Casa abierta al tiempo
**UNIVERSIDAD AUTÓNOMA METROPOLITANA**
**Unidad Cuajimalpa**

*Comunidad académica comprometida*
*con el desarrollo humano de la sociedad.*

DCCD.DTI.022.19
Mayo 20, 2019

**Dr. Octavio Mercado González**
Presidente del Consejo Divisional de la
División de Ciencias de la Comunicación y Diseño
Unidad Cuajimalpa
Presente

**Asunto:** Solicitud de Año Sabático del
Dr. Esaú Villatoro Tello

*Estimado Dr. Mercado:*

Con relación al asunto arriba referido, por este conducto me permito solicitarle se someta a consideración del Consejo Divisional la petición que el Dr. Esaú Villatoro Tello me ha enviado para disfrutar de un Año Sabático, que iniciaría en septiembre de 2019.

Para tal efecto, anexo copia de la documentación que me hizo llegar la Coordinación de Recursos Humanos de esta Unidad, en la cual se hace constar que el Esaú Villatoro Tello ha cumplido los requisitos de tiempo de antigüedad para disfrutar de tal año sabático. Anexo también el Plan de Trabajo que presenta el Esaú Villatoro Tello.

Se envían los documentos indicados y anexados en formato digital, vía correo electrónico.

Sin otro particular, envío a usted un cordial saludo.

**Atentamente,**

Casa abierta al tiempo

**Dr. Carlos Joel Rivero Moreno**
Jefe del Departamento de Tecnologías de la Información

Anexo: Lo mencionado.

c.c.p.: Dra. Gloria Angélica Martínez de la Peña – Secretaria del Consejo Divisional
Lic. Inés Andrea Zepeda Martínez – Oficina Técnica de Consejo Divisional.

*CJRM*

División Ciencias de la Comunicación y Diseño

**Unidad Cuajimalpa**
DCCD | Jefatura del Departamento de Tecnologías de la Información
Torre III, 5to. piso. Avenida Vasco de Quiroga 4871,
Colonia Santa Fe Cuajimalpa. Delegación Cuajimalpa de Morelos,
Tel. +52 (55) 5814-6557, C.P. 05348, México, D.F.
http://dccd.cua.uam.mx

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

FI-DRH-20 / 12182013

# SOLICITUD DE PERIODO SABÁTICO

MTRO. OCTAVIO MERCADO GONZÁLEZ

| FECHA DE ELABORACIÓN | DÍA | MES | AÑO |
|---|---|---|---|
| | 20 | 05 | 2019 |

DIRECTOR DE LA DIVISIÓN DE: CIENCIAS DE LA COMUNICACIÓN Y DISEÑO DE LA UNIDAD CUAJIMALPA

| APELLIDO PATERNO | APELLIDO MATERNO | NOMBRE (S) |
|---|---|---|
| VILLATORO | TELLO | ESAÚ |

CATEGORÍA Y NIVEL: PROFESOR TITULAR C

| UNIDAD | DIVISIÓN | |
|---|---|---|
| CUAJIMALPA | CIENCIAS DE LA COMUNICACIÓN Y DISEÑO | TECNOLOGÍAS DE LA INFORMACIÓN |

| FECHA DE INGRESO A LA UAM COMO PERSONAL ACADÉMICO | DÍA | MES | AÑO |
|---|---|---|---|
| | 01 | 06 | 2012 |

| | | DÍA | MES | AÑO | | DÍA | MES | AÑO | No. DE MESES |
|---|---|---|---|---|---|---|---|---|---|
| ÚLTIMO PERIODO SABÁTICO DISFRUTADO, EN SU CASO | **DEL** | | | | **AL** | | | | |
| FECHA DEL PERIODO SABÁTICO SOLICITADO: | **A PARTIR DEL** | 9 | 09 | 2019 | **AL** | 08 | 09 | 2020 | 12 |

(PARA SER LLENADO POR LA OFIC NA DEL CONSEJO DIVISIONAL)

APROBADO POR EL CONSEJO DIVISIONAL CON EL ACUERDO ____________________ DE LA SESIÓN ____________

**DOCUMENTOS QUE ACOMPAÑAN LA SOLICITUD:**

- CONSTANCIA OFICIAL DE SERVICIOS EN LA UNIVERSIDAD ✓
- PROGRAMA DE ACTIVIDADES ACADÉMICAS A DESARROLLAR ✓

INTERESADO

FIRMA

APROBACIÓN DEL CONSEJO DIVISIONAL (PRESIDENTE)

