

Towards Personalized Summaries in Spanish based on Learning Styles Theory

Uriel Ramírez¹, Yasmín Hernández², Alicia Martínez¹

¹ Tecnológico Nacional de México, CENIDET, Cuernavaca, Mexico

² Instituto Nacional de Electricidad y Energías Limpias, Gerencia de Tecnologías de la Información, Cuernavaca, Mexico
{uriel.ramirez18ce, amartinez}@cenidet.edu.mx,
yasmin.hernandez@ineel.mx

Abstract. Today, advances in information technologies have generated perhaps the largest and fastest exponential growing of electronic texts. On the Internet there are many electronic documents, such as books, technical documents, news articles, blogs, chats, emails and many other digital files. As a result, a user who wants to read and understand this information in a short time will find it a hard task. In this paper, we have conducted an important work in automatic text summarization. Also, we have considered the particular needs of readers. Thus, a model for personalized summarization based on learning styles theory is proposed.

Keywords: automatic text summarization, concept graph, domain specific summarization, multi-document summarization, natural language processing.

1 Introduction

At the present, advances in information technologies have generated perhaps the largest and fastest exponential growing of electronic texts. Previous studies indicate that the excessive handling of information could cause inefficiency in work, in addition to anxiety and stress [1]. Therefore, in many fields, domains and applications exist the necessity for developing tools that summarized information with different purposes. For example, in Education field the summaries are very needed.

The automatic generation of text summaries consists in taking a text and condensing it in a way sensitive to the needs of the reader [2]. Thus, a useful summary for a reader will be one that takes into account the particular characteristics of readers, such as: knowledge, interests, age, learning styles, etc.

The investigations carried out on Automatic Text Summarization started more than five decades ago with the previous work of Edmundson [3].

Thenceforth, several theories in linguistics and artificial intelligence have been proposed. Such as, superficial techniques [4,5,6], graph-based techniques [7,8,9,10], algebraic reduction techniques [11], statistics [12], etc. However, these researches are still improving. These researches are looking for methods to generate summaries similar to those generated by humans.

Learning styles theory states that every individual has a particular way to learn which includes strategies and preferences. This theory emphasizes that individuals perceive and process information in different ways. Accordingly, learning styles theory states that learning has more to do with a process focusing the learning style than with the intelligence of individuals. Several learning styles models have been proposed, Felder-Silverman learning Styles Model, which is a well-known and broadly used learning styles model [13,14].

In this research work, a model for personalized summarization base on a student model has been proposed. This model will represent the current state of a student. Also, the model includes his/her learning style, among other characteristics. For to do this, we have worked on model the affect and the learning styles of students in order to provide them with more adequate instruction [15,16,17]. Besides, we propose a framework based on conceptual graph for abstractive text summarization. The working language is Spanish.

The paper is organized as follows: Section 2 presents the Background on Automatic Text Summarization and learning styles models. Section 3 describes our approach for a model personalized summarization. Finally, Conclusions and future work are discussed in Section 4.

2 Background and Related Work

2.1 Automatic Text Summarization

The Automatic Text Summarization is the process of extracting or collecting important information from original text, and showing as summary [18]. The Summaries according its purpose can be classified as: generic, domain-specific and query-based on the other hand, the summaries according its output can be abstractive or extractive [19].

As it says, single-document summarization generates a summary from single document input. Conversely, multi-document summarization generates a summary from multiple documents input. Generally, these multiple documents discuss the same topic. Summarization of generic purpose summarizes all documents regardless of topics or domains [19]. While the domain-specific summary focuses on a domain of interest. For example, politics, finance, IT or weather news. The query-based summary deals with the information that is requested from users. In general, queries are sentences in natural language or keywords related to a particular topic [20].

Extractive summarization extracts relevant sentences from the original documents and link them to generate the summary with no changes in the original sentences. In the abstractive summarization, source documents are analyzed and understood using linguistic method to analyze and interpret the document [18,21].

Current research proposes several and diverse methods for automatic text summarization such as statistical [22], machine learning [23,24], text connectivity [25,26], conceptual graphs [27,28,29], algebraic reduction [30], clustering and probabilistic models [31,32,33] and methods adapted to the reader [34,35].

