

UNIVERSIDAD DE CASTILLA-LA MANCHA

BACHELOR IN COMPUTING ENGINEERING

Coursed intensification: Computer Engineering

BACHELOR DISSERTATION

MADTrack: Distributed System for Dataset and Artificial Intelligence Model Management

Miguel Ángel Ruiz Arreaza

Escuela
Superior
de Informática



Ciudad Real, Julio de 2025



**UNIVERSIDAD DE CASTILLA-LA MANCHA
ESCUELA SUPERIOR DE INFORMÁTICA**

Tutor's Department:

Coursed intensification: Computer Engineering

BACHELOR DISSERTATION

**MADTrack: Distributed System for Dataset and Artificial
Intelligence Model Management**

Author: Miguel Ángel Ruiz Arreaza

Tutor: José Luis Espinosa Aranda

Co-tutor: Pablo Tomás Toledano González

July, 2025

MADTrack: Distributed System for Dataset and Artificial Intelligence Model Management
© Miguel Ángel Ruiz Arreaza, 2025

The author may choose the license type they wish. This document is distributed under license CC BY-NC-SA 4.0. The full text of the license is available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>. The copy and distribution of this work is permitted in all parts of the world, without royalties and in any medium, as long as this notice is preserved. In addition, permission is granted to copy and distribute translations of this book from the English original to another language, provided that the copyright notice and this permission notice are preserved in all copies.



This text has been prepared with the L^AT_EX template for Final Dissertations in Computing Engineering for UCLM, published by [Jesús Salido](#) in the public Zenodo repository, DOI: [10.5281/zenodo.4561708](https://doi.org/10.5281/zenodo.4561708).

*To my family, close friends, fellow colleagues, classmates and professors
For their patience and for accompanying me and showing me the way through this exciting journey.*

MADTrack: DS for Dataset and AI Model Management

Miguel Ángel Ruiz Arreaza

Ciudad Real, Julio 2025

Abstract

«What»

Artificial Intelligence is on the rise, and many companies are striving to create a wide-ranging variety of models to help themselves and their customers in an equally wide range of specific tasks. This inevitably translates into multiple model trainings being performed a year and, in many cases, many datasets being created or modified in the same time interval. Such is the case of UBOTICA Technologies: a Space:AI company that puts faith in Computation *on the edge* to deliver AI solutions integrated in embedded systems incorporated in spatial modules, which come with limited space for storing data and computing power. The development of these solutions requires multiple training and deployment iterations, which over the years led to the current tangled mess of untraceable datasets and models, which makes it difficult to search for a specific AI training configuration.

«How»

MADTrack is a distributed configuration management system with the purpose of putting order to the configurational chaos previously mentioned. It will store, track and manage all changes within datasets and AI model configurations. The main restrictions over the development of the system are the limited storage space of the company for this resources (the management of the evolution of datasets and models has to be done efficiently), the distributed nature of the environment where the items are stored and managed and the need for the system to be integrated in a greater processing workflow. The development of the system will be divided in a series of prototypes with an iterative and incremental approach, following continuous testing policies.

«Conclusion»

The resulting system will consist of a distributed system following the client-server application, with a local or remote server attending requests from multiple clients sending requests by means of a software wrapper library which in turn will interact with other open-source technologies.

Keywords: Escuela Superior de Informática, UCLM, TFG, L^AT_EX, T_EXLive, AI, Deep Learning, Distributed System, Dataset, configuration management .

MADTrack: Distributed System for Dataset and Artificial Intelligence Model Management

Miguel Ángel Ruiz Arreaza
Ciudad Real, Julio 2025

Resumen

«Que»

La Inteligencia Artificial está en ascenso, y muchas empresas están trabajando en la elaboración de una inmensa variedad de modelos que las ayuden tanto a ellas como a sus clientes a realizar una variedad de tareas igualmente amplia. Esto provoca que se realicen muchas entrenamientos de modelos al año y, en muchos casos, se creen o modifiquen muchos conjuntos de datos en el mismo intervalo de tiempo. Este es el caso de UBOTICA Tecnologías: una empresa de Space:AI que apuesta por la computación *on the edge* para ofrecer soluciones de Inteligencia Artificial integradas en sistemas empotrados en módulos espaciales, que tienen un espacio de almacenaje y poder computacional limitado. El desarrollo de estas soluciones requiere de varias iteraciones de entrenamiento y despliegue, lo que ha llevado a un desorden caótico de conjuntos de datos y modelos difícilmente identificables, lo que dificulta la búsqueda de un modelo o configuración de entrenamiento específica.

«Como»

MADTrack es un sistema de gestión de configuración distribuido que tiene el propósito de ponerle orden al caos configuracional previamente mencionado. El sistema almacenará, rastreará y gestionará todas las modificaciones de conjuntos de datos y configuraciones de modelos IA. Las principales restricciones que tiene el desarrollo del sistema son el espacio de almacenamiento limitado de la empresa para estos recursos (la evolución de conjuntos de datos y modelos debe ser registrada de manera eficiente), la naturaleza distribuida del entorno donde se almacenan y gestionan los elementos y la necesidad de la integración del sistema en un gran flujo de procesamiento. El desarrollo del sistema se dividirá en una serie de prototipos con un enfoque iterativo y incremental, siguiendo políticas de pruebas continuas.

«Conclusiones»

El sistema resultante consistirá en un sistema distribuido que sigue la arquitectura cliente-servidor, con un servidor local o remoto atendiendo las solicitudes de varios clientes enviando solicitudes a través de una biblioteca Software del tipo Wrapper que a su vez se interactúe con otras tecnologías de software abiertas.

Palabras clave: Escuela Superior de Informática, UCLM, TFG, L^AT_EX, T_EXLive, AI, Deep Learning, Distributed System, Dataset, configuration management .

Agradecimientos

Aunque es un apartado opcional, haremos bueno el refrán «*es de bien nacidos, ser agradecidos*» si empleamos este espacio como un medio para agradecer a todos los que, de un modo u otro, han hecho posible que el trabajo realizado *llegue a buen puerto*. Esta sección es ideal para agradecer a directores, profesores, mentores, familiares, compañeros, amigos, etc.