NOMBRE Y FIRMA

T1 SUBDIRECCIÓN DE PERSONAL
T2 ÁREA DE RECURSOS HUMANOS DE UNIDAD
T3 CONSEJO DIVISIONAL
T4 INTERESADO

Casa abierta al tiempo
**UNIVERSIDAD AUTÓNOMA METROPOLITANA**
**Unidad Cuajimalpa**

*Comunidad académica comprometida*
*con el desarrollo humano de la sociedad.*

**RHC.080.2019**

**PERIODO SABÁTICO**

**14 de mayo 2019**

**DR. ESAU VILLATORO TELLO**
**DEPARTAMENTO DE TECNOLOGIAS DE LA INFORMACION**
**DIVISIÓN DE CIENCIAS DE LA COMUNICACIÓN Y DISEÑO**
**UNIDAD CUAJIMALPA**
**Presente.**

**Estimado Dr. Villatoro**

Conforme a su petición y de acuerdo a nuestros registros y a su trayectoria laboral dentro de nuestra Institución, usted inicia la acumulación de tiempo para el disfrute de periodo sabático, a partir del **1 de junio de 2012** y durante su estancia laboral en esta Universidad, no **ha disfrutado de ningún periodo sabático.**

Para esta fecha usted acumula para periodo sabático, seis años, once meses, trece días, de labores ininterrumpidas en su plaza académica al servicio de la Universidad, **por lo que puede solicitar y disfrutar de un periodo sabático por un tiempo máximo hasta de doce meses (un año).**

Sin otro particular, estoy a sus apreciables órdenes para cualquier aclaración al respecto.

**ATENTAMENTE**
**"CASA ABIERTA AL TIEMPO"**

Casa abierta al tiempo
...IDAD AUTÓNOMA METROPOLITANA
...AD CUAJIMALPA
...RSOS HUMANOS

**LIC. LUIS BECERRA CASTAÑEDA**
**COORDINADOR DE RECURSOS HUMANOS**

C.c.p. Mtro. Octavio Mercado González, Presidente del Consejo Divisional DCCD.
Dr. Carlos Joel Rivero Moreno, Jefe del Depto. de Tecnologías de la Información, DCCD.
Expediente

**Unidad Cuajimalpa**
Secretaría de Unidad
Torre III, 8to. piso. Avenida Vasco de Quiroga 4871, Colonia Santa Fe Cuajimalpa
Delegación Cuajimalpa de Morelos, CDMX., C.P. 05348,
Tel. 5814-6505 a 07; correo electrónico: cgarcia@correo.cua.uam.mx
www.cua.uam.mx

# Text Summarization of Spoken Documents

Research Project Proposal during Sabbatical leave

ESAÚ VILLATORO-TELLO, PH.D.
Universidad Autónoma Metropolitana - Unidad Cuajimalpa
May 2019

## Contents



**Abstract**

Given the Big Data era in what we live now, every day an enormous amount of multimodal data is generated across the world by distinct types of users. Accordingly, document summarization applications have gained popularity because of its ability for extracting vital information within a short time. Although plenty of research has been done by the NLP community in proposing novel summarization strategies, there is still a great room for improvement, especially for the task of summarizing spoken documents. As part of this project, we aim at investigating the pertinence of recent semantic encoding strategies in combination with multi-modal unsupervised learning strategies for effectively generating summaries from spoken documents.

**Keywords:** Document Summarization, Text Mining, Sentence Selection, Sentence Ranking, Unsupervised Learning.

## 1 Summary

Text summarization is the task concerning the automatic generation of document summaries, aiming at reducing document(s) in length and complexity while preserving essential information elements. The task can be classified in single document summarization (SDS) [Kupiec et al., 1999], or multiple-document summarization (MDS) [McKeown and Radev, 1999]. Additionally, methods can be categorized according to the summary's purpose into generic summarization and query-biased summarization. The former has no restrictions regarding the information that must be present at the final summary while the latter, must answer to particular user's information needs [Torres-Moreno, 2014].

