

---

**IRIS: Asistente virtual para la redacción personalizada  
de correos electrónicos**  
**IRIS: Virtual Assistant for Personalized Email Writing**

---



**Trabajo de Fin de Máster**  
**Curso 2020–2021**

**Autor**  
**Carlos Moreno Morera**

**Director**  
**Carlos Rodríguez Abellán**

**Máster en Data Science y Big Data**  
**U-Tad**



IRIS: Asistente virtual para la redacción  
personalizada de correos electrónicos  
IRIS: Virtual Assistant for Personalized  
Email Writing

**Trabajo de Fin de Máster en Data Science y Big Data**

**Autor**

**Carlos Moreno Morera**

**Director**

**Carlos Rodríguez Abellán**

**Convocatoria:** *Septiembre 2021*

**Calificación:** *Nota*

**Máster en Data Science y Big Data  
U-Tad**

**3 de septiembre de 2021**



# Dedicatoria

*A Pedro Pablo y Marco Antonio, por crear TeXiS  
e iluminar nuestro camino*



# Agradecimientos

A Guillermo, por el tiempo empleado en hacer estas plantillas. A Adrián, Enrique y Nacho, por sus comentarios para mejorar lo que hicimos. Y a Narciso, a quien no le ha hecho falta el Anillo Único para coordinarnos a todos.





# Resumen

## **IRIS: Asistente virtual para la redacción personalizada de correos electrónicos**

Un resumen en castellano de media página, incluyendo el título en castellano. A continuación, se escribirá una lista de no más de 10 palabras clave.

### **Palabras clave**

Máximo 10 palabras clave separadas por comas



# Abstract

## **IRIS: Virtual Assistant for Personalized Email Writing**

An abstract in English, half a page long, including the title in English. Below, a list with no more than 10 keywords.

## **Keywords**

10 keywords max., separated by commas.



# Índice

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>1</b>  |
| 1.1. Motivación . . . . .   | 2         |
| 1.2. Objetivos . . . . .  | 2         |
| 1.3. Plan de trabajo . . . . .  | 2         |
| 1.4. Explicaciones adicionales sobre el uso de esta plantilla . . . . . | 2         |
| 1.4.1. Texto de prueba . . . . .  | 2         |
| <b>2. Estado de la Cuestión</b>   | <b>7</b>  |
| 2.1. Generación de Lenguaje Natural . . . . .                           | 7         |
| 2.1.1. ¿Qué es la Generación de Lenguaje Natural? . . . . .             | 8         |
| <b>3. Descripción del Trabajo</b>                                       | <b>11</b> |
| 3.1. Enron corpus . . . . .   | 12        |
| <b>4. Conclusiones y Trabajo Futuro</b>                                 | <b>13</b> |
| <b>5. Introduction</b>  | <b>15</b> |
| <b>6. Conclusions and Future Work</b>                                   | <b>17</b> |
| <b>Bibliografía</b>   | <b>19</b> |
| <b>A. Título del Apéndice A</b>   | <b>23</b> |
| <b>B. Título del Apéndice B</b>   | <b>25</b> |



# Índice de figuras

|                                  |    |
|----------------------------------|----|
| 3.1. Ejemplo de imagen . . . . . | 11 |
|----------------------------------|----|





# Índice de tablas

|                                 |    |
|---------------------------------|----|
| 3.1. Tabla de ejemplo . . . . . | 11 |
|---------------------------------|----|



# Introducción

*“Frase célebre dicha por alguien inteligente”*

— Autor

El estudiante elaborará una memoria descriptiva del trabajo realizado, con una **extensión mínima recomendada de 50 páginas** incluyendo al menos una introducción, objetivos y plan de trabajo, resultados con una discusión crítica y razonada de los mismos, conclusiones y bibliografía empleada en la elaboración de la memoria.

La memoria se puede redactar en castellano o en inglés, pero en el primer caso la introducción y las conclusiones de la memoria tienen que traducirse también al inglés y aparecerán como capítulos **al final de la memoria**. En ambos casos, el título de la memoria aparecerá en castellano y en inglés.

