

SemantiQ AI

Sistema d'Anàlisi Intel·ligent de Documents

Lorenzo Cremonese
2º DAW
2024/25

I. Introducció

SemantiQ AI és un sistema d'anàlisi intel·ligent de documents que integra intel·ligència artificial per extreure informació i respondre preguntes basant-se en documents carregats pels usuaris. Mitjançant l'ús d'un sistema *Retrieval-Augmented Generation (RAG)*, el projecte combina la recuperació d'informació amb la generació de text per proporcionar respostes més precises i contextualitzades. L'objectiu principal és optimitzar la consulta i gestió de documents en entorns acadèmics i empresarials, facilitant l'accés a la informació de manera automatitzada.

1. Mòduls als que implica

Per al desenvolupament del projecte s'han aplicat coneixements de diferents mòduls del cicle formatiu **Desenvolupament d'Aplicacions Web**, cadascun amb un grau d'aplicació diferent:

- a. **Llenguatge de Marques i Gestió de la Informació:**
 - Ús d'HTML per estructurar dades i millorar la interoperabilitat del sistema.
- b. **Base de Dades:**
 - Implementació de MariaDB per a l'emmagatzematge de documents, metadades i registres de consultes.
- c. **Programació:**
 - Desenvolupament del backend amb Laravel per gestionar la interacció amb la base de dades i la IA.
- d. **Desplegament d'Aplicacions Web:**
 - Posada en marxa del sistema en un VPS per garantir escalabilitat i disponibilitat.
- e. **Desenvolupament Web en Entorn Client:**
 - Creació de la interfície amb React i Bootstrap per millorar l'experiència d'usuari.
- f. **Desenvolupament Web en Entorn Servidor:**
 - Implementació de l'API REST en Laravel per gestionar les peticions del frontend.
- g. **Disseny d'Interfícies Web:**
 - Creació d'una interfície intuïtiva que facilita la consulta i interacció amb el sistema.
- h. **Empresa i Iniciativa Emprenedora:**
 - Estudi del potencial de la solució en el mercat i possibles aplicacions en el sector empresarial.
 - Es tractaran aspectes relacionats amb l'ergonomia i els riscos laborals associats al desenvolupament d'aplicacions.

2. Tipus del Projecte

Aquest projecte té un enfocament **tècnic**, ja que proporciona una solució basada en tecnologies d'intel·ligència artificial i desenvolupament web. No es tracta d'un estudi d'una empresa ni de la seua millora, sinó d'una aplicació que ofereix una forma de gestionar i consultar informació documental de manera intel·ligent.

La solució té aplicacions en **entorns acadèmics, empresarials i tecnològics**, ja que permet millorar la cerca i gestió de grans volums de documentació mitjançant IA.

3. Descripció del Projecte

SemantiQ AI és una aplicació web que permet als usuaris carregar documents en diversos formats per analitzar-los mitjançant **intel·ligència artificial**. El sistema utilitza un model *RAG (Retrieval-Augmented Generation)* per millorar la qualitat de les respostes generades, combinant la recuperació d'informació amb la generació de text automàtica.

Un model *RAG (Retrieval-Augmented Generation)* és un tipus de model d'intel·ligència artificial que combina dues tècniques: la recuperació d'informació i la generació de text. El model cerca primer informació rellevant dins d'un conjunt de documents o bases de dades (recuperació d'informació) i, després, utilitza aquesta informació per generar respostes o resums contextuals (generació de text). Aquest enfocament permet millorar la precisió i rellevància de les respostes en comparació amb models purament generatius que no utilitzen informació externa.

El funcionament del projecte es pot dividir en diverses fases:

a. Pujada de documents:

- L'usuari carrega documents a la plataforma, que són emmagatzemats en una base de dades MariaDB.

b. Processament de la informació:

- Els documents són indexats per facilitar-ne la recuperació i consulta posterior.

c. Consulta per part de l'usuari:

- L'usuari introdueix preguntes relacionades amb el contingut dels documents carregats.

d. Generació de respostes mitjançant la IA:

- El sistema *RAG* cerca informació rellevant en els documents i genera una resposta contextualitzada amb l'ajuda de **OpenAI**.

e. Visualització de la resposta:

- La resposta és retornada a l'usuari a través d'una interfície interactiva.

Opcionalment, el projecte pot integrar:

- a. **Un sistema d'autenticació (OAuth2)** per restringir l'accés als documents.
- b. **Generació de gràfics i estadístiques** sobre l'ús del sistema i les consultes realitzades.

Amb aquest enfocament, **SemantiQ AI** facilita la gestió de la informació de manera més eficient i intel·ligent, millorant la productivitat en diferents àmbits.

II. Memòria explicativa

Aquest apartat té com a objectiu aprofundir en el procés de creació de **SemantiQ AI** des d'una perspectiva tècnica i metodològica. S'hi detallen les decisions preses durant el desenvolupament, les alternatives considerades i els motius que han portat a optar per una determinada arquitectura, tecnologies o eines. També s'inclou una anàlisi dels requisits de l'aplicació i una planificació temporal del projecte, així com els esquemes i models de disseny més rellevants per entendre el funcionament intern del sistema.

1. Anàlisi. Justificació

Anàlisi de necessitats i objectius:

El projecte SemantiQ AI naix de la necessitat creixent de gestionar grans volums d'informació documental en entorns acadèmics i empresarials. L'accés ràpid i intel·ligent a la informació és un recurs de valor inqüestionable, i les tecnologies d'intel·ligència artificial actuals ofereixen les eines adequades per abordar aquest repte. El principal objectiu és proporcionar una plataforma web que permeti consultar documents mitjançant preguntes formulades en llenguatge natural, generant respostes contextualitzades i precises.

Justificació de les solucions adoptades:

S'han estudiat diverses alternatives per implementar la funcionalitat principal del sistema:

- a. **Sistemes tradicionals de cerca (basats en paraules clau):** descartats per la seua manca de capacitat per entendre el context o generar respostes comprensibles.
- b. **Models de llenguatge sense RAG (purs):** considerats, però presentaven limitacions a l'hora de mantenir-se fidels al contingut real dels documents carregats.
- c. **Retrieval-Augmented Generation (RAG):** escollit per la seua capacitat de combinar la recuperació precisa d'informació amb la generació de respostes contextualitzades. Aquesta solució millora l'exactitud i la utilitat de les respostes.

Criteris tècnics i organitzatius:

a. Tecnologies web modernes:

- **Laravel** s'ha escollit per la seua robustesa, gran comunitat de suport, estructura clara MVC i la seua facilitat per integrar-se amb sistemes com Sanctum o APIs REST.
- **React** ofereix una interfície moderna, flexible, ràpida i mantinguda per una comunitat molt activa. És multiplataforma i compatible amb múltiples frameworks d'estil (com Bootstrap).

b. Emmagatzematge relacional amb MariaDB: MariaDB destaca per ser una base de dades relacional de codi obert molt compatible amb MySQL i àmpliament adoptada en entorns de producció.

c. Dockerització del sistema: Docker facilita la portabilitat entre entorns i assegura que el desplegament siga consistent i ràpid.

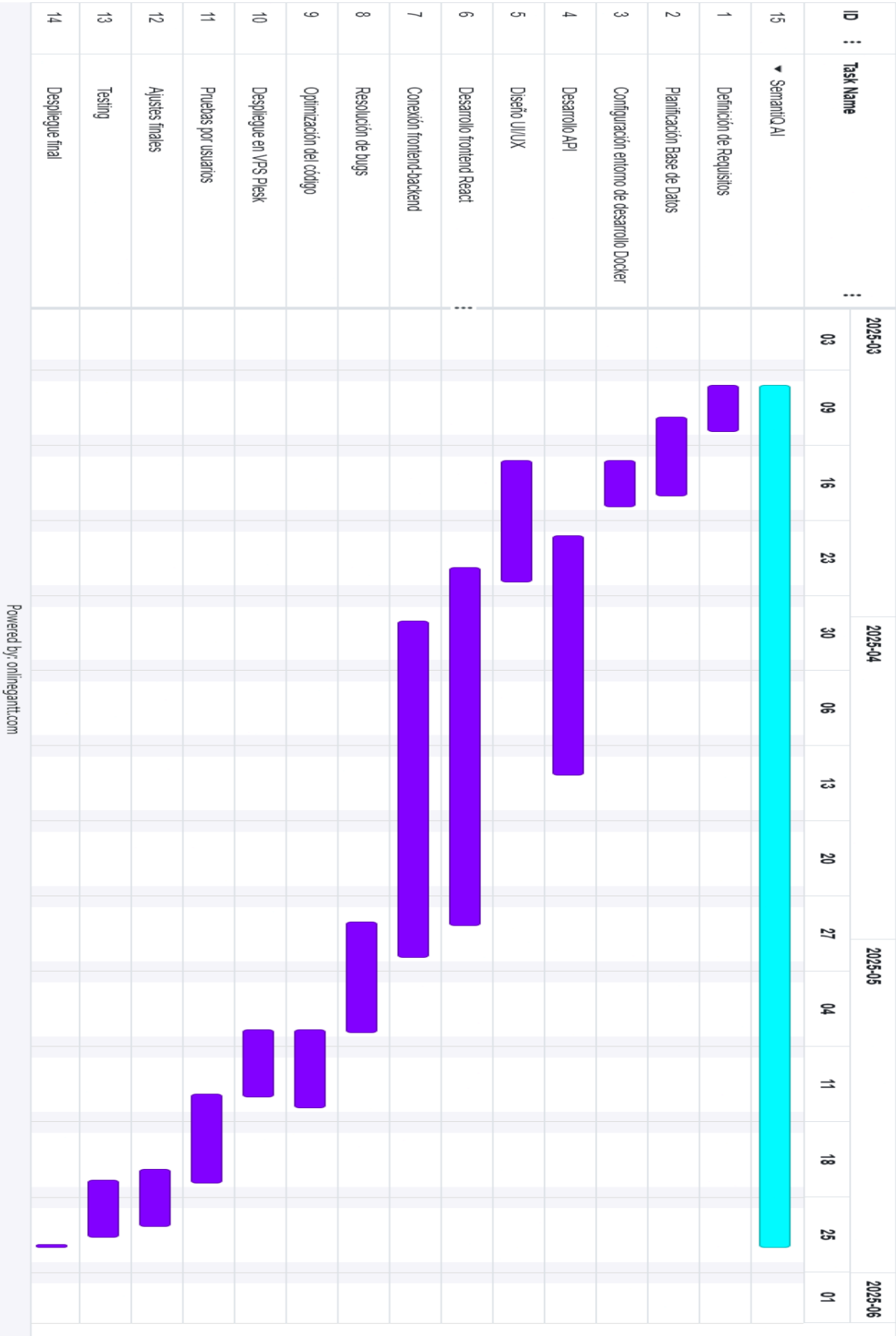
d. Desplegament en VPS (futur): considerat per oferir independència, escalabilitat i seguretat.

e. Criteris econòmics: s'han seleccionat tecnologies de codi obert i solucions autoalojades no sols per a reduir costos, sinó que també permeten una millora contínua gràcies a les actualitzacions comunitàries.

Estratègia de posicionament SEO

Per tal de millorar la visibilitat del projecte en entorns digitals, s'ha implementat una estratègia bàsica de posicionament SEO (Search Engine Optimization). Aquesta inclou l'ús de metadades adequades, estructures HTML semàntiques, rutes amigables per al navegador i l'ús de paraules clau relacionades amb la gestió intel·ligent de documents. També es preveu la integració amb eines com Google Search Console i Google Analytics per monitoritzar i optimitzar la presència en cercadors. Aquesta estratègia pot facilitar una major captació d'usuaris, especialment en l'àmbit educatiu i empresarial.

