

UNIVERSIDAD DE VALLADOLID  
MÁSTER UNIVERSITARIO  
**Ingeniería Informática**



TRABAJO FIN DE MÁSTER

**Herramientas y comparativas para el  
estudio de métodos de detección de URLs  
*phishing***

Realizado por **SERGIO AGUDELO BERNAL**



**Universidad de Valladolid**

**23 de julio de 2024**

Tutor: JESÚS MARÍA VEGAS HERNÁNDEZ



# Universidad de Valladolid



## Máster universitario en Ingeniería Informática

D. JESÚS MARÍA VEGAS HERNÁNDEZ, profesor del departamento de DEPARTAMENTO DEL TUTOR, área de AREA\_CONOCIMIENTO DEL TUTOR.

### **Expone:**

Que el alumno D. SERGIO AGUDELO BERNAL, ha realizado el Trabajo final de Máster en Ingeniería Informática titulado "HERRAMIENTAS Y COMPARATIVAS PARA EL ESTUDIO DE MÉTODOS DE DETECCIÓN DE URLs *PHISHING*".

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Valladolid, 23 de julio de 2024

Vº. Bº. del Tutor:

Jesús María Vegas Hernández



## **Resumen**

En este primer apartado se hace una **breve** presentación del tema que se aborda en el proyecto.

## **Descriptores**

Palabras separadas por comas que identifiquen el contenido del proyecto Ej: servidor web, buscador de vuelos, android ...

### **Abstract**

A **brief** presentation of the topic addressed in the project.

### **Keywords**

keywords separated by commas.

---

# Índice general

---

Índice general	III
Índice de figuras	V
Índice de tablas	VI
1. Introducción	1
2. Objetivos del proyecto	2
3. Conceptos teóricos	3
3.1. Secciones . . . . .	3
3.2. Referencias . . . . .	3
3.3. Imágenes . . . . .	3
3.4. Listas de ítems . . . . .	4
3.5. Tablas . . . . .	5
4. Técnicas y herramientas	6
5. Aspectos relevantes del desarrollo del proyecto	7
6. Trabajos relacionados	8
7. Conclusiones y Líneas de trabajo futuras	9
<b>Apéndices</b>	<b>10</b>
<b>Apéndice A Plan de Proyecto</b>	<b>11</b>
A.1. Introducción . . . . .	11
A.2. Planificación temporal . . . . .	11

A.3. Estudio de viabilidad . . . . .	11
<b>Apéndice B Especificación de Requisitos</b>	<b>12</b>
B.1. Introducción . . . . .	12
B.2. Objetivos generales . . . . .	12
B.3. Catalogo de requisitos . . . . .	12
B.4. Especificación de requisitos . . . . .	12
<b>Apéndice C Documento de Diseño</b>	<b>13</b>
C.1. Introducción . . . . .	13
C.2. Diseño de datos . . . . .	13
C.3. Diseño procedimental . . . . .	13
C.4. Diseño arquitectónico . . . . .	13
<b>Apéndice D Documentación del Programador</b>	<b>14</b>
D.1. Introducción . . . . .	14
D.2. Estructura de directorios . . . . .	14
D.3. Manual del programador . . . . .	14
D.4. Compilación, instalación y ejecución del proyecto . . . . .	14
D.5. Pruebas del sistema . . . . .	14
<b>Apéndice E Documentación de usuario</b>	<b>15</b>
E.1. Introducción . . . . .	15
E.2. Requisitos de usuarios . . . . .	15
E.3. Instalación . . . . .	15
E.4. Manual del usuario . . . . .	15
<b>Bibliografía</b>	<b>16</b>



---

# Índice de figuras

---

3.1. Autómata para una expresión vacía . . . . .	4
--	---

---

# Índice de tablas

---

3.1. Herramientas y tecnologías utilizadas en cada parte del proyecto . . . . .	5
---	---

---

# 1: Introducción

---

En un mundo cada vez más dependiente de las tecnologías de la información y comunicaciones, el acontecimiento de ciberataques trae mayores riesgos y posible impacto negativo en las sociedades. Como evidencia, 2023 fue un año en que ocurrieron importantes filtraciones de datos masivas, e incidentes que comprometieron el funcionamiento de entidades de ámbito nacional [3].