Además del cuerpo principal describiendo el trabajo realizado, la memoria contendrá los siguientes elementos, que no computarán para el cálculo de la extensión mínima del trabajo:

- un resumen en inglés de media página, incluyendo el título en inglés,
- ese mismo resumen en castellano, incluyendo el título en castellano,
- una lista de no más de 10 palabras clave en inglés,
- esa misma lista en castellano,
- un índice de contenidos, y
- una bibliografía.

La portada de la memoria deberá contener la siguiente información:

- "Máster en NOMBRE DEL MÁSTER, Facultad de Informática, Universidad Complutense de Madrid"
- Título
- Autor
- Director(es)
- Colaborador externo de dirección, si lo hay

- Curso académico
- Solo en la versión final: convocatoria y calificación obtenida

Para facilitar la escritura de la memoria siguiendo esta estructura, el estudiante podrá usar las plantillas en LaTeX o Word preparadas al efecto y publicadas en la página web del máster correspondiente.

Todo el material no original, ya sea texto o figuras, deberá ser convenientemente citado y referenciado. En el caso de material complementario se deben respetar las licencias y copyrights asociados al software y hardware que se emplee. En caso contrario no se autorizará la defensa, sin menoscabo de otras acciones que correspondan.

## 1.1. Motivación

Introducción al tema del TFM.

## 1.2. Objetivos

Descripción de los objetivos del trabajo.

## 1.3. Plan de trabajo

Aquí se describe el plan de trabajo a seguir para la consecución de los objetivos descritos en el apartado anterior.

## 1.4. Explicaciones adicionales sobre el uso de esta plantilla

Si quieres cambiar el **estilo del título** de los capítulos, edita `TeXiS\TeXiS_pream.tex` y comenta la línea `\usepackage[Lenny]{fncychap}` para dejar el estilo básico de L<sup>A</sup>T<sub>E</sub>X.

Si no te gusta que no haya **espacios entre párrafos** y quieres dejar un pequeño espacio en blanco, no metas saltos de línea (`\\`) al final de los párrafos. En su lugar, busca el comando `\setlength{\parskip}{0.2ex}` en `TeXiS\TeXiS_pream.tex` y aumenta el valor de `0,2ex` a, por ejemplo, `1ex`.

TFMTeXiS se ha elaborado a partir de la plantilla de TeXiS<sup>1</sup>, creada por Marco Antonio y Pedro Pablo Gómez Martín para escribir su tesis doctoral. Para explicaciones más extensas y detalladas sobre cómo usar esta plantilla, recomendamos la lectura del documento `TeXiS-Manual-1.0.pdf` que acompaña a esta plantilla.

El siguiente texto se genera con el comando `\lipsum[2-20]` que viene a continuación en el fichero `.tex`. El único propósito es mostrar el aspecto de las páginas usando esta plantilla. Quita este comando y, si quieres, comenta o elimina el paquete `lipsum` al final de `TeXiS\TeXiS_pream.tex`

### 1.4.1. Texto de prueba

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet,

---

<sup>1</sup><http://gaia.fdi.ucm.es/research/texis/>

tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a,

dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus

orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.





## Estado de la Cuestión

En el estado de la cuestión es donde aparecen gran parte de las referencias bibliográficas del trabajo. Una de las formas más cómodas de gestionar la bibliografía en  $\text{\LaTeX}$  es utilizando **bibtex**. Las entradas bibliográficas deben estar en un fichero con extensión *.bib* (con esta plantilla se proporciona el fichero *biblio.bib*, donde están las entradas referenciadas más abajo).

### 2.1. Generación de Lenguaje Natural

¿Y si fuera posible ahorrar todo este tiempo de escritura de correos electrónicos?

Para lograr este propósito es imprescindible profundizar en la rama de la Inteligencia Artificial conocida como *Generación de Lenguaje Natural* (cuyas siglas son *NLG* por su nombre en inglés *Natural Language Generation*). Un buen ejemplo de aplicación de las técnicas de generación automática de textos son los 100.000 libros que Philip M. Parker puso a la venta en la plataforma *Amazon.com* incluyendo títulos de temáticas tan variadas como *El libro oficial del paciente sobre la estenosis espinal* (Parker, 2002), *Perspectivas mundiales de 2009 a 2014 de los envases de 60 miligramos de Fromage Frais* (Parker, 2008a), *Perspectivas de 2007 a 2012 de las tapetes de nudo, alfombras de baño y conjuntos que miden 6 pies por 9 pies o menos en la India* (Parker, 2006) y *Tesaurus Quechua - Inglés* (Parker, 2008b).

