



TRABAJO FIN DE MÁSTER
MÁSTER EN INGENIERÍA INFORMÁTICA

Titulo del Proyecto

Subtitulo del Proyecto

Autor

Andrea Morales Garzón (alumno)

Directores

Nombre Apellido1 Apellido2 (tutor1)

Nombre Apellido1 Apellido2 (tutor2)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, 31 de marzo de 2020



Título del proyecto

Subtítulo del proyecto.

Autor

Nombre Apellido1 Apellido2 (alumno)

Directores

Nombre Apellido1 Apellido2 (tutor1)

Nombre Apellido1 Apellido2 (tutor2)

Título del Proyecto: Subtítulo del proyecto

Andrea Morales Garzón

Palabras clave: palabra_clave1, palabra_clave2, palabra_clave3,

Resumen

Poner aquí el resumen.

Project Title: Project Subtitle

Andrea, Morales Garzón

Keywords: Keyword1, Keyword2, Keyword3,

Abstract

Write here the abstract in English.

Yo, **Andrea Morales Garzón**, alumno de la titulación Máster en Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 77147632-C, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Máster en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Andrea Morales Garzón

Granada a 31 de marzo de 2020.

D. **Nombre Apellido1 Apellido2 (tutor1)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

D. **Nombre Apellido1 Apellido2 (tutor2)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***Título del proyecto, Subtítulo del proyecto***, ha sido realizado bajo su supervisión por **Andrea Morales Garzón**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 31 de marzo de 2020.

Los directores:

Nombre Apellido1 Apellido2 (tutor1) **Nombre Apellido1 Apellido2 (tutor2)**

Agradecimientos

Poner aquí agradecimientos...

Índice general

1. Introducción	1
1.1. Introducción	1
1.2. Concepto de Food Computing	2
1.3. Problemas a resolver con Food Computing	3
1.3.1. Reconocimiento de imágenes	3
1.3.2. Recomendación	3
1.3.3. Predicción	4
1.4. Nuestro problema a resolver	6
1.5. Objetivos	6
1.6. Contenido de la memoria en relación a los objetivos	6
2. Planificación	7
2.1. Planificación inicial	7
2.1.1. Implementación del modelo predictivo	7
2.1.2. Desarrollo de la aplicación	7
2.1.3. Redacción y desarrollo de la memoria	7
2.2. Planificación final	8
2.2.1. Implementación del modelo predictivo	8
2.2.2. Desarrollo de la aplicación	8
2.2.3. Redacción y desarrollo de la memoria	8
3. Revisión bibliográfica	9
3.1. Revisión bibliográfica	9
3.2. Tratamiento de datos textuales en nutrición	9
4. Herramienta a desarrollar	11
4.1. Descripción de la herramienta	11
4.2. Recetas	11
4.3. Adaptación de recetas a restricciones alimenticias	12
4.3.1. Detección de ingredientes	12
4.3.2. Restringir ingredientes sujetos a una restricción	12
4.3.3. Sustitución de ingredientes	12

5. Motor de sustitución de ingredientes	13
5.1. Descripción	13
5.2. Datos utilizados	13
5.2.1. Recopilación de recetas	13
5.2.2. Preprocesamiento de los datos textuales	13
5.3. Word Embedding	13
5.3.1. Entrenamiento del modelo	13
5.3.2. Funcionamiento	14
5.4. Mapping entre ingredientes	14
5.4.1. Funciones de distancia	14
5.4.2. Ejemplos	14
6. Capítulo 6	15
6.1. Capítulo 6	15
7. Pruebas	17
7.1. Pruebas	17
8. Conclusiones	19
8.1. Conclusiones	19

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

Resumen: Esta introducción abarca los límites del trabajo a desarrollar, partiendo de los estudios consultados en la literatura.

1.1. Introducción

La alimentación es esencial para el desarrollo de la vida humana. Nuestra alimentación, así como nuestros hábitos alimenticios, tienen un impacto directo con nuestra salud, así como con nuestra calidad de vida. Numerosos artículos de referencia muestran que existe una tendencia en la población en general, hacia mayores índices de obesidad, diabetes u otras enfermedades relacionadas con aspectos nutricionales, y estas tienen una clara relación con nuestros hábitos alimenticios. Factores como pueden ser los socio-económicos, han provocado que la población vaya evolucionando cada vez más hacia hábitos alimenticios poco saludables. Incluso en aquellas zonas donde la cultura gastronómica consta de características saludables (como puede ser la dieta mediterránea), hoy en día existe cada vez menos adherencia a este tipo de dietas, siendo sustituida por distintos modelos nutricionales menos recomendables como consecuencia de los cambios y la evolución social.