Planificació temporal:



2. Disseny del Sistema

Bases de dades:

S'ha dissenyat una estructura de base de dades relacional per MariaDB, amb les següents taules principals:

- a. **users:** id, name, email, password
- b. **files:** id, name, type, size, path, user_id
- c. **chunks:** id, file_id, content
- d. **messages:** id, user_id, question_id, message

Nota: la base de dades està optimitzada per a recuperar ràpidament informació rellevant per al sistema RAG, utilitzant el índex FULLTEXT en la columna 'content' de la taula 'chunks'.

Diagrama Entitar-Relació:

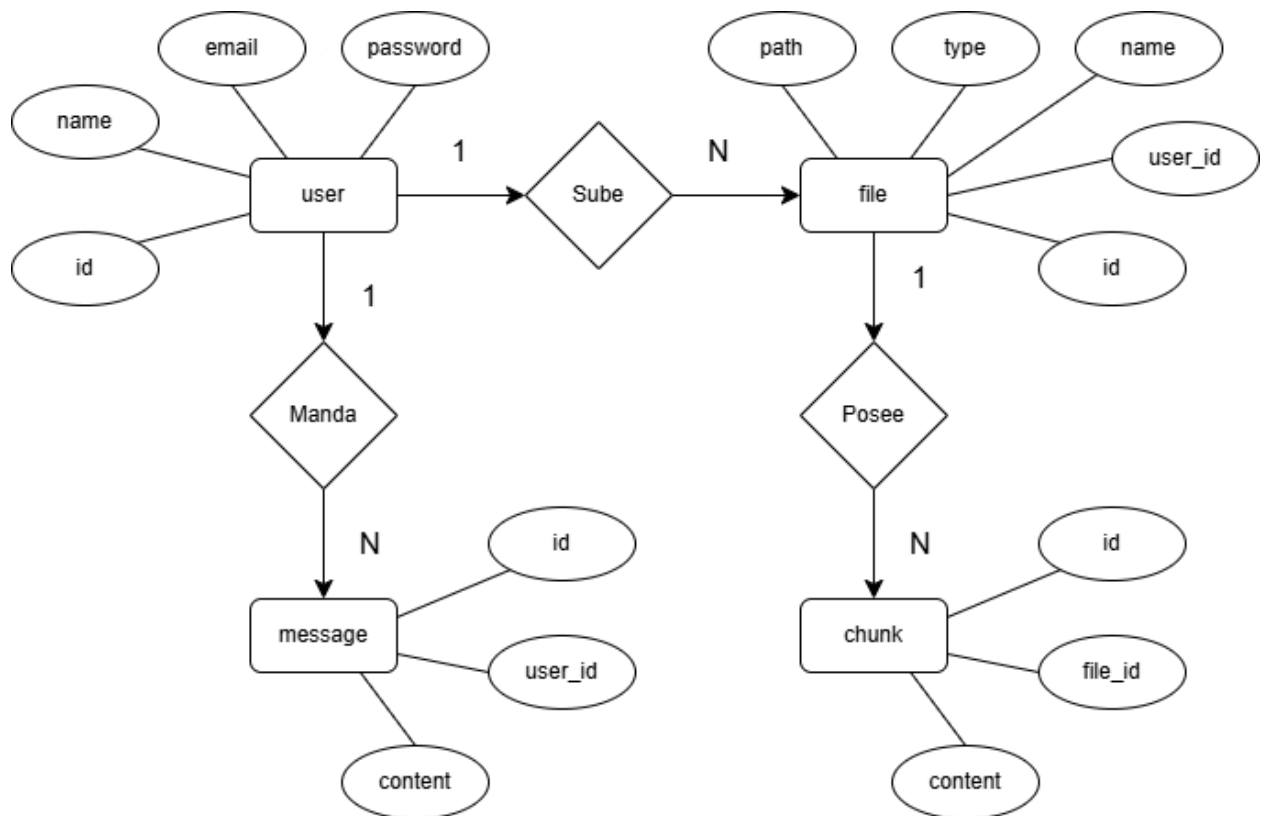
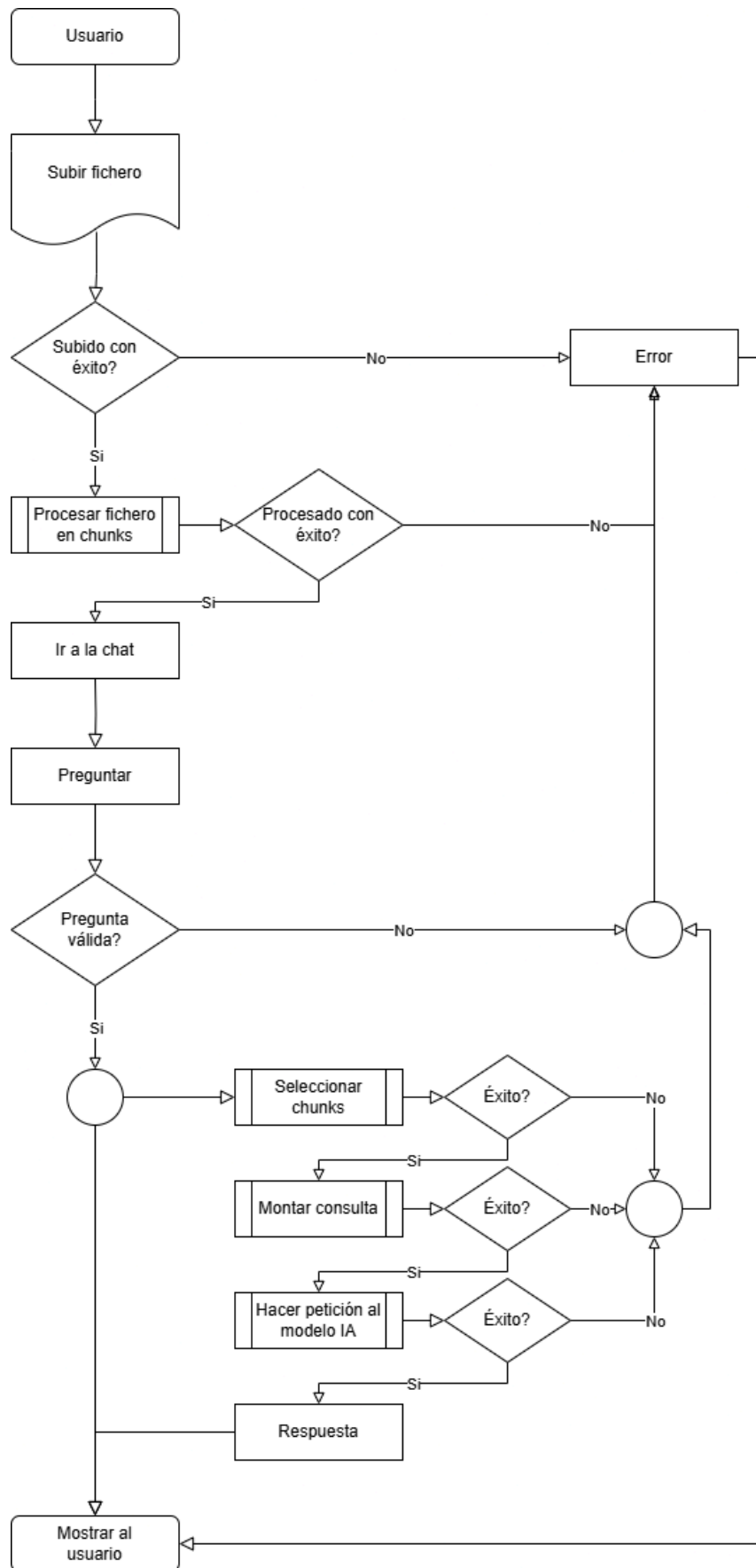


Diagrama de flux:



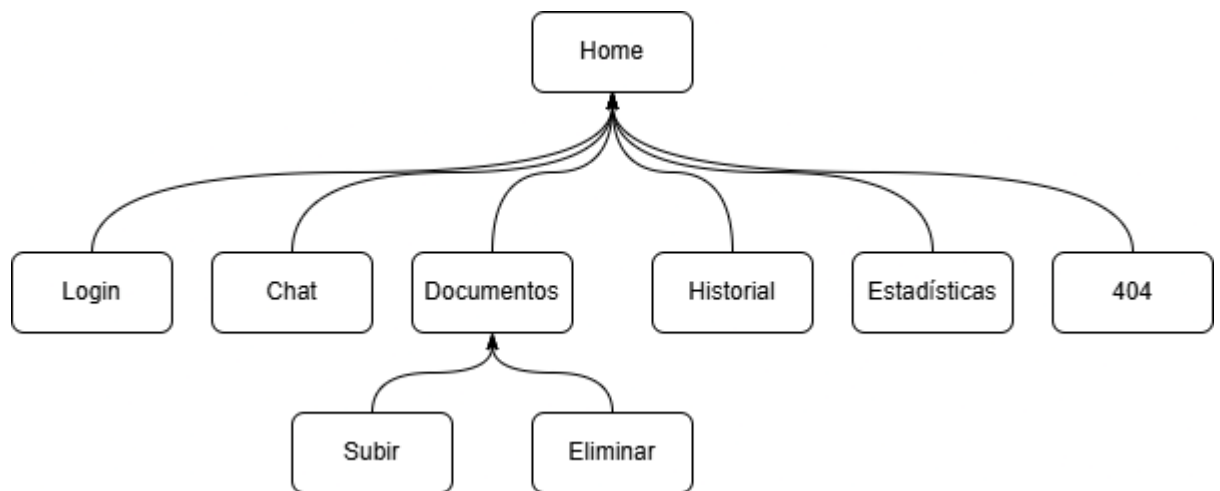
Mockups i UI

- **Paleta de colors:** Guia visual de la identitat gràfica del projecte.
- **Logo:** Element representatiu del projecte.

Pàgina	Descripció	Objectiu
Home	Pantalla d'inici amb informació general, accés a funcionalitats principals i invitació a pujar documents	Presentar el servei i facilitar l'accés ràpid a les funcionalitats de la aplicació
Documents	Secció per gestionar els arxius pujats	Permetre la gestió dels arxius pujats per l'usuari (afegir i eliminar)
Chat	Interfície central de consulta mitjançant IA, amb entrada de preguntes i àrea de resposta dinàmica	Interactuar amb la IA per fer preguntes i obtenir respostes contextualitzades
Historial	Llistat de preguntes realitzades i respostes obtingudes per dia	Mostrar les preguntes i respostes fetes anteriorment per dia
Login	Pantalla d'autenticació per a usuaris registrats	Autenticació d'usuaris registrats per accedir a funcionalitats personalitzades
Estadístiques	Dashboard interactiu amb gràfics d'ús, consultes més freqüents i anàlisi de comportament dels usuaris	Visualitzar mètriques sobre l'ús del sistema i patrons de consulta
404	Pàgina d'error personalitzada que manté la coherència visual del sistema	Indicar errors de navegació amb un estil coherent amb la resta de l'aplicació

Aquest conjunt de pantalles permet cobrir de manera integral totes les funcionalitats del sistema, garantint una experiència coherent i eficient. Es poden visualitzar els mockups en l'[Annex 1](#).

Mapa web



Tecnologies usades:

Component	Tecnologia	Versió
Backend	Laravel + PHP	Laravel 12, PHP 8.4
Base de dades	MariaDB	11.7
Contenidors	Docker	27
Frontend	React + Bootstrap	React 18.2, Bootstrap 5.3
IA	OpenAI API + sistema RAG propi	gpt-4o-mini
Autenticació	Laravel Sanctum, OAuth2	-
Estadístiques	Recharts	2.15

III. Anàlisi Econòmic

Viabilitat econòmica

Model de negoci:

SemantiQ AI es pot oferir sota un model freemium:

1. Versió gratuïta:

- a. Limitació de documents pujats per mes.
- b. Respostes limitades per dia.
- c. Sense accés a estadístiques ni funcionalitats avançades.