Resulta evidente que dicha cantidad de libros no pudieron ser escritos por Parker, sino que debió hacerse uso de técnicas de generación automática de textos. El algoritmo utilizado para dicho propósito, se engloba dentro de los métodos de generación conocidos como *text-to-text* (texto a texto en castellano), dado que este tipo de técnicas toman como entrada textos ya existentes (normalmente escritos a mano y no generados automáticamente) y producen un nuevo texto coherente como salida. Otras aplicaciones de este tipo de métodos son la traducción automática de un idioma a otro (Hutchins y Somers, 2009; Oettinger, 2013), el resumen automático de textos (Mani y Maybury, 2001; Nenkova y McKeown, 2011), la simplificación de textos complejos, ya sea para hacerlos más accesibles para un público de lectores de bajo nivel de alfabetización (Siddharthan, 2014; Bautista et al., 2011) o niños (Macdonald y Siddharthan, 2016), corrección automática de ortografía, gramática y texto (Kukich, 1992; Ng et al., 2014), generación automática de revisiones de artículos científicos (Bartoli et al., 2016), generación de paráfrasis dada una frase de entrada (Bannard y Callison-Burch, 2005), generación automática de preguntas con fines didácticos y educativos (Brown et al., 2005), generación automática de relatos dada una

descripción conceptual de la historia deseada (Gervás et al., 2004) o reescritura de textos (en concreto correos electrónicos) con estilo en función del destinatario (Moreno Morera, 2020).

Además de estos métodos *text-to-text*, existen los llamados *data-to-text* (datos a texto), en los cuales en lugar de recibir un texto como entrada, se genera el lenguaje a partir de datos. Estos pueden ser de todo tipo para dar lugar a informes o resúmenes como pueden ser de índole climatológica (Goldberg et al., 1994; Ramos-Soto et al., 2014), financiera (Plachouras et al., 2016), ingenieril, como por ejemplo el trabajo desarrollado por Yu et al. (2007) para generar resúmenes de datos recopilados por sensores en turbinas de gas, sanitaria (Hüske-Kraus, 2003; Banaee et al., 2013), como la investigación llevada a cabo por Portet et al. (2009) para obtener informes textuales a partir de datos de cuidados intensivos neonatales, o, incluso, deportivos (Theune et al., 2001; Chen y Mooney, 2008). Además de informes o resúmenes, también se utilizan los métodos *data-to-text* para otros propósitos como la composición de discursos narrativos para relatos de varios personajes a partir de partidas de ajedrez (Gervás, 2014), redacción de periódicos electrónicos a partir de datos de sensores (Molina et al., 2011), generación de texto que aborda problemas medioambientales como el seguimiento de la fauna (Siddharthan et al., 2012; Ponnampertuma et al., 2013), la información medioambiental personalizada (Wanner et al., 2015) y la mejora del compromiso de los ciudadanos científicos a través de los comentarios generados (Van der Wal et al., 2016) o producción de información interactiva sobre artefactos culturales (Stock et al., 2007), entre otros.

Debido a que el objetivo de este trabajo se centra en la generación de correos electrónicos a partir del asunto, exploraremos en detalle las técnicas de Generación de Lenguaje Natural y, en especial, los métodos *text-to-text*. Para profundizar en los algoritmos y arquitecturas empleados ante los problemas de tipo *data-to-text*, conviene consultar la investigación llevada a cabo por Gatt y Krahmer (2018), en la cual muestran el estado del arte de los trabajos realizados en este ámbito.