Entre las amenazas principales, el *phishing* se ha mantenido como el vector de acceso inicial más común por su inherente relación con la comunicación humana, y por la diversidad de métodos de ataque, los cuales se han vuelto más sofisticados ante la aparición de nuevos avances, como la IA generativa [2].

El medio más usual para transmitir ataques de *phishing* es a través de páginas web, y como mecanismo de detección se emplean desde métodos simples como listas negras, hasta los más sofisticados, que a partir de técnicas de *Machine Learning* o Inteligencia Artificial, realizan inferencias sobre la legitimidad del contenido, basándose en características de la página como (1) la URL; (2) el dominio; (3) metainformación del código fuente; (4) aspecto visual de la página; o (5) una combinación de las anteriores [1].

---

## 2: Objetivos del proyecto

---

Este apartado explica de forma precisa y concisa cuales son los objetivos que se persiguen con la realización del proyecto. Se puede distinguir entre los objetivos marcados por los requisitos del software a construir y los objetivos de carácter técnico que plantea a la hora de llevar a la práctica el proyecto.

---

## 3: Conceptos teóricos

---

En aquellos proyectos que necesiten para su comprensión y desarrollo de unos conceptos teóricos de una determinada materia o de un determinado dominio de conocimiento, debe existir un apartado que sintetice dichos conceptos.

Algunos conceptos teóricos de  $\text{\LaTeX}$ <sup>1</sup>.

### 3.1. Secciones

Las secciones se incluyen con el comando `section`.

#### Subsecciones

Además de secciones tenemos subsecciones.

#### Subsubsecciones

Y subsecciones.

### 3.2. Referencias

Las referencias se incluyen en el texto usando `cite [?]`. Para citar webs, artículos o libros `[?]`.

### 3.3. Imágenes

Se pueden incluir imágenes con los comandos standard de  $\text{\LaTeX}$ , pero esta plantilla dispone de comandos propios como por ejemplo el siguiente:

---

<sup>1</sup>Créditos a los proyectos de Álvaro López Cantero: Configurador de Presupuestos y Roberto Izquierdo Amo: PLQuiz



Figura 3.1: Autómata para una expresión vacía

### 3.4. Listas de items

Existen tres posibilidades:

- primer ítem.
- segundo ítem.

Herramientas	App	AngularJS	API REST	BD	Memoria
HTML5		X			
CSS3		X			
BOOTSTRAP		X			
JavaScript		X			
AngularJS		X			
Bower		X			
PHP			X		
Karma + Jasmine		X			
Slim framework			X		
Idiorm			X		
Composer			X		
JSON		X	X		
PhpStorm		X	X		
MySQL				X	
PhpMyAdmin				X	
Git + BitBucket		X	X	X	X
MikTeX					X
TeXMaker					X
Astah					X
Balsamiq Mockups		X			
VersionOne		X	X	X	X

Tabla 3.1: Herramientas y tecnologías utilizadas en cada parte del proyecto

1. primer item.
2. segundo item.

**Primer item** más información sobre el primer item.

**Segundo item** más información sobre el segundo item.

■

### 3.5. Tablas

Igualmente se pueden usar los comandos específicos de  $\text{\LaTeX}$  o bien usar alguno de los comandos de la plantilla.

---

## 4: Técnicas y herramientas

---

Esta parte de la memoria tiene como objetivo presentar las técnicas metodológicas y las herramientas de desarrollo que se han utilizado para llevar a cabo el proyecto. Si se han estudiado diferentes alternativas de metodologías, herramientas, bibliotecas se puede hacer un resumen de los aspectos más destacados de cada alternativa, incluyendo comparativas entre las distintas opciones y una justificación de las elecciones realizadas.

