Pliego
Desarrollo de un sistema inteligente AntiBots

Fecha: 10/08/2020

Introducción

Este documento presenta la propuesta tecnológica de PROVEEDOR para el y desarrollo de un sistema inteligente AntiBots dentro del proyecto SABOT que Hacce quiere presentar a la convocatoria de Red.es en la que se abordarían las tareas que se describen en el anexo técnico vinculadas a la empresa subcontratada.

Oferta

La duración estimada para la realización del proyecto es de **12 meses**.
El coste total de desarrollo del proyecto se muestra en la siguiente tabla:

1. VISIÓN GLOBAL

Visión global

En estasección se detallan el plan de trabajo y el cronograma del proyecto SABOT. Además, se incluye tanto un listado de los indicadores de evolución del proyecto o entregables y de los hitos principales a alcanzar como un diagrama con las relaciones entre los paquetes de trabajo a realizar.

El alcance del proyecto prevé el desarrollo y puesta en marcha de un prototipo funcional de SABOT, cumpliendo todos los objetivos mencionados en la Sección 1.4.A.2. El prototipo se validará mediante un caso de uso desplegado en una empresa del sector textil, demostrando así la flexibilidad de la solución y su viabilidad. Para el arranque contamos con un prototipo de UEBA dirigido a la seguridad empresarial que permite detectar anomalías de comportamiento.

El presente proyecto se plantea como **un proyecto de 12 meses de duración**. Para la ejecución del mismo se sigue el esquema de maduración de tecnología TRL (*Technology Readiness Level*), aplicado en la industria y aceptado internacionalmente. De esta forma el proyecto se ha dividido en 4 paquetes de trabajo técnicos (además de otro paquete transversal que reúne todas las acciones de dirección y coordinación del proyecto), dando cobertura a los niveles del 3 al 7 del esquema TRL (ver Figura 6).

En el **segundo paquete de trabajo (PT2)** se realizará un estado del arte y se definirán los requisitos técnicos y funcionales de la solución. Además, se realizará el diseño de la plataforma que dará soporte a las funcionalidades de la herramienta.

En el **tercer paquete de trabajo (PT3)** se desarrollará la plataforma que dará soporte a todo el sistema, de acuerdo al diseño realizado en el paquete de trabajo anterior. Además, se desplegarán los módulos de extracción, ingesta y almacenamiento de la plataforma. El desarrollo y despliegue del módulo de visualización y de actuación también se llevará a cabo dentro de este paquete de trabajo.

El **cuarto paquete de trabajo (PT4)** incluye las actividad central de SABOT, es decir, el desarrollo del módulo de IA que permitirá la detección temprana de *bots*.

Finalmente, en el **último paquete de trabajo (PT5)** se validará el sistema desarrollado mediante dos vías. Por un lado, se realizarán un conjunto de pruebas técnicas que permitan verificar la adecuación del sistema a los requisitos definidos. Por otra parte, se realizará una validación de la herramienta mediante el despliegue en un piloto, con el fin de obtener *feedback* que pueda ser de interés para el desarrollo de la herramienta y de probar su uso en entornos reales.

Adicionalmente, se llevará a cabo un **paquete de trabajo transversal (PT1)** que abarcará toda la

duración del proyecto, como actividad de gestión y coordinación. Esta actividad requiere acciones permanentes y una continua monitorización de las actividades del proyecto, así como los

diferentes factores externos que puedan afectar al mismo. Las dependencias entre los paquetes de trabajo pueden observarse en la Figura 9.