Although document summarization has been widely researched by the Natural Language Processing community for a long time now[1], one of the initial works that demonstrate the impact of machine learning strategies in solving the task is the work of [Kupiec et al., 1999], since then, the popularity of ML-based methods for generating extractive summaries has increased. *Extractive* summarization approaches focus on identifying and extracting the most salient document components in order to build a coherent summary [Kupiec et al., 1999], whereas *abstractive* summarization involves sentence compression and reformulation of contents [Knight and Marcu, 2000]. Currently, the main strategies fall into the extractive summarization category, given its simple but effective methodology for building summaries without requiring advanced post-processing techniques. In general terms, two big steps are required for generating an extractive summary: *i)* **sentence ranking**, and *ii)* **sentence selection** [Cao et al., 2015, Ren et al., 2017]. Sentence ranking aims at scoring sentences to measure its importance, while sentence selection searches for those sentences with highest relevance scores and low redundancy when generating the final summary until a budget constraint is met, e.g., the size of the summary in characters or bytes.

Nowadays, the vast majority of existing approaches have focused on developing novel strategies for summarizing (static) text documents[2]. However, due to the rapid development and maturity of multimedia technology, there are large volumes of audiovisual material containing valuable information, being spoken documents an essential source of information [Chen and Chen, 2008, Chen et al., 2013]. Contrary to text documents, spoken documents represent a more challenging scenario for summarization strategies. Spoken documents lack structure, and even though can be automatically transcribed, current recognition systems may introduce errors resulting in inaccurate sentences or paragraph boundaries as well as redundant information produced by the speaker disfluencies, fillers, off-topic expressions, and repetitions.

A key aspect of document summarization is how to encode the semantic information contained in documents effectively. Previous work has shown that sentence importance also depends on contextual relations, both, within-document, and cross-document relationships [Ren et al., 2017, Wan and Xiao, 2009]. However, while some of the previous work relies on purely lexical features [Wan and Xiao, 2009, Qiang et al., 2016], word embeddings such as Word2vec [Xiong and Ji, 2016], or very powerful and complex models, such as recurrent neural networks (RNNs) [Cao et al., 2015], CNNs [Zhang et al., 2017], attention models [Ren et al., 2017, Zhong et al., 2015], none of these previous work has evaluated the impact of simpler unsupervised sentence embeddings in capturing these semantic relationships. As part of this proposal, we argue that this aspect can be tackled through the use of sentence embeddings such as the method proposed in [Pagliardini et al., 2018], which is an unsupervised model, fast to train, and easily interpretable, that has demonstrated being more effective than models using deep learning architectures.

Besides, we want to evaluate the impact of multi-modal/noise-robust unsupervised strategies for the spoken document summarization. Mainly, given the nature of spoken documents, where features containing information cues about prosody/acoustics can be obtained, we want to evaluate the pertinence of multi-modal (speech and text) strategies for identifying relevant sentences [Xiong and Ji, 2016, Wan and Xiao, 2009].

[1] Early studies are dated back to the 1950s and 1960s.

[2] See §3 for a general description of the relevant literature in SDS and MDS.

## 2 Objectives

In general, the main goals of the sabbatical leave project would be the following:

1. Evaluate the impact of distinct sentence embeddings for encoding the semantic information contained in spoken documents' transcriptions.
2. Determine the pertinence of features extracted from the speech signal (prosody/acoustics) for identifying relevant information from spoken documents.
3. Assess the influence of a multi-modal (text and speech) approach based on unsupervised learning for summarizing spoken documents.

## 3 Relevant literature

Text summarization is a challenging task which has a long history dated back to 1950s. Since the early 2000s, machine learning has been widely applied to text summarization. The key idea of these approaches is to formulate summarization as a binary classification problem and train a classifier to distinguish relevant sentences or non-relevant sentences. Among the first works in proposing solving the task of document summarization as a supervised problem is the work of [Kupiec et al., 1999], where authors evaluate the impact of features related to the structure of the document, such as the position of the sentences, similarity with the title, length, etc. Similarly, under a supervised paradigm, in [Villatoro-Tello et al., 2006] authors exploit word-sequences for enriching sentences representation. More recently, deep learning has been applied to text summarization with promising results. For instance, in [Cao et al., 2015] authors design deep models for ranking sentences by using Recursive Neural Networks, in [Zhang et al., 2017] is proposed a Convolutional Neural Networks (CNN) with multiple viewpoints, and in [Ren et al., 2017] authors evaluated a Neural Attention Model able to incorporate contextual relations among sentences in the problem of sentence ranking.

