



UNIVERSIDAD DE JAÉN
EPSJ

Trabajo Fin de Grado

RECUPERACIÓN Y ANÁLISIS DE INFORMACIÓN MÉDICA A TRAVÉS DE REDES SOCIALES

Alumno: Andrés Ortega Rodríguez

Tutores: Prof. Dña. María Teresa Martín Valdivia
Prof. Dña. Salud María Jiménez Zafra

Dpto: Informática

Febrero, 2017



Universidad de Jaén
Escuela Politécnica Superior de Jaén
Departamento de Informática

Don M^a Teresa Martín Valdivia y Doña Salud M^a Jiménez Zafra, tutoras del Proyecto Fin de Carrera titulado: Recuperación y análisis de información médica a través de Redes Sociales, que presenta Andrés Ortega Rodríguez, autoriza su presentación para defensa y evaluación en la Escuela Politécnica Superior de Jaén.

Jaén, MES de 2013

El alumno:

Los tutores:

Andrés Ortega Rodríguez

Fdo: Dña. M^a Teresa Martín Valdivia
Fdo: Dña Salud M^a Jiménez Zafra

Índice

1.	INTRODUCCIÓN	7
1.1.	Introducción al proyecto	7
1.1.	Motivación	8
1.2.	Propósito	8
1.3.	Objetivos específicos.....	9
1.4.	Resultados esperados.	9
2.	TECNOLOGÍAS DE DESARROLLO.....	10
2.1.	Lenguaje de programación.....	10
2.1.1.	¿Qué es Python?	10
2.1.2.	Justificación de uso	11
2.2.	Base de datos	11
2.2.1.	SQLite	11
2.3.	Web Framework.....	11
2.3.1.	Flask.....	11
2.3.2.	Bootstrap.....	12
2.4.	BeautifulSoup	12
2.5.	Facebook API Graph	12
3.	DESARROLLO DEL PROYECTO	13
3.1.	Tareas fundamentales.....	14
3.2.	Planificación de tiempos	15
3.3.	Estimación de costes	16
3.3.1.	Costes hardware.....	17
3.3.2.	Costes Software.....	17
3.3.3.	Costes de personal	18
3.3.4.	Otros costes.....	19
3.3.5.	Coste total	19
3.4.	Metodología de trabajo	20
4.	PROCESO DE INVESTIGACIÓN	20
4.1.	El Cáncer.....	20

4.1.1.	Las cifras del Cáncer	21
4.1.2.	Incidencia del Cáncer	21
4.1.3.	Mortalidad del Cáncer	22
4.1.4.	5 cánceres más frecuentes en España	22
4.1.5.	La importancia de la investigación	22
4.2.	Recuperación y análisis de información sobre enfermedades publicada en Facebook 23	
4.2.1.	¿Por qué Facebook?	23
5.	DESARROLLO DEL PROTOTIPO	25
5.1.	Historias de usuario.....	25
5.2.	Estimación de las historias de usuario	28
5.3.	Iteración 1	30
5.3.1.	Priorización de las historias.....	30
5.3.2.	Definiendo la iteración	31
5.3.2.1.	Obtención de información	32
5.3.2.1.1.	Token de acceso	32
5.3.2.1.2.	Implementación del Scraper	34
5.3.2.1.3.	Elección de las fuentes de información.....	35
5.3.2.2.	Respaldo de la información.....	37
5.3.2.2.1.	Modelo Entidad Relación de la Base de Datos.....	38
5.3.2.2.2.	Controlador	41
5.3.3.	Final de la iteración	41
5.4.	Iteración 2	42
5.4.1.	Priorización de las historias.....	42
5.4.2.	Definiendo iteración.....	42
5.4.2.1.	Actualización de la información	43
5.4.2.2.	Mapa de localizaciones	44
5.4.2.2.1.	Modificar BBDD	44
5.4.2.2.2.	Obtener localización de los usuarios.....	45
5.4.2.2.3.	Implementación del Scraper	45
5.4.3.	Final de la iteración	46
5.5.	Iteración 3	47
5.5.1.	Priorización.....	47
5.5.2.	Definición de la iteración	47

5.5.2.1.	Diseño Web	48
5.5.2.1.	Clasificación por objetividad	49
5.5.2.1.1.	Normalización de los textos.	50
5.5.2.1.2.	Cálculo de la objetividad	50
5.5.2.2.	Clasificación por polaridad	51
5.5.2.2.1.	Enfoques para la clasificación de la polaridad.	51
5.5.2.2.2.	Cálculo de la polaridad	52
5.5.2.2.3.	Representación	52
5.5.3.	Final de la iteración	53
5.6.	Iteración 4	53
5.6.1.	Priorización.....	53
5.6.2.	Definición de la iteración	54
5.6.2.1.	Clasificación por me gusta.....	54
5.6.2.2.	Ver secciones.....	55
5.6.2.3.	Información de las secciones	55
5.6.2.3.1.1.	Cálculo de la tendencia	56
5.6.2.3.1.2.	Selección de tipos de cáncer	56
5.6.2.3.1.3.	Diseño de la vista.....	56
5.6.2.4.	Ver temas populares	57
5.6.2.5.	Resaltar palabras clave.....	58
5.6.2.6.	Descripción de términos.	59
5.6.2.1.	Buscador.....	60
5.6.3.	Final de la iteración	61
6.	Conclusiones.....	62
	Agradecimientos	62
7.	Bibliografía.	63
8.	Anexo A. Manual de instalación.....	65
9.	Anexo B. Manual de usuario.	68
10.	ANEXO C: ÍNDICE DE ILUSTRACIONES	75
11.	ANEXO C: ÍNDICE DE TABLAS.....	76

1. INTRODUCCIÓN

En este primer capítulo se abordan los aspectos más importantes que han originado la elaboración del proyecto *Recuperación y análisis de información médica a través de Redes Sociales*, el cual me ha permitido mediante su desarrollo e implementación adquirir conocimientos sobre técnicas Web scraping, técnicas de minería de opiniones y en general nuevas herramientas para el análisis y el tratamiento de la información en la red.

1.1. Introducción al proyecto

El cáncer es una enfermedad crítica que afecta a millones de personas y familias alrededor del mundo. En 2012 alrededor de 14,1 millones de casos nuevos de cáncer ocurrieron a nivel mundial (Ferlay et al., 2015). A veces, dada la incapacidad de la medicina actual para curar algunos tipos de cáncer en estados avanzados, es preferible renunciar al tratamiento curativo y aplicar un tratamiento paliativo que proporcione el menor grado posible de malestar y conduzca a una muerte digna. Debido a muchas razones como (los efectos secundarios de algunos tratamientos y la muerte de otros pacientes) los pacientes con cáncer tienden a verse afectados por trastornos emocionales graves, como la depresión (Cano, 2005).

Muchos pacientes con cáncer son usuarios de redes sociales y muchos de ellos participan en comunidades virtuales, donde intercambian mensajes de apoyo a otros pacientes y sentimientos o donde comentan el tratamiento que siguen para su cura. La mayoría de estas comunidades son de acceso público, y por tanto, fuentes útiles de información sobre el estado anímico de los pacientes. Basado en esto, los métodos de análisis de sentimientos pueden ser útiles para detectar automáticamente el estado positivo o negativo de los pacientes analizando sus mensajes.

1.1. Motivación

Muchos usuarios de Internet desean encontrar opiniones sobre un producto o servicio en el que están interesados, pero esta tarea de recopilación de opiniones no es sencilla, ya que en la mayoría de los casos las opiniones se ocultan en foros, blogs o redes sociales. Es difícil para un lector encontrar fuentes relevantes, extraer frases relacionadas con opiniones, leerlas, resumirlas y organizarlas de forma útil. Por lo tanto, se necesitan sistemas automatizados de análisis de opinión y de resumen. El análisis del sentimiento, también conocido como minería de opinión, surge de esta necesidad.

Mediante la minería de opiniones podemos extraer un significado a los mensajes escritos por los pacientes que están disponibles en la red, con el propósito de obtener una valoración sobre tratamientos y fármacos que se utilizan en la lucha contra el cáncer.

La creación de una herramienta útil para aquellas personas que padecen esta enfermedad y el interés por aprender nuevas tecnologías de extracción de información en las redes sociales; así como su posterior procesamiento, me han motivado en la elección de este proyecto, afrontando este comienzo con mucha ilusión.

1.2. Propósito

Este proyecto está especialmente orientado al desarrollo de una herramienta que facilite el análisis de la información publicada en Facebook sobre cáncer, clasificándola según su contenido, objetivo o subjetivo, y según su polaridad (positiva/negativa) en el caso de la información subjetiva.

La aplicación web facilitará a sus usuarios consultar opiniones de otros pacientes sobre los fármacos que están tomando o sus tratamientos, con el fin de que estos puedan encontrar información útil en dicha red social y a su vez sacar una conclusión del beneficio o perjuicio de los tratamientos.

1.3. Objetivos específicos

Los objetivos que se han fijado para la realización de este proyecto son:

- Estudiar del cáncer, como enfermedad y estudiar cómo afecta al estado de ánimo de los pacientes en las redes sociales
- Estudiar la API de Facebook para la recuperación de posts y comentarios de los usuarios.
- Aprender Python como nuevo lenguaje de programación.
- Aprender técnicas de Procesamiento del Lenguaje Natural, para realizar un análisis de sentimientos.
- Aprender técnicas de Scraping, para la obtención de información de sitios web.
- Estudiar las distintas estructuras y arquitecturas a utilizar para almacenar y gestionar la información
- Desarrollar una aplicación que facilite al usuario encontrar información acerca del cáncer, sus distintos tipos, fármacos y tratamientos a partir de los datos recabados.
- Desarrollar los módulos asociados así como la definir e implementar de la Base de Datos para el almacenamiento de la información necesaria.
- Redactar una memoria que recoja todo el trabajo desarrollado así como los manuales de instalación y usuario.

1.4. Resultados esperados.

Los resultados que se esperan obtener al realizar este proyecto son:

- Aplicación web que permita visualizar información acerca de los distintos tipos de cáncer, y conocer los tratamientos/fármacos existentes para combatirlo.
- Memoria final del proyecto, donde se redacta todo el proceso llevado a cabo.
- Manual de usuario que explica como ha de usarse la aplicación.

- Manual de instalación que recoge los pasos a seguir para poder ejecutar la aplicación.

2. TECNOLOGÍAS DE DESARROLLO

En este capítulo se plantean las herramientas que se emplearán en la creación de la aplicación.

2.1. Lenguaje de programación

La elección del lenguaje de programación para la confección de la aplicación está basada en tres de los objetivos iniciales del proyecto, estos son:

- Extracción de datos en la red.
- Técnicas de Procesamiento del Lenguaje Natural
- Creación de una interfaz web para la visualización de resultados

Basándonos en los objetivos propuestos anteriormente y con la intención de aprender un nuevo lenguaje de programación, se establece Python como lenguaje de programación que satisface los puntos citados.

2.1.1. ¿Qué es Python?

Python es un lenguaje de programación interpretado, orientado a objetos, que tiene una sintaxis simple de alto nivel, bastante fácil de aprender, modular y fácilmente legible.

Otras características son:

- El código fuente es procesado por un intérprete. No es necesario compilar antes de ejecutar el programa.
- Python es un lenguaje multiplataforma y puede ejecutarse en muchos sistemas operativos como Windows, Linux, MAC OS.
- Python es el idioma fácil para programadores y soporta el desarrollo de una amplia gama de aplicaciones, incluyendo manipulación de texto simple, navegadores, juegos.

2.1.2. Justificación de uso

- Python tiene grandes librerías para la extracción de datos como BeautifulSoup y Scrape.py. Estas librerías son utilizadas como herramientas de Scraping para obtener información de las páginas web rápidamente.
- Python posee la librería NLTK, una de las mas utilizadas para el Procesamiento del Lenguaje Natural o PLN, que a pesar de no poseer la implementación más eficiente, proporciona una gran cantidad de herramientas para el análisis de textos (Bird et al., 2009).
- Python tiene disponible una gran variedad de web-frameworks como Django, Pylons y Flask, que permiten la creación de aplicaciones web.

2.2. Base de datos

Los datos que se irán extrayendo de Internet deben ser almacenados para su posterior tratamiento o consulta. Se ha escogido una base de datos relacional para nuestro proyecto, en concreto se ha utilizado la librería SQLite.

2.2.1. SQLite

SQLite es un motor de base de datos SQL embebido, es uno de los motores de base de datos más populares en la actualidad, se considera muy rápido, eficiente (Bi, 2009). Sus bases de datos son fácilmente portables a otras plataformas sin necesidad de configuración previa. Además su código es libre y por tanto abierto a todo el mundo que lo necesite.

2.3. Web Framework

2.3.1. Flask

Para la creación de la página web, se utilizará el framework Flask. Este microframework escrito en Python está orientado a la creación de pequeñas aplicaciones con pocos requisitos. Flask utiliza el lenguaje de plantillas Jinja2 inspirado en Django.

2.3.2. **Bootstrap**

Para dar formato a los datos y desarrollar el aspecto visual de la web, se utilizará el framework Bootstrap.

Bootstrap es uno de los framework más populares para el desarrollo Front-end¹. Se trata de una colección de herramientas que facilita la creación de páginas web. Este framework posee multitud de plantillas basadas en CSS para la creación de formularios, botones, etiquetas y otros componentes pre-diseñados. Además gracias a su sistema de cuadrícula, la tarea de maquetación web resulta rápida y sencilla.

2.4. **BeautifulSoup**

Beautiful Soup es una librería de Python que proporciona métodos sencillos para la extracción de datos en archivos HTML y XML. Esta librería analiza los archivos y crea una estructura de árbol con todos los elementos del documento para extraer información de sitios web.

2.5. **Facebook API Graph**

Utilizaremos la herramienta principal que nos provee Facebook para obtener datos desde dentro y fuera de la plataforma de Facebook.

Esta API está basada en HTTP de bajo nivel y permite consultar datos, publicar nuevas historias, administrar anuncios, subir fotos y otras tareas. Funciona con cualquier lenguaje que tenga una librería HTTP, como cURL, urllib.

El nombre API Graph viene de la idea de 'gráfico social' - una representación de la información en Facebook compuesta por²:

- Nodos - básicamente "cosas", como un usuario, una foto, una página, un comentario

¹ <https://wappalyzer.com/categories/web-frameworks>

² <https://developers.facebook.com/docs/graph-api/overview>

- Bordes – son las conexiones entre esas "cosas", como las Fotos de una Página, o los comentarios de una Foto
- Campos - información acerca de esas "cosas", como el cumpleaños de una persona, o el nombre de una página

3. DESARROLLO DEL PROYECTO

Durante el desarrollo de la aplicación se va a seguir un proceso ágil, particularmente se ha optado por la metodología XP (programación extrema), este tipo de metodología a diferencia de las tradicionales se enfoca más en la adaptabilidad que en la previsibilidad.

Se cree que surgirán cambios de requisitos durante el desarrollo del proyecto, es por esto que saber adaptarse a los cambios en el transcurso del mismo es una aproximación más adecuada que intentar prever y controlar todos los posibles cambios de requisitos en la fases iniciales de desarrollo.

Las características principales de XP son:

- Desarrollo basado en iteraciones y de manera incremental.
- Pruebas frecuentes del código realizado.
- Programación por parejas.
- Refactorización del código, para aumentar su claridad y conservar su funcionamiento normal sin gran esfuerzo.
- Corregir los nuevos errores antes de añadir nuevas funcionalidades.
- Gran implicación del cliente en el equipo de desarrollo.
- Apuesta por el desarrollo de código simple, frente al complicado.

Debido a la naturaleza de nuestro proyecto, no será posible seguir la metodología XP completamente, aspectos como los roles en el proceso de desarrollo se realizarán sólo por una persona. Otro aspecto como la programación por parejas tampoco podrá llevarse a cabo.

3.1. Tareas fundamentales

A continuación se presenta un diagrama con las tareas principales que se van a realizar en el desarrollo del proyecto.