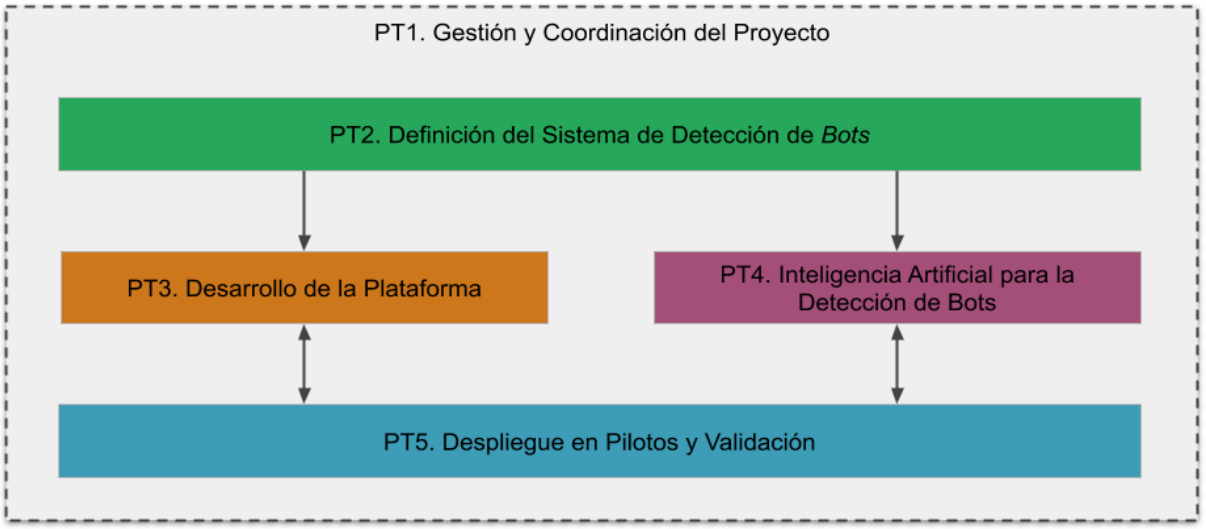

Figura 9. Diagrama PERT - Relación entre los paquetes de trabajo de SABOT

3.1.1. Cronograma

A continuación se presenta un diagrama Gantt exponiendo el tiempo de dedicación a los diferentes paquetes de trabajo y tareas durante el tiempo planificado para el proyecto.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
PT1. Gestión y coordinación del poyecto												
T1.1. Coordinación adminitrativa												
T1.2. Coordinación técnica												
PT2. Definición del sistema de detección de bots												
T2.1. Estado del arte												
T2.2. Definición de requisitos												
T2.3. Diseño del sistema												
PT3. Desarrollo de la plataforma												
T3.1. Extacción de características												
T3.2. Desarrollo e integración de la plataforma												
T3.3. Visualización y actuación												
T3.4. Despliegue de la plataforma												
PT4. Inteligencia Artificial para la detección de bots												
T4.1. Preprocesado de la información												
T4.2. Selección del algoritmo de IA												
T4.3. Impementación del algoritmo de IA												
T4.4. Evaluación del algoritmo de IA												
T4.5. Integración del módulo de IA												
PT5. Despliegue en pilotos y validación												
T5.1. Realización de pruebas												
T5.2. Despliegue en piloto												

Figura 10. Cronograma de SABOT

3.1.2. Indicadores de la evolución del proyecto

Los paquetes de trabajo definidos en el plan de trabajo del proyecto SABOT incluyen entregables documentales y en forma de prototipos/pruebas de concepto. El cumplimiento de estos entregables en las fechas estipuladas permitirá monitorizar la evolución del proyecto. A continuación se detallan los distintos hitos y entregables del proyecto en función del calendario de implementación definido, así como las relaciones entre los mismos.