2. Subscripció mensual (model SaaS):

- a. **Pla Bàsic:** 9,99€/mes Inclou fins a 100 documents pujats mensuals i consultes il·limitades. No inclou estadístiques avançades ni integració amb serveis externs.
- b. **Pla Pro:** 24,99€/mes Accés a estadístiques, prioritat en temps de resposta, integració amb altres eines (Drive, Dropbox).
- c. **Pla Empresarial (a mida):** Per a organitzacions amb grans volums de documentació i requisits especials (hosting dedicat, privacitat avançada, etc.).

Altres vies d'ingrés potencials:

- Llicències d'ús per a centres educatius o empreses.
- Implementacions personalitzades.
- Publicitat en la versió gratuïta (opcional).

Comparativa de competència

Producte	Preu mensual	Característiques principals	Limitacions principals	Valor afegit de SemantiQ AI
ChatPDF	Gratuït / 15 €	IA per fer preguntes sobre PDFs, càrrega ràpida, interfície simple	Només admet PDF, no permet integracions ni anàlisi d'ús	Suport a múltiples formats, backend personalitzat
NoteBookLM	9,99 €	Suport a múltiples formats, integracions amb altres eines	Limitat a documents acadèmics, no tan flexible com altres solucions	Integració amb diferents APIs
AskYourPDF	9,90 €	Consulta IA sobre PDF, interfície senzilla	Limitació de formats, publicitat	Sense anuncis, adaptable, sistema d'autenticació
SemantiQ AI	Gratuït / 9,99 € / 24,99 €	Multi-format, RAG + OpenAI, estadístiques	-	Arquitectura pròpia, escalable

Previsió d'ingressos (exemple primer any):

Tipus d'usuari	Nombre estimat	Preu mensual	Total anual
Usuaris bàsics (free)	300	0 €	0 €
Usuaris Pla Bàsic	100	9,99 €	11.988 €
Usuaris Pla Pro	20	24,99 €	5.997,6 €
Empreses (3 clients)	3	99 €/mes (mínim)	3.564 €
TOTAL	-	-	21.549,6 €

Aquesta previsió és optimitzable i pot créixer amb una bona estratègia de màrqueting i boca-orella.

Anàlisi de costos

Costos de desenvolupament (estimació pròpia):

Element	Temps estimat	Cost per hora	Total
Disseny i prototipat	20 h	25 €/h	500 €
Programació backend (Laravel)	60 h	25 €/h	1500 €
Integració amb IA / RAG	40 h	25 €/h	1000 €
Programació frontend (React + Bootstrap)	50 h	25 €/h	1250 €
Proves i optimització	20 h	25 €/h	500 €
Documentació i desplegament	20 h	25 €/h	500 €
Total hores treballades	200 h		5250 €

Elecció del hosting:

Proveïdor	Pla	Preu mensual	Espai Disc	RAM (GB)	CPU	Llargada de contracte	Altres característiques
IONOS	VPS L	10 €	50 GB SSD	2	1 vCPU	12 mesos	Còpia de seguretat, SSL inclòs, API d'integració
Hostinger	VPS 1	9,99 €	40 GB SSD	2	1 vCPU	12 mesos	Backup setmanal, accés a root, firewall gestionat
Arsys	VPS Starter	15 €	50 GB SSD	2	2 vCPUs	12 mesos	IP dedicat, suport 24/7, protecció contra DDoS
Hetzner	CX11	4,90 €	20 GB SSD	2	1 vCPU	12 mesos	IP dedicat, alta flexibilitat, protecció contra DDoS

Després d'analitzar diverses opcions d'hosting, es considera que el IONOS VPS L és una bona opció, amb 50 GB SSD i 2 GB de RAM a un preu competitiu. L'elecció final depèn del creixement dels usuaris i de les necessitats de capacitat. El preu baix de IONOS pot amagar limitacions de processament a mesura que creix la demanda, pel que cal valorar l'escalabilitat a llarg termini.

Costos operatius anuals (previsió mínima):

Recurs	Cost estimat
Hosting VPS bàsic (IONOS VPS L)	120 €/any
Domini	15 €/any
API OpenAI	A partir de 0,15 €/milió de tokens. El cost real dependrà del volum de consultes. Estimació mínima: 25 €/mes → 300 €/any
Certificat SSL (Let's Encrypt)	0 €
Eines de desenvolupament (Figma, GitHub, etc.)	0 € (plans gratuïts)
Total anual estimat	435 €

El cost de l'API de OpenAI és variable. El model utilitzat (GPT-4-turbo) té un preu de 0,15 € per cada milió de tokens processats, incloent entrada i resposta. Això vol dir que en entorns amb ús intensiu, aquest import pot augmentar significativament.

Altres costos possibles:

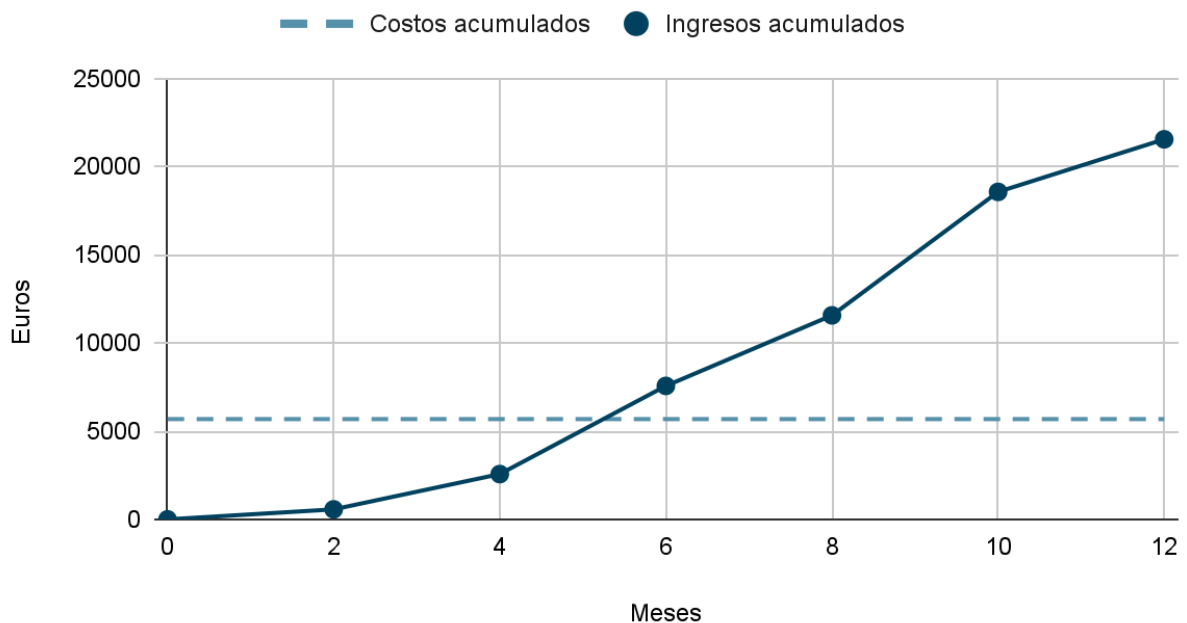
a. Llicències o serveis premium (opcional):

- Bases de dades vectorials avançades.
- Sistemes d'autenticació externa amb OAuth empresarial.
- Backups i serveis de monitorització.

S'ha prioritzat l'ús de tecnologia lliure i de codi obert per minimitzar despeses.

Conclusió econòmica:

Recuperación de la inversión y beneficios - Primer año



El projecte **SemantiQ AI** és viable econòmicament. Amb una inversió inicial moderada i costos recorrents baixos, té potencial per generar ingressos estables mitjançant un model *SaaS (Software as a Service)* escalable. A més, el mercat objectiu (educació, empreses tecnològiques, consultories, etc.) està en creixement i és sensible a solucions d'optimització intel·ligent de la informació.

IV. Ampliacions i/o millores

SemantiQ AI és un projecte amb una base sòlida però oberta a múltiples ampliacions i millores en futures versions. Aquestes opcions podrien incrementar-ne la funcionalitat, la usabilitat i el valor diferencial respecte d'altres productes similars.

a. Funcionalitats tècniques

- **Sistema de respostes multimodals:** integració d'entrada i consulta per veu o imatge (OCR), per tal de facilitar l'accessibilitat i ampliar l'ús en dispositius mòbils.
- **Base de dades vectorial dedicada:** per millorar l'eficiència en la recuperació semàntica i permetre consultes a gran escala amb menor latència.
- **Integració amb plataformes externes:** Google Drive, Dropbox, OneDrive, Moodle, etc. per facilitar la sincronització automàtica de documents.
- **API pròpia documentada:** per permetre que terceres aplicacions puguin fer consultes directament a SemantiQ AI.
- **Implementació d'API de models d'IA:** integració amb diversos models d'intel·ligència artificial externs, com ara OpenAI, Gemini, Copilot i altres, per ampliar la capacitat del sistema i millorar la qualitat de les respostes generades, oferint així més opcions i personalització per als usuaris.

b. Experiència d'usuari

- **Millores a l'historial:** filtres per data, document o tipus de consulta, exportació de consultes i respostes en PDF o CSV.
- **Sistema de notificacions:** alertes per finalització d'anàlisi o noves funcionalitats.

c. Seguretat i privacitat

- **Xifratge avançat de documents:** per a garantir la seguretat de la informació en entorns corporatius o educatius.
- **Controls de privacitat per document:** gestió de permisos per accedir a cada document en entorns compartits.

V. Prevenció de Riscos

Durant el desenvolupament del projecte SemantiQ AI s'han identificat diversos riscos relacionats tant amb el desenvolupament tècnic com amb l'entorn de treball físic i virtual dels desenvolupadors i usuaris. Per a minimitzar aquests riscos, s'ha elaborat un pla de prevenció que cobreix les següents àrees:

Riscos tècnics

- **Seguretat de dades:** Risc potencial de pèrdua o fuga de dades sensibles contingudes en els documents carregats.
 - **Pla de prevenció:** Implementació d'un sistema d'autenticació robust amb Laravel Sanctum i OAuth2, ús de connexions segures (HTTPS) i, en el futur, integració de xifrat avançat de documents per garantir privacitat i seguretat.
- **Integritat del sistema:** Possibles errors o interrupcions en la comunicació entre el frontend (React) i el backend (Laravel).
 - **Pla de prevenció:** Desenvolupament d'un sistema robust d'integració contínua, control de versions amb GitHub, així com la implementació de Docker per garantir estabilitat i facilitar la resolució ràpida de problemes.
- **Disponibilitat i escalabilitat del servei:** Risc de saturació o interrupció del servei per un augment inesperat del nombre d'usuaris.
 - **Pla de prevenció:** Desplegament en un servidor VPS escalable per gestionar adequadament càrregues de treball elevades.

Ergonomia del lloc de treball

- **Postura correcta:** Possibilitat de lesions o molèsties derivades de llargues sessions de treball en posicions inadequades.
 - **Pla de prevenció:** Mantenir una postura ergonòmica amb cadires ajustables, monitor a l'altura adequada dels ulls i pauses curtes cada hora per fer estiraments i descans visual.
- **Fatiga ocular i dolors de cap:** Fatiga ocular i dolors de cap derivats de llargs períodes davant del monitor.
 - **Pla de prevenció:** Utilització de ulleres amb filtres de llum blava per reduir l'impacte negatiu de la llum del monitor sobre la vista i disminuir l'aparició de dolors de cap.
- **Entorn de treball adequat:** Molèsties visuals o cansament causat per il·luminació inadequada o mala ventilació.
 - **Pla de prevenció:** Garantir una il·luminació òptima i adequada ventilació en l'espai de treball, així com fomentar un entorn net i ordenat.