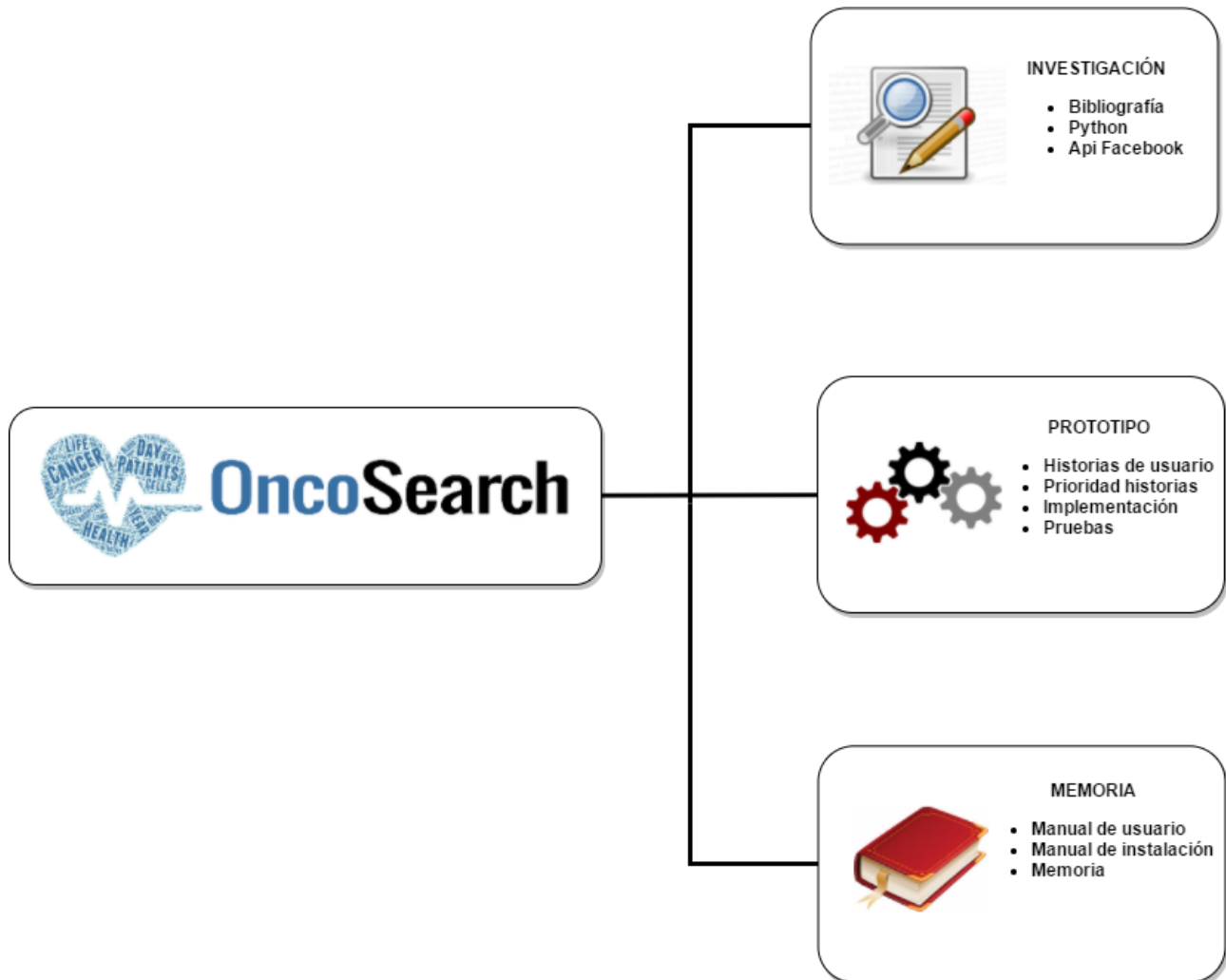


Ilustración 1: Tareas fundamentales

3.2. Planificación de tiempos

Se espera que mediante la siguiente planificación del proyecto, se obtenga una visión general y un control sobre las tareas a realizar, así como una estimación del tiempo que conllevará realizarlas.

Task Name	Duration	Start	Finish
☐ Desarrollo del proyecto	140 days	Mon 8/1/16	Fri 2/10/17
Inicio del proyecto	0 days	Mon 8/1/16	Mon 8/1/16
☐ Estudio de la bibliografía	30 days	Mon 8/1/16	Fri 9/9/16
Técnicas de PLN	10 days	Mon 8/1/16	Fri 8/12/16
Estudio de métodos Web Scraping	10 days	Mon 8/15/16	Fri 8/26/16
Estudio del Cáncer y tratamientos	10 days	Mon 8/29/16	Fri 9/9/16
☐ Estudio de Python	21 days	Mon 9/12/16	Mon 10/10/16
Estudio de manuales de Python	14 days	Mon 9/12/16	Thu 9/29/16
Aprendizaje de Flask	7 days	Fri 9/30/16	Mon 10/10/16
☐ Estudio de la API de Facebook	7 days	Tue 10/11/16	Wed 10/19/16
Estudio de los métodos de la API	7 days	Tue 10/11/16	Wed 10/19/16
☐ Desarrollo de la aplicación	65 days	Thu 10/20/16	Wed 1/18/17
Historias de usuario/ posibles nuevas historias	1 day	Thu 10/20/16	Thu 10/20/16
Priorizar historias de usuario	4 hrs	Fri 10/21/16	Fri 10/21/16
Definir iteración	4 hrs	Fri 10/21/16	Fri 10/21/16
Implementación de la iteración	60 days	Mon 10/24/16	Fri 1/13/17
Finalizar iteración y pruebas	3 days	Mon 1/16/17	Wed 1/18/17
☐ Memoria	140 days	Mon 8/1/16	Fri 2/10/17
Manual de usuario	4 days	Thu 1/19/17	Tue 1/24/17
Manual de instalación	4 days	Thu 1/26/17	Tue 1/31/17
Memoria	140 days	Mon 8/1/16	Fri 2/10/17

Tabla 1: Plan de tareas

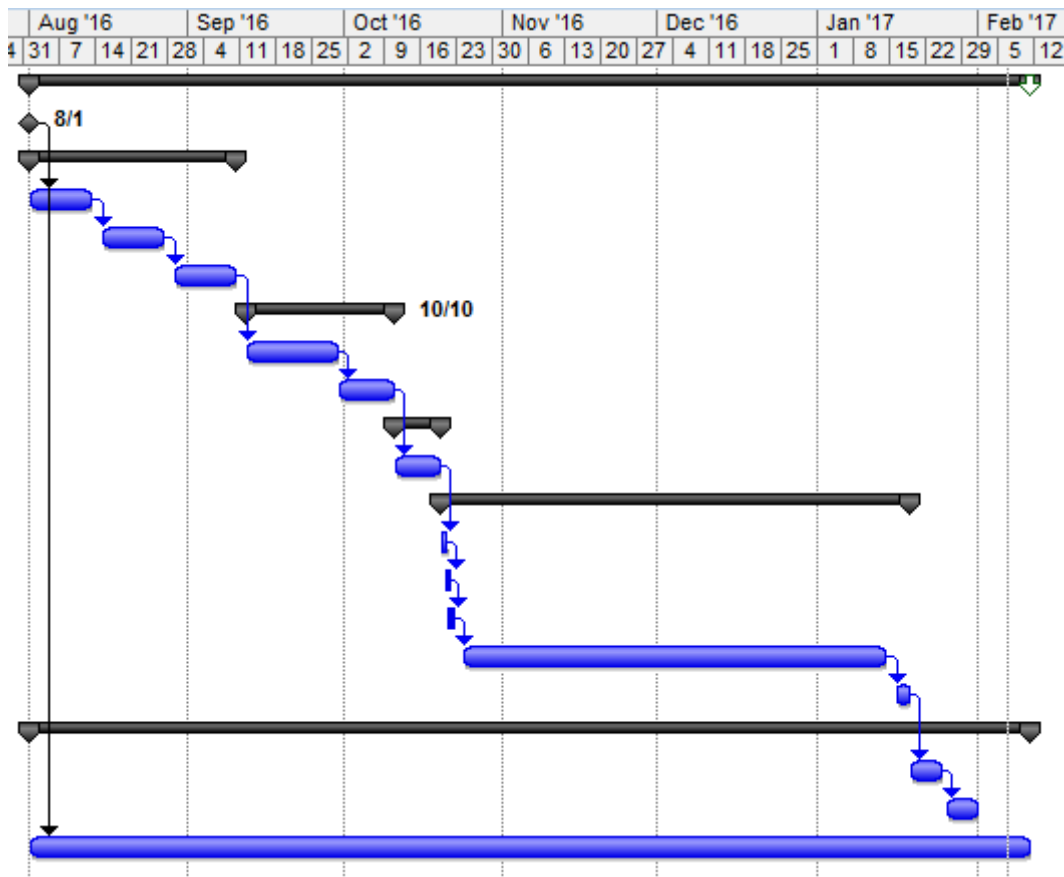


Ilustración 2: Diagrama de Gannt

El desarrollo de la aplicación sigue la dinámica de las metodologías ágiles, en la que el ciclo de trabajo queda dividido en iteraciones, estas iteraciones aún no ha sido definidas, por lo que hemos considerado la fase de desarrollo en el diagrama de Gannt como tarea única.

El proceso de elaboración del proyecto tendrá como fecha de comienzo el 1 de agosto de 2016 y como fecha de finalización, el 10 de febrero de 2017. Se establece una jornada completa de ocho horas de lunes a viernes sin contar fines de semana, con una duración total de 140 días .

3.3. Estimación de costes

La estimación de costes tiene como objetivo fundamental el control de las posibles desviaciones en costes y plazos, con el objetivo de evitarlas o al menos detectarlas a tiempo. Los costes se pueden dividir en costes hardware, costes software, costes de personal y otros costes.

3.3.1. Costes hardware

Se trata de los costes de aquellos dispositivos o equipos que han sido utilizados para el desarrollo y ejecución del proyecto. A continuación se detalla el equipo en cuestión.

- Dell inspiron 1464
 - Procesador: Intel Core i3
 - Memoria RAM 4 GB DDR3
 - Disco duro 500 GB
 - Tarjeta gráfica: Ati Mobility Radeon HD 4330

Este equipo tiene un valor de 500€.

Para calcular la amortización, se prevé que el ordenador tenga una vida útil de cinco años, y al final de este periodo su valor será 0€. Con un coste de 100€ al año o 8,33€ mensuales y una duración de 4,67 meses, el coste total es 38,91€.

3.3.2. Costes Software

Se trata de los costes asociados a las licencias softwares adquiridas para la elaboración y ejecución del proyecto. A continuación se detallan los programas con los que se ha construido la aplicación web.

Sistema Operativo: Windows 7	0€
PyCharm 2016.....	0€
Google Chrome.....	0€
Base de datos SQLite.	0€
Paquete Microsoft Office 2013.	149€
Visual Paradigm	0€
Gimp.....	0€

Para amortizar el coste de las licencias software, se estima que el Paquete Microsoft Office tenga una duración de tres años, con un coste de 4,13€ al mes y un total de 4,67 meses, el coste software es 19,28€.

3.3.3. Costes de personal

Se trata del sueldo del personal que se contrataría para desarrollar el proyecto. Este valor se va a calcular teniendo en cuenta que todo el proceso va a ser realizado por una persona, la cuál tomará distintos perfiles durante el transcurso del mismo.

En la siguiente tabla se recogen los sueldos sugeridos para los distintos perfiles de desarrollo.

Personal	Salario anual	Salario/hora
Analista	26.000€	13,88€
Programador	21.000€	11,21€
Técnico	16.000€	8,55€

Tabla 2: Salario mínimo

A continuación, se muestra las tareas a realizar junto con la duración en días / horas, además del rol del personal que realizará dicha tarea.

Tarea	Duración (días)	Duración (horas)	Personal
Estudio de la bibliografía			
Estudio técnicas de PLN	10	78	Analista
Estudio de métodos Web Scraping	10	78	Analista
Estudio del Cáncer y tratamientos	10	78	Analista
Estudio de manuales de Python	14	109.2	Analista
Estudio de la API de Facebook	7	54.6	Analista
Desarrollo de la aplicación			
Historias de usuario/ posibles nuevas historias	1	7.8	Programador
Priorizar historias de usuario		4	Programador

Definir iteración		4	Programador
Implementación de la iteración	60	468	Programador
Finalizar iteración y pruebas	3	23.4	Programador
Memoria	140	1092	Técnico
Manual de usuario	4	31.2	Técnico
Manual de instalación	4	31.2	Técnico

Tabla 3: Tiempos estimados y perfil del desarrollador

Total de horas junto con el salario total de cada trabajador:

Personal	Total de horas	Coste total
Analista	397.8	5.521€
Programador	507.2	5.685€
Técnico	1154.4	9870€
Total	2059.4	21.077€

Tabla 4: Personal y salario

3.3.4. Otros costes

Otro coste derivado del desarrollo del proyecto es, la conexión a Internet. Con un coste de 25€.

El coste total de la conexión a Internet es 116,75€, para 4,67 meses de duración.

3.3.5. Coste total

El coste total del proyecto viene dado por la suma de los costes hardware, costes software, costes de personal y otros costes.

Concepto	Coste
Costes hardware	38,91€
Costes software	19,28€
Costes de personal	21.077€
Otros costes	116,75
TOTAL	21.077€

Tabla 5: Coste del desarrollo del proyecto

3.4. Metodología de trabajo

- Estudio de los tipos de cáncer que más afectan a la sociedad.
- Estudio de los tratamientos y fármacos más usados por los pacientes de cáncer.
- Selección de las fuentes de dónde se extraerá la información.
- Revisión bibliográfica de las tecnologías disponibles.
- Estudio de las distintas técnicas de extracción de información y deberá hacerse un estudio comparativo que justifique las tecnologías empleadas para la solución del trabajo.
- Análisis de la información y aplicación de técnicas de Procesamiento de Lenguaje Natural para el tratamiento adecuado de la información
- Análisis y diseño de la Base de Datos necesaria para el almacenamiento de la información.
- Desarrollo de una aplicación donde mostrar el análisis realizado de la información extraída
- Documentación y pruebas para la experiencia de usuario.
- Pruebas de estabilidad y seguridad del prototipo.
- Generación de la memoria del trabajo realizado.

4. PROCESO DE INVESTIGACIÓN

4.1. El Cáncer

El cuerpo humano contiene miles de millones de células. A medida que las células viejas mueren, otras células se dividen para formar nuevas células. Este proceso ocurre en el cuerpo millones de veces al día.

Los cánceres comienzan cuando algunas células del cuerpo se vuelven anormales, se multiplican fuera de control y se hacen cargo de las células normales.

Estas células no se comportan de la misma manera que las células normales: pueden crecer y dividirse más rápido, o vivir más tiempo.

A medida que pasa el tiempo, se hacen más y más células anormales y comienzan a superar en número a las células normales en el área. Se multiplican fuera de control y forman un cáncer (Weinberg, 1996).

Los tratamientos contra el cáncer están diseñados para detener o ralentizar este proceso. Trabajan eliminando las células anormales, matándolas, o deteniéndolas en su proceso de crecimiento.

4.1.1. Las cifras del Cáncer.

Las cifras que se recogen a continuación han sido extraídas de la Sociedad Española de Oncología Médica (SEOM) que se basa en el estudio GLOBOCAN 2012 (Ferlay et al, 2015) realizado por parte de la Agencia Internacional para la Investigación del Cáncer (IARC). Dicho artículo ofrece información de como afecta el cáncer en el mundo y también en España, aportando datos de incidencia, prevalencia y mortalidad de los tipos de cáncer con más repercusión.

4.1.2. Incidencia del Cáncer

La incidencia de cáncer es el número de nuevos casos de cáncer que surgen en una población determinada durante un período de tiempo dado (normalmente un año). Puede expresarse como un número absoluto de casos dentro de toda la población por año o como una tasa por 100 000 personas al año. La tasa de incidencia de cáncer proporciona una aproximación del riesgo promedio de desarrollar un cáncer.

Año	Número estimado de nuevos casos	Hombre	Mujer	Ambos sexos
2012	Todas las edades	128.550	86.984	215.534
	< 65 años	46.202	39.225	85.427
	> = 65 años	82.348	47.759	130.107
2020	Todas las edades	148.998	97.715	246.713
	< 65 años	54.031	43.251	97.282
	> = 65 años	94.967	54.464	149.431

GLOBOCAN 2012 (IARC) - 14.1.2016

Las predicciones poblacionales fueron realizadas por el proyecto GLOBOCAN a partir de la revisión 2012, perspectivas población mundial, Naciones Unidas.

Tabla 6: Estimación de la incidencia del cáncer

4.1.3. Mortalidad del Cáncer

La mortalidad por cáncer es el número de muertes debidas al cáncer, que ocurre en una población determinada durante un período de tiempo dado (normalmente 1 año). Se puede expresar como un número absoluto de muertes en el conjunto de la población por año o como una tasa por 100 000 personas al año.

Año	Número estimado de casos de muerte por cáncer	Hombre	Mujer	Ambos sexos
2012		63579	39183	102762
	< 65 años	16679	9995	26674
	> = 65 años	46900	29188	76088
2015		67129	41261	108390
	< 65 años	17611	10399	28010
	> = 65 años	49518	30862	80380
Cambio demográfico		3550	2078	5628
< 65 años		932	404	1336
> = 65 años		2618	1674	4292

ESPAÑA. Todos los cánceres excluyendo cáncer de piel no melanoma

Fuente: Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC.

Tabla 7: Mortalidad del cáncer en España

4.1.4. 5 cánceres más frecuentes en España

	Hombre	Mujer	Ambos Sexos
1º	Próstata	Mama	Colorrectal
2º	Pulmón	Colorrectal	Próstata
3º	Colorrectal	Cuerpo de Útero	Pulmón
4º	Vejiga	Pulmón	Mama
5º	Estómago	Ovario	Vejiga

Fuente: Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC.

Tabla 8: Localizaciones más frecuentes del cáncer

4.1.5. La importancia de la investigación

Es crucial recordar que el cáncer no es una enfermedad, son más de doscientas. Todas estas enfermedades son diferentes y requieren diferentes enfoques para su tratamiento. Los tratamientos que se usan para algunos tipos de cáncer no funcionan para otros.

Sólo la investigación mundial mejorará nuestra comprensión del cáncer y nos ayudará a encontrar y desarrollar tratamientos mejores y más eficaces. Y sólo una búsqueda global de respuestas nos ayudará a lidiar con esta enfermedad.

4.2. Recuperación y análisis de información sobre enfermedades publicada en Facebook

Gran parte de la investigación existente sobre el procesamiento de la información se ha centrado en la extracción y recuperación de información objetiva, por ejemplo, búsqueda en la Web, clasificación, agrupación de textos y muchas otras tareas. Sin embargo, las opiniones son tan importantes que siempre que necesitamos tomar una decisión, queremos escuchar a nuestros amigos o familiares para así tener un conocimiento más completo y tomar una decisión más acertada. Esto no sólo ocurre con los clientes de un servicio o producto, sino que las propias empresas también están interesadas en encontrar opiniones de los consumidores y de sus competidores sobre sus productos (Jeff Zabin y Alex Jefferies, 2008).

Hace unos años, cuando una organización quería saber las opiniones o sentimientos del público en general sobre sus productos y servicios, esta realizaba encuestas de opinión a sus usuarios. Sin embargo, con la Web 2.0, especialmente con el crecimiento explosivo del contenido en redes sociales en los últimos años, el mundo se ha transformado.