2.2 Concept Graphs

Conceptual graphs are structures for knowledge representation based on first-order logic. Graphs are a natural, simple, and fine-grained semantic representation which can be used to describe texts [27].

Generally, a graph can be denoted in the form $G = (V, E)$, where V represents vertices in the graph and E represents edges between each vertex. In the context of text documents, vertices represent sentences and edges are the weight between two sentences. Consequently, in a graph representing a documents each sentence is a vertex and the weight between each vertex corresponds to the similarity between the two sentences [36].

Figure 1 presents a conceptual graph standing for sentence Manuel hit the piggy bank with a hammer. The concepts are *Manuel*, *Hit*, *Hammer* and *Piggy bank*. These concepts are connected by the relations *agent*, *inst* and *patient*.

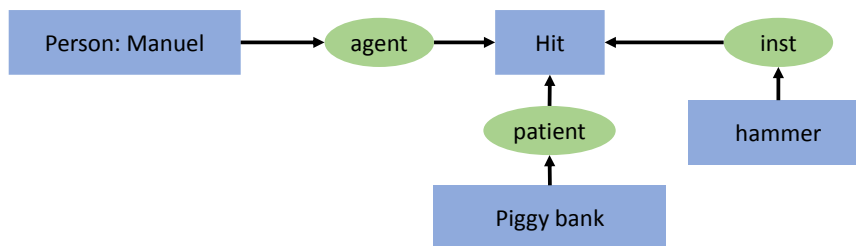


Fig. 1. Conceptual graph for sentence *Manuel hit the piggy bank with a hammer* [36].

2.3 Evaluation Measures for Text Summarization

An important part of the automatic text summarization is the evaluation of the process and results. There are two main tendencies: i) intrinsic evaluation which seeks to measure the quality of the summary, and ii) extrinsic evaluation which measures the performance of the summary in a particular task.

The intrinsic evaluation is based on a standard to be compared with the generated summary. The standard can be an existing dataset (text/summary) or an *ideal* summary generated by a human. On the other hand, extrinsic evaluation requires to select a task which use the summary, and to measure the impact of using the generated summary instead of the original text. In this evaluation, the problems are the selection of the task and the indicators for the measurement [37].

The most common measures to evaluate the quality of the summaries are i) the Measure Recall-Oriented Understudy of Gisting Evaluation (ROUGE) [38], and ii) Standard measures of information retrieval, such as precision, recall and F-measure. These measures are based on the comparison of n -grams between the automatic summary and a manually generated summary. There are several variants of ROUGE: ROUGE- n , ROUGE- L , ROUGE- SU .

2.4 Learning Styles Models

Learning theories describe how people learn concepts and abilities. Diverse learning theories have been proposed, all of them states different, and sometimes, opposite viewpoints. The learning styles theory states that individuals have a particular way to learn which includes strategies and preferences. This theory highlights that individuals perceive and process information in different ways. Consequently, learning styles theory states learning of individuals has more to do with a process focusing the learning style than with the intelligence of individuals[39,40].

Several learning styles models have been proposed, such as Felder-Silverman learning Styles Model [13,14]; which proposes sensing-intuitive, visual-verbal, active-reflective, and sequential-global as categorizations of learning styles. Learning theory of Kolb works on two levels: a four-stage cycle of learning and four separate learning styles. This theory is concerned with internal cognitive processes of learners.

In this sense, the Honey and Alonso questionnaire on learning styles (CHAEA, for the Spanish *Cuestionario de Honey y Alonso de Estilos de Aprendizaje*) has been used as a measurement tool to identify the different types of learning styles presented by students [41]. The instrument consists of 80 items presented in the form of questions, of which 20 correspond to learning styles: active, reflective, theoretical and pragmatic. The affirmative answers are added and the total score is obtained for each style, which is compared with the respective scales.

3 Model for Personalized Summarization

Currently, there are different types of summaries; these will depend on the purpose that the reader has to generate them. The automatic generation of abstracts can generate summaries without taking into account the reader for whom they are intended. This can result in poor understanding or a loss of readers' time. On the other hand, the summaries can be adapted to the peculiarities of the reader, such as: previous knowledge, areas of interest, information needs, learning styles, among other characteristics, to support a better understanding and also to comply with the required information. However, most of the work that has been done revolves around the generic summary [42].