- **Organització del lloc de treball:** Ineficiència per desorganització del material o de l'entorn immediat.
 - **Pla de prevenció:** Disposar l'espai de treball de manera ordenada i funcional, utilitzant suports ergonòmics per a teclats i monitors, i mantenint lliure l'àrea de treball d'elements innecessaris.

Ergonomia de l'aplicació

- **Disseny visual:** Fatiga visual o incomoditat causada per una interfície mal dissenyada.
 - **Pla de prevenció:** Utilització de tipografies clares, llegibles i d'una mida adequada. Selecció d'una paleta de colors que proporciona un contrast visual suficient per a la lectura còmoda i la reducció de la fatiga ocular.
- **Interacció usuari-sistema:** Confusió o dificultat en l'ús derivada d'una mala estructuració dels elements interactius.
 - **Pla de prevenció:** Implementació d'una interfície intuïtiva amb una estructura lògica i coherent, facilitant l'accés a funcionalitats essencials mitjançant mockups dissenyats amb React i Bootstrap per optimitzar l'experiència d'usuari en diferents dispositius (ordinador, tauleta i mòbil).

Aquest conjunt de mesures assegura tant la qualitat tècnica del projecte com un entorn saludable i una experiència òptima per als usuaris finals.

- VI. Avaluació del Projecte**
- VII. Bibliografia**

Annex 1



Il·lustració 1.1: Paleta de colors



Il·lustració 1.2: Logo



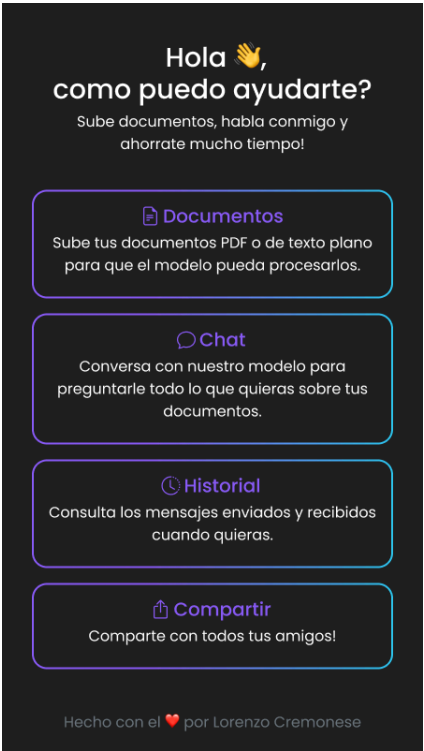
Il·lustració 1.3: Home Desktop



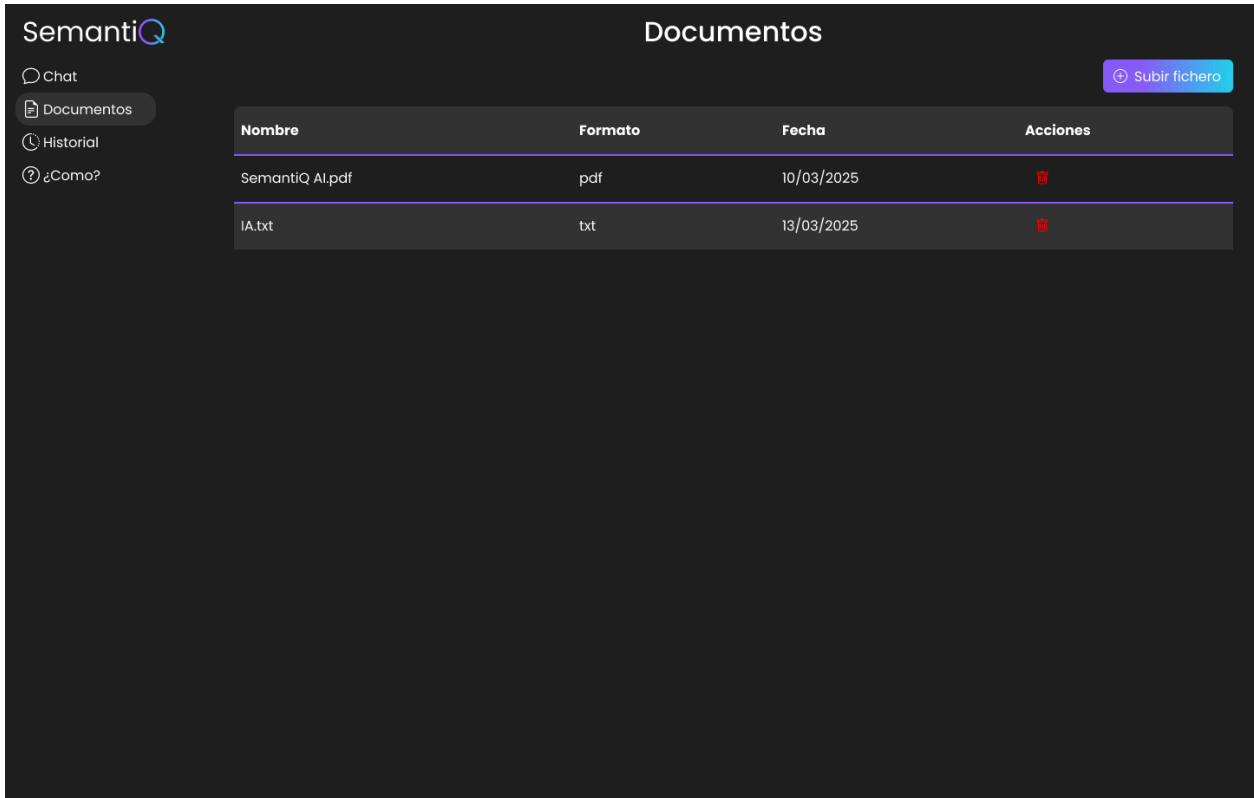
Il·lustració 1.4: Home Desktop Hover



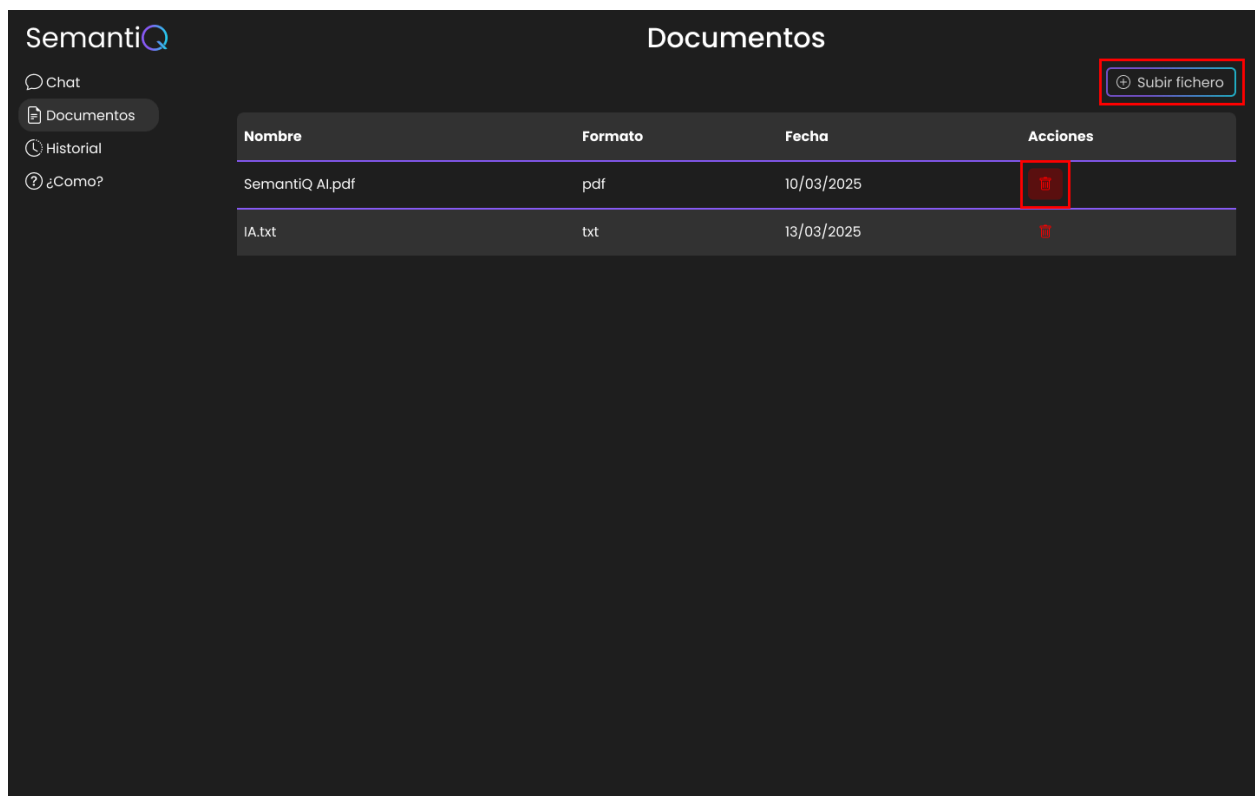
Il·lustració 1.5: Home Tablet



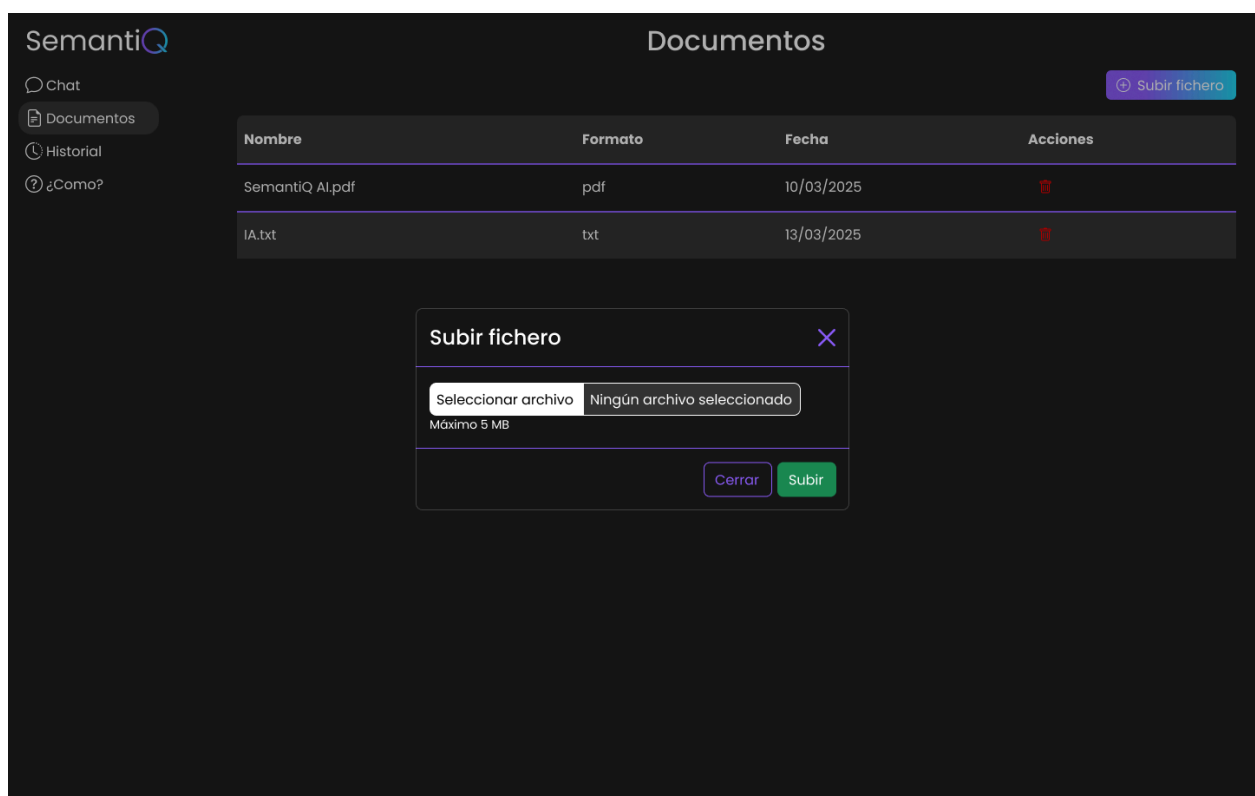
Il·lustració 1.6: Home Mobile



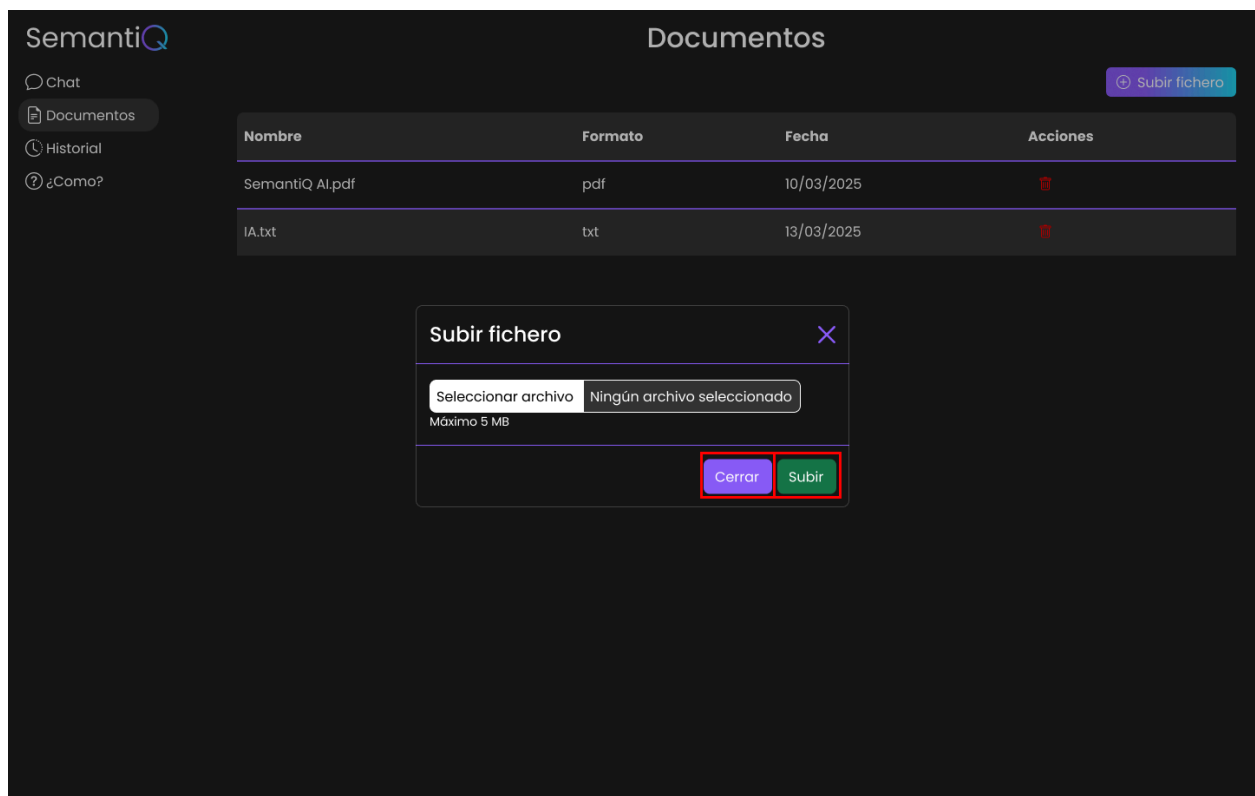
Il·lustració 1.7: Documentos Desktop



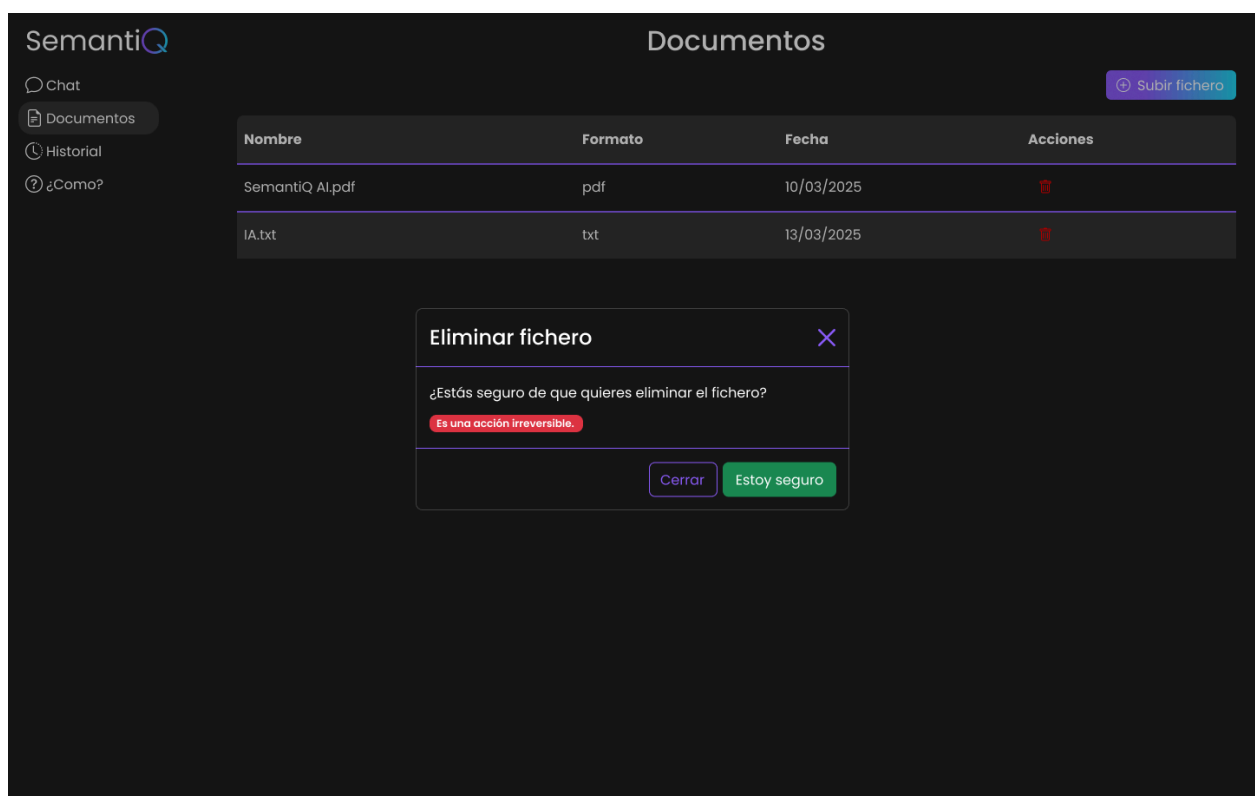
Il·lustració 1.8: Documentos Desktop Hover



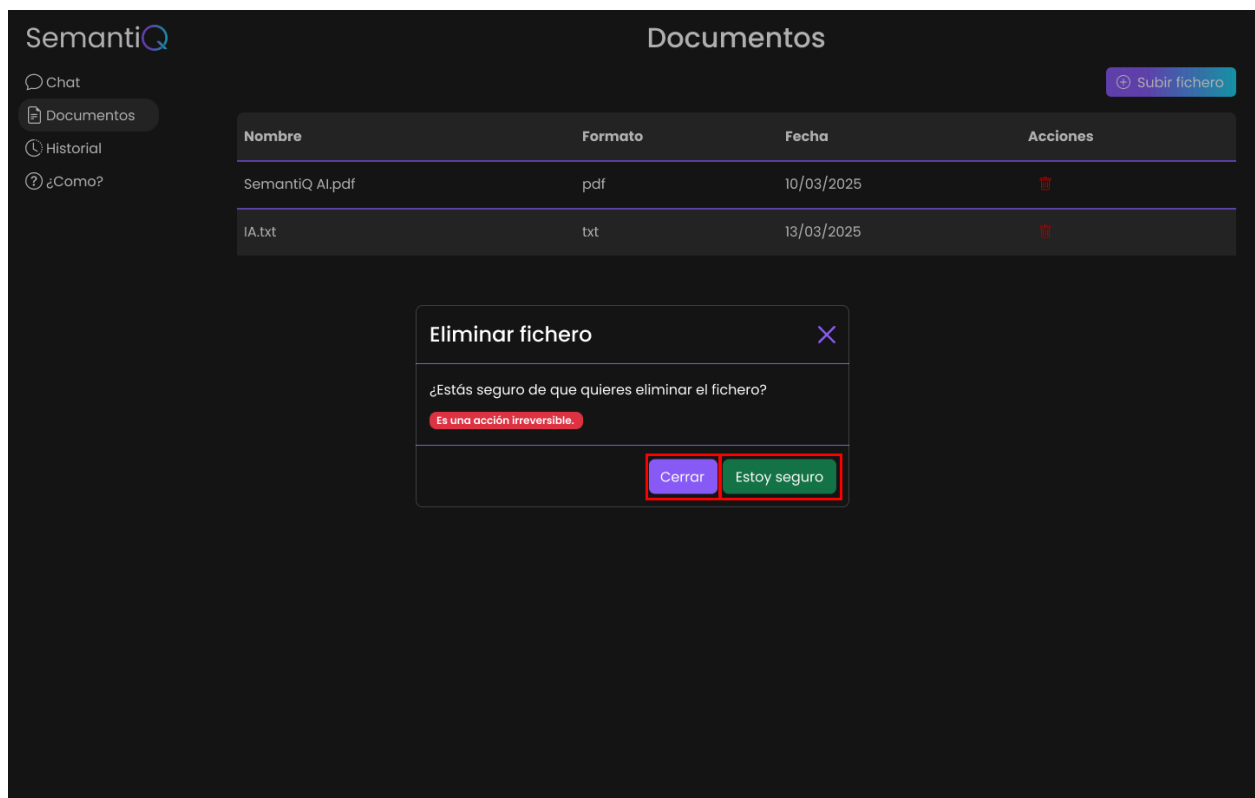
Il·lustració 1.9: Documentos Modal Subir



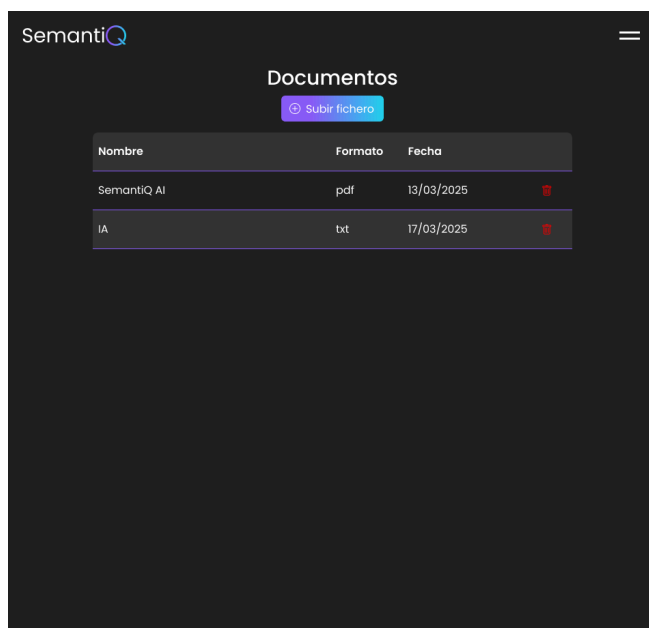
Il·lustració 1.10: Documentos Modal Subir Hover