ID	**Entregable/Hito**	**PT**	**Responsable**	**Tipo[1]**	**Fecha**
E1.1	Análisis de riesgos y plan de contingencia	PT1	Hacce	D	M3
E1.2	Informe final de gestión y coordinación del proyecto	PT1	Hacce	D	M12
E2.1	Estado del arte	PT2	Empresa Subcontratada	D	M2
E2.2	Análisis de requisitos y diseño del sistema	PT2	Hacce	D	M3
E3.1	Módulo de extracción de características	PT3	Hacce	P	M5
E3.2	Infraestructura y despliegue	PT3	Hacce	P	M11
E4.1	Algoritmo de detección de bots - versión inicial	PT4	Empresa Subcontratada	P	M8
E4.2	Algoritmo de detección de bots	PT4	Empresa Subcontratada	P	M11
E5.1	Resultados de las pruebas realizadas	PT5	Hacce	D	M12
E5.2	Descripción y resultados de los pilotos	PT5	Hacce	D	M12

1. D: documento; P: prototipo/demostrador

2. ESTRUCTURA DESAGREGADA DEL TRABAJO[2]

PT1. GESTIÓN Y COORDINACIÓN DEL PROYECTO			
Fecha inicio	M1	**Fecha fin**	M12
Duración	12 meses	**Participantes**	Hacce

Objetivo principal:

Este paquete de trabajo se centrará en la gestión y coordinación de todas las actividades del proyecto, asegurando que las diferentes tareas se ejecutan en plazo y de acuerdo con la planificación. La actividad permitirá por tanto establecer el marco organizativo para ejecutar el proyecto de acuerdo con la descripción del trabajo y de forma eficiente, coordinando los grupos de trabajo y la distribución de tareas entre los socios del consorcio. Además, se monitorizará el cumplimiento de los objetivos y tareas para asegurar la calidad de los resultados y, si fuese necesario, se propondrán las medidas correctoras necesarias. Por último, la actividad generará la información necesaria para la coordinación administrativa y gestionará la propiedad intelectual (IPR) que se pudiese derivar de los resultados del proyecto.

-

Descripción:

Tarea 1.1. Coordinación administrativa del proyecto (Hacce; M1-M12)

En esta tarea se realizará una gestión de las diferentes actividades del proyecto, asegurando que se ejecutan de acuerdo con la planificación establecida. Se llevará a cabo la realización de un análisis pormenorizado, identificando los principales riesgos que puedan impedir o retrasar la ejecución del proyecto, así como las estrategias a adoptar para mitigar los mismos. Además, se coordinarán también los grupos de trabajo y la distribución de tareas.

Tarea 1.2. Coordinación técnica (Hacce; M1-M12)

En esta tarea se llevará a cabo la gestión técnica y ejecución del proyecto. Seguirá de cerca el desarrollo del proyecto, comprobando la calidad de los resultados y gestionando todos los riesgos que aparezcan por parte de las actividades técnicas.

2. Utilizar la tabla tantas veces como paquetes de trabajo se definan.

- **Condiciones para el arranque:** ninguna.
- **Hitos principales:** Seguimiento de proyecto general
- **Recursos:**
- **Perfiles implicados (esfuerzos):** Especialista Gestion Senior (1750h)

Entregables:

E1.1. Análisis de riesgos y plan de contingencia (Hacce, M3): identificación de los principales riesgos que podrían afectar a la ejecución del proyecto y definición de un plan de contingencia. **E1.2. Informe final de gestión y coordinación del proyecto** (Hacce, M12)**:** documento con el resultado de las actividades de gestión y coordinación del proyecto.

PT2. DEFINICIÓN DEL SISTEMA DE DETECCIÓN DE BOTS

Fecha inicio	M1	**Fecha fin**	M3
Duración	3 meses	**Participantes**	Hacce, Empresa Subcontratada

Objetivo principal:

El objetivo que se pretende alcanzar con esta actividad es doble. Por un lado, determinar el estado del arte en los ámbitos que conciernen al proyecto y determinar las tecnologías de referencia a tener en cuenta en su desarrollo, para que puedan servir de base para la toma de decisiones en la fase de diseño. Por otro lado, definir los requisitos iniciales de SABOT en base a la experiencia de Hacce, así como el diseño de la plataforma que dará soporte a la solución. En esta actividad también será necesaria la contribución de Empresa Subcontratada para definir los requisitos y el diseño de la plataforma en base a su experiencia en el ámbito de la Inteligencia Artificial y ciberseguridad.