### 2.1.1. ¿Qué es la Generación de Lenguaje Natural?

Dado que tanto los sistemas *text-to-text* como *data-to-text* y todas sus aplicaciones mencionadas anteriormente pertenecen a la rama de Generación de Lenguaje Natural, esta no debe definirse en función de la entrada del sistema, sino en la salida. Según Reiter y Dale (2000) la NLG es la conceptualización del “campo de la inteligencia artificial y la lingüística computacional que se centra en los sistemas informáticos que son capaces de producir textos comprensibles en inglés u otra lengua humana. [...] Como área de investigación, la NLG presenta una perspectiva única ante problemas fundamentales de la inteligencia artificial, la ciencia cognitiva y la interacción. Estos incluyen cuestiones como por ejemplo cómo deben ser representados y cómo debe razonarse con la lingüística y el dominio del conocimiento, qué significa que un texto esté correctamente redactado y cómo es la mejor forma de comunicar información entre las computadoras y los usuarios.” Por lo tanto, la Generación de Lenguaje Natural se puede definir como el ámbito que engloba el estudio de la producción de lenguaje no artificial, así como el diseño e implementación de algoritmos y sistemas computacionales cuyo resultado debe ser un texto que imite la forma en que los humanos se comunican verbalmente (Vicente et al., 2015), ya sea oralmente o por escrito (del Socorro Bernardos, 2007). Es decir, independientemente de la entrada recibida, se precisa el significado de NLG a partir de la salida esperada por el problema planteado. Tanto es así, que, como hemos visto, la entrada del sistema puede variar excesivamente (McDonald, 1993): desde textos (que son precisamente los sistemas *text-to-text*) hasta

datos de todo tipo como partidas de ajedrez (Gervás, 2014), pictogramas (González Álvarez y López Pulido, 2019) e, incluso, vídeos (Thomason et al., 2014). Sin embargo, autores como Dušek et al. (2020) acotan la definición de los sistemas de NLG estableciendo que la entrada deben ser representaciones semánticas, obviando así la primera tarea de la arquitectura propuesta por Reiter y Dale (2000) conocida como macro planificación o determinación del contenido (se explicará en la sección **poner n<sup>o</sup>**), que es precisamente el punto en el que se generan dichas representaciones semánticas.

Cabe destacar que, aunque desde un principio hayamos diferenciado entre métodos text-to-text y data-to-text, ni los límites entre las dos aproximaciones ni la pertenencia de algunas técnicas a ellas se encuentran claramente definidos. Un ejemplo de ello podemos encontrarlo en la generación automática de resúmenes de textos. En principio se caracterizaría claramente como un sistema text-to-text. No obstante, al hacer frente a este problema se han desarrollado soluciones con las conocidas técnicas abstractivas (Genest y Lapalme, 2011), que, como explican Hahn y Mani (2000), a diferencia de los métodos de extracción evitan recoger las frases completas y se limitan a tomar unidades semánticas. Este tipo de técnicas usadas, por ejemplo, en la obtención de opiniones de reseñas para la posterior generación de frases nuevas (Labbé y Portet, 2012), también provienen de problemas data-to-text. A la inversa, un sistema data-to-text puede hacer uso de técnicas que principalmente son utilizadas en los casos de uso text-to-text (McIntyre y Lapata, 2009; Kondadadi et al., 2013). Por otro lado, podría parecer que los métodos de *deep learning* (Goodfellow et al., 2016) deben ser mayoritariamente utilizados en los problemas data-to-text. Sin embargo, gracias al trabajo de Mikolov et al. (2013), se han desarrollado extensamente esta clase de soluciones para la NLG con gran variedad de arquitecturas como las redes neuronales recurrentes (Cho et al., 2014; Tang et al., 2016) muy a menudo combinadas con la memoria a corto plazo o LSTM (Chen et al., 2016).



# Capítulo 3

## Descripción del Trabajo

Aquí comienza la descripción del trabajo realizado. Se deben incluir tantos capítulos como sea necesario para describir de la manera más completa posible el trabajo que se ha llevado a cabo. Como muestra la figura 3.1, está todo por hacer.



Figura 3.1: Ejemplo de imagen

Si te sirve de utilidad, puedes incluir tablas para mostrar resultados, tal como se ve en la tabla 3.1.

| Col 1 | Col 2 | Col 3 |
|-------|-------|-------|
| 3     | 3.01  | 3.50  |
| 6     | 2.12  | 4.40  |
| 1     | 3.79  | 5.00  |
| 2     | 4.88  | 5.30  |
| 4     | 3.50  | 2.90  |
| 5     | 7.40  | 4.70  |

Tabla 3.1: Tabla de ejemplo

### 3.1. Enron corpus

Para llevar a cabo este trabajo, se ha elegido el corpus conocido como Enron<sup>1</sup>, dado que los correos electrónicos que contiene pertenecieron a la empresa con el mismo nombre. Precisamente se hicieron públicos tras una investigación legal llevada a cabo a esta compañía por parte de la Comisión Federal de Regulación de la Energía<sup>2</sup> de Estados Unidos.