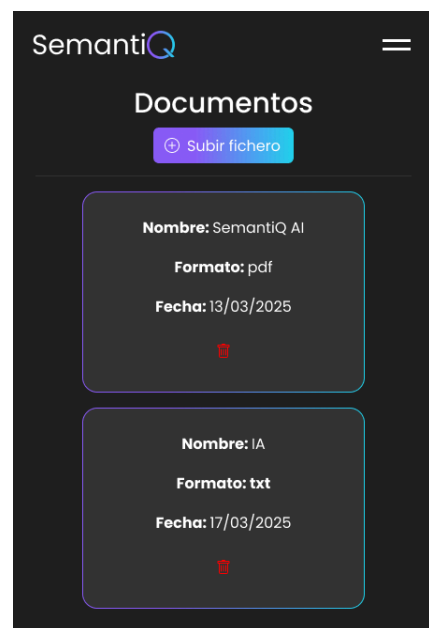
Il·lustració 1.11: Documentos Modal Eliminar



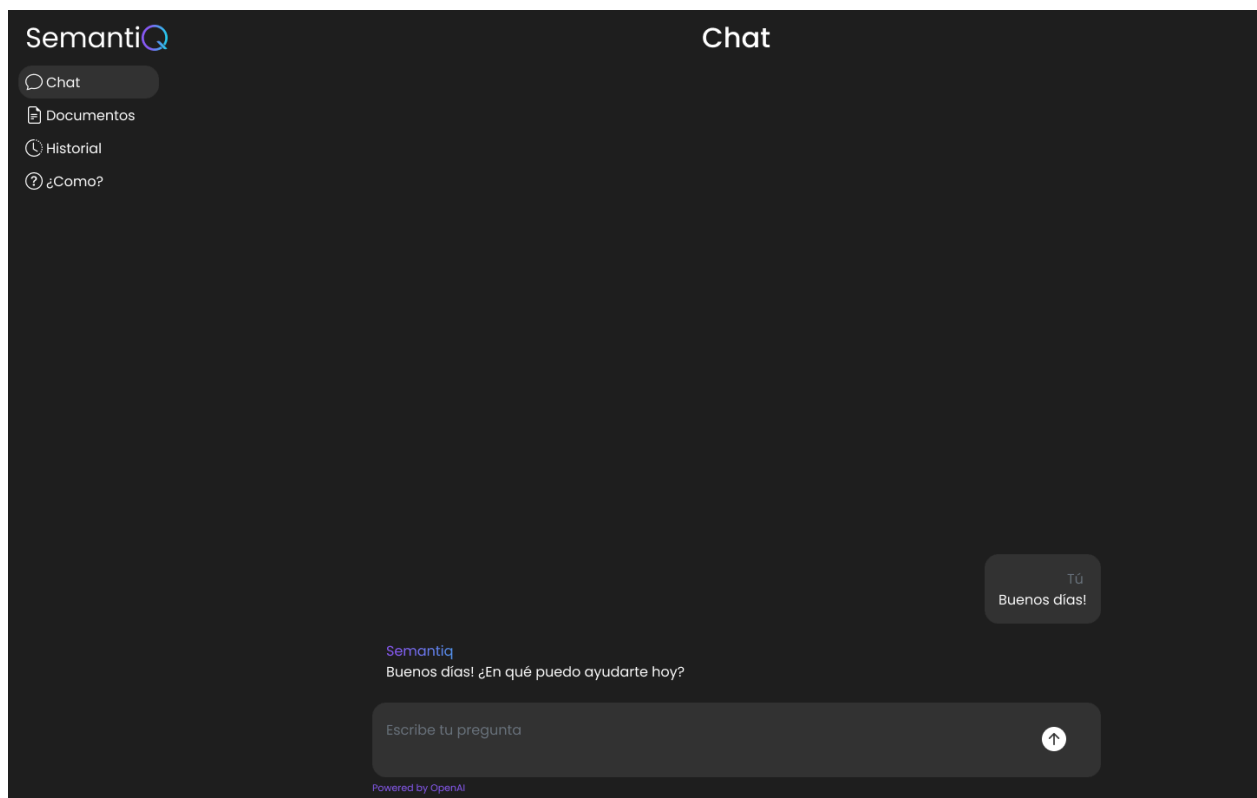
Il·lustració 1.12: Documentos Modal Eliminar Hover