Descripción:

Tarea 2.1. Análisis de soluciones existentes (Empresa Subcontratada, Hacce; M1-M2)

En esta actividad se realizará un estado del arte en relación a las tecnologías del proyecto que incluye la búsqueda de soluciones comerciales existente para la detección de bots así como el estudio y validación de las contribuciones más académicas relacionadas con la temática. Además, se realizará un análisis de las herramientas open source existentes que se podrían integrar en cada uno de los módulos de SABOT, con el fin de complementar su funcionalidad.

Tarea 2.2. Definición de requisitos (Hacce, Empresa Subcontratada; M1-M2)

En esta actividad se realizará un análisis de los requisitos técnicos y funcionales del sistema desarrollado en el proyecto y de cada uno de los módulos que la componen.

Tarea 2.3. Diseño del sistema (Hacce, Empresa Subcontratada; M2-M3)

En esta actividad se realizará un diseño de la plataforma del proyecto, así como de cada uno de sus módulos individuales. Además, se realizará una selección de las tecnologías que se utilizarán como base para el desarrollo de la solución.

- **Condiciones para el arranque:** ninguna.
- **Hitos principales:** diseño del sistema.
- **Recursos:**
- **Perfiles implicados (esfuerzos):** Investigador-Ingeniero(150h), Especialista UX-Producto (450), Especialista Web (450h), Analista Senior (450h)

Entregables:

E2.1. Estado del arte (Empresa Subcontratada, M2): estado del arte de la temática del proyecto, incluyendo el análisis de las soluciones *open source* existentes que podrían ser integradas en la herramienta.
E2.2. Análisis de requisitos y diseño del sistema (Hacce, M3): descripción de los requisitos técnicos y funcionales de la plataforma, y diseño técnico de la misma.

PT3. DESARROLLO DE LA PLATAFORMA

Fecha inicio	M4	**Fecha fin**	M11
Duración	8 meses	**Participantes**	Hacce

Objetivo principal:

El objetivo de este paquete de trabajo es el desarrollo completo de la plataforma SABOT, una vez diseñada en el PT2. Esto incluye la preparación de la infraestructura que dará soporte a la plataforma, el despliegue automático de los módulos que forman parte de la herramienta, la extracción e ingesta de información de navegación así como el almacenamiento eficiente de la información recopilada. Se desarrollará también en este paquete de trabajo una interfaz gráfica responsiva que permitirá interactuar con la herramienta, visualizar métricas e información sobre la detección de bots y las alertas generadas.

Descripción:

Tarea 3.1. Extracción de características (Hacce; M4 - M5)

El objetivo de esta tarea es proporcionar un módulo que permita extraer información de navegación necesaria para alimentar a los modelos de Inteligencia Artificial del PT4. Se desarrollará un agente *Javascript* cuyo código se incluirá en todas las páginas, que para cada interacción con la *web* recopilando todos los datos requeridos por el módulo de Inteligencia Artificial y los enviará al módulo de almacenamiento. La solución implementada deberá ser lo suficientemente ligera como para no suponer una carga apreciable para el navegador, teniendo como requisito ser totalmente transparente para el usuario de la página.

Tarea 3.2. Desarrollo e integración de la plataforma (Hacce; M6-M9)

Esta tarea se centra en desarrollar la plataforma que permita integrar los diferentes módulos de SABOT. Se plantea la creación de una infraestructura *Cloud* y *Big Data* que almacene los datos recopilados en la tarea T3.1. de forma eficiente, y que de soporte a la ingesta posterior en tiempo real por el módulo de Inteligencia Artificial. El módulo de almacenamiento debe permitir realizar consultas eficientes sobre las grandes cantidades de datos que se van a almacenar. Además, se prevé la integración de un soporte de almacenamiento como base de datos estándar que permita almacenar los diferentes parámetros de control de acceso y configuración del módulo de visualización.