As previous work, we have worked with intelligent training environments for electrical industry where instruction is presented according to particular needs of students [18,19]. In this sense, we have a student model which represents previous knowledge, affective state, personality and learning styles of students [20]. In these training environments, students have to read many handbooks, regulation documents and technical documents. We have notice that in some specific training situations, students do not want to read. They prefer to conduct the practices in a trial and error way. For this reason, we think if students have a personalized summary, they could grasp some aspects that could help them to follow their exercises or lessons. For example, a personalized summary could be useful before to start a lesson to decide which topics to learn, or before to take a quiz or to conduct a practice to review the required material. However, we want to focus the particular needs of students considering their individual characteristics in summarization process.

Additionally, in the academic field, graduate students need to read many research works in order to conduct their work and in order to have success in their studies. In this context personalized summaries are also needed.

Considering these two different domains. We propose a model for personalized summarization in intelligent learning environment. The summarization model will generate abstractive, rather than extractive summaries, for Spanish multi-document source. Fig. 2 shows the architecture of the intelligent learning environment with the model for personalized summarization.

In a high level, the intelligent learning environment consists of student model, tutor module, expert module and domain knowledge. In order to generate the personalized summaries, the model for text summarization was integrated. The personalization process is based on a student model.

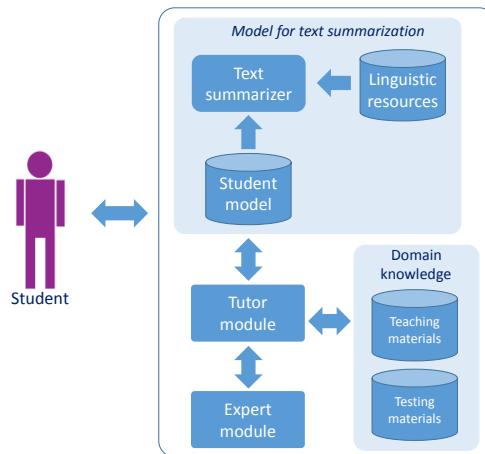


Fig. 2. Architecture of the intelligent learning environment with the model for personalized summarization.

The student model is built with base on the student interaction and its performance in lessons, practices and quizzes. The student model contains following information about students: previous knowledge, affective state, personality, learning styles, age and schooling.

In an initial approach, we wanted to provide students with an adaptive summary according to their learning style. Our proposal for learning styles is based on the CHAEA instrument [43].

The summarization model also integrates linguistic resources built taking into account knowledge about specific topics.

This proposal will be applied in the electrical domain, and some topics on Computer Science. We will use techniques based on conceptual graphs for representing the sentences, entities and their relationships.

Our proposed method aims to generate a summary from a set of documents. However, if the documents are not related, a non-representative or incoherent summary could be obtained. Therefore, we propose first grouping related input documents, and then generating the summary.

The method of summarizing consists on several tasks such as: pre-processing, transformation, synthesis and sentences selection. Fig. 3 shows our proposed method for summarization. This proposed method includes four stages: i) Preprocessing stage, ii) Transformation stage, iii) Synthesis stage, and iv) Sentences selection stage.

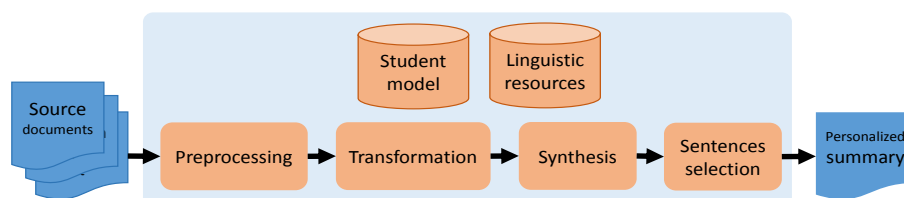


Fig. 3. Proposal for the summarization model including preprocessing, transformation, synthesis and sentences selection.