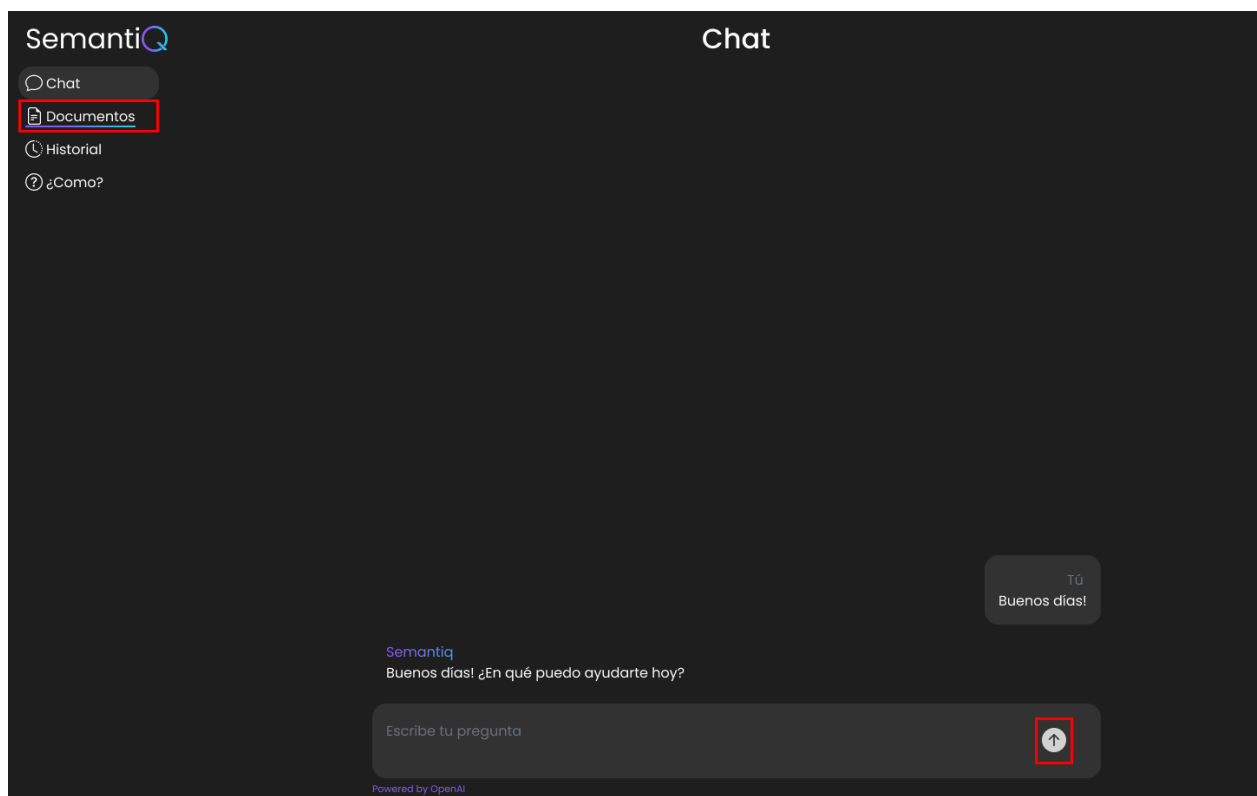
Il·lustració 1.13: Documentos Tablet



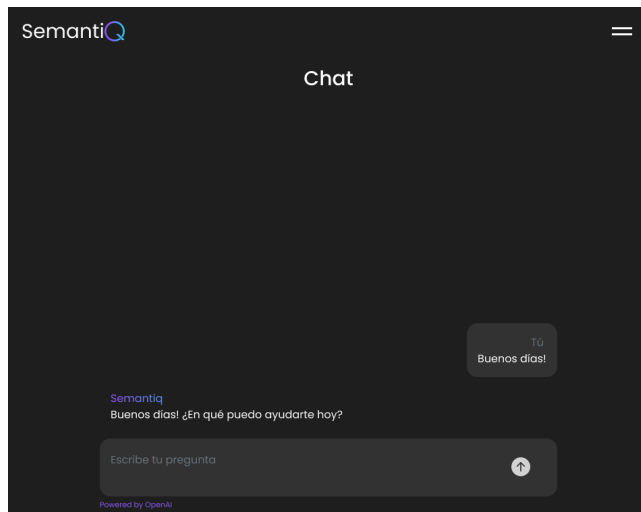
Il·lustració 1.14: Documentos Mobile



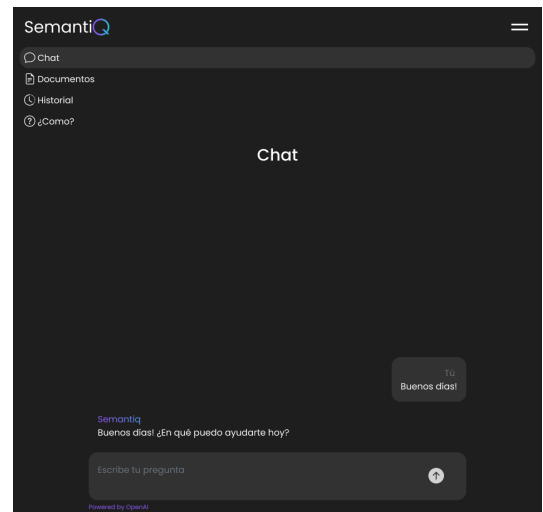
Il·lustració 1.15: Chat Desktop



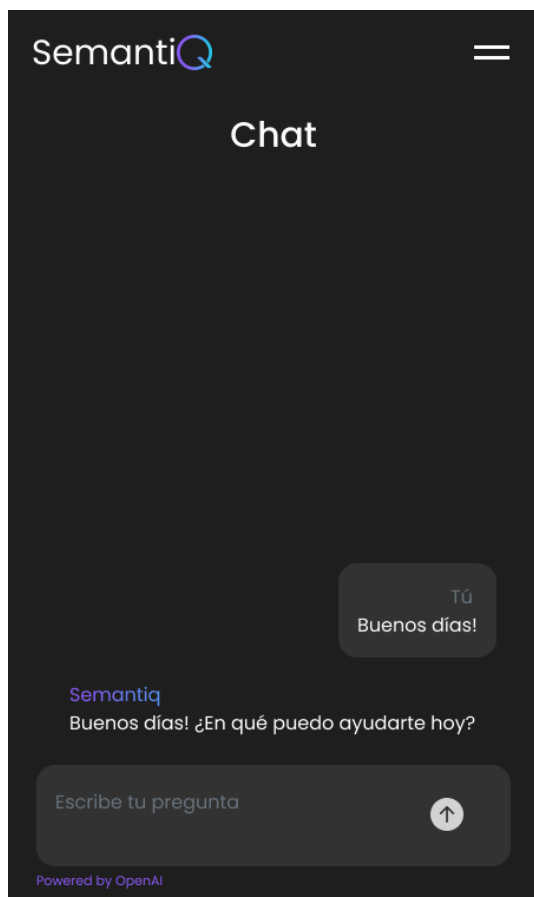
Il·lustració 1.16: Chat Menu Hover



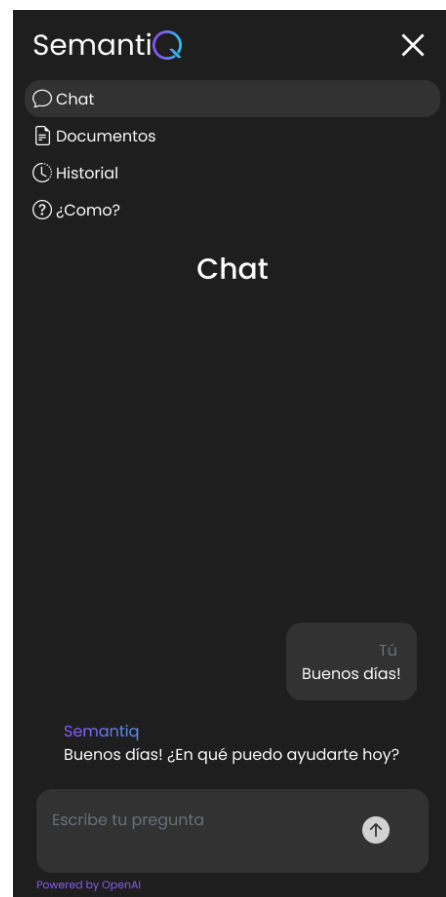
Il·lustració 1.17: Chat Tablet



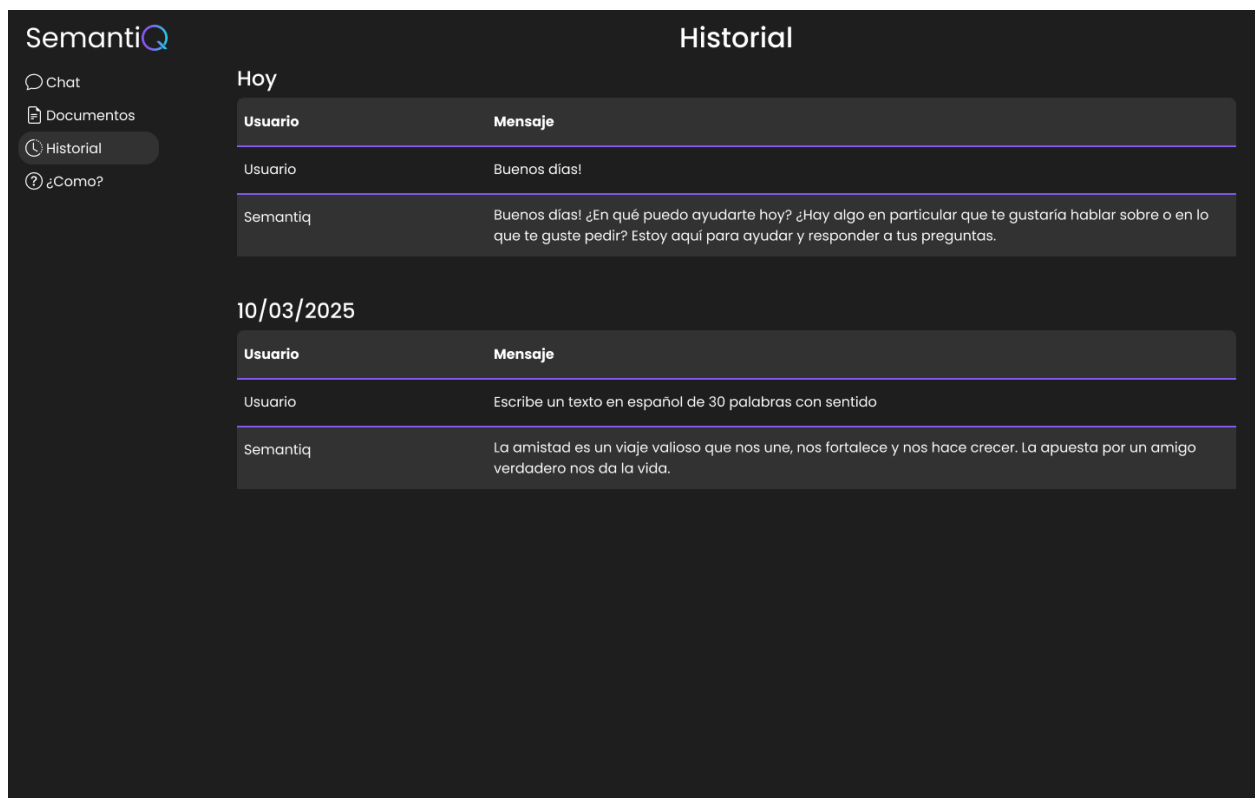
Il·lustració 1.18: Chat Tablet Menu Open