Por otra parte, la llamada *occidentalización* de las dietas es cada vez más frecuente, sobretudo en la gente joven, siendo el origen, entre muchos otros factores, de un aumento (y en crecimiento) de la obesidad en la población en distintas zonas globales.

Estos problemas, unidos al auge del uso de la tecnología en la vida diaria, han sido uno de los puntos de partida del desarrollo de aplicaciones relacionadas con la nutrición, así como con el bienestar, dando lugar a desmesuradas cantidades de datos que, con las técnicas adecuadas y un correcto tratamiento y análisis de los datos, puede dar lugar a mejor comprensión de las tendencias y el funcionamiento en este área.

1.2. Concepto de Food Computing

Desde que se comenzó a desarrollar y abrir un campo en el mundo de la alimentación y la nutrición, se ha tratado este amplio campo en diversas áreas de estudio, como pueden ser la salud, la nutrición y el bienestar, el consumo de comida, aspectos socioeconómicos y su impacto, el impacto medioambiental o la multiculturalidad.

Sin embargo, en un principio, estas áreas previamente mencionadas se centraron en los datos disponibles en ese momento, los cuáles se podían resumir en cuestionarios, recopilaciones de recetas, menús, y otro tipo de metodologías a pequeña escala. Con el desarrollo de la tecnología en los últimos años, ha dado lugar a un fenómeno alimenticio, centrado en los hábitos y costumbres alimenticias. El mundo totalmente-conectado, el Internet de las cosas, las redes sociales, y la *necesidad* de los usuarios a transmitir y compartir actividades tan rutinarias como la alimentación que siguen día a día, así como la reciente digitalización de las empresas en este sector ha dado lugar a la generación de una cantidad ingente de datos referentes al mundo de la alimentación de todo tipo de naturaleza, desde las dietas que siguen usuarios de todo el mundo, hasta los restaurantes que visitan, preferencias alimenticias, consecuencias en la salud, enfermedades relacionadas con nuestros hábitos alimenticios, etc.

El desarrollo, así como la tendencia de hoy en día de tecnologías como *Big Data* y su aplicación en la interpretación y comprensión de cantidades desmesuradas de datos también han llegado a este sector, ya que el llamado *Food data* conlleva de manera intrínseca una cantidad de conocimiento que, de otra manera, podría ser complejo o incluso imposible de conseguir. En este contexto, se introduce el concepto de *Food Computing*, y abarca la totalidad de técnicas y modalidades que abarca, desde la adquisición de datos dentro de este contexto, como el análisis, técnicas y sistemas que los utilizan con el objetivo de desarrollar herramientas que mejoren la calidad de vida de la población, así como entender de manera más profunda el comportamiento humano en lo que a esta área concierne. Abarca por tanto, más campos allá de la informática, como puede ser la medicina, la nutrición, o incluso la psicología.

En términos generales, aquellas tareas llevadas a cabo dentro del cuadro del Food Computing, parten de un procedimiento de extracción de información, cuyo análisis permite su aplicación en tareas como pueden ser labores de reconocimiento, percepción, recomendación, predicción, monitorización o incluso recuperación de información. Estas tareas de Food Computing, que se verán más en detalle en la siguiente sección, tienen presencia en distintas áreas, como puede ser la salud, la cultura, la agricultura, biología, medicina, etc.

1.3. Problemas a resolver con Food Computing

1.3.1. Reconocimiento de imágenes

Centrado principalmente en tareas relativas al reconocimiento de alimentos (los llamados *item food*, así como de la categoría o de los ingredientes de los mismos. Este campo está centrado principalmente en imágenes de alimentos.

1.3.2. Recomendación

Los sistemas de recomendación forman una de las áreas más explotadas en el campo de la alimentación por los últimos años. Conlleva el desafío de información compleja y polifacética, y esto mismo es lo que la diferencia de las tareas de recomendación centradas en otro tipo de áreas, donde puedan existir distintos estándares o componentes más objetivas y, por tanto, sencillas de calcular e interpretar. Principalmente, han abarcado dos vías principales que merece la pena destacar. Por una parte, la recomendación basada en las preferencias del usuario en cuestión, teniendo en cuenta para ello sus gustos, sus rutinas, así como patrones en su alimentación que a simple vista puedan ser más complicados de detectar. Por otra parte, se le suma a estos sistemas de recomendación la inmersión del mundo de la nutrición y vida sano, dando lugar a sistemas de recomendación basados en el usuario, pero que aún así conllevan una fuerte componente saludable.