Preprocessing stage: this stage has aims analyzing the collection of documents, such as: segmentation of the text, extraction of tokens, elimination of empty words, identification of the lexical root of the words and morph-syntactic analysis. The segmentation of text consists on defragmenting the text in paragraphs and sentences. In the extraction of tokens, each sentence is composed of tokens that are not more than each of the parts of the sentence, that is, words, numbers, punctuation marks, etc.

In addition, disambiguation mechanism to discern between different meanings is also necessary to apply in this proposed method. Therefore, it is necessary to translate the lexicon of the document into concepts of the knowledge base of the domain. The disambiguation algorithm as linguistic resources must take into account the domain of the text to be summarized. The disambiguation algorithm will analyze the input sentences; and, as result, a list of concepts or meanings of the knowledge base will be obtained.

Transformation stage: Preprocessing and transformation is based on the model proposed by Miranda, Gelbukh, and Sidorov [37]. For the transformation, this model uses linguistic resources in English language, whereas our model will use linguistic resources in Spanish language, and a user model. In this stage, the representation in the form of a graph will be constructed for each sentence of the document. Capturing the semantic structure and the relationships between terms.

Synthesis stage: The synthesis method is based on a set of operations in Conceptual graphs (GCs): generalization, union or association, weighting and pruning [37].

The generalization operation combines two GCs according to their common elements. For example, the following graphs can be read as G1: Peter buys a crocodile (Peter buys a crocodile) and G2: Mary buys a bird (Mary buys a bird). The GCs are compared and, subsequently, the minimum common concepts to unite them are determined. The hierarchy for crocodile and bird, Animal is the minimum common concept between both concepts; and Person is the least common concept for Peter and Mary. So G3 is the resulting graph after combining the two graphs. G3 can be read as "Two persons buy two animals" (Two people buy two animals).

The union is the operation unites two related concepts of two GCs. This operation supports and improves the results of the weighting process.

The weighting is the operation that filters the most important nodes of the graph. To determine the importance of the nodes, the HITS algorithm is used. HITS is an iterative algorithm that takes into account the degree of input and the degree of output of the node to determine its importance.

Pruning is the operation that is applied to reduce graphs. This operation takes into account the results of the weighting, the verbal patterns to remove the irrelevant nodes and the compression rate or threshold to establish how many nodes should be included in the resulting summary.

Sentences selection stage. This stage aims to select the sentences that will be part of the final summary. Our hypothesis is that knowing the model of the student it will be possible to create a personalized summary. The summary will be constructed from a set of rules and algorithms that will be based on the student's learning style. Subsequently, the sentences are selected according to the order of the documents in which they appear first, and then in the order in which they appear inside the document, to avoid inconsistencies. As a result, the collection of documents is summarized.

In order to evaluate our model, the resulting summary will be evaluated with ROUGE metrics [38]. We use ROUGE metrics because they are automatic metrics which do not need human judges to measure the quality of the summaries and because they are standard metrics used by several research projects.

4 Conclusions and Future Work

The exponential growing of digital documents has motivated the development of technologies as the natural language processing. A field of this area is the automatic text summarization, which is the process of extracting or collecting important information from original text, and showing as summary. Despite there have been diverse efforts, automatic text summarization is in an initial stage. Most researches are interested in to generate a standard summary where particular needs of readers are not considered.

We propose a model of a personalized summarization according on a student model. This due to the model intends integrate to an intelligent learning environment.

For the time being, we propose to personalize summaries based on the learning styles theory. However, we want examine other ways to personalize the summary; such as, schooling level, age, among other characteristics.

An important contribution of our model will be the linguistic resources. Which will be constructed in the Spanish language. Currently, there is no great variety of resources for the Spanish language.

In this paper, we presented our proposal method for generating Personalized Summaries in Spanish based on Learning Styles Theory. Although this work is in an initial state, we are working on the generation of the user's model.

Acknowledgments. This research work has been partially funded by European Commission and CONACYT, through the SmartSDK project.