Estos agradecimientos pueden ser tan personales como deseas e incluir anécdotas y chascarrillos, pero recuerda que *no deberían ocupar más de una página*.

Miguel Ángel Ruiz Arreaza
Ciudad Real, 2025

Notación y acrónimos

NOTACION

(Texto aclaratorio -*suprime*-). Ejemplo de lista con notación (o nomenclatura) empleada en la memoria del TFG. Debes editarla según las necesidades de tu trabajo intenta que sea informativa y evita que incorpore información obvia.¹

- A, B, C, D : Variables lógicas
 f, g, h : Funciones lógicas
 \cdot : Producto lógico (AND), a menudo se omitirá como en AB en lugar de $A \cdot B$
 $+$: Suma aritmética o lógica (OR) dependiendo del contexto
 \oplus : OR exclusivo (XOR)
 \bar{A} o A' : Operador NOT o negación

LISTA DE ACRÓNIMOS

(Texto aclaratorio -*suprime*-). Ejemplo de lista *ordenada alfabéticamente* con los acrónimos empleados en el texto. Se pueden omitir aquellos acrónimos que son reconocidos en el contexto académico (p. ej., PhD), aunque aquí se han incluido a efectos ilustrativos.

¹Se incluye únicamente con propósito de ilustración, ya que el documento no emplea la notación aquí mostrada.

Índice general

Abstract	III
Resumen	IV
Agradecimientos	V
Notación y acrónimos	VI
Índice de figuras	IX
Índice de tablas	X
Listings	XI
List of Algorithms	XII
1. Introduction	1
1.1. Motivation	1
1.2. Problems and Solutions	2
1.3. Document Structure	3
2. Objective	4
2.1. Main Objective	4
2.2. Specific Objectives	4
3. State Of The Art	6
3.1. Datasets	6
4. Desarrollo y resultados	7
4.1. Requisitos y análisis del sistema	7
4.2. Diseño del sistema	7
4.3. Implementación del sistema	7
4.4. Pruebas del sistema	7
4.5. Despliegue	8
5. Conclusiones	9
5.1. Objetivos alcanzados	9
5.2. Justificación de competencias adquiridas	9
5.3. Trabajos derivados y futuros	9

5.4. Valoración personal	9
Bibliografía	10
A. Sobre la Bibliografía	12
B. Breve introducción a L^AT_EX	14
B.1. Listas	14
B.2. Tablas	15
B.3. Ecuaciones matemáticas	15
B.4. Figuras	16
B.5. Listados de código fuente	20
B.6. Menús, paths y teclas con el paquete menukeys	21

Índice de figuras

B.1. Ejemplo de figura	16
B.2. Ejemplo de subfiguras	17
B.3. Gráfico girado	18
B.4. Gráfico apaisado	19

Índice de tablas

B.1. Ejemplo de uso de la macro <code>cline</code>	15
B.2. Ejemplo de tabla con especificación de anchura de columna	15

Listings

B.1. Código fuente en Java	20
B.2. Ejemplo de código fuente en lenguaje C	20
B.3. Ejemplo de script en Matlab	20

List of Algorithms

B.1. Cómo escribir algoritmos	21
---	----

CAPÍTULO 1

Introduction

With the passage of time, the development of Artificial Intelligence (AI) has become of increasing interest due to the powerful tools it can provide to any organisation [2]. Since the development of *The Bombe*, the machine that was able to decode the *Enigma* machine in 1939, passing through a whole set of ups and downs and even a silent winter before its comeback in 2015 thanks to Deep Learning, Artificial Intelligence is being applied in a number of fields, sometimes even reaching the point of having the potential of threatening human safety and raising awareness of the need of regulations for its use.

The development of these models has been made possible thanks to the contribution of a number of organisations, such as Google, Microsoft and OpenAI, which invested an objectively significant amount of time and money (reaching a total investment of 24.0 billion dollars in 2018 [2]) to develop systems as broad as the GPT-4 model, which is able to generate text from both imaged and textual inputs [1], and is capable of helping professionals in solving doubts in a large number of fields.

With time, the development of these models became increasingly complex and difficult, and this would be reflected in the number of iterations required for a model to reach the desired level of accuracy in its predictions. Moreover, the application of these techniques on areas where little data was available made this development even harder. Some companies even started to employ their own resources to gather new data to train their models, which led to a model having more iterations as the available data grew. Summarizing, a model (and even a dataset) can become complex structures that have many evolutionary stages within their lifecycle.

Many organisations, aware of the increasing complexity of the lifecycle of datasets and models, saw a market opportunity to develop systems that would manage these lifecycles in an organized and efficient way, producing the AI and dataset configuration management tools, such as MLFlow [9], DVC, which are open source and available to everyone, as well as premium tools such as Neptune.ai, which are Platform as a Service (PaaS) that provide even further monitoring capabilities.

The aim of this Final Degree Project is to tackle this issue in a specific particular case. In the coming subsections, the reasons that motivated the elaboration of this project, along with the challenges it faces and how they will be solved, will be thoroughly described. Finally, an overview of the structure of the document will be given, so as to give readers the necessary information to follow the development documentation of the project.

1.1. MOTIVATION

It is of common knowledge that Artificial Intelligence has gained significant importance in the 2020s. The development of tools such as the GPT models have headed a revolution in the way humans solve both simple and complex tasks. A revolution that was made possible with the intervention of

multiple companies and organisations, and a considerable amount of money and time invested in the development of these models [2].

UBOTICA Technologies is a pioneer company that develops **AI Computer Vision (CV)** solutions. This means that, as a company, they use Artificial Intelligence techniques to extract information from images. These solutions are then integrated in embedded systems with limited capabilities, and that are part of a bigger, more complex system. The domains where these solutions are used are mainly in the space industry with their new CogniSAT-6 project [6], and recently even in the culinary industry. The development of the various solutions of the company requires multiple iterations where the models are trained, validated and tested, either in existing datasets, or in new ones. Moreover, the CogniSAT-6 system has the capability for creating new datasets out of self-taken images, which may be periodically added to the existing datasets, or even used to replace the existing ones.