Las redes sociales han cambiado drásticamente la forma en que la gente expresa sus puntos de vista y opiniones. Ahora se pueden publicar comentarios de productos y expresar opiniones sobre estos en los foros de Internet, grupos de discusión y blogs. Este comportamiento posibilita nuevas fuentes de información. Hoy en día si una persona quiere comprar un producto, esta ya no se limita a preguntar a amigos y familiares, también revisa las opiniones del producto en Internet.

4.2.1. ¿Por qué Facebook?

Se ha elegido Facebook como red social para la recuperación de información porque la mayoría de trabajos sobre minería de opinión se han centrado principalmente en otras redes sociales como Twitter, y desde mi punto de vista es posible clasificar con mayor precisión el sentimiento en los mensajes de la red de Facebook (Ahkter et al, 2010).

Aspectos más importantes para la elección de Facebook como fuente de información:

- De acuerdo con el estudio realizado por IABSpain³, los españoles dedican de media 4 horas y 23 minutos semanales en la red social Facebook, que se alza en tercera posición entre las redes más visitadas

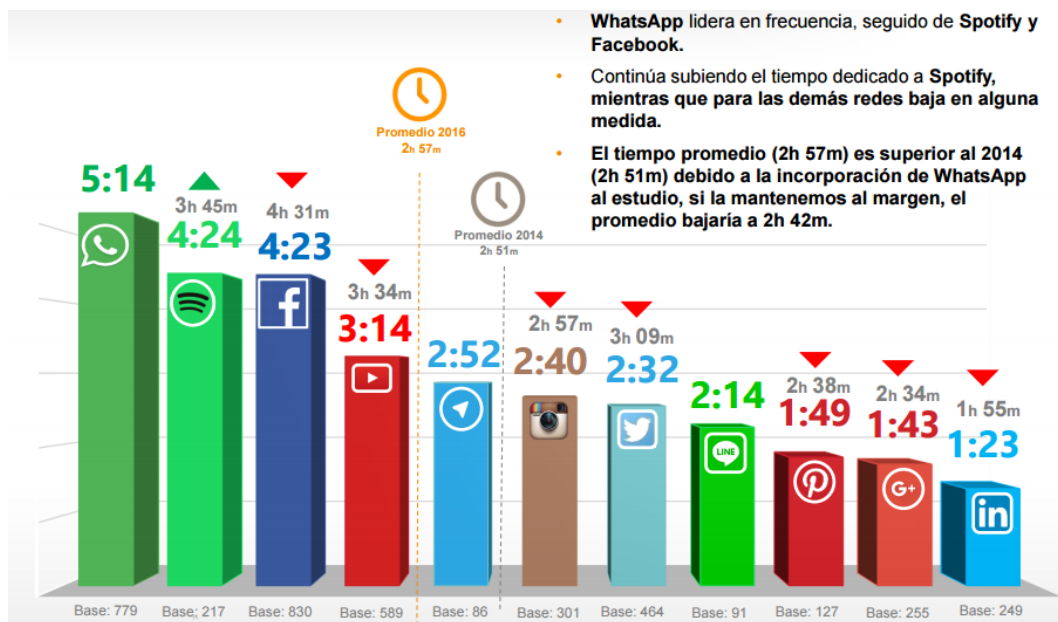


Ilustración 3: Redes sociales y frecuencia de uso

- En Facebook existen comunidades públicas de pacientes con cáncer, que discuten sobre sus experiencias, sentimientos, síntomas...
- Los mensajes que los usuarios escriben en Facebook son más fáciles de clasificar que los tweets porque estos no están restringidos sólo a 140 caracteres esto permite a los usuarios una mejor redacción y por tanto, un retrato más preciso de sus emociones.

³http://www.iabspain.net/wp-content/uploads/downloads/2016/04/IAB_EstudioRedesSociales_2016_VCorta.pdf

5. DESARROLLO DEL PROTOTIPO

5.1. Historias de usuario

Como parte central de muchas metodologías de desarrollo ágil, las historias de usuarios definen los requisitos que describen un proyecto software. Las historias de usuarios son priorizadas por el cliente (o el propietario del producto) para indicar cuáles son las más importantes para el sistema. Estas se desglosarán en tareas y serán estimadas por los desarrolladores.

A continuación se presentan las historias de usuario.

Historia: Obtención de información	
Como	Desarrollador
Quiero	Extraer comentarios y posts de grupos relacionados con la enfermedad del cáncer en Facebook
Para	Realizar mediante técnicas de PLN un análisis de sentimiento

Tabla 9: Historia Obtención de información

Historia: Respaldo de la información	
Como	Desarrollador
Quiero	Una base de datos donde se almacenen todos los datos recabados de Internet
Para	Poder tratar, analizar o consultar dichos datos

Tabla 10: Historia Respaldo de la información

Historia: Clasificación por objetividad	
Como	Usuario de la aplicación
Quiero	Poder organizar los posts y comentarios según su objetividad
Para	Poder recopilar comentarios según sean opiniones o hechos

Tabla 11: Historia Clasificación por objetividad

Historia: Clasificación por polaridad	
Como	Usuario de la aplicación
Quiero	Poder organizar los posts y comentarios según su polaridad
Para	Poder conocer qué textos son positivos o negativos

Tabla 12: Historia Clasificación por polaridad

Historia: Diseño Web	
Como	Usuario de la aplicación
Quiero	Una aplicación web con un diseño fácil, intuitivo y profesional
Para	Poder navegar cómodamente por la aplicación

Tabla 13: Historia Diseño Web

Historia: Clasificación por me gusta	
Como	Usuario de la aplicación
Quiero	Poder organizar los posts y comentarios por número de likes, wow, love, angry y sad
Para	Poder ordenar los textos según estos parámetros

Tabla 14: Historia Clasificación por me gusta

Historia: Ver secciones	
Como	Usuario de la aplicación
Quiero	Una sección que muestre: • Tipos de cáncer • Tratamientos • Fármacos
Para	Poder consultar información acerca de estas secciones

Tabla 15: Historia Ver secciones

Historia: Información de las secciones	
Como:	Usuario de la aplicación
Quiero:	Una tabla que muestre la polaridad, nº de menciones y tendencia asociada a cada tema de cada sección (Tipos de cáncer/Tratamientos/Fármacos)
Para:	Poder consultar información sobre ellos.

Tabla 16: Historia Información de las secciones

Historia: Ver temas populares	
Como:	Usuario de la aplicación
Quiero:	Ver una gráfica que muestre los temas a los que más se hace referencia en los comentarios de Facebook
Para:	Poder conocer cuáles son los temas más populares.

Tabla 17: Historia Ver temas populares

Historia: Buscador	
Como:	Usuario de la aplicación
Quiero:	Poder buscar sobre otros temas que no aparezcan definidos en las secciones por defecto
Para:	Poder consultar información sobre ellos

Tabla 18: Historia Buscador

Historia: Actualización información	
Como:	Desarrollador
Quiero:	Descargar los posts y comentarios recientes realizados en Facebook, sin sobrescribir los ya existentes en la BBDD
Para:	Mantener la base de datos actualizada

Tabla 19: Historia Actualización información

Historia: Mapa de localizaciones	
Como:	Usuario
Quiero:	Un mapa de España, en el que se reflejen los lugares en los que se habla de forma objetiva/subjetiva, con un sentimiento positivo/negativo en el caso de información subjetiva y también las zonas con mayor número de seguidores
Para:	Conocer más sobre el cáncer a nivel de España

Tabla 20: Historia Mapa de localizaciones

Historia: Resaltar palabras clave	
Como:	Usuario
Quiero:	Que la aplicación resalte las palabras de los comentarios asociadas con un sentimiento positivo o negativo
Para:	Saber qué palabras del comentario se han detectado como positivas o negativas.

Tabla 21: Historia Resaltar palabras clave

Historia: Descripción de términos	
Como:	Usuario
Quiero:	Que la aplicación extraiga información de Wikipedia
Para:	Obtener una descripción del tema que estoy consultando

Tabla 22: Historia Descripción de términos

5.2. Estimación de las historias de usuario

Para realizar la estimación de la dificultad de cada tarea utilizaremos la técnica del planning poker (Mahnič et al, 2012).

El planning poker, es una herramienta de estimación, basada en el consenso y utilizada para cuantificar la dificultad del trabajo que hay que realizar. En el planning poker, los miembros del grupo hacen estimaciones jugando cartas numeradas boca abajo. Al ocultar la estimación del coste de una tarea, el grupo puede evitar el verse influenciado por la estimación

realizada por la primera persona del grupo. Tras esto las cartas son reveladas, y se discuten la estimaciones realizadas.

El grupo estuvo formado por cuatro personas, las cuales me ayudaron con la estimación de las tareas, el procedimiento lo realizamos con la aplicación *Scrum Poker Cards*, la cual utiliza la escala: 0, ½, 1, 2, 3, 5, 8, 13, 20, 40, 100.

Historias de usuario	Puntos de Historia
Obtención de información	13
Respaldo de la información	8
Clasificación por objetividad	5
Clasificación por polaridad	8
Diseño Web	8
Clasificación por me gusta	2
Ver secciones	2
Información de las secciones	3
Ver temas populares	5
Buscador	2
Actualización información	2
Mapa de localizaciones	13
Resaltar palabras clave	2
Descripción de términos	2

Tabla 23: Estimación de puntos de historia

Se han obtenido un total de 75 puntos de historia para el desarrollo de la aplicación. A continuación se calcula la velocidad de desarrollo.

Como vimos anteriormente, se estima que el desarrollo de la aplicación tendrá un periodo de dos meses, confiando en que esto sea así y con iteraciones de dos semanas, podemos calcular la velocidad. Teniendo un total de 75 puntos de historia, dividido entre en número de iteraciones (en este caso 4 iteraciones) obtenemos un resultado de 19 puntos de historia / iteración.

A continuación se un presenta un diagrama Burndown (Cho, 2008), propio de la metodología *SCRUM*, pero que utilizaremos en cada una de nuestras iteraciones para representar gráficamente el tiempo que falta para completar nuestro trabajo.

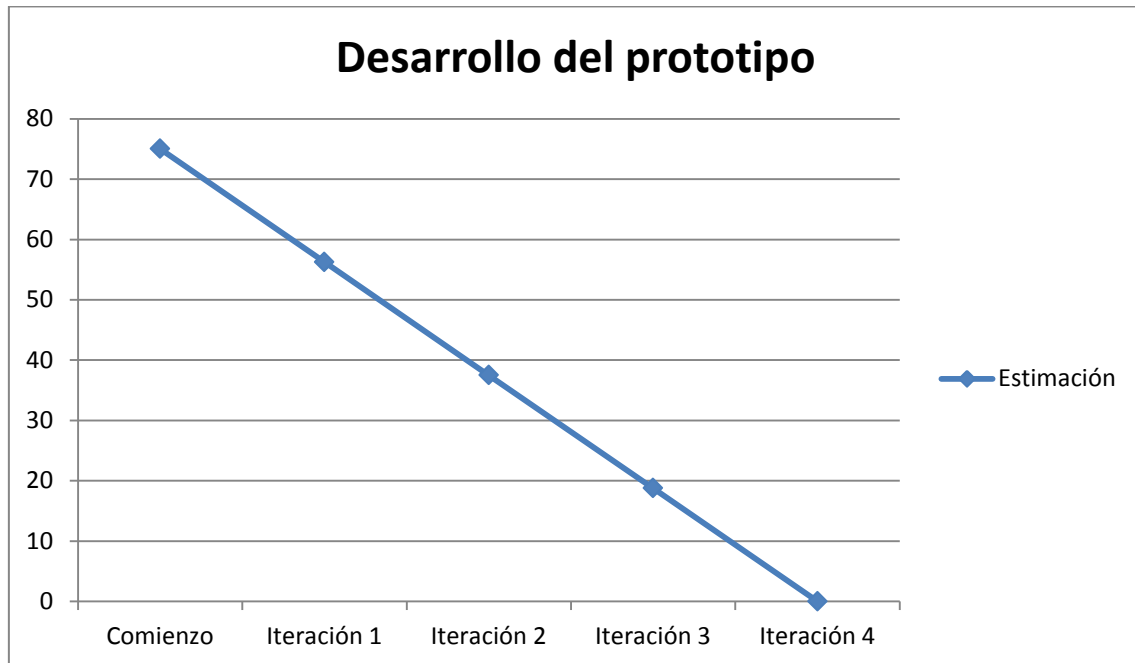


Ilustración 4: Diagrama BurnDown

5.3. Iteración 1

Nuestro proceso de desarrollo sigue la metodología XP, dividida en iteraciones de dos semanas con un total de 4 iteraciones.

5.3.1. Priorización de las historias.

Tras haber planteado las historias de usuario con su dificultad mediante el planning poker y haber calculado la velocidad de desarrollo, procedemos a ordenar las tareas según su importancia y dependencia con el resto de tareas, es decir, en primer lugar tendremos que desarrollar aquellas funcionalidades fundamentales que permiten avanzar en el proyecto.

Historias de usuario	Puntos de historia
Obtención de información	13
Respaldo de la información	8

Diseño Web	8
Mapa de localizaciones	13
Ver secciones	2
Clasificación por objetividad	5
Clasificación por polaridad	8
Clasificación por me gusta	2
Información de las secciones	3
Buscador	2
Ver temas populares	5
Resaltar palabras clave	2
Actualización de la información	2
Descripción de términos	2

Tabla 24: Priorización de las historias de usuario

5.3.2. Definiendo la iteración

Sabiendo que nuestra velocidad estimada es 19 puntos de historia/iteración procedemos a elegir las primeras historias de usuario como tareas de nuestra iteración.

Historias de usuario	Dificultad
Obtención de información	13
Respaldo de la información	8

Tabla 25: Historias de usuario para la primera iteración

Estas historias de usuario las vamos a segmentar en subtarefas que debemos realizar para completar la historia.

Obtención de información
Creación token de acceso
Facebook Scraper
Elección de las fuentes

Tabla 26: Subtarefas de la historia Obtención de información

Respaldo de la información
Modelo Entidad Relación de la BBDD
Controlador

Tabla 27: Subtareas de la historia Respaldo de la información

5.3.2.1. Obtención de información

Historia: Obtención de información	
Como:	Desarrollador
Quiero:	Extraer comentarios y post de grupos relacionados con la enfermedad del cáncer en Facebook.
Para:	Realizar mediante técnicas de PLN un análisis de sentimiento.

Tabla 28: Historia Obtención de información

Como se comentó anteriormente la información se extraerá de la red social Facebook. Para esto, se implementará un Scraper que hará uso de la API de Facebook para obtener la información. Existen varias librerías para Python que agilizan este proceso, pero se optó por realizar un Scraper propio, para poder controlar y ajustar todos los parámetros.

5.3.2.1.1. Token de acceso

Para poder realizar llamadas a la API, vamos a necesitar un token de acceso. Los tokens de acceso son cadenas de caracteres que identifican a un usuario, aplicación o página, permitiendo realizar consultas sobre la API Graph.

Para obtener un token de acceso podemos acceder a:

<https://developers.facebook.com/tools/debug/accesstoken/>

Por defecto los tokens de acceso tienen una duración máxima de una hora, puesto que resultaría engorroso estar cambiando el token de acceso cada poco tiempo necesitaremos generar un token de larga duración.

Pulsando sobre *Extend Acces Token*, generaremos un token con una duración máxima de 30 días, tras esto deberemos renovarlo para poder seguir haciendo llamadas a la API.

Ahora que ya tenemos acceso a la API, se elige una página de la que se quiera extraer información, por ejemplo: Asociación Española Contra el Cáncer.

Para extraer el identificador de la página, debemos fijarnos en su URL.

<https://www.facebook.com/unidoscontraelcancer/>

A partir de esta URL podemos obtener el nombre de la página **unidoscontraelcancer**.

Ahora solo falta definir los parámetros que queremos consultar, algunos campos interesantes son:

Campo	Descripción
id	Identificador de la página
name	Nombre de la página
about	Descripción de la página
category	Categoría a la que pertenece
general_info	Información general de la página
link	URL de la página de Facebook
location	Localización
picture	Foto de perfil
overall_star_rating	Número de estrellas de la página
rating_count	Número de valoraciones de la página
talking_about_count	Número de personas hablando sobre esta página

Tabla 29: Campos relevantes de las páginas de Facebook

Un ejemplo de consulta que podemos realizar puede ser:

276999106249?fields=id,name,about,category,general_info,link,location{city},picture{url},overall_star_rating,rating_count,talking_about_count

Que nos devuelve el siguiente resultado:

The screenshot shows the Facebook Graph API Explorer interface. At the top, there's a 'Token de acceso' field and a 'GET' method dropdown. The URL bar shows the query: `/v2.8 / 276999106249?fields=id,name,about,category,general_info,link,location{city},picture{url},overall_star_rating,rating_count,talking_about_count`. On the left, a list of fields is shown with checkboxes: `id`, `name`, `about`, `category`, `general_info`, `link`, `location`, `picture`, `url`, `overall_star_rating`, `rating_count`, and `talking_about_count`. The right pane displays the JSON response for the node 276999106249, which is the 'Asociación Española Contra el Cáncer' (AECC).

```
{
  "id": "276999106249",
  "name": "Asociación Española Contra el Cáncer",
  "about": "Página oficial de facebook de la Asociación Española Contra el Cáncer (aecc). Más de 60 años comprometido",
  "category": "Non-Profit Organization",
  "link": "https://www.facebook.com/unidoscontraelcancer/",
  "location": {
    "city": "Madrid"
  },
  "picture": {
    "data": {
      "url": "https://scontent.xx.fbcdn.net/v/t1.0-1/p50x50/563404_10150684883126250_499006687_n.jpg?oh=b67495c7e45fi"
    }
  },
  "overall_star_rating": 4.7,
  "rating_count": 6152,
  "talking_about_count": 7401
}
```

Ilustración 5: Ejemplo consulta API Graph

5.3.2.1.2. Implementación del Scraper

Implementación simplificada para extraer datos de una página.

```
def scrapping(self, pageID, access_token):
    base='https://graph.facebook.com/v2.8/'

    params='?fields=id%2Cname%2Cabout%2Clink%2Cfan_count&access_token='
    url=base+pageID+params+access_token
    try:
        req = urllib2.Request(url)
        res = self.opener.open(req)
    except urllib2.HTTPError, e:
        print e.msg
        print url
        return 1
    except urllib2.URLError, e:
        print e.reason[1]
        print url
        return 1
    return json.loads(res.read())
```

5.3.2.1.3. Elección de las fuentes de información

No todas las páginas o grupos de Facebook tiene información relevante, es por esto que necesitamos filtrar la información para quedarnos sólo con aquellas páginas que están relacionadas con el cáncer. Para ello realizamos un primer análisis de los grupos relacionados con el cáncer mediante el comando:

```
search?type=page&q=cancer&fields=name,about
```

Esta búsqueda nos devolverá los grupos que posean la palabra cáncer en su nombre junto con la descripción de la página.

