (2007) "Textual energy of associative memories: performants applications of Enertex algorithm in text summarization and topic segmentation", Mexican International Conference on Artificial Intelligence (MICA I '07), Aguascalientes: Springer-Verlag, pp. 861–871.
- [31] Svore, K., Vanderwende, L. & Burges, C. J. (2007). "Enhancing single-document summarization by combining RankNet and third-party sources", Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLPCoNLL), Prague, Czech Republic, pp. 448–457.
- [32] Saggion, H. (2008). "SUMMA: A Robust and Adaptable Summarization Tool", Traitement Automatique des Langues, vol. 49, no. 2, pp. 103–125.
- [33] Filippova, K. (2010). "Multi-sentence compression: finding shortest paths in word graphs", International Conference on Computational Linguistics (COLING '10), Beijing, China.
- [34] Nenkova, A. & McKeown, K. (2011). "Automatic summarization", Foundations and Trends in Information Retrieval, vol. 5, nos. 2–3, pp. 103–233.
- [35] Torres-Moreno, J.-M. (2012). "Artex is another text summarizer", CoRR, vol. abs/1210.3312.
- [36] Litvak, M. & Vanetik, N. (2014). "Multi-document summarization using tensor decomposition", Computación y Sistemas (CYS), vol. 18, no. 3.
- [37] Köksal A. (1981) Tümüyle Özdevimli Deneysel Bir Belge Dizinleme ve Erişim Dizgesi: TÜRDER. In the Proceedings of 3. Ulusal Bilişim Kurultayı, Ankara, Turkey, ss. 37-44.
- [38] Oflazer K. and Kuruöz, İ. 1994. Tagging and morphological disambiguation of Turkish text, Proceedings of the Fourth Conference on Applied Natural Language Processing, October 13-15, Stuttgart, Germany.
- [39] Tür G., Hakkani-Tür D. and Oflazer, K. 2003. A statistical information extraction system for Turkish. Natural Language Engineering, 9, pp. 181-210.
- [40] Bilgin O., Çetinoğlu Ö. and Oflazer K. 2004. Building a wordnet for Turkish. Romanian Journal of Information Science and Technology, 7 (1-2), pp. 163-72.
- [41] Karakaya, K. M. and Güvenir, H. A. 2004. ARG: A Tool for Automatic Report Generation, Istanbul University - Journal of Electrical & Electronics Engineering, Vol. 4, No. 2, pp. 1101-9.
- [42] Amasyalı, M. F. and Diri, B. 2006. Automatic Turkish text categorization in terms of author, genre and gender. NLDB'06 Proceedings of the 11th international conference on Applications of Natural Language to Information Systems, pp. 221-6.
- [43] Ercan, G. 2006. Automated Text Summarization and Keyphrase Extraction. Unpublished MSc thesis, Bilkent University.
- [44] Ercan, G. and Çiçekli, İ. 2007. Using lexical chains for Keyword Extraction. Information Processing and Management, 43: 1705-14.
- [45] Kutlu, M., Çığır, C. and Çiçekli, İ. 2010. Generic Text Summarization in Turkish. The Computer Journal, 53: 8, pp. 1315-23.
- [46] Özsoy, M. G., Çiçekli, İ. and Alpaslan, F. N. 2010. Text summarization of Turkish texts using latent semantic analysis. Proceedings of the 23rd International Conference on Computational Linguistics, COLING'10, pp. 869-76.
- [47] Uzun-Per, M. 2011. Developing a Concept Extraction System for Turkish. Unpublished MSc. Thesis, Boğaziçi University.
- [48] Demir, Ş.D., El-Kahlout, İ., Ünal, E. and Kaya, H. 2012. Turkish Paraphrase Corpus. Proceedings of the Eight International Conference on Language Resources and Evaluation LREC'12. pp. 4087-91.