This continuous rise of the available models and datasets, paired with a lack of a real control protocol over the new and improved versions of an AI model or dataset, has led often to chaotic situations where an abnormal amount of time is taken on looking for the desired dataset, and accessing the necessary training configuration that produced an specific result.

The aim of MADTrack is to develop a system that establishes a real configuration management basis for these datasets and AI models. A system that can integrate both new and existing datasets and models in a distributed, remote environment, and that will make the best use of the available resources the company dedicates to this management.

1.2. PROBLEMS AND SOLUTIONS

The main problem to be tackled in this Final Degree Project is the lack of a real configuration management protocol over the produced datasets and models, the difficulty at not only determining the correct configuration, but also bringing it to the environment where it is going to be used, and the need for establishing it from the methods used by the users so as to make the system familiar to them.

Furthermore, another problem resides within the protocols and deployment requirements of the system, which may need special authentication protocols from the environment where the system is going to be deployed, as well as from the environment where the models and datasets are stored (which in turn may also differ from the one where the system is running in). All of this may result in the need to develop a distributed system, where many components interact with each other to satisfy a complex need. Moreover, the system must be easy to integrate in bigger workflows, so the study of the available pipelines the company uses and the possible integration points of the system within them may be taken as another problem.

Finally, the last problem is the limitations of the company to dedicate resources for the storage of all the data produced by the model and dataset lifecycles. The developed system must, hence, make an efficient use of the resources available, so as to maximize the amount of data that could be stored and managed. Also, the system must be able to adapt to the needs of a growing company, where many requests may be taken concurrently.

For the resolution of all these three problems, the system will be formed by a server that will satisfy requests and perform the necessary storage and fetching operations to bring both the datasets and the models to the users, whilst the client will be a library that enables the integration of the necessary items into the configuration management, sending the evolutionary changes of a dataset or model to the server that stores them.

1.3. DOCUMENT STRUCTURE

According to the steps necessary to fully describe the elaboration process of this Final Degree Project, the following structure has been decided:

1. **Introduction.** The domain and main problematics are described, as well as what is the solutions the project will establish to these problematics.
2. **Objective.** The main objective, as well as the specific objectives of the project, are enumerated and detailed.
3. **State-of-the-art research.** This chapter contains the results of an extensive research and study of all the concepts and technologies relevant to the development of the project.
4. **Methodology.** Description of the development framework and methodology to be used, and its specific application to the development of the project.
5. **Desarrollo y resultados.** En este capítulo se explica cómo se han llevado a cabo las fases del trabajo cumpliendo el plan previsto y enumerando los resultados obtenidos.
6. **Conclusions.** A summary of the achieved results will be made, as well as the proposal of the future work to be carried out on this project.
7. **Bibliografía.** Lista de las referencias bibliográficas que se hayan citado en el texto. Recuerda que no debes incluir fuentes de información relacionada que no hayas citado explícitamente en la memoria.
8. **Anexos.** Contenidos auxiliares que complementan del trabajo, como manuales de uso, diagramas, figuras, tablas, listados de código, etcétera.

CAPÍTULO 2

Objective

The main aim of this chapter is to define the main objective of the project, and also enumerate and detail the specific objectives that will be necessary to fulfill so as to achieve this main objective. The definition of the objective and the specific objectives has the purpose of providing a better explanation of the work that will be carried out within the scope of the project, and to provide an indicator of its good progress.

2.1. MAIN OBJECTIVE

The main aim of the project is to produce a system capable to track the configuration of datasets and AI models, and is easy to integrate in the internal workflows and pipelines of the end user (UBOTICA Technologies). The system will also be characterized by its distributed nature, its scalability (it will make efficient use of the resources, so that an increase on the resources or number of users will make a minimal impact on the system's performance), security (the system must be prepared for possible attacks, specially from injection and buffer overflow attacks) and robustness (Upon the case of minor failures, the system must not go down and provide adequate and meaningful logging).

The system will provide user-friendly mechanisms for accessing the dataset and model configuration database, so the users can make operations on this configuration directly from their codespaces.

2.2. SPECIFIC OBJECTIVES

The aforementioned main objective can be divided into a set of partial objectives, also referred to as subobjectives. This final project can be divided into – subobjectives that will mark the development progress of the project, which could be in turn considered as finished when all of the subobjectives have been completed, and the resulting system satisfies the specifications of the main objective.

- *Development and deployment planning of a server able to track the configuration of AI models and datasets.*

A planning will be made regarding the deployment details and the necessary infrastructure to be able to hold the Tracking server that will satisfy the requests from the other components of the system. These details involve the necessary hardware requirements (Memory, CPU cores, network configuration, available ports) and software requirements (dependencies and entrypoint scripts) that will be used to design and develop the tracking server, which will be deployed in the future inside of the company's intranet infrastructure. It is also necessary to specify how this server will interact with the rest of components of the system and when should it be deployed so as to ensure its well-functioning.

- *Development of a library module that manages the configuration management of datasets. Registering changes on their contents.*

Aside from the tracking server, multiple library components will be necessary to manage the configuration of datasets and models. Some of these components will be developed under a module that will handle issues regarding the configuration management of datasets. These components will focus on providing code mechanisms that enable the integration of new datasets into the system, as well as providing the necessary means to bring the datasets in a specific evolutionary stage to the users.

- *Development of a library module that manages the configuration management of AI models, and facilitates the search of models according to their performance.*

Other components of the library will be gathered within a module focused on managing the configuration of Artificial Intelligence models. These components will interact with the tracking server in order to track the parameters and metrics produced by the experimental runs of the models performed by the company, and register the final models and any other file meaningful to these inside a database.

CAPÍTULO 3

State Of The Art

This chapter aims to put into context the different concepts and technologies relevant to the development of the project. The following subsections provide an overview of these technologies, as well as the related concepts they handle, so as to contextualize the upcoming work. In the coming sections, concepts about Artificial Intelligence models, datasets and configuration management will be presented. Then, a thorough overview of the available technologies on the market to solve the problems described within Chapter 2 will be given.