No se pretende que este apartado se convierta en un capítulo de un libro dedicado a cada una de las alternativas, sino comentar los aspectos más destacados de cada opción, con un repaso somero a los fundamentos esenciales y referencias bibliográficas para que el lector pueda ampliar su conocimiento sobre el tema.



---

## 5: Aspectos relevantes del desarrollo del proyecto

---

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto, comentados por los autores del mismo. Debe incluir desde la exposición del ciclo de vida utilizado, hasta los detalles de mayor relevancia de las fases de análisis, diseño e implementación. Se busca que no sea una mera operación de copiar y pegar diagramas y extractos del código fuente, sino que realmente se justifiquen los caminos de solución que se han tomado, especialmente aquellos que no sean triviales. Puede ser el lugar más adecuado para documentar los aspectos más interesantes del diseño y de la implementación, con un mayor hincapié en aspectos tales como el tipo de arquitectura elegido, los índices de las tablas de la base de datos, normalización y desnormalización, distribución en ficheros<sup>3</sup>, reglas de negocio dentro de las bases de datos (EDVHV GH GDWRV DFWLYDV), aspectos de desarrollo relacionados con el WWW... Este apartado, debe convertirse en el resumen de la experiencia práctica del proyecto, y por sí mismo justifica que la memoria se convierta en un documento útil, fuente de referencia para los autores, los tutores y futuros alumnos.

---

## **6: Trabajos relacionados**

---

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final de máster no parece tan obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.

---

## **7: Conclusiones y Líneas de trabajo futuras**

---

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

# Apéndices

## *Apéndice A*

---

# **Plan de Proyecto**

---

Este apéndice presentará el plan de proyecto elaborado para la realización del trabajo. En el caso de trabajos que supongan el desarrollo de software, será sustituido por el Plan de Desarrollo de Software.

### **A.1. Introducción**

### **A.2. Planificación temporal**

### **A.3. Estudio de viabilidad**

Viabilidad económica

Viabilidad legal

## *Apéndice B*

---

# **Especificación de Requisitos**

---

Si el TFM comporta el desarrollo de software, en este apéndice se reunirá la Especificación de Requisitos del mismo.

### **B.1. Introducción**

### **B.2. Objetivos generales**

### **B.3. Catalogo de requisitos**

### **B.4. Especificación de requisitos**

## *Apéndice C*

---

# **Documento de Diseño**

---

Si el TFM comporta el desarrollo de software, en este apéndice se reunirá el Documento de Diseño asociado al mismo.

### **C.1. Introducción**

### **C.2. Diseño de datos**

### **C.3. Diseño procedimental**

### **C.4. Diseño arquitectónico**

## *Apéndice D*

---

# **Documentación del Programador**

---

En este apéndice se incluye la documentación necesaria para que el programador de aplicaciones pueda comprender la estructura de la solución software aportada y para poder modificarla.

### **D.1. Introducción**

### **D.2. Estructura de directorios**

### **D.3. Manual del programador**

### **D.4. Compilación, instalación y ejecución del proyecto**

### **D.5. Pruebas del sistema**



## *Apéndice E*

---

# **Documentación de usuario**

---

En este apéndice se incluye la documentación necesaria para que el usuario sepa cómo debe instalar y usar el sistema desarrollado.

### **E.1. Introducción**

### **E.2. Requisitos de usuarios**

### **E.3. Instalación**

### **E.4. Manual del usuario**

---

## Bibliografía

---

- [1] CASTAÑO, F., FIDALGO, E., ALEGRE, E., CHAVES, D., AND SANCHEZ-PANIAGUA, M. State of the art: Content-based and hybrid phishing detection, 2021.
- [2] EUROPEAN UNION AGENCY FOR CYBERSECURITY. ENISA Threat Landscape. Report/Study, European Union Agency for Cybersecurity, 2023.
- [3] MOORE, M. Top Cybersecurity Threats in 2023 - University of San Diego Online Degrees, 2024. [Internet; accedido a 26/04/2024].