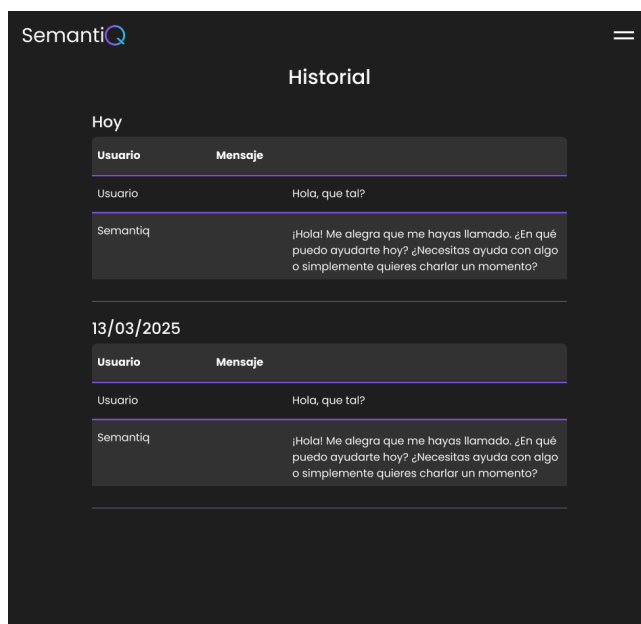
Il·lustració 1.19: Chat Mobile



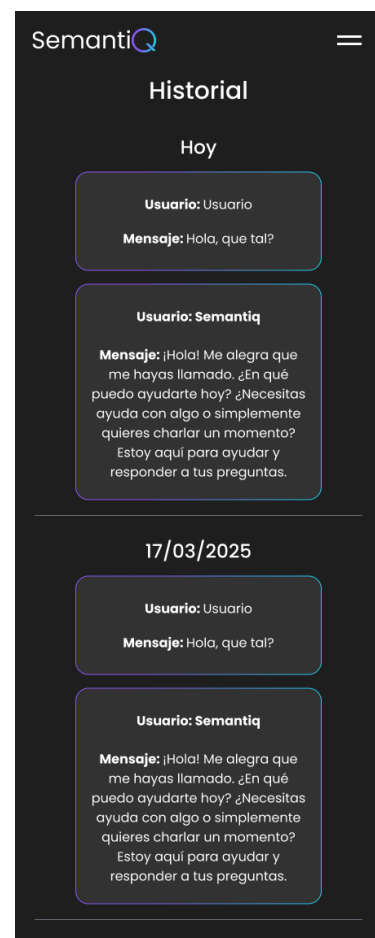
Il·lustració 1.20 Chat Mobile Menu Open



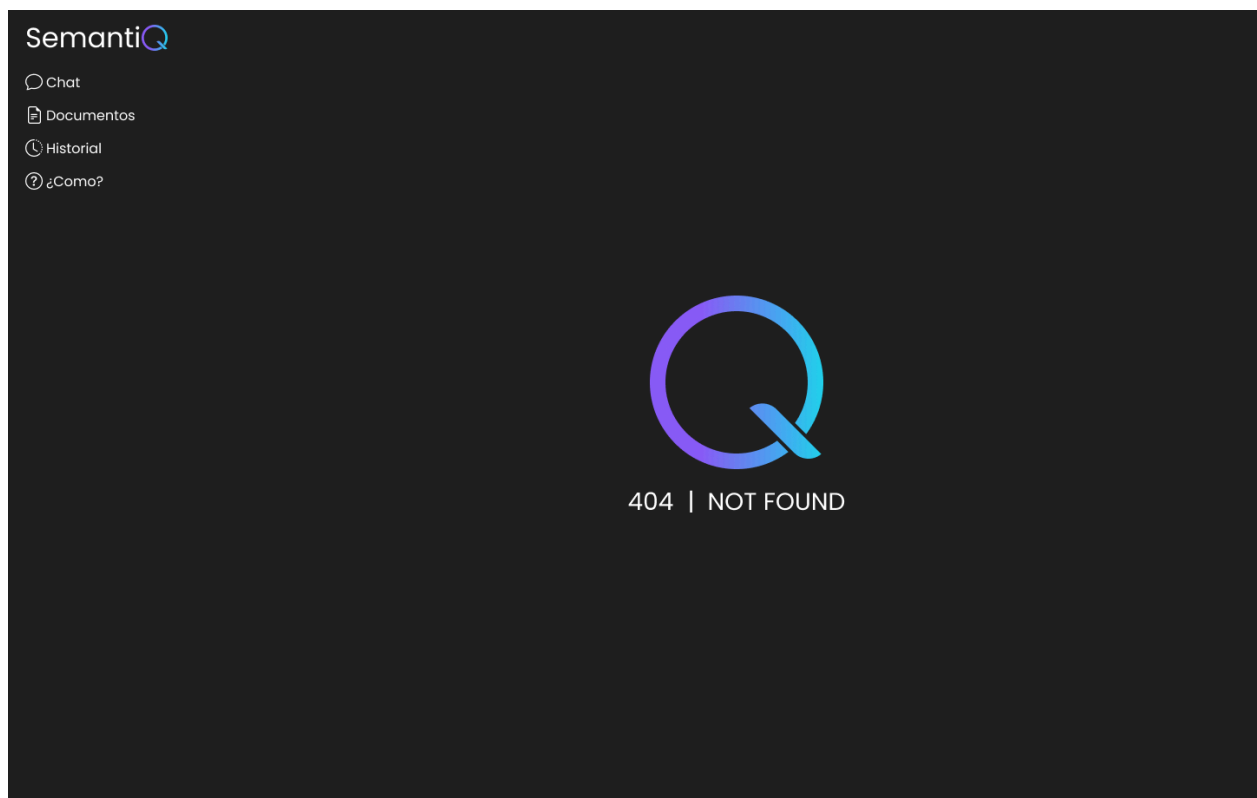
Il·lustració 1.21: Historial Desktop



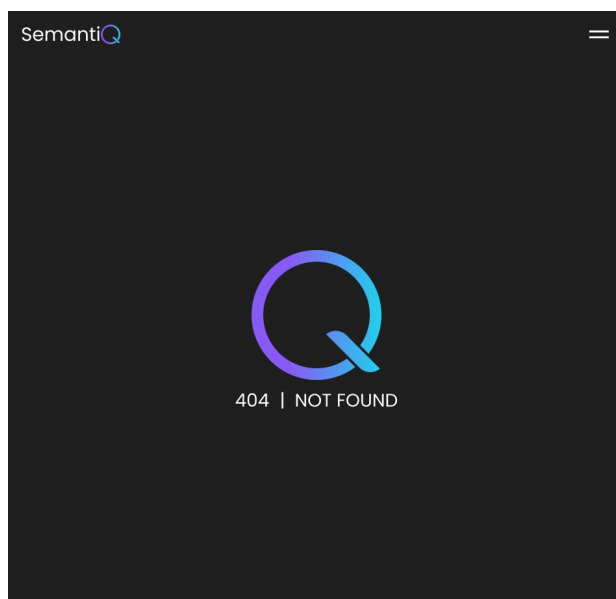
Il·lustració 1.22: Historial Tablet



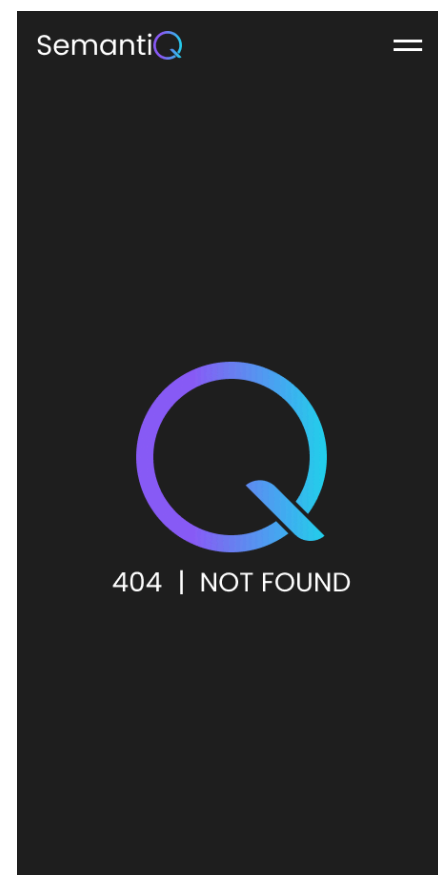
Il·lustració 1.23: Historial Mobile



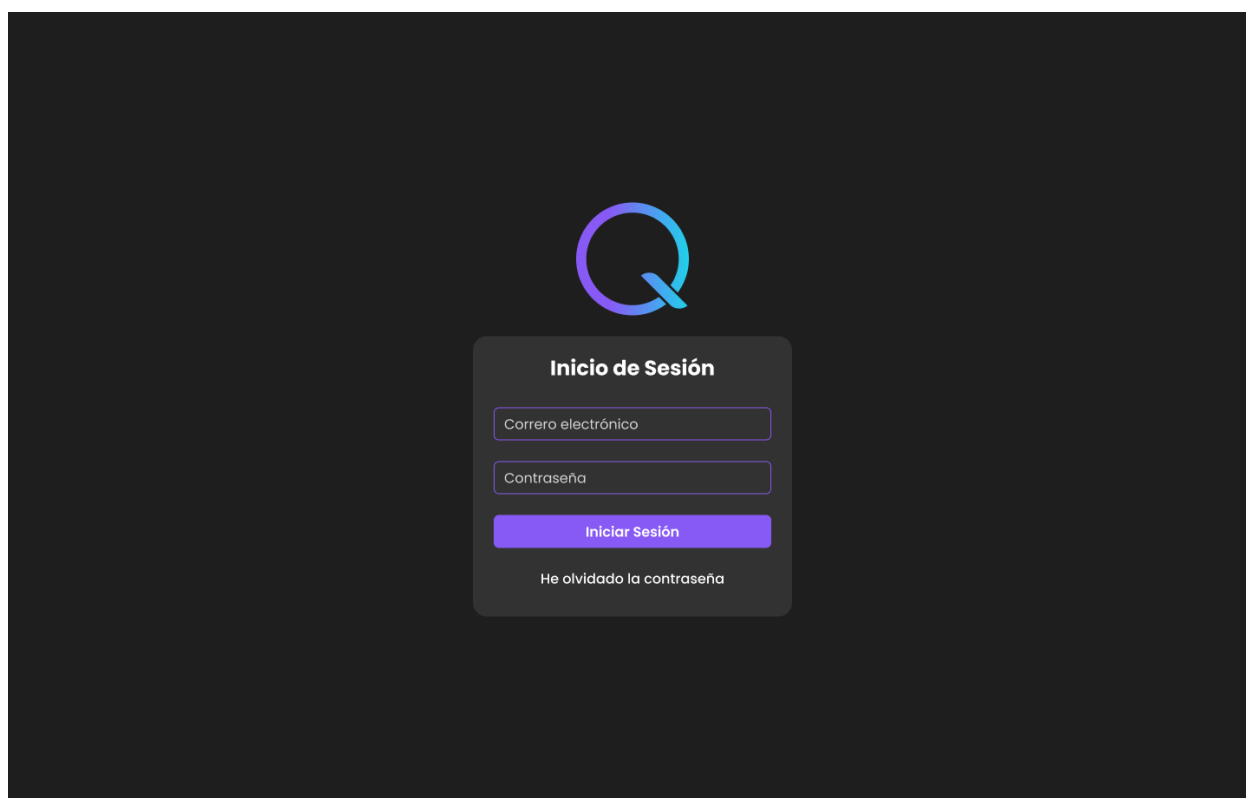
Il·lustració 1.24: 404 Desktop



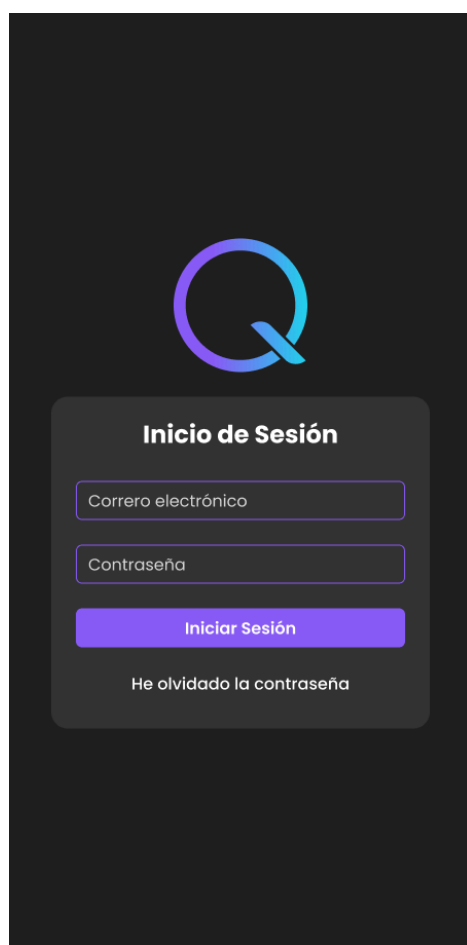
Il·lustració 1.25: 404 Tablet



Il·lustració 1.26: 404 Mobile



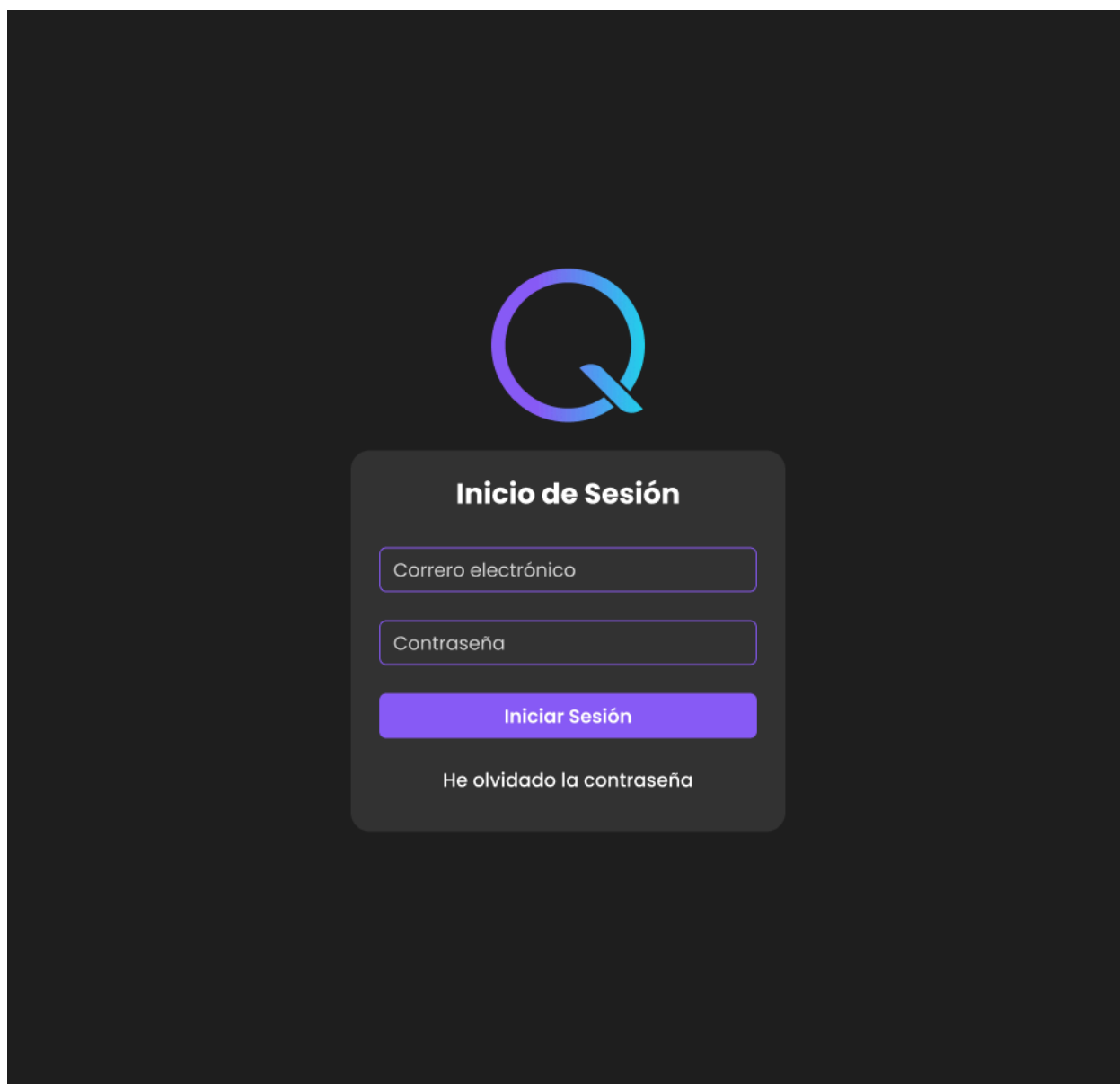
Il·lustració 1.27: Login Desktop



Il·lustració 1.28: Login Mobile



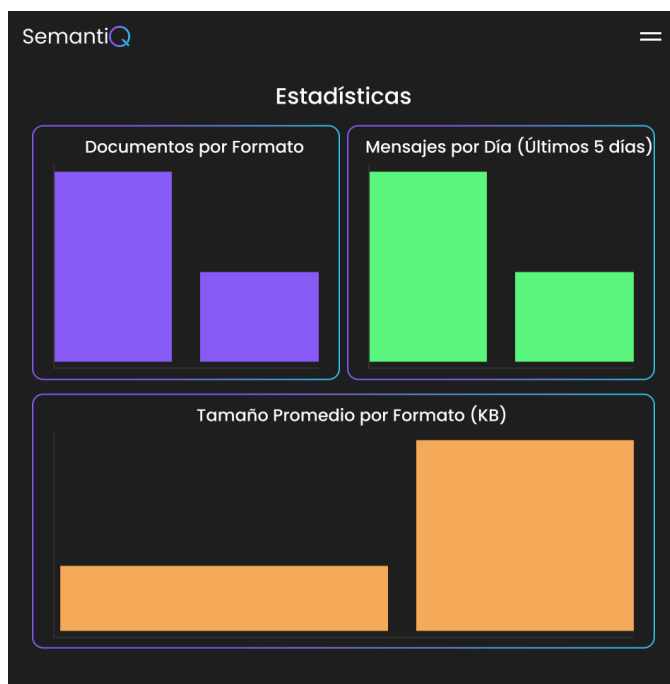
Il·lustració 1.29: Login Mobile Hover



Il·lustració 1.30: Login Tablet



Il·lustració 1.31: Estadísticas Desktop



Il·lustració 1.32: Estadísticas Tablet



Il·lustració 1.33: Estadísticas Mobile