3.1. DATASETS

In a statistical context, and according to a recent article by IBM workers [5], datasets are collections of organized data. The organization structure comes in a wide-ranging variety of formats, being the most common ones **Comma Separated Values (CSV)** and **JavaScript Object Notation (JSON)**. The data contained within them can be collected from a number of sources, e.g. customer interactions, data obtained from distributed IoT devices (In the case of the end users, the satellites equipped with camera devices for taking pictures), or even public social media.

These arrangements of data are of great value for statistical data analysis, as well as for **Machine Learning (ML)** and other **AI** techniques, which require large amounts of accessible and reliable data to achieve a satisfactory performance. The quality of the data contained within a dataset

CAPÍTULO 4

Desarrollo y resultados

En este capítulo se describirá las diferentes fases del ciclo de desarrollo de software de acuerdo al plan de gestión expuesto en el capítulo ??.

4.1. REQUISITOS Y ANÁLISIS DEL SISTEMA

En esta sección se presentarán los objetivos y el catálogo de requisitos del proyecto: funcionales, no funcionales, de información, reglas de negocio, etc. Una vez catalogados los requisitos del sistema se procederá con su análisis empleando el lenguaje de modelado UML. El análisis mencionado incluirá:

- **Modelo conceptual.** En él se identifican clases, atributos, relaciones, etc.
- **Modelo de casos de uso.** Representan las interacciones entre los actores y el sistema bajo estudio.
- **Modelo de interfaz de usuario.** Puede consistir en un prototipo de baja fidelidad de la interfaz de usuario.

4.2. DISEÑO DEL SISTEMA

En esta sección se define la arquitectura lógica general del sistema. Para describirla se puede emplear un modelo como C4 [4], el cual permite representar la arquitectura de un sistema software mediante varios diagramas a distintos niveles de abstracción.

4.3. IMPLEMENTACIÓN DEL SISTEMA

En este apartado se describirá la organización del código fuente y *scripts*, describiendo el propósito de los distintos ficheros y su distribución en paquetes y directorios. Puede ser conveniente incluir alguna porción significativa de código fuente que sea de especial relevancia por su funcionalidad.

En el desarrollo de proyectos software cobra especial importancia el empleo de sistemas de control de versiones junto a repositorios en línea. Estas herramientas se convierten en esenciales para disponer de un registro histórico del desarrollo que también puede ayudar a evaluar el trabajo realizado. Por este motivo, en la memoria del proyecto se debe indicar la dirección URL de los repositorios empleados (p. ej., Github).

4.4. PRUEBAS DEL SISTEMA

En esta sección se describirá el plan de pruebas del sistema incluyendo todos los tipos llevados a cabo. El desarrollo de la sección debería tratar los aspectos siguientes:

- **Estrategia.** Donde se indica el alcance de las pruebas y los procedimientos.

- **Pruebas unitarias.** Destinadas a localizar errores en cada nuevo módulo software desarrollado antes de su integración con el resto del sistema.
- **Pruebas de integración.** Su objetivo es localizar errores en subsistemas completos analizando la interacción entre varios módulos de software.
- **Pruebas de sistema.** Contempla las pruebas funcionales con las que se realiza el análisis del buen funcionamiento de la implementación de los casos de uso del sistema. Además, en estas pruebas se comprueba el funcionamiento respecto a los requisitos no funcionales: eficiencia, seguridad, etcétera.
- **Pruebas de aceptación.** Su intención es demostrar, con la participación del cliente, que el producto está listo para su puesta en funcionamiento en el entorno producción.

4.5. DESPLIEGUE

Esta sección recoge la arquitectura física propuesta del sistema, las instrucciones para su despliegue, operación y mantenimiento del nivel de servicio. Es muy importante que todas las justificaciones aportadas se sustenten no solo en juicios de valor sino en evidencias tangibles como: historiales de actividad, repositorios de código y documentación, porciones de código, trazas de ejecución, capturas de pantalla, demos, etcétera.

CAPÍTULO 5

Conclusiones

5.1. OBJETIVOS ALCANZADOS

En este capítulo se realizará un juicio crítico y discusión sobre el objetivo general y objetivos secundarios alcanzados durante el desarrollo del trabajo.

5.2. JUSTIFICACIÓN DE COMPETENCIAS ADQUIRIDAS

Es muy importante recordar que según la normativa vigente en la ESI-UCLM, el capítulo de conclusiones debe incluir *obligatoriamente* un apartado destinado a justificar la aplicación en el TFG de las competencias específicas (una o más) adquiridas en la tecnología específica cursada, como se indica a continuación:

En el TFG se han aplicado las competencias correspondientes a la Tecnología Específica de *[poner lo que corresponda]*:

Código de la competencia 1: *[Texto de la competencia 1]*. Explicación de cómo se ha aplicado en el TFG.

...(otras más si las hubiera).

5.3. TRABAJOS DERIVADOS Y FUTUROS

Si es pertinente se puede incluir información sobre trabajos derivados como publicaciones o ponencias en preparación, así como trabajos futuros (*solo si estos están iniciados o planificados en el momento que se redacta el texto*).

Se recomienda reflexionar sobre la conveniencia de inclusión de una lista de posibles mejoras, ya que puede transmitir la impresión de que el trabajo se encuentra en un estado incompleto o inacabado.

5.4. VALORACIÓN PERSONAL

En esta sección final se realizará un rápido análisis de las lecciones aprendidas en las que se pueden incluir tanto buenas prácticas adquiridas (tecnológicas y procedimentales) como cualquier otro aspecto de interés. También se puede resumir cuantitativamente el tiempo y esfuerzo dedicados al proyecto a lo largo de su desarrollo.