Está muy ligada a las tareas de predicción, sobretodo, para la parte relativa a la recomendación en función de las preferencias del usuario, ya que suele llevar asociado tareas predictivas a partir de datos procedentes de redes sociales (en su mayoría). Al hablar de recomendación en el contexto de Food Computing, podemos diferenciar cuatro tipos principales [Trattner2017]:

- *Recomendación basada en el contenido.* En este caso concreto, los sistemas de recomendación en la literatura se basan principalmente en sistemas de recomendación basados en recetas, en función de la puntuación y opinión que los usuarios tienen de los ingredientes que las forman [Freyne:2010:IFP:1719970.1720021][a3e625bf40904a799c3b8e35929388b7].
- *Recomendación basada en filtrado colaborativo.* Técnicas clásicas basadas en filtrado colaborativo también han sido empleadas en este sector, como puede ser el uso del coeficiente de Pearson con matrices de puntuaciones [Freyne:2010:IFP:1719970.1720021] para la generación de vecindarios, así como otras más sofisticadas, como puede ser el uso de matrices ponderadas en [Griffiths'gibbsampling].

- *Recomendación híbrida.* Se han propuesto sistemas de recomendación híbrida, principalmente orientados a la tarea de recomendación de recetas. En [Freyne:2010:IFP:1719970.1720021] se propuso un sistema de recomendación combinando métodos de filtrado colaborativo con técnicas basadas en el contenido. Otro sistema de recomendación híbrido conocido fue el propuesto por [10.1007/978-3-319-02432-5˙19], que obtuvo resultados con muy buenas aproximaciones, combinando técnicas de SVD tanto con usuarios como con alimentos.
- *Recomendación dependiente del contexto,* fundamentalmente en base al género, tiempo, aficiones, localización u otros aspectos culturales relacionados con los usuarios. También hay distintos sistemas de recomendación que tienen en cuenta distintas combinaciones de sabores, o incluso patrones de combinación de ingredientes en las distintas recetas, y otros basados en otro tipo de datos que en principio puedan parecer menos relevantes, como la rutina diaria de usuarios obtenida a partir de redes sociales como *Twitter*. Al fin y al cabo, se centran en la idea de que para que un sistema de recomendación sea efectivo, deben tenerse en cuenta los patrones en la alimentación de los usuarios, pero también otras cuestiones culturales relacionadas con la vida diaria de los usuarios[Rokicki2016PlateAP].

Recomendación de dietas saludables

Por otra parte, la incorporación de aspectos saludables en los sistemas de recomendación de dietas tiene especial importancia, y, debido a la gran cantidad de literatura centrada en la incorporación de este factor a los sistemas de recomendación, merece ser diferenciada en una sección independiente. En los últimos años, se ha producido un auge en el desarrollo de sistemas de recomendación centrados en la generación de dietas personalizadas teniendo como requisito que sean saludables[Trattner2017].

1.3.3. Predicción

Como es de esperar, la gran cantidad de datos que se produce en este área, posiciona un foco de atracción por parte de técnicas predictivas en este área. Las redes sociales han tenido una gran presencia en las labores de predicción en este área, debido a la gran cantidad de información generada. En [Fried2015][Abbar2015] se utilizaron datos de usuarios de Twitter con el objetivo de intentar predecir valores de sobrepeso y diabetes en la población.

Otro campo de estudio en el que se han centrado diferentes labores de predicción ha sido en los aspectos multiculturales relativos al mundo cu-

linario, analizando datos de distintas zonas geográficas, así como a nivel mundial [Min2018][Sajadmanesh2019]. También se han realizado patrones de combinación de ingredientes en distintas regiones [Bossard2014], así como para la búsqueda de equivalencias de cocina regional de unas zonas geográficas a otras totalmente dispares[10.3389/fct.2018.00014].

Por otra parte, las Redes Neuronales Artificiales también tienen presencia en el mundo culinario, abarcando principalmente tres vías, que se podrían resumir en *predicción de parámetros, clasificación y estudios de calidad*.