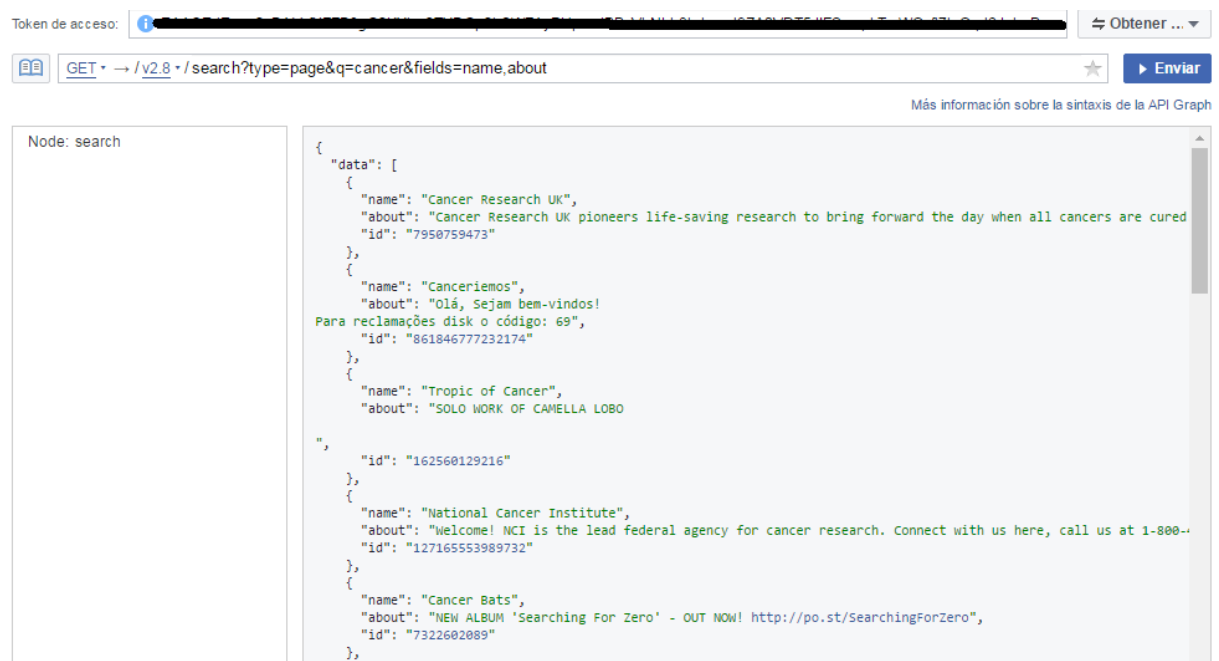


Ilustración 6: Búsqueda de páginas sobre la API Graph

Como podemos observar muchas de estas páginas están en inglés o están relacionadas con el horóscopo. Debido a que el proyecto está orientado al análisis de sentimientos en español, deberemos filtrar los grupos por idioma y eliminar aquellas páginas no relacionadas con la enfermedad del cáncer.

Para conocer si una página es española podemos lanzar un detector de idioma sobre el campo About de dicho grupo. En este proyecto se utilizará la librería langdetect 1.0.7⁴.

Los grupos que finalmente se han seleccionado para la extracción de comentarios y posts son:

De carácter general:

Name	talking_about_count	Fan_count
Asociación Española Contra el Cáncer	10166	314434
VIDA Contra El Cáncer	2222	65339
CRIS contra el Cancer	1862	109891
Vencer el Cáncer	555	6162
Instituto Nacional del Cáncer de EE. UU.	1090	54027
Grupo Español de Pacientes con Cáncer	170	8561

Tabla 30: Grupos sobre el cáncer de Facebook

Cáncer colorrectal:

Name	talking_about_count	Fan_count
Red de contacto para personas con cáncer de colon	136	1252
La Alianza Contra el Cáncer del Colon	6	618

Tabla 31: Grupos sobre el cáncer colorrectal de Facebook

Cáncer próstata:

Name	talking_about_count	Fan_count
APRECAP - Asoc. Pro Prevención Cáncer de Próstata	3	1 1324

Tabla 32: Grupos sobre el cáncer de próstata en Facebook

Cáncer pulmón:

Name	talking_about_count	Fan_count
Asociación Española de Afectados de Cáncer de Pulmón	263	3414

⁴ <https://pypi.python.org/pypi/langdetect>

Amigos de los enfermos de cancer de pulmon	56	1116
ICAPEM - Asociación para la Investigación del Cáncer de Pulmón en Mujeres	1	240

Tabla 33: Grupos sobre el cáncer de pulmón en Facebook

Cáncer de mama:

Name	talking_about_count	Fan_count
FUCAM AC Cáncer de Mama	3538	43956
Mujeres Afectadas de Cáncer de Mama de Aspe - MACMA	362	1163
Cáncer de mama	213	25223
Vencer el Cáncer	555	6162
Ajicam (asociación jiennense cáncer de mama)	471	2156

Tabla 34: Grupos sobre el cáncer de mama en Facebook

Cáncer de vejiga:

Name	talking_about_count	Fan_count
Cáncer De Vejiga	3	28

Tabla 35: Grupo sobre el cáncer de vejiga en Facebook

Posteriormente se agregaron grupos asociados a los cánceres de útero, tiroides, páncreas, leucemia. Actualmente la aplicación extrae información de 25 grupos distintos.

5.3.2.2. Respaldo de la información.

Historia: Respaldo de la información	
Como:	Desarrollador
Quiero:	Una base de datos donde se almacene todos los datos recabados de Internet.
Para:	Poder tratar, analizar o consultar dichos datos.

Tabla 36: Historia Respaldo de la información

Una vez montado el Scraper de Facebook, la siguiente tarea fundamental es diseñar e implementar una base de datos en la que almacenar toda la información necesaria para nuestra aplicación.

En la tabla Post, se guardarán todos los posts realizados en las distintas páginas. Además se incluirá la polaridad y objetividad del mensaje escrito.

Post		
 id_post	TEXT	Clave del post
id_page	TEXT	Clave de la página
created_time	TEXT	Fecha de creación
author	TEXT	Autor del post
link	TEXT	Link del post
message	TEXT	Texto del post
polarity	REAL	Polaridad del texto
objective	REAL	Porcentaje objetividad

Tabla 38: Campos tabla Post

En la tabla Comment, se guardarán todos los comentarios de los Posts escritos en las páginas. Además, se incluirá la polaridad y objetividad del mensaje escrito.

Comment		
 id_comment	TEXT	Clave del comentario
id_post	TEXT	Clave del post
created_time	TEXT	Fecha de creación
author	TEXT	Autor del comentario
message	TEXT	Texto del comentario
like_count	INTEGER	Número de likes
polarity	REAL	Polaridad del texto
objective	REAL	Porcentaje objetividad

Tabla 39: Campos tabla Comment

La tabla Cancer_type nos permitirá saber el tipo de cáncer de una página.

Cancer_type		
 id_type	INTEGER	Identificador tipo de cáncer
name	TEXT	Nombre del tipo de cáncer
description	TEXT	Descripción del tipo de cáncer

Tabla 40: Campos tabla Cancer_type

En la tabla Popularity se guardarán aquellos campos asociados con la popularidad de la página de Facebook como son: número de seguidores, número de personas hablando sobre la página y número de estrellas.

Popularity		
 id_page	TEXT	Clave de página
 time	TEXT	Día y hora
fan_count	INTEGER	Número de seguidores de la página
talking_about_count	INTEGER	Número de personas hablando sobre esta página
star_rating	REAL	Número de estrellas de la página

Tabla 41: Campos tabla Popularity

En la siguiente tabla se almacenarán la polaridad y objetividad asociada a una página de Facebook.

Optimism		
 id_page	TEXT	Clave de página
 time	TEXT	Día y hora
polarity	REAL	Polaridad general de la página
objective	REAL	Porcentaje de objetividad de la página

Tabla 42: Campos tabla Optimism

En esta tabla se almacenarán los temas más populares asociados a una página de Facebook, junto con la fecha en la que se hizo la comprobación.

Key_words		
 word	TEXT	tema
 id_page	TEXT	Clave de página
 time	REAL	Fecha de la comprobación
word_ocurrency	REAL	Número de ocurrencias de la palabra

Tabla 43: Campos de la tabla Key_words

5.3.2.2.2. Controlador

Se ha implementado un controlador, que llama al Scraper de Facebook y descarga la información en la base de datos.

5.3.3. Final de la iteración

Con una velocidad estimada de 19 PH (Puntos de Historia) / Iteración se ha conseguido completar las dos primeras historias de usuario, con un total de 21PH.

Una vez refactorizado el código y comprobado el funcionamiento de la base de datos implementada se ha obtenido los siguiente resultados:

Número de grupos	Número de comentarios	Número de posts
25	122.835	19.063

Tabla 44: Valores actuales de la base de datos

Tras finalizar la primera iteración, el diagrama de Burndown queda de la siguiente manera:

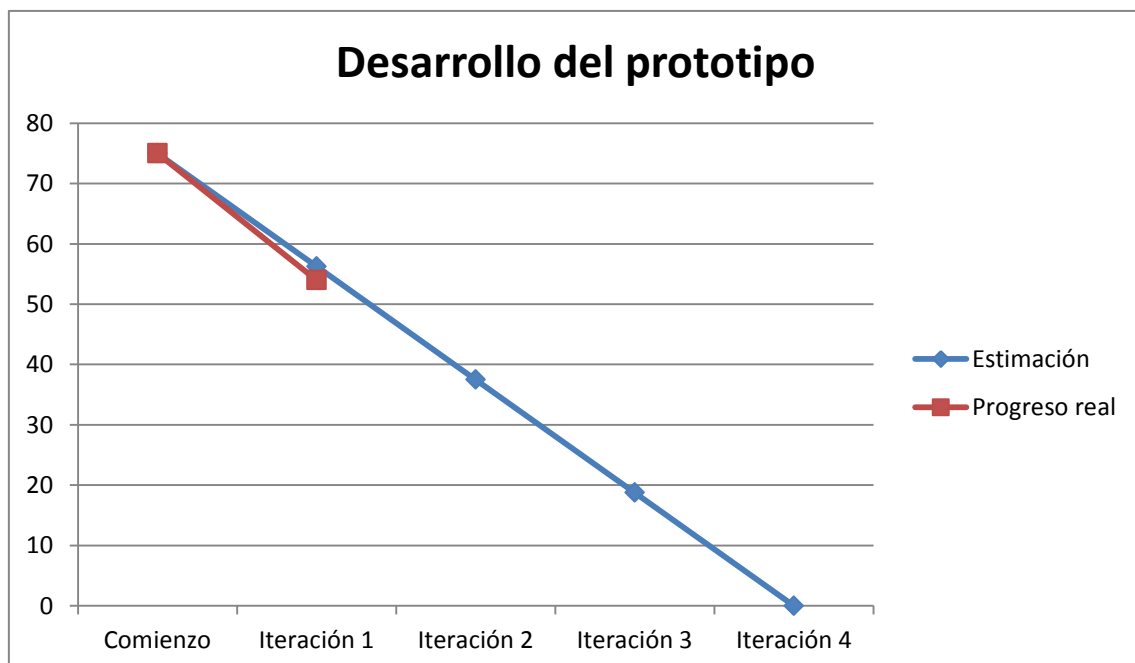


Ilustración 8: Diagrama Burndown primera iteración

5.4. Iteración 2

En esta segunda iteración se elegirá la siguiente historia de usuario acorde con el orden de prioridades.

5.4.1. Priorización de las historias

En un análisis previo de las tareas restantes, se cree conveniente que una vez implementada la base de datos es fundamental que se pueda ir actualizando, recopilando más posts y comentarios recientes de Facebook cada vez que se hagan peticiones a la API de Facebook sin sobrescribir los ya existentes; es por esto que se ha reajustado el orden de las tareas, colocando “Actualización de la información” en tercer lugar.

Historias de usuario	Puntos de historia
Obtención de información	13
Respaldo de la información	8
Actualización de la información	3
Mapa de localizaciones	13
Diseño Web	8
Ver secciones	1
Clasificación por objetividad	5
Clasificación por polaridad	8
Clasificación por me gusta	2
Información de las secciones	3
Buscador	2
Ver temas populares	5
Resaltar palabras clave	2
Descripción de términos	2

Tabla 45: Priorización historias de usuario segunda iteración

5.4.2. Definiendo iteración

Con una velocidad de 19 PH / Iteración se estima poder completar las historias “Actualización de la información” y “Mapa de localizaciones” . Tras esto, se comenzará la historia “Diseño Web”.

5.4.2.1. Actualización de la información

Historia: Actualización información	
Como:	Desarrollador
Quiero:	Descargar los post y comentarios recientes realizados en Facebook, sin sobrescribir los ya existentes en la BBDD.
Para:	Mantener la base de datos actualizada.

Tabla 46: Historia Actualización información

Se ha actualizado el controlador de la base de datos para que cuando realicemos nuevas llamadas al Scraper de Facebook podamos actualizar la información de nuestra base de datos sin sobrescribir los datos.

Para realizar dicha operación, se comprueba el último post escrito en una página Facebook, y se compara con la fecha del último post asociado a esa página almacenado en la base de datos; si existen nuevos posts estos se descargarán automáticamente.

Un aspecto a tener en cuenta es el formato en el que Facebook guarda la fecha de creación de un post o comentario. Esta fecha tiene el formato ISO-8601 y como podemos comprobar en el ejemplo siguiente, no resulta sencilla de entender.

```
"created_time": "2017-02-03T10:24:25+0000"
```

Por consiguiente, se definió una función que convirtiera la fecha con formato ISO-8601, al formato AA/MM/DD HH/MM/SS que resultará mas claro para el usuario de la aplicación.

5.4.2.2. Mapa de localizaciones

Historia: Mapa de localizaciones	
Como:	Usuario
Quiero:	Un mapa de España, en el que se reflejen los lugares en los que se habla de forma objetiva/subjetiva, con un sentimiento positivo/negativo en el caso de información subjetiva y también las zonas con mayor número de seguidores
Para:	Conocer más sobre el cáncer a nivel de España

Tabla 47: Historia Mapa de localizaciones

Subtareas en las que se desglosa:

Mapa de localizaciones
Modificar BBDD
Obtener localización de los usuarios
Implementar Scraper
Dibujar Mapa de España
Implementar Filtros

Tabla 48: Subtareas de la historia Mapa de localizaciones

5.4.2.2.1. Modificar BBDD

Se ha creado una nueva tabla en la base de datos llamada Location.

Location		
Id_profile	INTEGER	Identificador de usuario
coordinates	TEXT	Coordenadas de localización
name_location	TEXT	Nombre del lugar

Tabla 49: Campos de la tabla Location

En esta tabla se almacenarán los IDs de los usuarios que han escrito un post o un comentario, así como aquellos usuarios que le hayan dado a *me gusta* a un post.

5.4.2.2.2. Obtener localización de los usuarios

Mediante la API Graph no se puede obtener directamente la localización de los usuarios. Para solventar este problema, en primer lugar, se extraerán todos los IDs de los usuarios que han escrito o reaccionado a un comentario/post, y seguidamente se implementará un scraper que obtenga la localización asociada a cada usuario.

La obtención del identificador de usuario es tarea sencilla puesto que, cuando realizamos una llamada a la API de Facebook preguntando por quién escribió un post, esta nos devuelve su identificador.

Una vez hemos conseguido los IDs y los tenemos almacenados en la base de datos, podemos proceder a la obtención de la localización de cada uno de ellos.

5.4.2.2.3. Implementación del Scraper

La manera de obtener la localización de un perfil de Facebook, **si este es público**, es dirigirnos a su perfil de usuario y extraer el campo de “Ciudad Actual” o “Localidad Natal”.

Obtenemos el ID de usuario, por ejemplo: 10101640953722381, correspondiente a *Mark Zuckerberg*.

Ahora podemos acceder a su perfil de usuario escribiendo en el navegador:

<https://www.facebook.com/10101640953722381>

Pero para poder acceder desde nuestra aplicación de Python al perfil, tendremos que usar `/app_scoped_user_id/`.

```
url="https://www.facebook.com/app_scoped_user_id/" + ID,
```

A partir de aquí utilizando la librería BeautifulSoup podemos extraer el valor del campo localización:

```
soup.findAll('div', attrs={'title': 'Ciudad actual'})[0].find('a').text
```

Y finalmente obtener las coordenadas del lugar, con la ayuda de la librería *Geocoder*.

5.4.3. Final de la iteración

Con una velocidad estimada de 19 PH / Iteración se ha conseguido completar sólo la primera historia de usuario, puesto que, tras haber conseguido extraer la localización a un conjunto de personas, Facebook ha limitado el número de peticiones que se pueden hacer a través de *app_scoped_user_id*.