Bibliografía

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [2] Jasmin Praful Bharadiya, Reji Kurien Thomas, and Farhan Ahmed. Rise of artificial intelligence in business and industry. *Journal of Engineering Research and Reports*, 25(3):85–103, 2023.
- [3] Alexánder Borbón and Walter Mora. *Edición de textos científicos con L^AT_EX. Composición, diseño editorial, gráficos, Inkscape, Tikz y presentaciones Beamer*. Instituto Tecnológico de Costa Rica, 2 edition, 2021.
- [4] Simon Brown. The C4 model for visualising software architecture. URL: <https://c4model.com/>, 2022. Último acceso: 19 feb. 2024.
- [5] IBM. What is a dataset?, 2024.
- [6] David Rijlaarsdam, Tom Hendrix, Pablo T Toledano González, Alberto Velasco-Mata, Léonie Buckley, Juan Puig Miquel, Oriol Aragon Casaled, and Aubrey Dunne. The next era for earth observation spacecraft: An overview of cognisat-6. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.
- [7] Servicio de publicaciones UCLM. Propiedad intelectual. Documentos elaborados por el Grupo de Gestión del Conocimiento y Propiedad Intelectual de la Universidad de Castilla-La Mancha. URL: <https://e.uclm.es/servicios/doc/?id=UCLMDOCID-12-739>. Último acceso: feb. 2024.
- [8] Universidad de Cantabria. Cómo usar imágenes en trabajos. Artículo técnico disponible en URL: https://web.unican.es/buc/Documents/Formacion/guia_imagenes.pdf, 2018. Último acceso: sep. 2021.
- [9] Matei Zaharia, Andrew Chen, Aaron Davidson, Ali Ghodsi, Sue Ann Hong, Andy Konwinski, Siddharth Murching, Tomas Nykodym, Paul Ogilvie, Mani Parkhe, et al. Accelerating the machine learning lifecycle with mlflow. *IEEE Data Eng. Bull.*, 41(4):39–45, 2018.

APPENDICES

APÉNDICE A

Sobre la Bibliografía

En los anexos se incluirá, de modo opcional, material suplementario que podrá consistir en manuales de usuario, listados seleccionados de código fuente, esquemas, planos y en general aquel contenido que complementa a la memoria. Se recomienda que no sean excesivamente voluminosos, aunque su extensión no está sometida a la regulación por normativa, ya que esta afecta únicamente al texto principal de la memoria.

En esta plantilla hemos decidido incluir dos anexos. En el primero de ellos se hacen algunos comentarios adicionales sobre la bibliografía. En el segundo se aporta una breve introducción a L^AT_EX cuya información puede servir de ejemplo de inclusión de ciertos elementos en la preparación de la memoria del TFG.

Todo el material de terceros se debe citar convenientemente sin contravenir los términos de las licencias de uso y distribución de dicho material. Esto se extiende al uso de diagramas y cualquier elemento gráfico. El incumplimiento de la legislación vigente en materia de protección de la propiedad intelectual es responsabilidad exclusiva del autor, independientemente de la cesión de derechos que este haya convenido.¹

La sección de *Bibliografía*, que si se prefiere se puede titular *Referencias*, incluirá un listado ordenado preferentemente por orden alfabético (primer apellido del autor principal), con todas las obras citadas en el texto. En la lista de referencias se especificará para cada obra:

- autores,
- título,
- editorial, y
- año de publicación.

Este formato se conseguirá en L^AT_EX mediante el uso del estilo estándar plain o cualquier otro derivado con estilo de citación numérica. En algunas titulaciones se obliga a una ordenación por orden de cita en el texto. Este resultado se obtiene con BibT_EX mediante los estilos estándar ieeetr (estilo para los IEEE *transactions*) y unsrt (estilo *unsorted*).²

Es muy importante tener presente que en esta sección solo se debe incluir las referencias bibliográficas citadas expresamente en el documento. Si se desea incluir fuentes consultadas, pero no citadas, se puede confeccionar con ellas una sección denominada *Material de consulta*, aunque estas referencias se pueden incluir opcionalmente a lo largo del documento como notas a pie de página.

En las titulaciones técnicas se empleará estilo de citación numérico con el número de la referencia entre corchetes. La cita podrá incluir el número de página concreto de la referencia que se desea citar.

¹<https://www.uclm.es/areas/biblioteca/encuentra-informacion/perfiles/alumno/antiplagio>

²<https://www.wikiwand.com/es/articles/BibTeX>

El uso correcto de la citación implica dejar claro al lector cuál es el texto, material o idea citado. .

Cuando se desee incluir referencias a páginas genéricas de Internet sin mención expresa a un artículo con título y autor definido, dichas referencias se pueden incluir como notas al pie de página o como un apartado de fuentes de consulta dedicado a *Direcciones de Internet*. Por el contrario, los documentos electrónicos publicados en Internet se pueden incluir en la sección de Bibliografía empleando el tipo de entrada `misc` en el fichero `.bib` con el comando `url` (como se muestra en la bibliografía de ejemplo distribuida con este documento). Observarás que el campo `note` se emplea para añadir información adicional como la fecha de la última consulta de fuentes publicadas en Internet, y para la inclusión del DOI de algunas obras para su rápida recuperación.³

³Esta estrategia necesita adaptación cuando se emplea un estilo de citación autor-año (p.ej., `natbib`).

APÉNDICE B

Breve introducción a L^AT_EX

El contenido del trabajo final de estudios se organiza en capítulos que se subdividen en secciones. Con L^AT_EX este tipo de organización se realiza de modo inmediato mediante la generación automática de los estilos correspondientes a los títulos de cada sección y su inclusión en la tabla de contenidos. Los ajustes relativos a la generación del formato y estilos asociados a secciones del documento se realizan con el paquete `titlesec` empleado en esta plantilla.

En las secciones siguientes se comenta la inclusión con L^AT_EX de distintos elementos de organización de información junto a ejemplos que facilitan su utilización en la memoria del trabajo.¹

B.1. LISTAS

Existen dos tipos de listas: enumeraciones y listas con viñetas. En el primer tipo, los elementos de la lista se preceden de una clave numérica o alfabética, mientras que en el segundo tipo se emplea una viñeta. En ambos casos los elementos se pueden anidar para crear una jerarquía entre ellos. En L^AT_EX se recomienda la inclusión del paquete `enumitem` (incluido en esta plantilla) que permite personalizar fácilmente las listas de un documento. A continuación se muestran algunos ejemplos de las posibilidades.