References

1. Klingberg, T.: Information Overload and the limits of working memory. Stockholm, Sweden (2009)
2. Steinberger, J., Ježek, K.: Evaluation measures for text summarization. *Comput. Informatics* 28(2), pp. 251–275 (2009)
3. Edmundson, H.P.: *New Methods in Automatic Extracting* (1969)
4. Hahn, U., Mani, I.: Challenges of automatic summarization. *Computer* 33(11), pp. 29–36, Long Beach, Calif. (2000)
5. Ulrich, J., Carenini, G., Murray, G., Ng, R.: Regression-Based Summarization of Email Conversations. *Intelligence*, pp. 334–337 (2009)
6. Wan, S., McKeown, K.: Generating overview summaries of ongoing email thread discussions. In: *Proc. 20th Int. Conf. Comput. Linguist. - COLING '04*, p. 549–es (2004)
7. Plaza-Morales, L.: *Uso de Grafos Semánticos en la Generación Automática de Resúmenes y Estudio de su Aplicación en Distintos Dominios: Biomedicina, Periodismo y Turismo*. p. 352 (2011)
8. Mihalcea, R., Tarau, P.: TextRank: Bringing order into texts. In: *Proc. EMNLP*, vol. 85, pp. 404–411 (2004)
9. Brin, S., Page, L.: Reprint of: The anatomy of a large-scale hypertextual web search engine. *Comput. Networks* 56(18), pp. 3825–3833 (2012)
10. Mittal, N., Agarwal, B., Vijay, N., Gupta, A., Upadhyay, N.K.: Semantic Enhanced Text Summarization. Vol. 1, no. 1, pp. 26–29 (2014)
11. Vicente, M., Lloret, E.: The University of Alicante at MultiLing 2015 : approach, results and further insights. No. September, pp. 250–259 (2015)
12. Aries, A., Zegour, D.E., Hidouci, K.W.: AllSummarizer system at MultiLing 2015 : Multilingual single and multi-document summarization. In: *Proc. MultiLing 2015*, vol. 2015, no. September, pp. 237–244 (2015)
13. Graf, S., Kinshuk: Advanced Adaptivity in Learning Management Systems by Considering Learning Styles. In: *IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol.*, vol. 3, p. 235–238 (2009)
14. Felder, R., Silverman, L.: Learning and teaching styles in engineering education. *Eng. Educ.*, vol. 78, no. June, pp. 674–681 (1988)
15. Model, L.S.: Learning styles theory for intelligent learning environments - Adapting the Instruction. In: *Proc. 3rd Int. Conf. Comput. Support. Educ.*, pp. 456–459 (2011)
16. Hernández, Y., Pérez-Ramírez, M., Zatarain-Cabada, R., Barrón-Estrada, L., Alor-Hernández, G.: Designing empathetic animated agents for a b-learning training environment within the electrical domain. *Spec. Issue Intell. Affect. Learn. Environ. New Trends Challenges Educ. Technol. Soc.* 19(2) (2016)
17. Hernández, Y., Arroyo-Figueroa, G., Sucar, L.E.: A Model of Affect and Learning for Intelligent Tutors. *Univers. Comput. Sci.*, 21(7), pp. 912–934 (2015)
18. Gaikwad, D.K., Mahender, C.N.: A Review Paper on Text Summarization. *Int. J. Adv. Res. Comput. Commun. Eng.* 5(3), pp. 154–160 (2016)
19. Kumar, Y.J., Goh, O.S., Basiron, H., Choon, N.H., Suppiah, P.C.: A review on automatic text summarization approaches. *J. Comput. Sci.* 12(4), pp. 178–190 (2016)
20. Vilca, G.C.V.: *Generación automática de resúmenes abstractivos mono documento utilizando análisis semántico y del discurso*. Pontificia Universidad Católica del Perú (2017)
21. Acero, I., Alcojor, M., Díaz, A., Gómez, J.M., Maña, M.: Generación automática de resúmenes personalizados. *Proces. del Leng. Nat.* 27(33), pp. 281–290 (2001)
22. Jeong, H., Ko, Y., Seo, J.: Efficient keyword extraction and text summarization for reading articles on smart phone., vol. 34, pp. 779–794 (2015)
23. Wong, K.F., Wu, M., Li, W.: Extractive Summarization Using Supervised and Semi-