Tarea 3.3. Visualización y actuación (Hacce; M8-M11)

El módulo de visualización y actuación presentará métricas de todos los indicadores relevantes manejados por el sistema, en forma de gráficos e informes. Contará además con una sección de actuación, en la que se podrán definir umbrales o criterios que permitan neutralizar en tiempo real las amenazas detectadas, generando de manera automática reglas en el cortafuegos de la página monitorizada para bloquear el tráfico que la plataforma haya identificado como proveniente de *bots* maliciosos.

Tarea 3.4. Despliegue de la plataforma (Hacce; M10-M11)

Se plantea el despliegue de la infraestructura utilizando contenedores (basados en *Docker* u otras tecnologías similares) que darán soporte a los diferentes módulos desarrollados en el proyecto..

Esta tarea incluye la definición de los ficheros de configuración necesarios para automatizar el despliegue de los diferentes módulos.

- **Condiciones para el arranque:** definición de los datos que necesitará el módulo de inteligencia artificial, modelo de datos del módulo de almacenamiento y tener definido el sistema de detección (PT2).
- **Hitos principales:** diseño del sistema, recursos de procesado y almacenamiento Cloud, provisión de la plataforma, despliegue de la plataforma
- **Recursos:** servidor web con una réplica de la página desplegada en el entorno de producción
- **Perfiles implicados (esfuerzos):** Especialista FullStack (1050), Desarrollador Backend Sistemas (1050), Desarrollador BackEnd (800), Desarrollador BackEnd (1200), Especialista
  Front UX&QA (1050)

Entregables:

E3.1. Módulo de extracción de características (Hacce, M5)**:** prototipo del módulo de extracción de características procedentes de la navegación *web*.

E3.2. Infraestructura y despliegue (Hacce, M11)**:** configuración de infraestructura y automatización de despliegue de la plataforma.

PT4. INTELIGENCIA ARTIFICIAL PARA LA DETECCIÓN DE BOTS			
Fecha inicio	M4	**Fecha fin**	M11
Duración	8 meses	**Participantes**	Empresa Subcontratada, Hacce

Objetivo principal:

El objetivo principal de este paquete de trabajo es el desarrollo del módulo de inteligencia artificial de SABOT. Por un lado, se realizará la selección del algoritmo utilizar para la detección de *bots*, partiendo del documento de estado del arte del entregable E1 y llevando a cabo las pruebas necesarias. Además, se realizarán las tareas de preprocesado de los datos extraídos en la tarea T3.1, que nos permitan identificar y generar las variables que puedan ser relevantes para la detección de *bots* en la plataforma *web*.

Descripción:

Tarea 4.1. Preprocesado de la información (Empresa Subcontratada, Hacce; M4-M6)

En esta tarea se realizarán diferentes actividades de análisis y de preprocesamiento sobre los datos extraídos en la tarea T3.1. El objetivo es prepararlos y optimizarlos para que puedan ser utilizados por el algoritmo de IA que seleccionaremos a continuación. Dado que los datos pueden proceder de diferentes fuentes, se realizará en primer lugar un proceso de normalización de los mismos con el fin de convertirlos a un formato estándar. A continuación, se analizará el contenido procurando identificar aquellas variables (o generando otras nuevas a partir de ellas) que resulten más significativas para modelar el comportamiento de un usuario en la plataforma *web*, y que por tanto resulten relevantes para determinar si se trata de un humano o un *bot*.

Tarea 4.2. Selección del algoritmo de IA (Empresa Subcontratada; M5-M10)

Partiendo como base del estado del arte alcanzado en el entregable E1, en esta tarea se seleccionará el algoritmo a utilizar para la detección de bots. Para ello se plantea la aplicación de técnicas UEBA (*User and Entity Behaviour Analytics*), modelando el comportamiento de los usuarios de la plataforma *web* con el fin de poder determinar si se trata de un humano o de un *bot*. Se tomarán como base las características extraídas en la tarea T3.1 y se aplicarán técnicas de selección de variables que nos permitan decidir qué características son las más relevantes para obtener los mejores resultados de detección.