Multitud de parámetros, sobretodo relativos a los alimentos, se han estudiado desde el punto de vista de la predicción. En concreto, se ha hecho especial hincapié en la evaluación de propiedades de alimentos, sobretodo parámetros relacionados con componentes bioactivos y características psicoquímicas de los mismos. En [Correa2018] se recopilan algunos de los estudios llevados a cabo en este ámbito, como la evaluación de las propiedades antioxidantes de los aceites o el determinar la tasa de fermentación de las semillas de cacao, en función de medidas de aminoácidos y de cambios de color en los alimentos.

Respecto al estudio de la calidad de los alimentos, estos han tenido un gran protagonismo en las tareas predictivas realizadas en este campo. Se han realizado estudios de predicción realizados en base a monitorización con tratamientos térmicos, como vía a determinar estados de buena calidad en productos alimenticios. También se pueden consultar estudios acerca de la conductividad térmica de productos de panadería en Nigeria, así como de patrones en base al nivel de deshidratación de frutas, estableciendo una relación entre ésta y la pérdida de agua, que lleva a la solidificación del alimento. También se han realizado estudios que buscan predecir la relación entre la carga de bacterias y la concentración de las mismas, en determinados vegetales, que estudia este tipo de relaciones en el tomate y en las hojas de lechuga [Correa2018].

Otro campo explotado desde estas técnicas ha sido el de la seguridad alimentaria. En [Correa2018] se presenta un estudio que busca determinar el tiempo de caducidad de determinados productos, así como también se ha buscado predecir el estado ideal de refrigeración de platos cocinados, o incluso la calidad de los alimentos en función de la conservación en frío de los mismos. En este área, también entran estudios relativos a la detección de enfermedades en plantas (en vista a la detección de brotes). En [towards-food-security-ia] estudian la detección de la enfermedad *Blast* en la hoja del arroz a partir de procesamiento de imágenes y *Back Propagation*.

Estas técnicas también han cobrado protagonismo en lo que concierne a tareas de **clasificación**. En [Correa2018] se propone una metodología para clasificar distintos tipos de aceite, de vinagre, o incluso de variedades de

queso, mientras que en otras vías se han estudiado clasificadores de grupos alimenticios, a partir de parámetros como pueden ser la claridad, el color, el grado de fermentación o incluso la acidez. Asimismo, las Series Temporales también tienen cabida en este sector, aunque principalmente han tenido presencia desde un punto de vista médico. En [bellaci-biomedical] se hace uso de estas técnicas para interpretar datos procedentes de la monitorización de pacientes con diabetes, intentando evaluar la salida obtenida a partir de un tipo de terapia concreta. Otros estudios relacionadas con series temporales en el sector alimentario se centran en el estudio de los precios u otras características económicas dentro del mundo de la comida [Zou20072913].

Sin embargo, a pesar de los grandes avances que se han hecho en los últimos años en este área, la complejidad intrínseca en los datos, así como las relaciones entre ellos, hacen que realmente este tipo de tareas tenga éxito en entornos muy controlados y restringidos, y para una cantidad concreta de items. En escenarios reales, la variedad de platos y de factores a considerar es tan grande que, junto con la falta de estandarización en este campo, hace inevitable el tener que acotar el problema, reduciendo así las probabilidades de éxito que tendría su aplicación en otras fuentes externas.

1.4. Nuestro problema a resolver

1.5. Objetivos

1. *Hacer estado del arte en predicción*
2. *Elaborar algoritmo de predicción*
3. *Buscar datos*
4. *Experimentación*

1.6. Contenido de la memoria en relación a los objetivos

Capítulo 2

Planificación

Resumen: Este capítulo presenta

2.1. Planificación inicial

2.1.1. Implementación del modelo predictivo

Tarea	Duración	Fecha tope
Diseño del modelo		
Implementación		
Pruebas y mejoras		

2.1.2. Desarrollo de la aplicación

Tarea	Duración	Fecha tope
Diseño de la aplicación		
Implementación: interfaz gráfica		
Implementación: inclusión del motor predictivo		

2.1.3. Redacción y desarrollo de la memoria

Tarea	Duración	Fecha tope

2.2. Planificación final

2.2.1. Implementación del modelo predictivo

Tarea	Duración	Fecha tope
Diseño del modelo		
Implementación		
Pruebas y mejoras		

2.2.2. Desarrollo de la aplicación

Tarea	Duración	Fecha tope
Diseño de la aplicación		
Implementación: interfaz gráfica		
Implementación: inclusión del motor predictivo		

2.2.3. Redacción y desarrollo de la memoria

Tarea	Duración	Fecha tope

Capítulo 3

Revisión bibliográfica

Resumen: Este capítulo presenta

3.1. Revisión bibliográfica

¿Todo esto en la introducción? ¿O extenderme aquí?