Although the supervised approach is prevalent, its main drawback is the availability of large datasets for training classification models. Paper [Radev et al., 2004] represents some of the initial research works in proposing an unsupervised method for document summarization based on centroids of clusters. Besides, several graph-based methods exist for unsupervised extractive summarization. This type of techniques produces a similarity graph, in which each node represents a sentence and edges represent sentences similarity computed through traditional text similarity measures. In [Erkan and Radev, 2004] a method named Lexrank is proposed, which ranks sentences employing the well-known PageRank algorithm. Other graph-based methods are presented in [Villatoro-Tello et al., 2009, Luna-Tlatelpa et al., 2017] where sentence centrality is used to determine its relevance. In [Wan and Xiao, 2009] a mani-fold-ranking algorithm is employed for producing the summary. Similarly, [Xiong and Ji, 2016] proposes a hypergraph-based Vertex-reinforced random walk framework for learning the word-topic distributions in sentences. In [Qiang et al., 2016] authors proposed a pattern-based model which exploits closed patterns to extract most salient information elements from a document.

As to the development of spoken document summarization, quite methods have been proposed, employing both supervised and unsupervised techniques. In [Christensen et al., 2003] evaluate the impact of lexical and structural features for summarizing

English spoken documents. In [Zhang and Fung, 2007, Chen et al., 2013] authors showed the relevance of acoustic and structural features for summarizing automatically generated speech transcriptions. In a similar line of research in [Chen and Chen, 2008, Wang and Cardie, 2012, Kim and Kim, 2016] authors investigate the use of latent topics for representing the information and produce extractive summaries of spoken documents.

As preliminary observations, we noticed that current unsupervised strategies could be enriched using novel sentence embedding strategies. Additionally, employing multi-modal approaches for detecting sentences relevance in spoken documents can be investigated.

## 4 Methodology

In order to achieve the objectives proposed in this project, we plan to work according to the following methodology.

1. Evaluate the impact of several sentence embedding strategies for encoding the semantics of automatically generated transcriptions, specifically Sent2Vec [Pagliardini et al., 2018]. For this, we plan to compare traditional approaches such as CBOW, Skipgram, FastSent, etc. Comparisons against traditional topic detection strategies are considered, such as LSA, LDA, or matrix co-factorization [Nguyen et al., 2017].
2. Incorporate several acoustic features into an unsupervised sentence ranking approach to determine its relevance on detecting silent portions of information from a speech signal. We plan to investigate the pertinence of basic features such as the pitch values. Additional features can be incorporated such as the duration of sentences [Chen et al., 2013].
3. Implement and evaluate the impact of Multi-modal unsupervised learning algorithms. Specifically, we aim at evaluating the manifold-ranking algorithm as proposed in [Xiong and Ji, 2016] for fusing two distinct modalities (text and speech) in the problem of spoken document summarization.

### 4.1 Data

Throughout this project, we plan to work with the data available for the SM2 project. This project addresses several challenges related to the majority of financial institutions in Switzerland and abroad. According to the project's description, we will be testing our proposed approaches on records of all voice and text communications with clients that had a trading purpose.

Hence, the main domain of the data is financial and is expected that we will not have too much training/development data. In addition, if necessary, we plan to build and evaluate our proposal with out-of-domain data.

## 5 Expected results

As the main result of this collaboration project, we expect to develop novel strategies for spoken documents summarization. Our main idea is to prove the impact of recent developments on sentence embedding for accurately encode sentences semantics in spoken documents. Additionally, given the nature of the data, we want to determine how relevant are the set of features that can be extracted from the speech signal in solving the posed task. From this set of experiments is expected to have at least one conference paper published.

As additional results, is expected to improve the collaboration between the IDIAP and UAM-Mexico in future research projects.

# References


**Cao et al., 2015.** Cao, Z., Wei, F., Dong, L., Li, S., and Zhou, M. (2015). Ranking with recursive neural networks and its application to multi-document summarization. In *Twenty-ninth AAAI conference on artificial intelligence*.

**Chen and Chen, 2008.** Chen, B. and Chen, Y.-T. (2008). Extractive spoken document summarization for information retrieval. *Pattern Recognition Letters*, 29(4):426–437.

**Chen et al., 2013.** Chen, B., Lin, S.-H., Chang, Y.-M., and Liu, J.-W. (2013). Extractive speech summarization using evaluation metric-related training criteria. *Information Processing & Management*, 49(1):1–12.

**Christensen et al., 2003.** Christensen, H., Gotoh, Y., Kolluru, B., and Renals, S. (2003). Are extractive text summarisation techniques portable to broadcast news? In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No. 03EX721)*, pages 489–494. IEEE.