Enron corpus contiene 517.401 correos electrónicos de 150 usuarios distintos. Además de la ventaja de la gran cantidad de elementos pertenecientes a este dataset, también ha sido elegido por encontrar diversos trabajos sobre este mismo conjunto de e-mails, como el llevado a cabo por Klimt y Yang (2004).

---

<sup>1</sup><http://www-2.cs.cmu.edu/~enron/>

<sup>2</sup><https://www.ferc.gov/>

## Capítulo 4

# Conclusiones y Trabajo Futuro

Conclusiones del trabajo y líneas de trabajo futuro.

Antes de la entrega de actas de cada convocatoria, en el plazo que se indica en el calendario de los trabajos de fin de máster, el estudiante entregará en el Campus Virtual la versión final de la memoria en PDF. En la portada de la misma deberán figurar, como se ha señalado anteriormente, la convocatoria y la calificación obtenida. Asimismo, el estudiante también entregará todo el material que tenga concedido en préstamo a lo largo del curso.





# Chapter 5

## Introduction

Introduction to the subject area. This chapter contains the translation of Chapter 1.



# Chapter 6

## Conclusions and Future Work

Conclusions and future lines of work. This chapter contains the translation of Chapter 4.



# Bibliografía

*Y así, del mucho leer y del poco dormir, se le  
secó el cerebro de manera que vino a perder el  
juicio.*

*(modificar en Cascaras\bibliografia.tex)*

Miguel de Cervantes Saavedra

- BANAEI, H., AHMED, M. U. y LOUTFI, A. Towards nlg for physiological data monitoring with body area networks. En *14th European Workshop on Natural Language Generation, Sofia, Bulgaria, August 8-9, 2013*, páginas 193–197. 2013.
- BANNARD, C. y CALLISON-BURCH, C. Paraphrasing with bilingual parallel corpora. En *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, páginas 597–604. 2005.
- BARTOLI, A., DE LORENZO, A., MEDVET, E. y TARLAO, F. Your paper has been accepted, rejected, or whatever: Automatic generation of scientific paper reviews. En *International Conference on Availability, Reliability, and Security*, páginas 19–28. Springer, 2016.
- BAUTISTA, S., LEÓN, C., HERVÁS, R. y GERVÁS, P. Empirical identification of text simplification strategies for reading-impaired people. En *Everyday Technology for Independence and Care*, páginas 567–574. IOS Press, 2011.
- BROWN, J., FRISHKOFF, G. y ESKENAZI, M. Automatic question generation for vocabulary assessment. En *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, páginas 819–826. 2005.
- CHEN, D. L. y MOONEY, R. J. Learning to sportscast: a test of grounded language acquisition. En *Proceedings of the 25th international conference on Machine learning*, páginas 128–135. 2008.
- CHEN, Q., ZHU, X., LING, Z., WEI, S., JIANG, H. y INKPEN, D. Enhanced lstm for natural language inference. *arXiv preprint arXiv:1609.06038*, 2016.
- CHO, K., VAN MERRIËNBOER, B., GULCEHRE, C., BAHDANAU, D., BOUGARES, F., SCHWENK, H. y BENGIO, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