Por tanto damos a la historia “Mapa de localizaciones” por finalizada, sin haberse completado, ya que debido a la limitación de Facebook, no es posible extraer todas las localizaciones de los usuarios.

Tras finalizar la segunda iteración, el diagrama de Burndown queda de la siguiente manera:

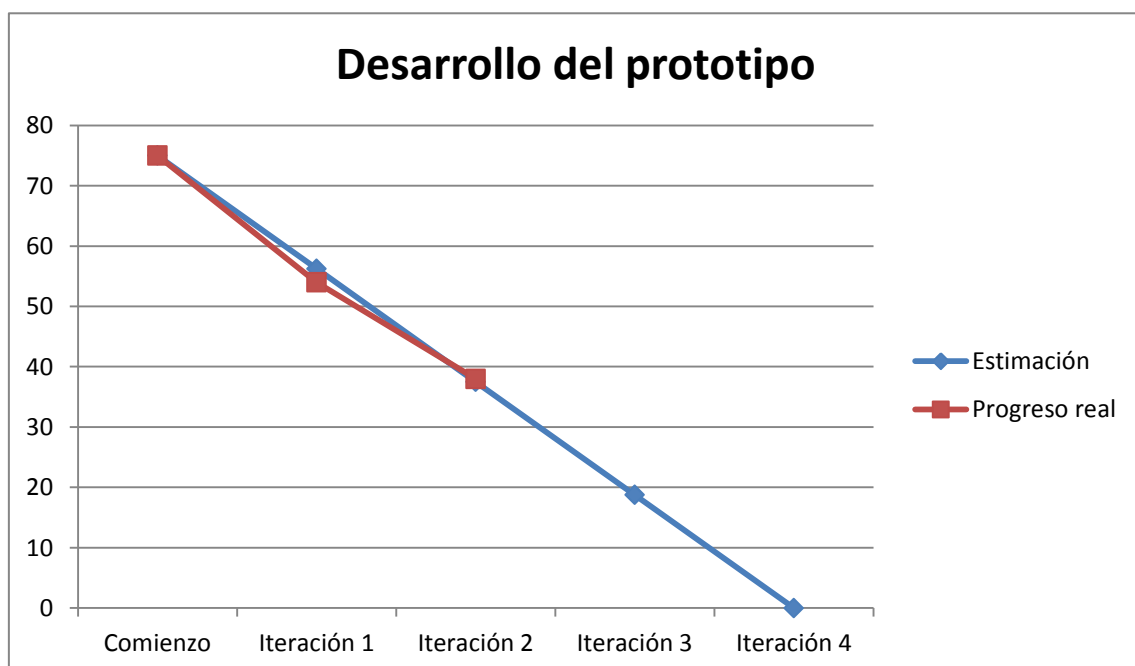


Ilustración 9: Diagrama Burndown de la segunda iteración

5.5. Iteración 3

5.5.1. Priorización

Se realiza un nuevo ajuste de las prioridades, y se establece la historia “Clasificación por objetividad” por detrás de “Clasificación por polaridad” ya que para poder clasificar la polaridad de un mensaje, previamente hay que determinar si este es subjetivo o si por el contrario es objetivo.

Historias de usuario	Puntos de historia
Obtención de información	13
Respaldo de la información	8
Actualización de la información	3
Mapa de localizaciones	13
Diseño Web	8
Clasificación por polaridad	8
Clasificación por objetividad	5
Ver secciones	1
Clasificación por me gusta	2
Información de las secciones	3
Buscador	2
Ver temas populares	5
Resaltar palabras clave	2
Descripción de términos	2

Tabla 50: Priorización de las historias de usuario tercera iteración

5.5.2. Definición de la iteración

En esta tercera iteración se pretende comenzar con el diseño de la página web y a su vez comenzar la implementación de los algoritmos para el cálculo de la polaridad.

Tareas que se estiman completar:

Historia	PH
Diseño Web	8
Clasificación por polaridad	8
Clasificación por objetividad	5

Tabla 51: Historias a desarrollar en la tercera iteración

Con un total de 21 PH.

5.5.2.1. Diseño Web

Historia: Diseño Web	
Como	Usuario de la aplicación
Quiero	Una aplicación web con un diseño fácil, intuitivo y profesional.
Para	Poder navegar cómodamente por la aplicación.

Tabla 52: Historia Diseño Web

Utilizando el framework Flask de Python se desarrollará la web que servirá como interfaz al usuario para poder visualizar los datos.

Debido a la naturaleza dinámica de la página web los datos que se visualizan cambiarán según la información obtenida. La estructura de la página web se va a realizar mediante el lenguaje de plantillas *Jinja2* (Ronacher, 2008)

Tendremos una plantilla con la estructura base del HTML, y a su vez, esta plantilla llamará a otras dependiendo de la información que se quiera mostrar. Por ahora se ha implementado la plantilla base y la cabecera de la web.

Además, nuestro prototipo seguirá la arquitectura Modelo Vista Controlador o MVC, la cual separa la lógica de la aplicación de la interfaz de usuario (González, y Romero, 2012).

- Modelo: representa la lógica de la aplicación.
- Vista: es la interfaz que permite al usuario visualizar los datos e interactuar con la aplicación a través del controlador.
- Controlador: permite la comunicación entre la Vista y el Modelo.

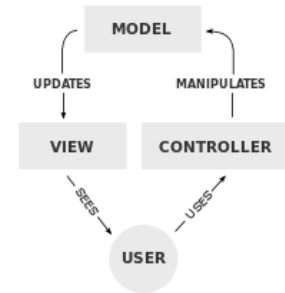


Ilustración 10: Modelo-vista-controlador

5.5.2.1. Clasificación por objetividad

Historia: Clasificación por objetividad	
Como	Usuario de la aplicación
Quiero	Poder organizar los posts y comentarios según su objetividad.
Para	Poder recopilar comentarios según sean opiniones o hechos.

Tabla 53: Historia: Clasificación por objetividad

Los textos pueden clasificarse principalmente en dos tipos: hechos y opiniones:

Los hechos son expresiones objetivas sobre entidades, eventos y sus propiedades.

“El presidente del gobierno español, Mariano Rajoy, visita la ciudad de Jaén con motivo de la inauguración del nuevo museo íbero.”

Las opiniones son usualmente expresiones subjetivas que describen los sentimientos, apreciaciones de las personas hacia entidades, eventos y sus propiedades.

“Este verano he disfrutado mucho yendo con mis amigos de vacaciones, la ciudad que visitamos estaba muy limpia, y la gente fue muy hospitalaria”

5.5.2.1.1. Normalización de los textos.

Los datos sin procesar adquiridos de Facebook mediante Facebook API Graph necesitan ser preprocesados antes de lanzar un análisis completo. Los pasos de normalización de la información empleados en la aplicación son: tokenización, eliminar palabras vacías (stopwords), convertir letras mayúsculas en minúsculas, eliminar signos de puntuación y la eliminación de tildes.

- **La tokenización** se utiliza para dividir una oración en palabras, frases, símbolos u otros elementos significativos. Para ello empleamos la función `word_tokenize` de la librería NLTK .
- **Filtrado de stopwords:** Existen palabras que son irrelevantes en el análisis de sentimientos, como preposiciones, artículos, entre otras. Estas palabras son eliminadas de las publicaciones porque pueden generar resultados imprecisos durante la clasificación. La librería NLTK también dispone de un diccionario de stop words de la lengua española, `stopwords.words("spanish")`.
- **La conversión de letras mayúsculas en minúsculas, así como la eliminación de signos de puntuación y tildes**, son necesarias para la obtención de un texto normalizado.

```
def remove_acents(self, s):  
    return ''.join((c for c in unicodedata.normalize('NFD', s) if  
unicodedata.category(c) != 'Mn'))
```

Una vez hemos normalizado el texto, procedemos a determinar si es objetivo o subjetivo y a calcular su polaridad en el caso de que sea subjetivo.

5.5.2.1.2. Cálculo de la objetividad

Para determinar si un texto es objetivo o subjetivo se utilizará la siguiente heurística. Se calculará el número de palabras indicadoras de opinión presentes en el texto, es decir, una vez normalizado el texto y segmentado en palabras comprobaremos si estas se encuentran dentro del lexicon iSOL (*improved Spanish Opinion Lexicon*). A continuación se calculará el porcentaje de palabras que expresan opinión respecto al total de palabras, de manera que

un texto será clasificado como subjetivo si al menos dos de cada diez palabras son indicadoras de opinión y como objetivo en caso contrario.

5.5.2.2. Clasificación por polaridad

Historia: Clasificación por polaridad	
Como	Usuario de la aplicación
Quiero	Poder organizar los posts y comentarios según su polaridad
Para	Poder conocer que textos son positivos o negativos

Tabla 54: Historia Clasificación por polaridad

5.5.2.2.1. Enfoques para la clasificación de la polaridad.

Existen dos enfoques diferentes que se pueden utilizar para calcular la polaridad de un mensaje: el basado en la orientación semántica de las palabras y el basado en técnicas de aprendizaje automático. El primero de ellos consiste en el uso de un lexicón (diccionario de términos clasificados con un valor de sentimiento) para localizar palabras con carga sentimental. El segundo de ellos, enfoque de aprendizaje automático, consiste en entrenar un clasificador en base a una colección de textos etiquetados a priori con el fin de clasificar cualquier texto de entrada a partir del reconocimiento de patrones. Ambos enfoques tienen ventajas y deficiencias:

Los enfoques de aprendizaje automático obtienen mejores resultados en dominios acotados. Sin embargo, para ser implementados, requieren un conjunto de mensajes ya etiquetados (cuanto mayor sea el conjunto, más preciso serán) para que puedan aprender patrones que distingan mensajes positivos, neutros y negativos.

Los enfoques basados en un orientación semántica requieren un diccionario de palabras, cada una de ellas anotadas con su orientación semántica (polaridad emocional positiva / negativa). Este enfoque no depende un conjunto de datos previamente etiquetados. Uno de sus inconvenientes es que existen pocos recursos disponibles y son dependientes del idioma.

En nuestro proyecto utilizaremos un enfoque basado en un lexicón, más concretamente *improved Spanish Opinion Lexicon* o iSOL, que nos proporciona una lista de palabras indicadoras de opinión. La lista está constituida por 2.509 palabras positivas y por 5.626 negativas.

5.5.2.2.2. Cálculo de la polaridad

Para determinar la polaridad de los textos se utilizará la siguiente heurística. Se calculará el número de palabras positivas y negativas presentes en el texto, es decir, una vez normalizado el texto y segmentado en palabras comprobaremos si estas se encuentran dentro de la lista de palabras positivas o negativas de iSOL. A continuación se calculará el porcentaje de palabras positivas respecto al total de palabras que indican opinión, de manera que una opinión será clasificada como positiva si este porcentaje es igual o superior al 50%. Por ejemplo si una opinión tiene diez palabras indicadoras de opinión y siete de ellas son positivas, esta será clasificada como positiva.

5.5.2.2.3. Representación

A continuación se muestra la escala de polaridad que se va a utilizar en el prototipo.

Porcentaje de positividad	Representación
Polaridad $\leq 30\%$	muy negativo
$30\% \geq \text{Polaridad} \leq 40\%$	negativo
$40\% \geq \text{Polaridad} \leq 50\%$	neutro
$50\% \geq \text{Polaridad} \leq 70\%$	positivo
Polaridad $\geq 70\%$	muy positivo

Tabla 55: Representación de la polaridad

5.5.3. Final de la iteración

Hemos completado todas las tareas previstas en esta iteración, con un total de 21PH.

Tras refactorizar el código y hacer distintas pruebas de clasificación de texto según su polaridad y objetividad se ha finalizado la tercera iteración. El diagrama de Burndown queda de la siguiente manera

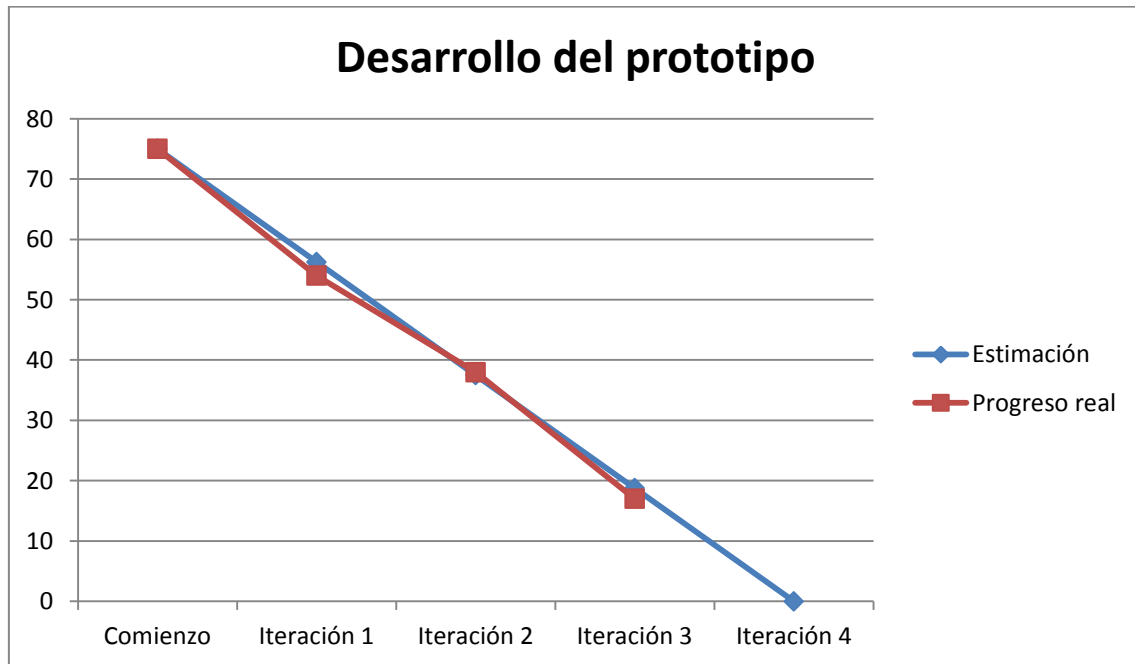


Ilustración 11: Diagrama Burndown tercera iteración

5.6. Iteración 4

En esta última iteración se espera terminar con el resto de tareas restantes para finalizar el prototipo.

5.6.1. Priorización

Historias de usuario	Puntos de historia
Obtención de información	13
Respaldo de la información	8
Actualización de la información	3
Mapa de localizaciones	13

Diseño Web	8
Clasificación por polaridad	8
Clasificación por objectividad	5
Ver secciones	1
Clasificación por me gusta	2
Información de las secciones	3
Buscador	2
Ver temas populares	5
Resaltar palabras clave	2
Descripción de términos	2

Tabla 56: Priorización de las historias de usuario de la cuarta iteración

5.6.2. Definición de la iteración

5.6.2.1. Clasificación por me gusta

Historia: Clasificación por me gusta	
Como	Usuario de la aplicación
Quiero	Poder organizar los posts y comentarios por número de likes, wow, love, angry y sad
Para	Poder ordenar los textos según estos parámetros

Tabla 57: Historia: Clasificación por me gusta

Esta tarea es muy sencilla, puesto que tenemos almacenado el número de likes de cada post y comentario en nuestra base de datos. Simplemente haremos una consulta que nos devuelva los textos ordenados por número de *likes*.

5.6.2.2. Ver secciones

Historia: Ver secciones	
Como	Usuario de la aplicación
Quiero	Una seccion que muestre: • Tipos de cáncer • Tratamientos • Fármacos.
Para	Poder consultar información acerca de estas secciones.

Tabla 58: Historia Ver secciones

Se continua con el diseño de la página web y se crean las secciones citadas.



Ilustración 12: Visualización de las secciones

5.6.2.3. Información de las secciones

Historia: Información de las secciones	
Como:	Usuario de la aplicación
Quiero:	Una tabla que muestre la polaridad, nº menciones y tendencia asociada a cada tema de cada sección (Tipos de cáncer/Tratamientos/Fármacos)
Para:	Poder consultar información sobre ellos.

Tabla 59: Historia Información de las secciones

Esta historia de usuario se puede dividir en las siguientes funciones:

Información de las secciones
Cálculo de la tendencia
Selección de tipos de cáncer por defecto
Diseño de la vista

Tabla 60: Subtareas de la historia Información de las secciones

5.6.2.3.1.1. Cálculo de la tendencia

La tendencia indica el crecimiento o decrecimiento de opiniones negativas/positivas en los últimos siete días. Para ello realizamos la sumatoria del porcentaje de positividad de los comentarios almacenados en la base de datos, con fecha creación entre el día actual y siete días atrás, a este valor se le resta el porcentaje de comentarios negativos.

El valor se indica en la página web con una flecha ascendente o descendente.

5.6.2.3.1.2. Selección de tipos de cáncer

Los tipos de cáncer que se mostrarán son aquellos que más afectan a la sociedad española, según el estudio realizado en capítulos anteriores.

5.6.2.3.1.3. Diseño de la vista

A continuación se muestra la interfaz web que se ha desarrollado para visualizar los datos.



Tipos de Cáncer Tratamientos Fármacos

Realice búsquedas sobre otros temas.

	Topic	Menciones	Sentimiento	Tendencia
	Cáncer de mama	1993	negativo	↑
	Cáncer de colon	432	negativo	↑
	Leucemia	1215	positivo	↑
	Cáncer de pulmón	618	negativo	—
	Cáncer de tiroides	526	neutro	—
	Cáncer de próstata	150	muy negativo	—

Ilustración 13: Visualización de la página web

5.6.2.4. Ver temas populares

Historia: Ver temas populares	
Como:	Usuario de la aplicación
Quiero:	Ver una gráfica que muestre los temas a los que más se hace referencia en los comentarios de Facebook.
Para:	Poder conocer cuáles son los temas más populares.