B.1.1. Ejemplo de lista con viñetas personalizadas

- pera
- ☛ manzana
- ❶ naranja

B.1.2. Ejemplo de lista condensada

Este tipo de lista presenta una separación mínima entre ítems, en varias columnas y configuración de la etiqueta.

- | | |
|-------------|--------------|
| (1) pera | (4) patata |
| (2) manzana | (5) calabaza |
| (3) naranja | (6) fresa |

Además del texto, los documentos pueden incluir elementos que enriquecen su contenido facilitando su exposición y comprensión. En las secciones siguientes tratamos brevemente dichos elementos.

¹Las explicaciones de este anexo forman parte del contenido del curso «L^AT_EX esencial para preparación de TFG y otros documentos académicos» de la ESI-UCLM.

B.2. TABLAS

A continuación se incluyen algunos ejemplos de tablas elaboradas con L^AT_EX mediante el empleo de paquetes dedicados. Para la realización de tablas más complejas se recomienda la consulta de manuales [3] y el empleo de asistentes o herramientas en línea.²

Observa que el título de las tablas se ubica en la parte superior de la tabla. Puesto que el contenido de la tabla es texto, tiene sentido leer primero el título para contextualizar el contenido de la tabla antes de su lectura.

Tabla B.1: Ejemplo de uso de la macro `cline`

7C0	hexadecimal
3700	octal
11111000000	binario
1984	decimal

Ejemplo de tabla en la que se controla el ancho de la celda.

Tabla B.2: Ejemplo de tabla con especificación de anchura de columna

Día	Temp Mín (°C)	Temp Máx (°C)	Previsión
Lunes	11	22	Día claro y muy soleado. Sin embargo, la brisa de la tarde puede hacer que las temperaturas desciendan
Martes	9	19	Nuboso con chubascos en muchas regiones. En Cataluña claro con posibilidad de bancos nubosos al norte de la región
Miércoles	10	21	La lluvia continuará por la mañana, pero las condiciones climáticas mejorarán considerablemente por la tarde

B.3. ECUACIONES MATEMÁTICAS

La composición de ecuaciones requiere el uso de comandos especializados. Por tanto, para facilitar dicha tarea se aconseja el empleo de programas especializados como MathType o asistentes como el incluido en editores como T_EXstudio³ o herramientas en línea.⁴ Es muy simple incluir fórmulas matemáticas sencillas en el mismo texto en el que se escribe. Por ejemplo, $h^2 = a^2 + b^2$ que podría ser la ecuación representativa del teorema de Pitágoras (ver también ec. B.1).

²<https://www.tablesgenerator.com/>

³<https://www.texstudio.org/>

⁴<https://latex.codecogs.com/>, <http://www.sciweavers.org/free-online-latex-equation-editor>

Las fórmulas también se pueden separar del texto para que aparezcan destacadas, así:

$$c^2 = \int (a^2 + b^2) \cdot dx$$

Pero si se desea, las ecuaciones pueden ser numeradas de forma automática e incluso utilizar referencias cruzadas a ellas:

$$h^2 = b^2 + c^2 \tag{B.1}$$

B.4. FIGURAS

A diferencia de lo que sucede en las tablas, el título de las figuras aparece en la parte inferior de estas. Para la inclusión de las figuras se debe tener en cuenta que su contenido se encuentra en un fichero individual con el formato y resolución apropiados para garantizar la calidad del resultado final.

En esta sección se añaden ejemplos de muestra para la inclusión de figuras simples y otras compuestas de subfiguras mediante el empleo del paquete `subcaption`.



Figura B.1: Fotografía a color (Fuente: J. Salido, CC BY-NC-ND)

Ejemplo de figura compuesta por dos subfiguras incluidas mediante paquete `subcaption`. A través del uso de etiquetas (`\label`) es posible incluir referencias cruzadas a subfiguras como la fotografía en blanco y negro de la Fig. B.2b.



(a) Fotografía a color



(b) Fotografía en blanco y negro

Figura B.2: Ejemplo de inclusión de subfiguras en un mismo entorno (Fuente: J. Salido, CC BY SA)

En los trabajos académicos la inclusión de imágenes y figuras que no son propiedad del autor suscitan bastante controversia, ya que con frecuencia se incumple inadvertidamente la ley vigente de propiedad intelectual. Respecto a este hecho se recomienda, tanto al alumnado como al profesorado encargado de la tutorización, consultar documentación informativa sobre el uso correcto de figuras en documentos académicos [7, 8]. Entre las «incorrecciones» más habituales en los documentos académicos, se observa:

- *Abuso del derecho de cita.* Se produce al incluir, con fines exclusivamente decorativos o ilustrativos de la explicación, una figura sujeta a derechos de uso restringido invocando el derecho de cita (incluso con correcta atribución de la obra).
- *Incorrecta atribución de la obra.* Es habitual confundir al autor de la obra con la fuente de origen de la misma. La fuente es precisa cuando se cita la obra original. Sin embargo, la licencia de muchas obras exige la atribución al autor y la inclusión de la licencia bajo la que se distribuye o hace uso de la misma. Véase como ejemplo cómo se realiza una correcta atribución en las Fig. B.1 y B.2 mencionando al autor y la licencia Creative-Commons⁵ bajo la que se rige el uso de la imagen y el mecanismo de título alternativo para que dicha atribución no aparezca en el índice de figuras usando título opcional.
- *Supresión de los detalles de la licencia de uso.* Al incluir obras de terceros debemos tener presente los términos de distribución de la misma e incluirlos junto a la atribución de autoría.

La inclusión de material de *dominio público*, sin restricciones de uso o con permiso, hace innecesaria la atribución al autor, pero se recomienda incluir una nota de agradecimiento.⁶

Cuando se presenta la necesidad de incluir un gráfico demasiado grande para el tamaño de la página, una opción muy apropiada es la impresión del gráfico en modo girado en una página aparte. Este efecto se consigue con el entorno `sidewaysfigure` proporcionado por el paquete `rotating`. La Fig. B.3 muestra un ejemplo del entorno citado con un gráfico PDF.