3.2. Tratamiento de datos textuales en nutrición

- Ontologías
- Aplicación de datos textuales en Food Computing
- Identificación de items automática (aplicado a la nutrición)
- Fuentes de datos principales, BD de referencia...

Capítulo 4

Herramienta a desarrollar

Resumen: Este capítulo presenta

4.1. Descripción de la herramienta

Aplicación en la que puedes seleccionar una receta, aplicar una restricción y obtener sustitutos de ingredientes aptos para restricciones alimenticias concretas.

4.2. Recetas

- Origen de las recetas, carga de las recetas en la aplicación, y almacenamiento.
- Descripción de las recetas. ¿Qué atributos tiene? ¿Composición nutricional, etc?
- ¿Cómo se procesa su información? ¿Cómo se obtienen los ingredientes? ¿Y sus cantidades y demás?

Recordatorio: he escogido USDA porque me permite obtener la cantidad en gramos de ingredientes (normalmente dados en cup, tablespoons...)

4.3. Adaptación de recetas a restricciones alimenticias

4.3.1. Detección de ingredientes

- Obtener ingredientes a partir de las recetas
- Determinar la cantidad y la medida de ingrediente para poder obtener su equivalente en gramos (permite cálculos de valores nutricionales)
- Determinar el texto que se corresponde únicamente con el ingrediente
- **Aplicar preprocesamiento a los campos textuales correspondientes con la descripción del ingrediente (esto no está implementado pero habría que mirarlo).**

4.3.2. Restringir ingredientes sujetos a una restricción

Aplicar la restricción a los diferentes ingredientes de la base de datos de ingredientes que estemos utilizando. De esta forma se limita la cantidad de posibles alimentos que se pueden mapear.

Concretar cuando se sepa concretamente la restricción a aplicar.

4.3.3. Sustitución de ingredientes

Realizar el mapping hacia el subset de ingredientes resultante de la sección previa. Para la identificación de este ingrediente, se utilizará una herramienta de mapping de alimentos desarrollada con técnicas predictivas, la cuál se explica detalladamente en el siguiente capítulo.

Para ello, se buscará, en la base de datos, el ingrediente más similar al proporcionado, basándonos en sus campos textuales. Con la herramienta desarrollada, calcularemos la distancia entre representaciones numéricas de las descripciones textuales.

Capítulo 5

Motor de sustitución de ingredientes

Resumen: Este capítulo presenta

5.1. Descripción

¿Dónde hablar de los n-gramas que se implementó al principio?
¿Separar word embedding y cálculo de la distancia en dos capítulos?

5.2. Datos utilizados

5.2.1. Recopilación de recetas

5.2.2. Preprocesamiento de los datos textuales

- Campos utilizados
- Limpieza de los datos
- Aplicación de bigramas

5.3. Word Embedding

5.3.1. Entrenamiento del modelo

Tunning de los parámetros

5.3.2. Funcionamiento

Mostrar el funcionamiento obteniendo las palabras más similares (del vocabulario) a una dada

5.4. Mapping entre ingredientes

Se calcula la distancia entre las representaciones numéricas de un ingrediente dado, respecto a un conjunto de ingredientes que cumplan con las propiedades que se han indicado previamente. Se devolverá como sustituto el ingrediente más similar que cumpla las restricciones previamente establecidas. Para conocer la similitud entre un ingrediente y otro, se han implementado distintas medidas de distancia.

5.4.1. Funciones de distancia

Para ello, se aplicarán medidas de distancia.... ¿Explico aquí todas las medidas de distancia que hemos ido implementando y sus pruebas?

5.4.2. Ejemplos

Capítulo 6

Capítulo 6

Resumen: Este capítulo presenta

6.1. Capítulo 6

no blanco

Capítulo 7

Pruebas

Resumen: Este capítulo recoge los principales aspectos concluidos del desarrollo del documento.

7.1. Pruebas

Capítulo 8

Conclusiones

Resumen: Este capítulo recoge los principales aspectos concluidos del desarrollo del documento.

8.1. Conclusiones