- DUŠEK, O., NOVIKOVA, J. y RIESER, V. Evaluating the state-of-the-art of end-to-end natural language generation: The e2e nlg challenge. *Computer Speech & Language*, vol. 59, páginas 123–156, 2020.
- GATT, A. y KRAHMER, E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, vol. 61, páginas 65–170, 2018.
- GENEST, P.-E. y LAPALME, G. Framework for abstractive summarization using text-to-text generation. En *Proceedings of the workshop on monolingual text-to-text generation*, páginas 64–73. 2011.
- GERVÁS, P. Composing narrative discourse for stories of many characters: a case study over a chess game. *Literary and Linguistic Computing*, vol. 29(4), páginas 511–531, 2014.
- GERVÁS, P., DÍAZ-AGUDO, B., PEINADO, F. y HERVÁS, R. Story plot generation based on cbr. En *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, páginas 33–46. Springer, 2004.
- GOLDBERG, E., DRIEDGER, N. y KITTREDGE, R. I. Using natural-language processing to produce weather forecasts. *IEEE Expert*, vol. 9(2), páginas 45–53, 1994.
- GONZÁLEZ ÁLVAREZ, S. y LÓPEZ PULIDO, J. M. Traductor de pictogramas a texto. 2019.
- GOODFELLOW, I., BENGIO, Y. y COURVILLE, A. *Deep learning*. MIT press, 2016.
- HAHN, U. y MANI, I. The challenges of automatic summarization. *Computer*, vol. 33(11), páginas 29–36, 2000.
- HÜSKE-KRAUS, D. Text generation in clinical medicine—a review. *Methods of information in medicine*, vol. 42(01), páginas 51–60, 2003.
- HUTCHINS, W. J. y SOMERS, H. L. *An introduction to machine translation*. 2009.
- KLIMT, B. y YANG, Y. Introducing the enron corpus. En *CEAS*. 2004.
- KONDADADI, R., HOWALD, B. y SCHILDER, F. A statistical nlg framework for aggregated planning and realization. En *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, páginas 1406–1415. 2013.
- KUKICH, K. Techniques for automatically correcting words in text. *Acm Computing Surveys (CSUR)*, vol. 24(4), páginas 377–439, 1992.
- LABBÉ, C. y PORTET, F. Towards an abstractive opinion summarisation of multiple reviews in the tourism domain. En *The First International Workshop on Sentiment Discovery from Affective Data (SDAD 2012)*, páginas 87–94. 2012.
- MACDONALD, I. y SIDDHARTHAN, A. Summarising news stories for children. ACL, 2016.
- MANI, I. y MAYBURY, M. T. Automatic summarization. 2001.
- MCDONALD, D. D. Issues in the choice of a source for natural language generation. *Computational Linguistics*, vol. 19(1), páginas 191–197, 1993.

- MCINTYRE, N. y LAPATA, M. Learning to tell tales: A data-driven approach to story generation. En *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, páginas 217–225. 2009.
- MIKOLOV, T., CHEN, K., CORRADO, G. y DEAN, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- MOLINA, M., STENT, A. y PARODI, E. Generating automated news to explain the meaning of sensor data. En *International Symposium on Intelligent Data Analysis*, páginas 282–293. Springer, 2011.
- MORENO MORERA, C. Un modelo de análisis estilométrico de correos electrónicos para la redacción personalizada basada en el destinatario. 2020.
- NENKOVA, A. y MCKEOWN, K. *Automatic summarization*. Now Publishers Inc, 2011.
- NG, H. T., WU, S. M., BRISCOE, T., HADIWINOTO, C., SUSANTO, R. H. y BRYANT, C. The conll-2014 shared task on grammatical error correction. En *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, páginas 1–14. 2014.
- OETTINGER, A. G. *Automatic language translation*. Harvard University Press, 2013.
- PARKER, P. M. *The Official Patient's Sourcebook on Spinal Stenosis*. Icon Group International Incorporated, 2002.
- PARKER, P. M. *The 2007-2012 Outlook for Tufted Washable Scatter Rugs, Bathmats, and Sets That Measure 6-Foot by 9-Foot or Smaller in India*. Icon Group International Incorporated, 2006.
- PARKER, P. M. *The 2009-2014 World Outlook for 60-milligram Containers of Fromage Frais*. Icon Group International Incorporated, 2008a.
- PARKER, P. M. *Webster's Quechua - English Thesaurus Dictionary*. Icon Group International Incorporated, 2008b.
- PLACHOURAS, V., SMILEY, C., BRETZ, H., TAYLOR, O., LEIDNER, J. L., SONG, D. y SCHILDER, F. Interacting with financial data using natural language. En *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, páginas 1121–1124. 2016.
- PONNAMPERUMA, K., SIDDHARTHAN, A., ZENG, C., MELLISH, C. y VAN DER WAL, R. Tag2blog: Narrative generation from satellite tag data. En *Proceedings of the 51st annual meeting of the association for computational linguistics: System demonstrations*, páginas 169–174. 2013.
- PORTET, F., REITER, E., GATT, A., HUNTER, J., SRIPADA, S., FREER, Y. y SYKES, C. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, vol. 173(7-8), páginas 789–816, 2009.
- RAMOS-SOTO, A., BUGARIN, A. J., BARRO, S. y TABOADA, J. Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data. *IEEE Transactions on Fuzzy Systems*, vol. 23(1), páginas 44–57, 2014.