Tabla 61: Historia Ver temas populares

Para conocer los temas más populares de una sección, se utiliza el la clase Counter del módulo Collections de Python.

Un Counter es una colección no ordenada donde los elementos se almacenan como claves de diccionario y su total se almacena como valores de diccionario. Se permite que los conteos sean cualquier valor entero incluyendo cero o números negativos.

Una vez normalizado un texto y segmentado en tokens, podemos conocer el número de veces que se repite cada token o palabra en el texto. Utilizando la función *most_common()* obtenemos las palabras que más se han utilizado.

El diagrama se va a realizar con la librería Pygal de Python, la cual nos permite crear todo tipo de gráficas y diagramas.

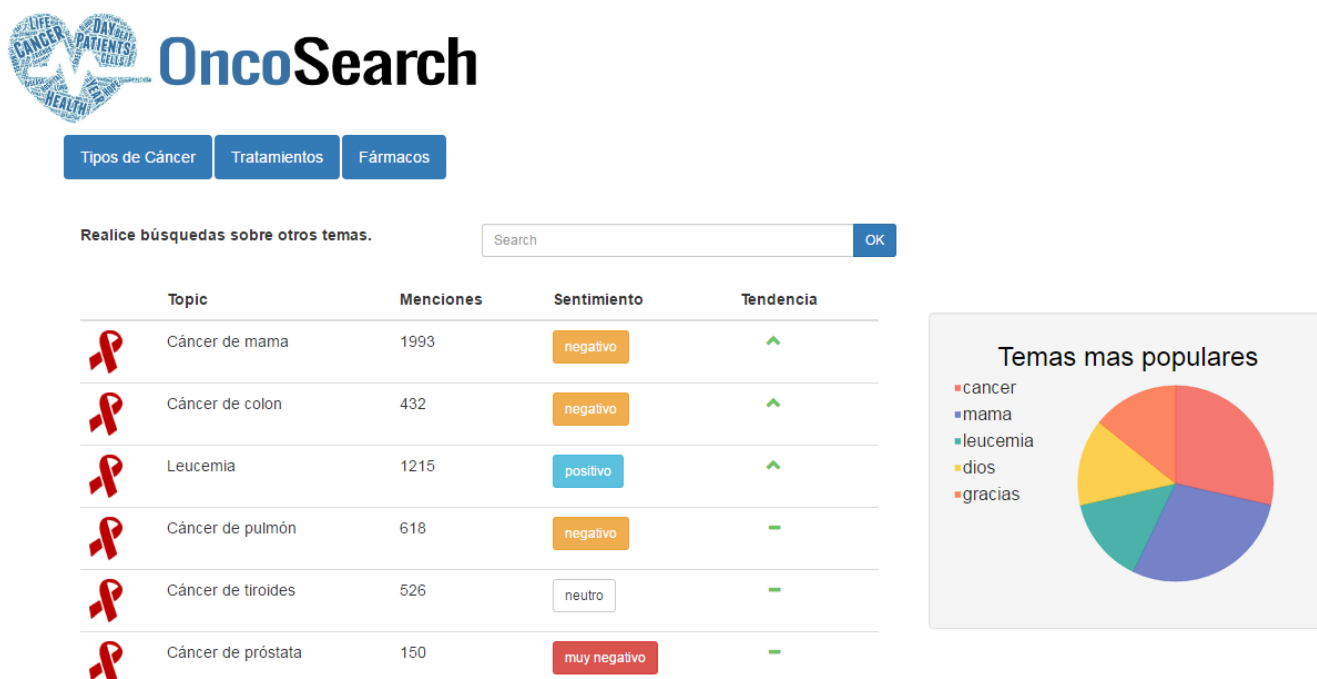


Ilustración 14: Vista de la página web con el diagrama de temas populares

5.6.2.5. Resaltar palabras clave

Historia: Resaltar palabras clave	
Como:	Usuario
Quiero:	Que la aplicación resalte las palabras de los comentarios asociadas con un sentimiento positivo o negativo
Para:	Saber qué palabras del comentario se han detectado como positivas o negativas

Tabla 62: Historia Resaltar palabras clave

Esta funcionalidad permite al usuario conocer aquellas palabras que indican una opinión.



Tipos de Cáncer

Tratamientos

Fármacos

Descripción

Comentarios

Posts

Likes

Positivos

Negativos

Objetivos

Subjetivos

Anterior

Siguiente

muy positivo

2013-11-28 17:22:34

holaaa mucho **animo** para ti, yo este verano sali de un **cancer** de colon **derecho** me hicieron una colectomia y todo ha ido **super** ya llevo mi segundo control y no ha habido **novedades**, solo buscate un **buen** cirujano coloproctologo y seposita tu **confianza** y fe en el ... muchas fuerzas para ti..un **abrazo** gigantr desde chile.

Visitar

muy positivo

2014-05-02 10:54:59

yo estoi operada de **cancer** de colon pero tengo una **amiga** mui querida para mi que si tienene omejor tenia lucemia y esta muy **bien** esta muy **contenta** los tratamientos desde il principio le fueron **estupendamente** asi tienes que tener mucha fuerza el **animo** fundamental todo nuestro apollo y carino

Visitar

muy positivo

2015-09-18 17:50:20

yo padeci **cancer** de colon hace 7 anos me dieron 6 meses quimio y estoy **fenomenal** me hacen una revision anual y hasta ahora todo me sale **bien** asi que lorena dile a tu madre que piense en **positivo** y ya veras que todo va ir **fenomenal** y que **disfrute** de la vida

Visitar

Ilustración 15: Vista de los comentarios con las palabras resaltadas según su polaridad

Las palabras positivas se marcan como verdes, y las negativas como rojas.

5.6.2.6. Descripción de términos.

Historia: Descripción de términos	
Como:	Usuario
Quiero:	Que la aplicación extraiga información de Wikipedia
Para:	Obtener una descripción del tema que estoy consultando

Tabla 63: Descripción de términos

En esta historia de usuario se pretende obtener una descripción del tema que se está consultando para esto se creará una función que permite obtener de Wikipedia esta información.

Aquí se muestra el Scraper que se ha implementado:

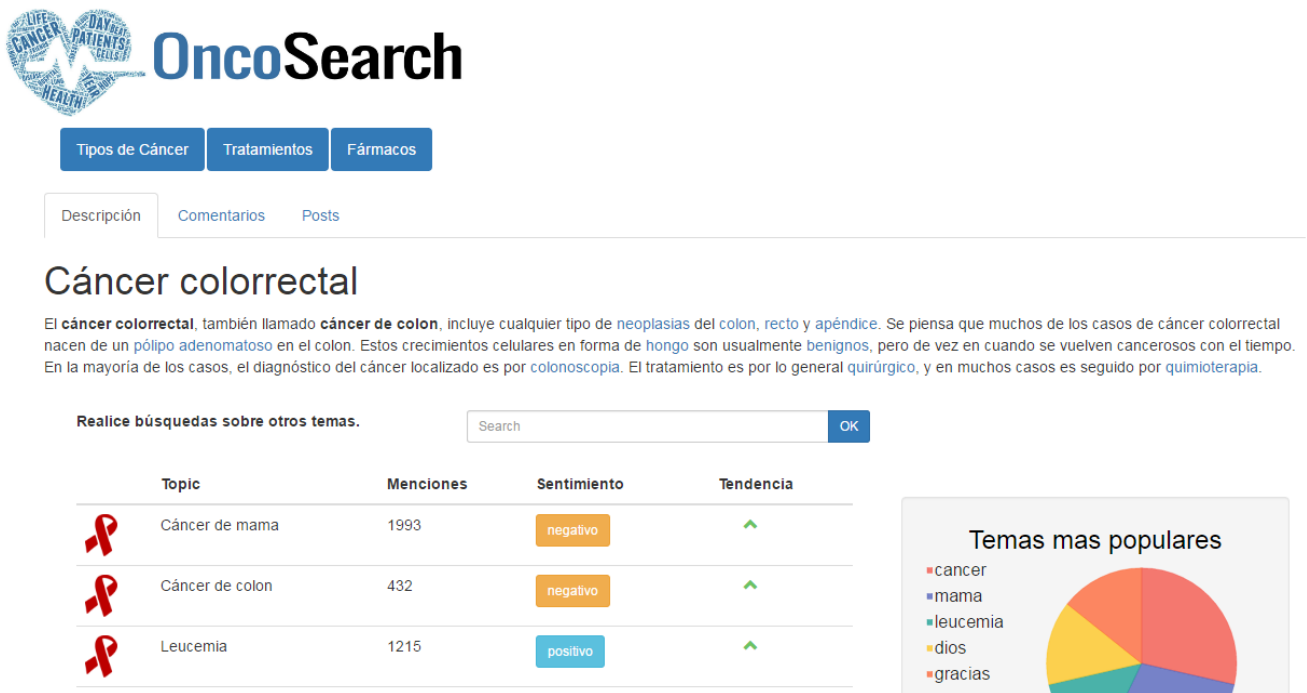
```
def about_wiki(self, term):
    url = "https://es.wikipedia.org/wiki/" + term.encode('utf-8')
    req = urllib2.Request(url)
    req.add_header('User-Agent', 'Mozilla/5.0')
    html = "No existe información sobre este campo".decode('utf-8')
    try:
        res = urllib2.urlopen(req)
        soup = BeautifulSoup(res, 'html.parser')
        h1 = soup.find('h1')
```

```

        about = soup.find('p')
        html = str(h1).decode('utf-8') + str(about).decode('utf-
8')
    except urllib2.HTTPError, e:
        print e.msg
    except urllib2.URLError, e:
        print e.reason[1]
    return html

```

Vista de la descripción del tema seleccionado.



OncoSearch

Tipos de Cáncer | Tratamientos | Fármacos

Descripción | Comentarios | Posts

Cáncer colorrectal

El **cáncer colorrectal**, también llamado **cáncer de colon**, incluye cualquier tipo de **neoplasias del colon, recto y apéndice**. Se piensa que muchos de los casos de cáncer colorrectal nacen de un **pólipo adenomatoso** en el colon. Estos crecimientos celulares en forma de **hongo** son usualmente **benignos**, pero de vez en cuando se vuelven cancerosos con el tiempo. En la mayoría de los casos, el diagnóstico del cáncer localizado es por **colonoscopia**. El tratamiento es por lo general **quirúrgico**, y en muchos casos es seguido por **quimioterapia**.

Realice búsquedas sobre otros temas.

Topic	Menciones	Sentimiento	Tendencia
 Cáncer de mama	1993	negativo	↑
 Cáncer de colon	432	negativo	↑
 Leucemia	1215	positivo	↑

Temas mas populares

- cancer
- mama
- leucemia
- dios
- gracias



Ilustración 16: Vista de la descripción del tema

5.6.2.1. Buscador

Historia: Buscador	
Como:	Usuario de la aplicación
Quiero:	Poder buscar sobre otros temas que no aparezcan definidos en las secciones por defecto.
Para:	Poder consultar información sobre ellos.

Tabla 64: Historia Buscador

El buscador se ha implementado con una consulta SQL que devuelve aquellos post o comentarios que contienen el término buscado.

La aplicación muestra el porcentaje de positividad, así como número de menciones y su tendencia.



Tipos de Cáncer

Tratamientos

Fármacos

Descripción

Comentarios

Posts

Neoplasia

Neoplasia es el término que se utiliza en **medicina** para designar una masa anormal de **tejido**. Se produce porque las **células** que lo constituyen se multiplican a un ritmo superior a lo normal. Las neoplasias pueden ser benignas cuando se extienden solo localmente y malignas cuando se comportan de forma agresiva, comprimen los tejidos próximos y se diseminan a distancia.^[1]

Realice búsquedas sobre otros temas.

Search

OK

Topic	Menciones	Sentimiento	Tendencia
 neoplasia	21	negativo	—

Ilustración 17: Ejemplo de búsqueda

5.6.3. Final de la iteración

Tras finalizar la tercera iteración, el diagrama de Burndown queda de la siguiente manera:

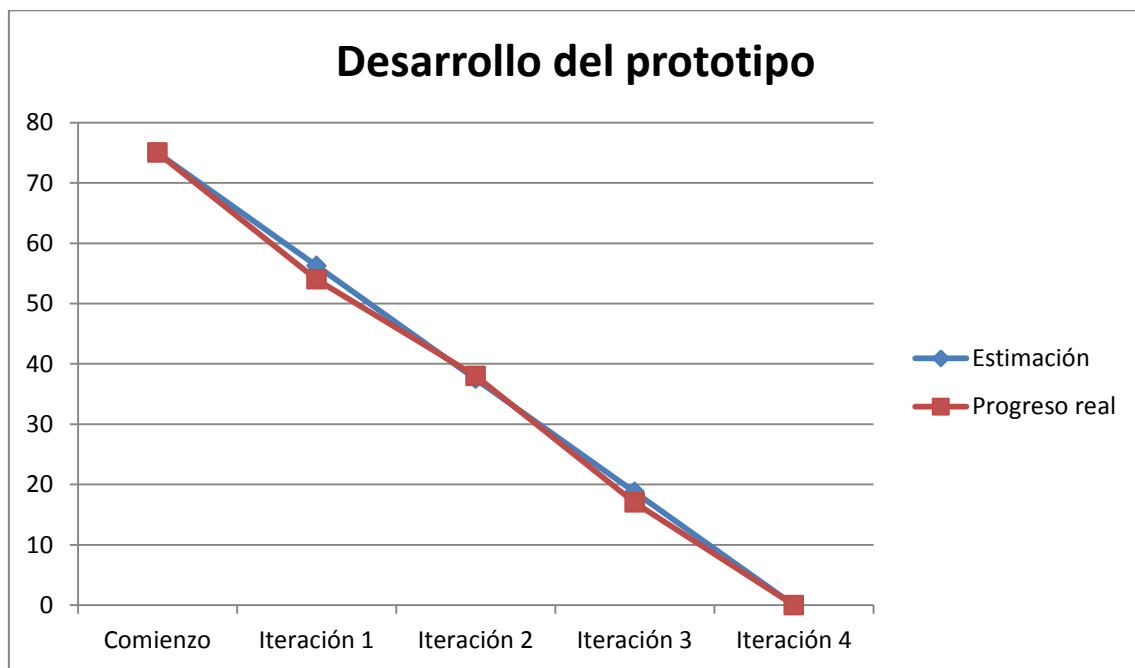


Ilustración 18: Diagrama Burndown cuarta iteración

6. Conclusiones

Debido a la necesidad real de los pacientes del cáncer de conocer más opiniones de otros usuarios de Internet afectados por esta enfermedad se ha implementado una aplicación web que extrae comentarios de una de las redes sociales más populares actualmente, Facebook. La aplicación se ha desarrollado con la finalidad de proporcionar a sus usuarios una herramienta la cual les permita buscar opiniones y comentarios sobre distintos temas en el ámbito del cáncer.

En referencia a la elección de una metodología ágil, más concretamente XP (eXtreme Programming) basada en el desarrollo por iteraciones para la elaboración del prototipo, ha permitido adaptarse mejor a los cambios que han ido surgiendo en el transcurso del proyecto. Seguir el proceso de esta metodología ha mejorado el resultado final de la aplicación obtenida ya que la refactorización y las pruebas frecuentes han servido para aumentar la calidad del código.

En lo que concierne a trabajos futuros se propone continuar con el sistema de análisis de sentimientos. El sistema de clasificación por polaridad puede mejorarse con la identificación de la negación y el procesamiento de emoticonos. Se pretende aumentar las fuentes de extracción de información para enriquecer la actual base de datos. Finalmente agregar un comparador de páginas de Facebook que permita entre otras cosas conocer qué páginas aportan más datos objetivos y cuáles más opiniones.

Agradecimientos

Por último dar mis agradecimientos a Salud María Jiménez Zafra por sus sugerencias y consejos durante todo el transcurso del proyecto que me han ayudado a mejorar este de manera significativa. Agradecer a María Teresa Martín Valdivia por ofrecerme la posibilidad de investigar y conocer más sobre el análisis de sentimientos. Doy las gracias a mi amigo Alberto Vico por las valiosas discusiones que me han ayudado a aclarar conceptos prácticos. También quiero expresar mi gratitud a mi familia y a Elif Çengel que a pesar de la distancia me ha dado su apoyo incondicional en todo momento.

7. Bibliografía.

Ahkter, J. K., & Soria, S. (2010). Sentiment analysis: Facebook status messages. *Unpublished master's thesis, Stanford, CA*.

Bi, C. (2009). Research and application of SQLite embedded database technology. *wseas transactions on computers*, 8(1), 83-92.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."

Cano, A. (2005). Control emocional, estilo represivo de afrontamiento y cáncer: ansiedad y cáncer. *Psicooncología*, 2(1), 71-80.

Cho, J. (2008). Issues and Challenges of agile software development with SCRUM. *Issues in Information Systems*, 9(2), 188-195.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., ... & Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*, 136(5), E359-E386.

González, Y. D., & Romero, Y. F. (2012). Patrón Modelo-Vista-Controlador. *Revista Telemática*, 11(1), 47-57.

Jeff Zabin and Alex Jefferies. Social media monitoring and analysis: Generating consumer insights from online conversation. Aberdeen Group Benchmark Report, January 2008.

Mahnič, V., & Hovelja, T. (2012). On using planning poker for estimating user stories. *Journal of Systems and Software*, 85(9), 2086-2095.

Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, 40(18), 7250-7257.

Ronacher, A. (2008). Jinja2 Documentation.

Van Rossum, G. (2009). What's New In Python 3.0. *What's New In Python 3.0—Python V3. 0.1 Documentation*.

Weinberg, R. A. (1996). How cancer arises. *Scientific American*, 275(3), 62-71.

8. Anexo A. Manual de instalación.

En este anexo se explica los pasos a seguir para instalar las herramientas que se emplearán en el desarrollo de la aplicación.