⁵<https://creativecommons.org>

⁶Incluyendo un texto como: «Por cortesía de ...»

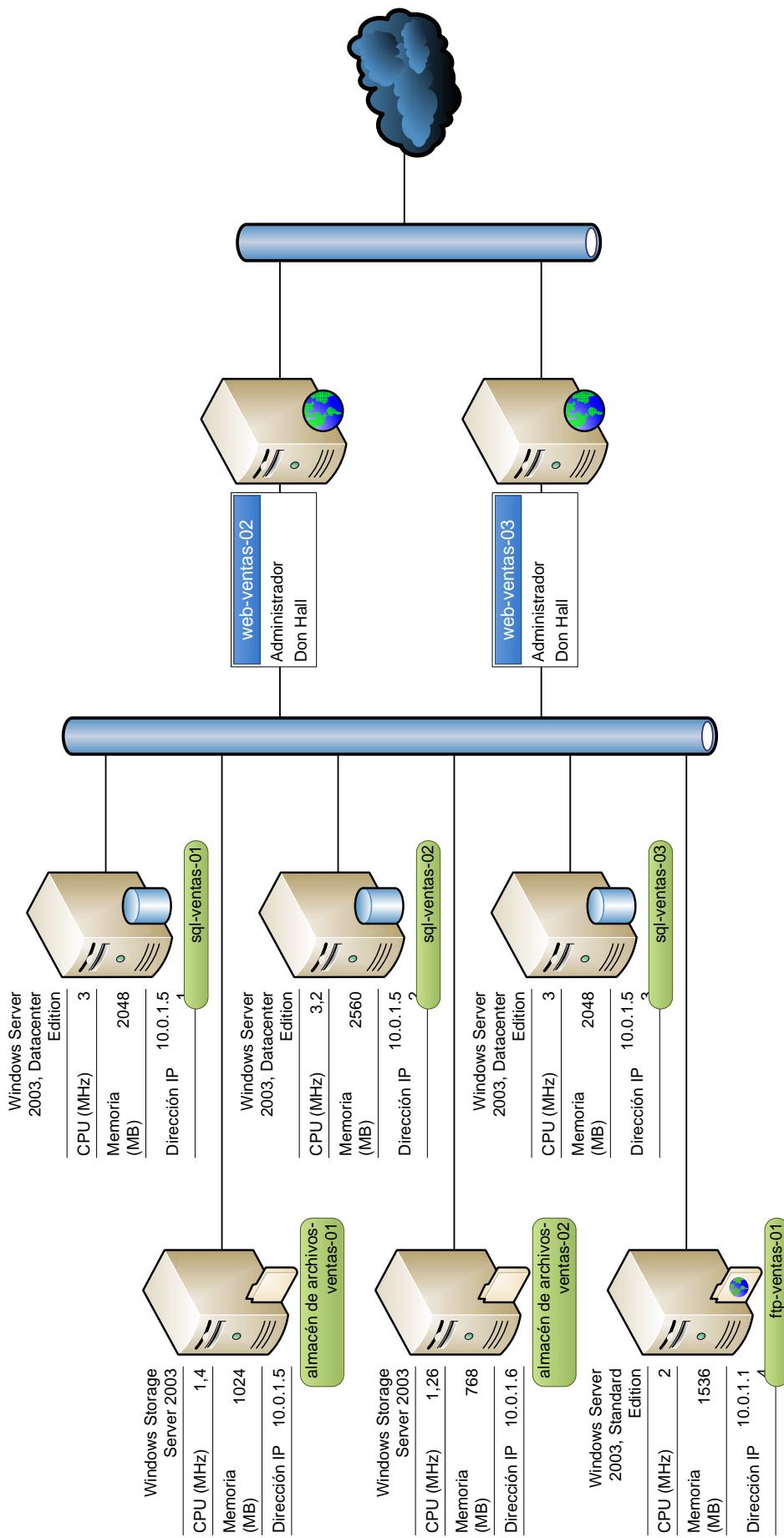


Figura B.3: Figura vectorial con impresión girada

También es posible imprimir una página en formato apaisado cuando contiene una figura muy ancha. Este efecto se consigue con el paquete `pdfescape` y el entorno `landscape` proporcionado. Además, en este caso se han suprimido tanto la cabecera como el pie de página. La figura B.4 se muestra apaisada a modo de ejemplo.