Tarea 4.3. Implementación del algoritmo de IA (Empresa Subcontratada; M9-M11)

En esta tarea se implementará el algoritmo de IA seleccionado en la tarea anterior (T4.2) que permita realizar detecciones avanzadas de bots. Este algoritmo será desarrollado en un lenguaje de programación como R o Python. Ambos lenguajes son muy versátiles y extensibles mediante librerías o paquetes.

Tarea 4.4. Evaluación del algoritmo de IA (Empresa Subcontratada; M10-M11)

Esta tarea tiene como fin determinar la capacidad del algoritmo implementado en la tarea T4.3 para detectar la presencia de *bots* en la plataforma web. Se utilizarán para ello distintas métricas como pueden ser la *accuracy*, Área Bajo la Curva (AUC), *recall*, *F1 score*, etc.

Tarea 4.5. Integración del módulo de IA (Empresa Subcontratada; M10-M11)
En esta tarea se llevará a cabo la integración del módulo de IA en la plataforma. Además, se definirán y realizarán las pruebas oportunas que permitan verificar el correcto funcionamiento del módulo integrado.

- **Condiciones para el arranque:** tener definido el sistema de detección (PT2).
- **Hitos principales:** algoritmo de detección de *bots*.
- **Recursos:**
- **Perfiles implicados (esfuerzos):** Investigador-Ingeniero (900h), Desarrollador senior (900)

Entregables:

E4.1. Algoritmo de detección de *bots* - versión inicial (Empresa Subcontratada, M8)**:** primera versión de la implementación del algoritmo para la detección automática de *bots*

E4.2. Algoritmo de detección de *bots* (Empresa Subcontratada, M11)**:** implementación final del algoritmo para la detección automática de *bots*

PT5. DESPLIEGUE EN PILOTOS Y VALIDACIÓN			
Fecha inicio	M4	**Fecha fin**	M12
Duración	9 meses	**Participantes**	Hacce

Objetivo principal:

El objetivo de esta actividad será la verificación y validación de la solución SABOT, con el fin de demostrar la versatilidad y flexibilidad de la herramienta desarrollada. Para ello se identificarán los posibles aspectos de mejora de la solución, así como cualquier requisito o funcionalidad que no se haya identificado inicialmente pero que se debería incluir como parte de la solución final.
La validación se realizará por medio del despliegue en un piloto con una empresa del sector textil, 4Elementos. El *feedback* obtenido en este piloto se utilizará para mejorar el sistema, orientándose así hacia las necesidades reales de los clientes.

Descripción:

Tarea 5.1. Realización de pruebas (Hacce; M4-M12)

En esta tarea se incluye la definición y realización de las pruebas que permitirán verificar el correcto funcionamiento de la solución SABOT.

Tarea 5.2. Despliegue en piloto (Hacce; M11-M12)

Esta actividad incluye la implantación de la herramienta SABOT en un piloto. Este piloto ya está identificado y pertenece al sector textil, en concreto se trata de un negocio con comercio electrónico centrado en la venta de zapatillas deportivas de marca.

- **Condiciones para el arranque:** tener definido el sistema de detección (PT2).
- **Hitos principales:** piloto desplegado.
- **Recursos:** Servidores web de plataforma para despliegues.
- **Perfiles implicados (esfuerzos):** Especialista Sistemas (300h), Investigador-Ingeniero (150)

Entregables:
E5.1 Resultados de las pruebas realizadas (Hacce, M12): documento en el que se indican las pruebas realizadas sobre la plataforma y los resultados obtenidos de las mismas. **E5.2 Descripción y resultados de los pilotos** (Hacce, M12): documento en donde se describe el piloto realizado y los resultados de su implantación.