El lenguaje de programación utilizado ha sido Python, más concretamente la versión 2.7.13. Esta versión se puede descargar desde la página <https://www.python.org/downloads/>. Es importante saber que la nueva versión Python 3.0 es incompatible con las versiones 2.x. El lenguaje es principalmente el mismo, pero cambian muchas de las funcionalidades (Van Rossum, 2009).

Una vez se ha instalado Python se procede a la instalación del IDE (Integrated Development Environment) en este proyecto se optó por utilizar PyCharm. Puede ser descargado a partir del siguiente enlace: <https://www.jetbrains.com/pycharm/download>.

En la elaboración del prototipo se ha hecho uso de algunas librerías tales como:

- **Sqlite:** Proporciona una base de datos embebida que no requiere un servidor independiente y permite acceder a la base de datos utilizando una variante de SQL.
- **Datetime:** Se ha utilizado para convertir la fecha con el formato ISO-8601 a un formato más legible.
- **Flask:** Microframework utilizado para crear la aplicación web.
- **Urllib2:** Librería que nos permite leer los ficheros HTML de las páginas web.
- **Json:** Se usa para tratar los objetos JSON devueltos por la API Graph de Facebook.
- **Counter:** Subclase de la librería *collections* que permite contar elementos. Utilizada en el gráfico de *Temas populares*.
- **BeautifulSoup:** Esta librería de Python facilita la extracción de datos de archivos HTML y XML.
- **Pygal:** Permite la creación de gráficos con una gran variedad de estilos.

PyCharm proporciona una herramienta para la instalación de estas librerías. Para ello debemos acceder al *Project Interpreter* dentro de la opciones de configuración del proyecto.

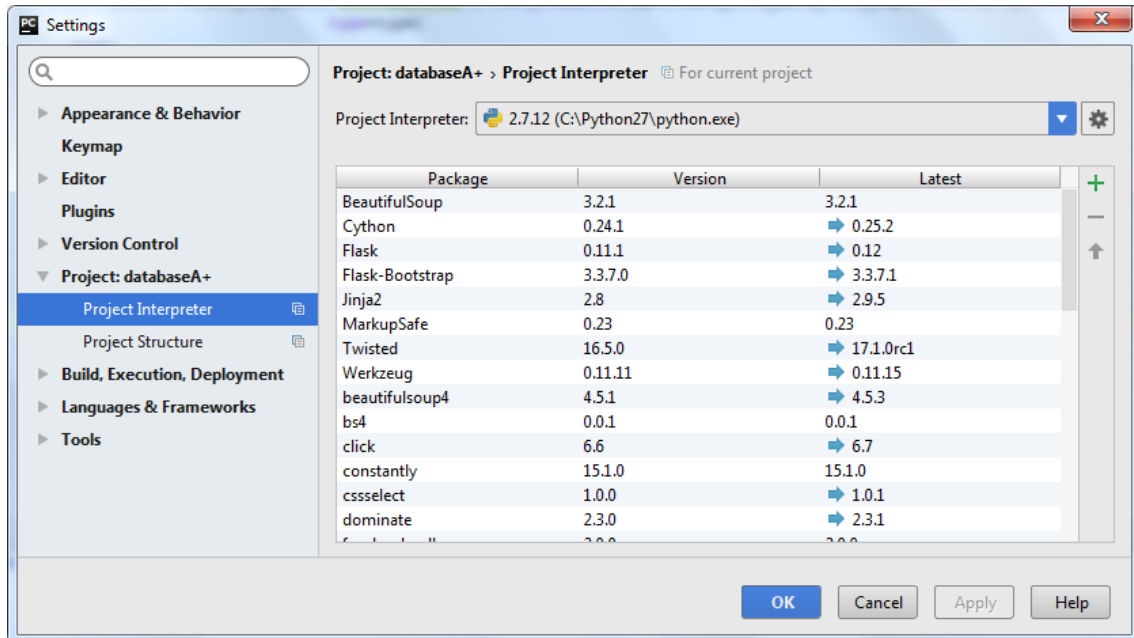


Ilustración 19: Project Interpreter PyCharm

Pulsando sobre *Install* podemos añadir un nuevo paquete.

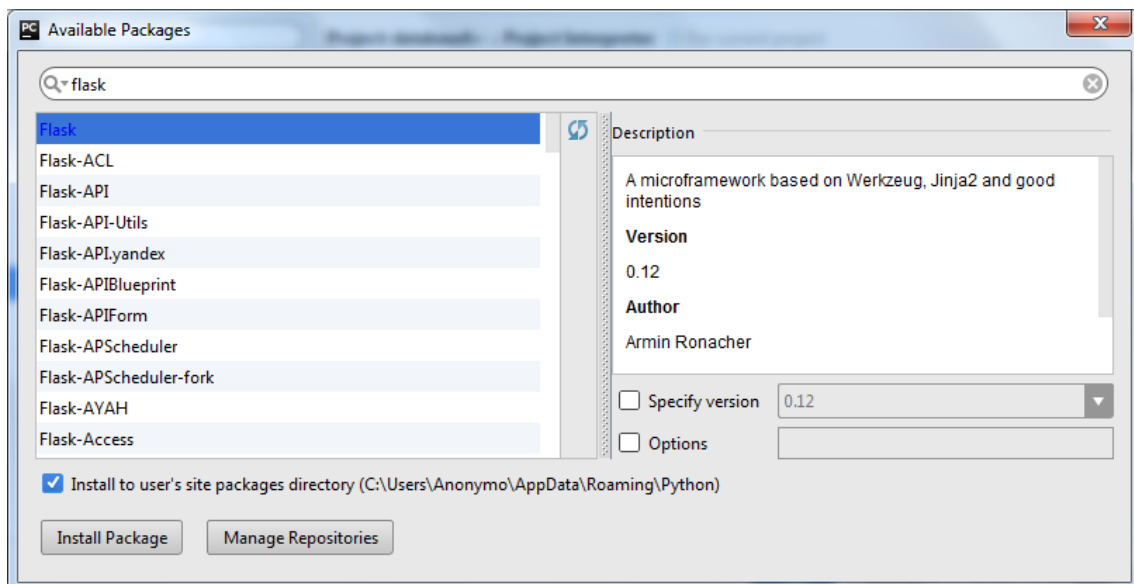
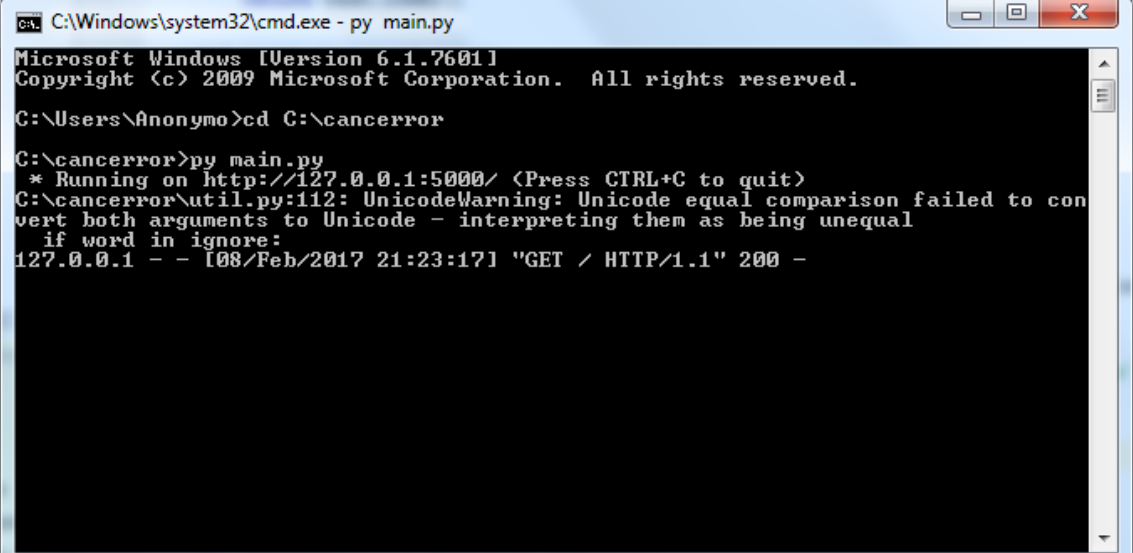


Ilustración 20: Paquetes disponibles para instalar

Finalmente se hace click en *Install Package*.

Podemos lanzar la aplicación desde PyCharm pulsando sobre la opción *Run 'main'* o desde la consola de comandos de Windows.

Desde la consola, nos dirigimos a la carpeta del proyecto y mediante el comando `py main.py` se inicia la ejecución de la aplicación.



```
C:\Windows\system32\cmd.exe - py main.py
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\Anonymo>cd C:\cancerror

C:\cancerror>py main.py
* Running on http://127.0.0.1:5000/ <Press CTRL+C to quit>
C:\cancerror\util.py:112: UnicodeWarning: Unicode equal comparison failed to con
vert both arguments to Unicode - interpreting them as being unequal
  if word in ignore:
127.0.0.1 - - [08/Feb/2017 21:23:17] "GET / HTTP/1.1" 200 -
```

Ilustración 21: Ejecución de la aplicación desde la consola

Por último abrimos el navegador y nos dirigimos a <http://127.0.0.1:5000/>.

9. Anexo B. Manual de usuario.

En este anexo se explica el funcionamiento del prototipo que se ha desarrollado para este proyecto.

Al lanzar la aplicación se muestra el panel principal como vemos en la siguiente ilustración.

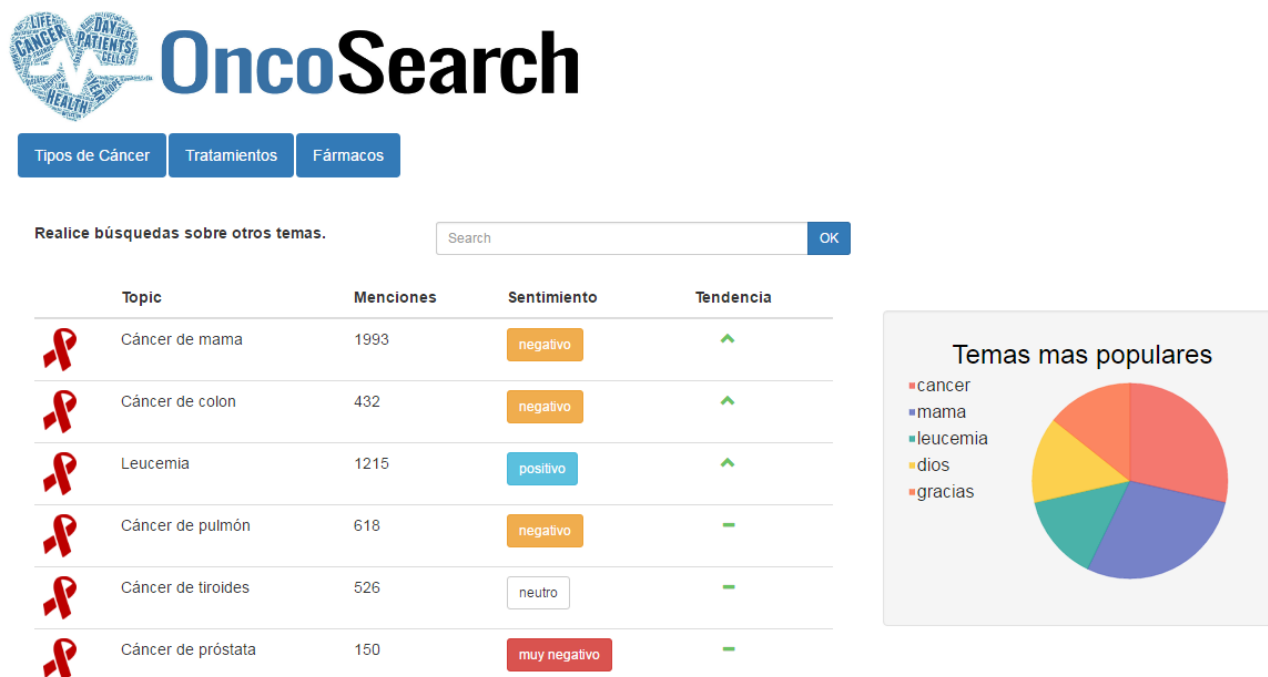


Ilustración 22: Vista principal del prototipo

En la parte superior de la página web se encuentra el logo de la aplicación mediante el cual se podrá volver al inicio desde cualquier apartado.

En el menú principal tenemos tres secciones (*Tipos de Cáncer*, *Tratamientos*, *Fármacos*) pulsando sobre cada una de ellas accedemos a la información de cada apartado.

Al ejecutar la aplicación nos encontramos con la sección *Tipos de Cáncer*, en este apartado se encuentran los cánceres más prevalentes en la sociedad española según los datos recopilados en el capítulo 4.1.

La tabla muestra por filas, el tipo de cáncer junto con el número de menciones en la base de datos, el sentimiento (*opiniones positivas o negativas sobre este cáncer*, representando con el código de colores explicado en el

apartado 5.9.2.4 Representación) y su tendencia (si las últimas referencias sobre este cáncer son positivas o negativas).

En la parte derecha de la pantalla tenemos los temas más referenciados de cada una de las secciones, que se representan mediante un gráfico circular. Este gráfico se actualiza en función de los datos de cada apartado (*Tipos de cáncer, Tratamientos, Fármacos*)

En el apartado *Tratamientos* tenemos los procedimientos más utilizados en la lucha contra el cáncer.

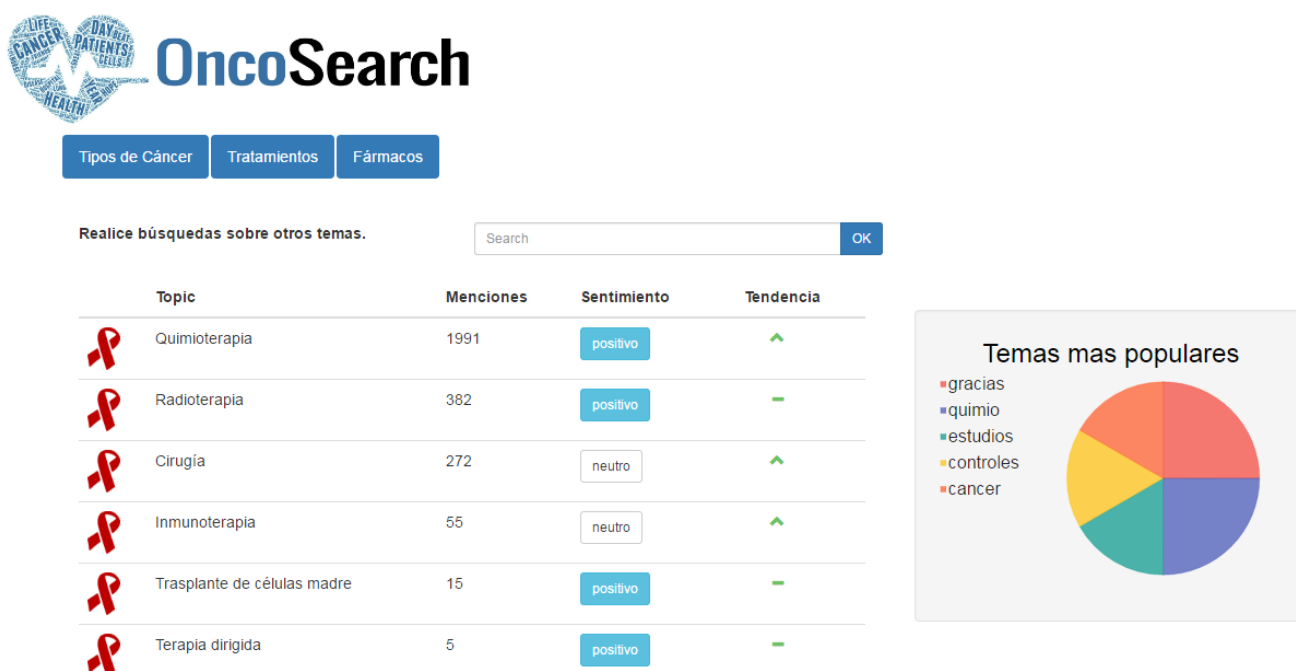


Ilustración 23: Apartado de Tratamientos

Finalmente el apartado *Fármacos* se muestra algunos de los medicamentos que los pacientes utilizan en el tratamiento del cáncer.



Tipos de Cáncer Tratamientos Fármacos

Realice búsquedas sobre otros temas.

	Topic	Menciones	Sentimiento	Tendencia
	tamoxifeno	181	positivo	—
	Letrozol	32	neutro	—
	levotiroxina	18	neutro	—
	XELODA	11	positivo	—
	Exemestano	10	muy positivo	—
	Zoladex	12	muy positivo	—

Ilustración 24: Apartado de Fármacos

Como podemos observar en este apartado no se muestra el gráfico de temas populares, porque no ha habido actividad sobre estos temas en los últimos días.

Realice búsquedas sobre otros temas.

Ilustración 25: Vista del buscador

Mediante el uso del buscador se puede consultar otros temas que no han sido definidos por defecto en las secciones anteriores.



Realice búsquedas sobre otros temas.

Search

OK

	Topic	Menciones	Sentimiento	Tendencia
	Biopsia	160	neutro	↑

Ilustración 26: Ejemplo de búsqueda 1

Otro ejemplo:



Realice búsquedas sobre otros temas.