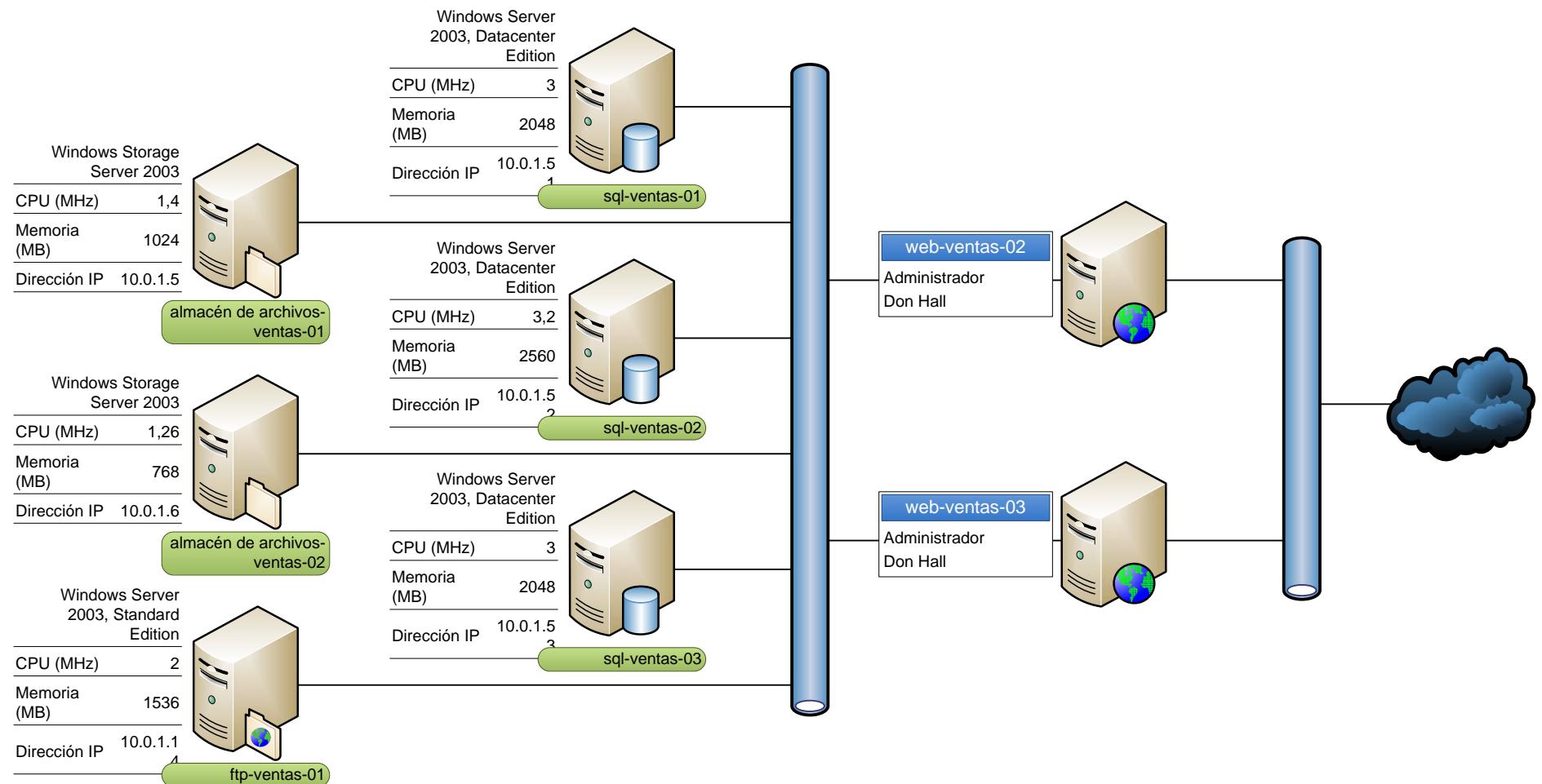


Figura B.4: Figura vectorial con vista en página apaisada

B.5. LISTADOS DE CÓDIGO FUENTE

En los textos científicos relacionados con las TIC⁷ (Tecnologías de la Información y Comunicaciones) suelen aparecer porciones de código en las que se explica alguna función o característica relevante del trabajo que se expone.

La inclusión de porciones de código fuente se puede formatear de modo sencillo en LATEX mediante el uso del paquete `listings`. A continuación, se muestran varios ejemplos de porciones de código correspondientes a distintos lenguajes de programación.

Listing B.1: Ejemplo de código fuente en lenguaje Java

```

1 // @author www.javadb.com
2 public class Main {
3 // Este método convierte un String a un vector de bytes
4
5 public void convertStringToByteArray() {
6
7 String stringToConvert = "This\u00a0String\u00a0is\u00a015";
8 byte[] theByteArray = stringToConvert.getBytes();
9 System.out.println(theByteArray.length);
10 }
11
12 public static void main(String[] args) {
13 new Main().convertStringToByteArray();
14 }
15 }
```

Listing B.2: Ejemplo de código fuente en lenguaje C

```

1 // Este código se ha incluido tal cual está en el fichero LATEX
2 #include <stdio.h>
3
4 int main(int argc, char* argv[]) {
5     puts("¡Hola\u00a0mundo!");
6 }
```

Listing B.3: Ejemplo de script en Matlab

```

1 function f = fibonacci(n)
2 % FIBONACCI Fibonacci sequence
3 % f = FIBONACCI(n) generates the first n Fibonacci numbers.
4 % Copyright 2014 Cleve Moler
5 % Copyright 2014 The MathWorks, Inc.
6
7 f = zeros(n,1);
8 f(1) = 1;
9 f(2) = 2;
10 for k = 3:n
11     f(k) = f(k-1) + f(k-2);
12 end
```

Algunas veces lo que se quiere ilustrar es un algoritmo o método con el que se resuelve un problema abstrandose del lenguaje de implementación. El paquete `algorithm2e` proporciona un entorno `algorithm` para la impresión apropiada de algoritmos, tratándolos como objetos flotantes y con mucha flexibilidad de personalización, como se observa en el algoritmo B.1 del ejemplo.

⁷Por supuesto, en un TFG (Trabajo Fin de Grado) o tesis de un centro superior de Informática.

Algorithm B.1: Cómo escribir algoritmos

```

Datos :este texto
Resultado:como escribir algoritmos con LATEX2e
1 inicialización;
2 while no es el fin del documento do
3   leer actual;
4   if comprendido then
5     ir a la siguiente sección;
6     la sección actual es esta;
7   else
8     ir al principio de la sección actual;
9   end
10 end

```

B.6. MENÚS, PATHS Y TECLAS CON EL PAQUETE MENUKEYS

Cada vez es más usual que los trabajos en ingeniería exijan el uso de software. Para poder especificar de modo elegante el uso de menús, pulsaciones de teclas y directorios, se recomienda el uso del paquete `menukeys`.⁸ Este paquete nos permite especificar el acceso a un menú, por ejemplo:

Herramientas > Órdenes > PDFLaTeX

También un conjunto de teclas. Por ejemplo: `Ctrl + ↑ + T`

O un directorio: `█ C: ▶ user ▶ LaTeX ▶ Ejemplos`

Aunque este paquete permite muchas opciones de configuración de los estilos aplicados, esto no es necesario para obtener unos resultados muy elegantes.

⁸<https://osl.ugr.es/CTAN/macros/latex/contrib/menukeys/menukeys.pdf>