**Erkan and Radev, 2004.** Erkan, G. and Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479.

**Kim and Kim, 2016.** Kim, H. H. and Kim, Y. H. (2016). Generic speech summarization of transcribed lecture videos: Using tags and their semantic relations. *J. Assoc. Inf. Sci. Technol.*, 67(2):366–379.

**Knight and Marcu, 2000.** Knight, K. and Marcu, D. (2000). Statistics-based summarization-step one: Sentence compression. *AAAI/IAAI*, 2000:703–710.

**Kupiec et al., 1999.** Kupiec, J., Pedersen, J., and Chen, F. (1999). A trainable document summarizer. *Advances in Automatic Summarization*, pages 55–60.

**Luna-Tlatelpa et al., 2017.** Luna-Tlatelpa, L., Villatoro-Tello, E., de-la Rosa, G. R., and Rivero-Moreno, C. J. (2017). Query-based multi-document summarization through distributional representations. *Research in Computing Science (Spanish)*, 134:127–139.

**McKeown and Radev, 1999.** McKeown, K. and Radev, D. R. (1999). Generating summaries of multiple news articles. *Advances in automatic text summarization*, pages 381–389.

**Nguyen et al., 2017.** Nguyen, M.-T., Cuong, T. V., Hoai, N. X., and Nguyen, M.-L. (2017). Utilizing user posts to enrich web document summarization with matrix co-factorization. In *Proceedings of the Eighth International Symposium on Information and Communication Technology*, pages 70–77. ACM.

**Pagliardini et al., 2018.** Pagliardini, M., Gupta, P., and Jaggi, M. (2018). Unsupervised learning of sentence embeddings using compositional n-gram features. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational*

*Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 528–540.

**Qiang et al., 2016.** Qiang, J.-P., Chen, P., Ding, W., Xie, F., and Wu, X. (2016). Multi-document summarization using closed patterns. *Knowledge-Based Systems*, 99:28–38.

**Radev et al., 2004.** Radev, D. R., Jing, H., Styś, M., and Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938.

**Ren et al., 2017.** Ren, P., Chen, Z., Ren, Z., Wei, F., Ma, J., and de Rijke, M. (2017). Leveraging contextual sentence relations for extractive summarization using a neural attention model. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 95–104. ACM.

**Torres-Moreno, 2014.** Torres-Moreno, J.-M. (2014). *Automatic text summarization*. John Wiley & Sons.

**Villatoro-Tello et al., 2006.** Villatoro-Tello, E., Villaseñor-Pineda, L., and Montes-y Gómez, M. (2006). Using word sequences for text summarization. In Sojka, P., Kopeček, I., and Pala, K., editors, *Text, Speech and Dialogue*, pages 293–300, Berlin, Heidelberg. Springer Berlin Heidelberg.

**Villatoro-Tello et al., 2009.** Villatoro-Tello, E., Villasenor-Pineda, L., Montes-y Gómez, M., and Pinto-Avendano, D. (2009). Multi-document summarization based on locally relevant sentences. In *2009 Eighth Mexican International Conference on Artificial Intelligence*, pages 87–91. IEEE.

**Wan and Xiao, 2009.** Wan, X. and Xiao, J. (2009). Graph-based multi-modality learning for topic-focused multi-document summarization. In *Twenty-First International Joint Conference on Artificial Intelligence*.

**Wang and Cardie, 2012.** Wang, L. and Cardie, C. (2012). Unsupervised topic modeling approaches to decision summarization in spoken meetings. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '12, pages 40–49, Stroudsburg, PA, USA. Association for Computational Linguistics.

**Xiong and Ji, 2016.** Xiong, S. and Ji, D. (2016). Query-focused multi-document summarization using hypergraph-based ranking. *Information Processing & Management*, 52(4):670–681.

**Zhang and Fung, 2007.** Zhang, J. and Fung, P. (2007). Speech summarization without lexical features for mandarin broadcast news. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pages 213–216. Association for Computational Linguistics.

**Zhang et al., 2017.** Zhang, Y., Er, M. J., Zhao, R., and Pratama, M. (2017). Multiview convolutional neural networks for multidocument extractive summarization. *IEEE transactions on cybernetics*, 47(10):3230–3242.

**Zhong et al., 2015.** Zhong, S.-h., Liu, Y., Li, B., and Long, J. (2015). Query-oriented unsupervised multi-document summarization via deep learning model. *Expert systems with applications*, 42(21):8146–8155.