Search

OK

	Topic	Menciones	Sentimiento	Tendencia
	neoplasia	21	negativo	↓

Ilustración 27: Ejemplo de búsqueda 2

Haciendo click sobre el tema que se ha buscado o sobre cualquier otro tema de los apartados anteriores se accede al submenú (*Descripción, Comentarios, Posts*)



Descripción Comentarios Posts

Neoplasia

Neoplasia es el término que se utiliza en [medicina](#) para designar una masa anormal de [tejido](#). Se produce porque las [células](#) que lo constituyen se multiplican a un ritmo superior a lo normal. Las neoplasias pueden ser benignas cuando se extienden solo localmente y malignas cuando se comportan de forma agresiva, comprimen los tejidos próximos y se diseminan a distancia.^[1]

Realice búsquedas sobre otros temas.

Search

Topic	Menciones	Sentimiento	Tendencia
 neoplasia	21	negativo	-

Ilustración 28: Vista de la descripción de un tema

La descripción del tema que se ha seleccionado se extrae de Wikipedia mediante un Scraper. En el caso de que no exista definición para este tema, aparece el mensaje “No existe descripción para este término”.

Al acceder a la pestaña de Comentarios o Posts, se puede ver los comentarios que han realizados los usuarios en Facebook.



Descripción Comentarios Posts

Likes Positivos Negativos Objetivos Subjetivos

Anterior Siguiente

neutro 
2017-01-26 20:53:28

el termino «cancer» es generico y designa un **amplio** grupo de enfermedades que pueden **afectar** a cualquier parte del cuerpo ; tambien se habla de «tumores malignos» o «neoplasias» . una característica del **cancer** es la multiplicacion **rapida** de celulas anormales que se extienden mas alla de los **limites** normales y pueden **invadir** partes adyacentes del cuerpo y propagarse a otros organos , proceso **conocido** como metastasis . las metastasis son la **principal** causa de **muerte** por **cancer** .

Visitar

muy negativo 
2016-11-29 21:11:40

la **enfermedad** llamada **cancer** , y preferiria llamarla neoplasia porque la palabra **cancer** los ignorantes la han utilizado como sinonimo de lo **peor** y la **muerte** , hace parte del sinnumero de enfermedades que padecemos los **humanos** , tambien **mala** la desnutricion , la diabetes , el epoc , las neumonias.en conclusion estoy planteando que veamos las neoplasias como una **enfermedad** mas para que las personas que las padecen no entren en **panico** o **temor** .

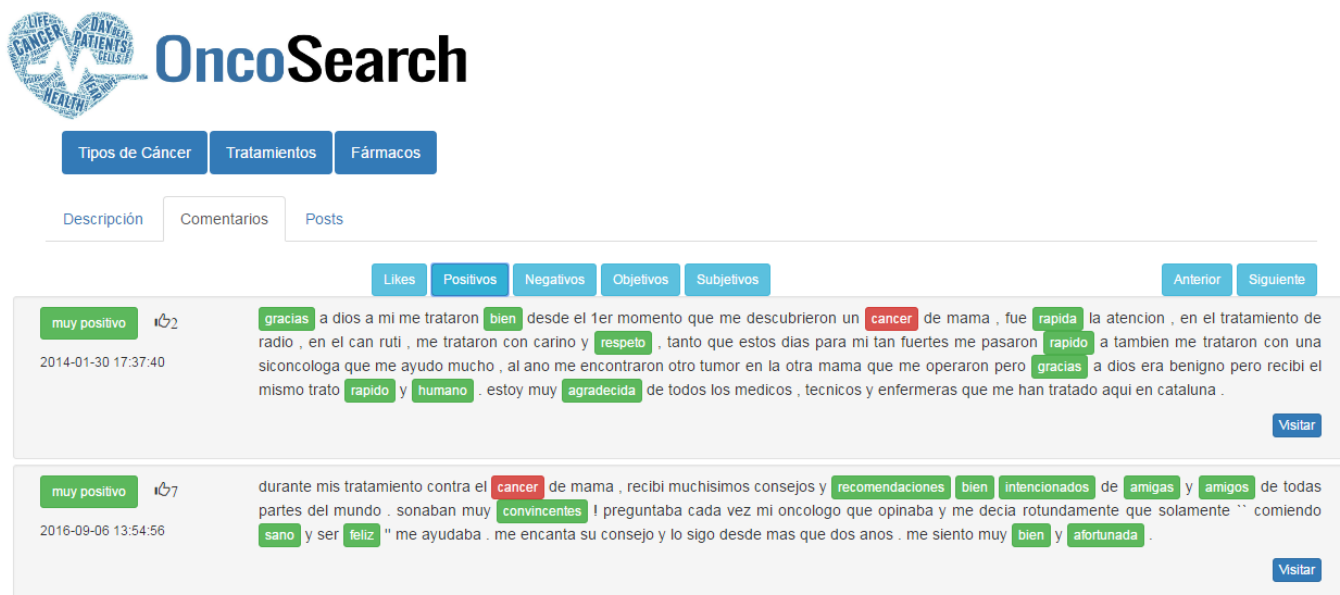
Visitar

Ilustración 29: Paginador de comentarios y posts

En este apartado se dispone de un paginador que muestra los comentarios de 5 en 5, con el uso de los botones *Anterior* y *Siguiente* podemos ver el resto de comentarios o posts realizados.

Además el paginador dispone de filtros de ordenación:

- Likes: Ordena los textos por número de *me gusta*.
- Positivos: Ordena los textos por polaridad mostrando primero aquellas opiniones con un sentimiento positivo.



The OncoSearch interface displays a list of comments sorted by positivity. The top navigation bar includes 'Tipos de Cáncer', 'Tratamientos', and 'Fármacos'. Below this, tabs for 'Descripción', 'Comentarios', and 'Posts' are visible. A secondary filter bar shows 'Likes', 'Positivos', 'Negativos', 'Objetivos', and 'Subjetivos', with 'Positivos' selected. The comments are displayed in a list with a 'Visitar' button next to each. The first comment is dated 2014-01-30 17:37:40 and is labeled 'muy positivo'. The text of the comment is: 'gracias a dios a mi me trataron bien desde el 1er momento que me descubrieron un cancer de mama , fue rapida la atencion , en el tratamiento de radio , en el can ruti , me trataron con carino y respeto , tanto que estos dias para mi tan fuertes me pasaron rapido a tambien me trataron con una siconcologa que me ayudo mucho , al ano me encontraron otro tumor en la otra mama que me operaron pero gracias a dios era benigno pero reci el mismo trato rapido y humano . estoy muy agradecida de todos los medicos , tecnicos y enfermeras que me han tratado aqui en cataluna .'. The second comment is dated 2016-09-06 13:54:56 and is also labeled 'muy positivo'. The text of the comment is: 'durante mis tratamiento contra el cancer de mama , recibi muchisimos consejos y recomendaciones bien intencionados de amigas y amigos de todas partes del mundo , sonaban muy convincentes . I preguntaba cada vez mi oncologo que opinaba y me decia rotundamente que solamente `` comiendo sano y ser feliz `` me ayudaba . me encanta su consejo y lo sigo desde mas que dos anos . me siento muy bien y afortunada .'. Both comments have a 'Visitar' button next to them.

Ilustración 30: Ordenación de comentarios por positividad

- Negativos: Ordena los textos por polaridad mostrando primero aquellas opiniones con sentimiento negativo.



The OncoSearch interface displays a list of comments sorted by negativity. The top navigation bar includes 'Tipos de Cáncer', 'Tratamientos', and 'Fármacos'. Below this, tabs for 'Descripción', 'Comentarios', and 'Posts' are visible. A secondary filter bar shows 'Likes', 'Positivos', 'Negativos', 'Objetivos', and 'Subjetivos', with 'Negativos' selected. The comments are displayed in a list with a 'Visitar' button next to each. The first comment is dated 2015-02-25 17:07:15 and is labeled 'muy negativo'. The text of the comment is: 'ahora va a hacer 3 anos de la muerte de mi hija de 4 anos por smd . esta enfermedad nos pillo por sorpresa ya que mi hija era una nina con mucha vitalidad y de repente empeece a notarla cansada y con muchos hematomas , empezo a tener otitis que jamas habia tenido , sus defensas ya estaban fatal . asi empezo un calvario para ella que duro 10 meses puesto que el trasplante de medula fue con cordon y no funciona . es una enfermedad muy complicada para ninis y apenas se da entre ellos pero toca y claro que toca . ojala con la investigacion podamos acabar con todos los tipos de cancer de sangre . yo a dia de hoy soy socia y espero que entre todos podamos hacer mucho mas .'. The second comment is dated 2016-10-08 17:10:42 and is also labeled 'muy negativo'. The text of the comment is: 'pues mi padre muno sufriendo . los 2 meses q duro su enfermedad estuvo con cuidados paliativos , que agradecemos de corazon , pero llegado el momento de su muerte , y firmada la orden por su medico para sedarlo , en el caso de q ya fuera definitivo , nadie y digo nadie , vino a ponerle nada a mi padre hasta q murio afixiado y con lagrimas en los ojos . ojala llegue el dia en que podamos decidir cuando terminar con nuestro sufrimiento y ninguna persona tenga que pasar lo que paso mi padre y nosotros por estar presentes'. Both comments have a 'Visitar' button next to them.

Ilustración 31: Ordenación de comentarios por negatividad

- Objetivos: Ordena los textos según la objetividad de cada texto.



Tipos de Cáncer
Tratamientos
Fármacos

Descripción
Comentarios
Posts

Likes
Positivos
Negativos
Objetivos
Subjetivos
Anterior
Siguiente

2015-02-28 16:45:27

yo llevaba desde los 26 años con un mioma , que decían que se eliminaba solo y cada tres años hemorragias fuertes , anemias como caballo , y seguían diciendo que ese mioma sangrante no afectaba para nada y va hacer el 5/3/2014 que me han tenido que vaciar por que era un **cancer** de utero que se pegaba en las paredes y se podía extender algún órgano y un cirujano de alcázar de Henares que tenía la última palabra para poderlo operar en su día decía que no era para operar . hizo un raspado donde extendió más el **problema** medí arreglarlo en su día .

Visitar

2013-10-20 00:33:45

decirle a Beatriz Cuevas Arnaiz que el día del **cancer** de utero es el 26 de septiembre , el de próstata el 11 de junio , el de pulmón el 17 de noviembre , el de leucemia el 22 de septiembre y el de melanoma el 23 de mayo

Visitar

Ilustración 32: Ordenación de comentarios por objetividad

- Subjetivos: Ordena los textos según el índice de subjetividad de cada texto.



Tipos de Cáncer
Tratamientos
Fármacos

Descripción
Comentarios
Posts

Likes
Positivos
Negativos
Objetivos
Subjetivos
Anterior
Siguiente

muy positivo

2015-11-27 08:44:35

hacen más de 20 años tuve **cancer** de utero , fue diagnosticado a tiempo y con una operación me extirparon el útero **gracias** a dios , tengo 55 años , en ese momento estaba sola y lejos de mi familia y ya este próximo 6 de enero van hacer 5 años que perdí a mi hermano **mayor** de **cancer** en el riñón izquierdo fue detectado demasiado tarde y **gracias** a dios no sufrí mucho , lo más **triste** es que de igual forma lejos de la familia , pero nos ayudó mucha gente entre ellos familiares y **amigos** para poder repatriar el cadáver de mi hermano , para mí lo más fuerte era pensar en mi madre que lo vio ir y lo vio llegar en su ferretó sabes ? , yo a esto me voy si te han o te van a operar y luego te van hacer tratamiento de quimioterapia ten fe y mucha esperanza en dios y la virgen **santa** , porque es una oportunidad más de vida , lo que muchas otras personas no lo han tenido , recuerda que mientras hay vida hay esperanza y la vida es hermosa con todos sus problemas vale ? , así que arriba corazón de **oro** , mi reina querida , arriba esos **ánimos** y fuerza mi **valiente** guerrera y luchadora que fíjate te bendiga y la virgen sí te proteja con su manto , salud , fuerza y vida para ti madre querida !!!!!!!
 , suerte y muchas **bendiciones** !!!!!

Visitar

Ilustración 33: Ordenación de comentarios por subjetividad

Cada fila está dividida en dos columnas, a la izquierda tenemos la información acerca de la polaridad, número de *me gusta* y fecha de creación. A la derecha de esta columna tenemos el texto, resaltando las palabras indicadoras de opinión (*verde para las positivas, rojo para las negativas*). Por último podemos acceder al post original realizado en Facebook haciendo click sobre el botón *Visitar*.

10. ANEXO C: ÍNDICE DE ILUSTRACIONES

Ilustración 1: Tareas fundamentales	14
Ilustración 2: Diagrama de Gannt.....	16
Ilustración 3: Redes sociales y frecuencia de uso	24
Ilustración 4: Diagrama BurnDown	30
Ilustración 5: Ejemplo consulta API Graph	34
Ilustración 6: Búsqueda de páginas sobre la API Graph	35
Ilustración 7: Modelo entidad relación	38
Ilustración 8: Diagrama Burndown primera iteración	41
Ilustración 9: Diagrama Burndown de la segunda iteración	46
Ilustración 10: Modelo-vista-controlador.....	49
Ilustración 11: Diagrama Burndown tercera iteración	53
Ilustración 12: Visualización de las secciones.....	55
Ilustración 13: Visualización de la página web	57
Ilustración 14: Vista de la página web con el diagrama de temas populares ...	58
Ilustración 15: Vista de los comentarios con las palabras resaltadas según su polaridad.....	59
Ilustración 16: Vista de la descripción del tema.....	60
Ilustración 17: Ejemplo de búsqueda	61
Ilustración 18: Diagrama Burndown cuarta iteración	61
Ilustración 19: Project Interpreter PyCarm	66
Ilustración 20: Paquetes disponibles para instalar	66
Ilustración 21: Ejecución de la aplicación desde la consola	67
Ilustración 22: Vista principal del prototipo	68
Ilustración 23: Apartado de Tratamientos.....	69
Ilustración 24: Apartado de Fármacos.....	70
Ilustración 25: Vista del buscador.....	70
Ilustración 26: Ejemplo de búsqueda 1	71
Ilustración 27: Ejemplo de búsqueda 2	71
Ilustración 28: Vista de la descripción de un tema	72
Ilustración 29: Paginador de comentarios y posts	72
Ilustración 30: Ordenación de comentarios por positividad	73
Ilustración 31: Ordenación de comentarios por negatividad.....	73
Ilustración 32: Ordenación de comentarios por objetividad	74
Ilustración 33: Ordenación de comentarios por subjetividad	74

11. ANEXO C: ÍNDICE DE TABLAS

Tabla 1: Plan de tareas	15
Tabla 2: Salario mínimo	18
Tabla 3: Tiempos estimados y perfil del desarrollador	19
Tabla 4: Personal y salario	19
Tabla 5: Coste del desarrollo del proyecto	19
Tabla 6: Estimación de la incidencia del cáncer	21
Tabla 7: Mortalidad del cáncer en España	22
Tabla 8: Localizaciones más frecuentes del cáncer	22
Tabla 9: Historia Obtención de información	25
Tabla 10: Historia Respaldo de la información	25
Tabla 11: Historia Clasificación por objetividad	25
Tabla 12: Historia Clasificación por polaridad	26
Tabla 13: Historia Diseño Web	26
Tabla 14: Historia Clasificación por me gusta	26
Tabla 15: Historia Ver secciones	26
Tabla 16: Historia Información de las secciones	27
Tabla 17: Historia Ver temas populares	27
Tabla 18: Historia Buscador	27
Tabla 19: Historia Actualización información	27
Tabla 20: Historia Mapa de localizaciones	28
Tabla 21: Historia Resaltar palabras clave	28
Tabla 22: Historia Descripción de términos	28
Tabla 23: Estimación de puntos de historia	29
Tabla 24: Priorización de las historias de usuario	31
Tabla 25: Historias de usuario para la primera iteración	31
Tabla 26: Subtareas de la historia Obtención de información	31
Tabla 27: Subtareas de la historia Respaldo de la información	32
Tabla 28: Historia Obtención de información	32
Tabla 29: Campos relevantes de las páginas de Facebook	33
Tabla 30: Grupos sobre el cáncer de Facebook	36
Tabla 31: Grupos sobre el cáncer colorrectal de Facebook	36
Tabla 32: Grupos sobre el cáncer de próstata en Facebook	36
Tabla 33: Grupos sobre el cáncer de pulmón en Facebook	37
Tabla 34: Grupos sobre el cáncer de mama en Facebook	37
Tabla 35: Grupo sobre el cáncer de vejiga en Facebook	37
Tabla 36: Historia Respaldo de la información	37
Tabla 37: Campos tabla Page	38
Tabla 38: Campos tabla Post	39
Tabla 39: Campos tabla Comment	39
Tabla 40: Campos tabla Cancer_type	39
Tabla 41: Campos tabla Popularity	40
Tabla 42: Campos tabla Optimism	40
Tabla 43: Campos de la tabla Key_words	40
Tabla 44: Valores actuales de la base de datos	41
Tabla 45: Priorización historias de usuario segunda iteración	42
Tabla 46: Historia Actualización información	43
Tabla 47: Historia Mapa de localizaciones	44
Tabla 48: Subtareas de la historia Mapa de localizaciones	44
Tabla 49: Campos de la tabla Location	44

Tabla 50: Priorización de las historias de usuario tercera iteración.....	47
Tabla 51: Historias a desarrollar en la tercera iteración	48
Tabla 52: Historia Diseño Web.....	48
Tabla 53: Historia: Clasificación por objetividad	49
Tabla 54: Historia Clasificación por polaridad	51
Tabla 55: Representación de la polaridad	52
Tabla 56: Priorización de las historias de usuario de la cuarta iteración	54
Tabla 57: Historia: Clasificación por me gusta	54
Tabla 58: Historia Ver secciones.....	55
Tabla 59: Historia Información de las secciones	55
Tabla 60: Subtarear de la historia Información de las secciones.....	56
Tabla 61: Historia Ver temas populares	57
Tabla 62: Historia Resaltar palabras clave	58
Tabla 63: Descripción de términos	59
Tabla 64: Historia Buscador	60