- REITER, E. y DALE, R. *Building Natural Language Generation Systems*. Cambridge University Press, 2000. ISBN 9780521620369.
- SIDDHARTHAN, A. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, vol. 165(2), páginas 259–298, 2014.
- SIDDHARTHAN, A., GREEN, M. J., VAN DEEMTER, K., MELLISH, C. S. y VAN DER WAL, R. Blogging birds: Generating narratives about reintroduced species to promote public engagement. En *Proceedings of the 7th International Natural Language Generation Conference (INLG 2012)*. ACL Anthology, 2012.
- DEL SOCORRO BERNARDOS, M. ¿qué es la generación de lenguaje natural? una visión general sobre el proceso de generación. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, vol. 11(34), páginas 105–128, 2007.
- STOCK, O., ZANCANARO, M., Busetta, P., CALLAWAY, C., KRÜGER, A., KRUPPA, M., KUFLIK, T., NOT, E. y ROCCHI, C. Adaptive, intelligent presentation of information for the museum visitor in peach. *User Modeling and User-Adapted Interaction*, vol. 17(3), páginas 257–304, 2007.
- TANG, J., YANG, Y., CARTON, S., ZHANG, M. y MEI, Q. Context-aware natural language generation with recurrent neural networks. *arXiv preprint arXiv:1611.09900*, 2016.
- THEUNE, M., KLABBERS, E., DE PIJPER, J.-R., KRAHMER, E. y ODIJK, J. From data to speech: a general approach. *Natural Language Engineering*, vol. 7(1), páginas 47–86, 2001.
- THOMASON, J., VENUGOPALAN, S., GUADARRAMA, S., SAENKO, K. y MOONEY, R. Integrating language and vision to generate natural language descriptions of videos in the wild. En *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, páginas 1218–1227. 2014.
- VICENTE, M., BARROS, C., PEREGRINO, F. S., AGULLÓ, F. y LLORET, E. La generación de lenguaje natural: análisis del estado actual. *Computación y Sistemas*, vol. 19(4), páginas 721–756, 2015.
- VAN DER WAL, R., SHARMA, N., MELLISH, C., ROBINSON, A. y SIDDHARTHAN, A. The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology*, vol. 30(3), páginas 550–561, 2016.
- WANNER, L., BOSCH, H., BOUAYAD-AGHA, N., CASAMAYOR, G., ERTL, T., HILBRING, D., JOHANSSON, L., KARATZAS, K., KARPPINEN, A., KOMPATSIARIS, I. ET AL. Getting the environmental information across: from the web to the user. *Expert Systems*, vol. 32(3), páginas 405–432, 2015.
- YU, J., REITER, E., HUNTER, J. y MELLISH, C. Choosing the content of textual summaries of large time-series data sets. *Natural Language Engineering*, vol. 13(1), páginas 25–49, 2007.



Apéndice **A**

Título del Apéndice A

Contenido del apéndice



|          |          |
|----------|----------|
| Apéndice | <b>B</b> |
|----------|----------|

Título del Apéndice B



Este texto se puede encontrar en el fichero Cascaras/fin.tex. Si deseas eliminarlo, basta con comentar la línea correspondiente al final del fichero TFMTeXiS.tex.

*–¿Qué te parece desto, Sancho? – Dijo Don Quijote –  
Bien podrán los encantadores quitarme la ventura,  
pero el esfuerzo y el ánimo, será imposible.*

*Segunda parte del Ingenioso Caballero  
Don Quijote de la Mancha  
Miguel de Cervantes*

*–Buena está – dijo Sancho –; firmela vuestra merced.  
–No es menester firmarla – dijo Don Quijote–,  
sino solamente poner mi rúbrica.*

*Primera parte del Ingenioso Caballero  
Don Quijote de la Mancha  
Miguel de Cervantes*