- supervised Learning. In: Proc. 22nd Int. Conf. Comput. Linguist. 1. Assoc. Comput. Linguist. 2008., no. August, pp. 985–992 (2008)
24. Fattah, M.A.: A hybrid machine learning model for multi-document summarization. *Appl. Intell.* 40(4), pp. 592–600 (2014)
 25. Pal, A.R., Maiti, P.K., Saha, D.: An approach to automatic text summarization using simplified lesk algorithm and wordnet. *Int. J. Control Theory Comput. Model.* 3(4), pp. 15–23 (2013)
 26. Atkinson, J., Munoz, R.: Rhetorics-based multi-document summarization. *Expert Syst. Appl.* 40(11), pp. 4346–4352 (2013)
 27. Miranda-Jiménez, S. *et al.*: Conceptual Graphs as Framework for Summarizing Short Texts. *Int. J. Concept. Struct. Smart Appl.* 2(2), pp. 55–75 (2014)
 28. Martínez-Romo, J., Araujo, L., Fernández, A.D.: SemGraph : Extracting Keyphrases Following a Novel Semantic Graph-Based Approach. *J. Assoc. Inf. Sci. Technol.* 67(1), pp. 71–82 (2016)
 29. Chatterjee, N., Sahoo, P.K.: Random Indexing and Modified Random Indexing based approach for extractive text summarization. *Comput. Speech Lang.* 29(1), pp. 32–44 (2015)
 30. Lee, J.H., Park, S., Ahn, C.M., Kim, D.: Automatic generic document summarization based on non-negative matrix factorization. *Inf. Process. Manag.* 45(1), pp. 20–34 (2009)
 31. Mei, J.P., Chen, L.: SumCR: A new subtopic-based extractive approach for text summarization. *Knowl. Inf. Syst.* 31(3), pp. 527–545 (2012)
 32. Wang, M., Tang, X.: Extract summarization using concept-obtained and hybrid parallel genetic algorithm. In: Proc. - Int. Conf. Nat. Comput., no. Icncc, pp. 662–664 (2012)
 33. Nagesh, K., Murty, M.N.: Obtaining single document summaries using latent dirichlet allocation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7666 LNCS, no. PART 4, pp. 66–74 (2012)
 34. Kumar, C., Pingali, P., Varma, V.: Generating Personalized Summaries Using Publicly Available Web Documents. In: 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 103–106 (2008)
 35. Díaz, A., Gervás, P.: User-model based personalized summarization. *Inf. Process. Manag.* 43(6), pp. 1715–1734 (2007)
 36. Rao, P.R.K., Lalitha Devi, S.: Enhancing multi-document summarization using concepts. *Sadhana - Acad. Proc. Eng. Sci.* 43(2), pp. 1–11 (2018)
 37. Miranda, S., Gelbukh, A., Sidorov, G.: Generación de resúmenes por medio de síntesis de grafos conceptuales. *Rev. Signos* 47(86), pp. 463–485 (2014)
 38. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Proc. Work. text Summ. branches out, no. 1, pp. 25–26 (2004)
 39. Funderstanding: Nothing found for About_Learning Cfm. Engaging kids, learning theories. [Online]. Available: https://www.funderstanding.com/about_learning.cfm. [Accessed: 12-Sep-2018].
 40. Kazu, I.Y.: The Effect of Learning Styles on Education and the Teaching Process. *J. Soc. Sci.* 5(2), pp. 85–94 (2009)
 41. Alonso, C.M., Gallego, D.J., Honey, P.: Los estilos de aprendizaje. *Procedimientos de diagnóstico y mejora. Procedimientos diagnóstico y Mejor.* Bilbao Mensajero (1994)
 42. Nenkova, A., McKeown, K.: Automatic Summarization. *Found. Trends® Inf. Retr.* 5(2), pp. 103–233 (2011)
 43. Alonso, C.M., Gallego, J., Honey, P.: CHAEA. Cuestionario Honey-Alonso de Estilos de Aprendizaje. Universidad de Deusto. Instituto de Ciencias de la Educación (ICE), 2017. [Online]. Available: <http://diged.usac.edu.gt/sfpu/cuestionario/chaea>.