



Apellidos, nome /Apellidos, nombre:

Otero Freijeiro, David

Título:

Plataforma para la etiquetación asistida de casos de riesgo temprano en internet

Mención cursada:

Ingeniería del Software

Resumo / Resumen:

Según la OMS, en una de sus últimas publicaciones sobre la salud mental de los adolescentes, una de cada 6 personas que padecen trastornos mentales, tiene entre 10 y 19 años. Esto nos lleva a que un 16% de personas que son adolescentes entre todos los que padecen algún trastorno. En muchos casos, estas enfermedades desembocan en suicidios. Por tanto, es fundamental la detección temprana de estas. Es sabido que el lenguaje es un indicador significativo del estado emocional de las personas y, por lo tanto, ayuda en la detección de estas enfermedades. Una forma de detectar estas dolencias es el desarrollo de herramientas que analicen el lenguaje que estas personas utilizan en redes sociales. Debido a la importancia de estas herramientas, es necesario evaluar la eficacia de las mismas. Para los sistemas de búsqueda en otros escenarios (web, noticias, patentes, artículos científicos, etc), esta evaluación se hace, típicamente, con colecciones de prueba. Estas colecciones están formadas por un conjunto de documentos, un conjunto de tópicos que representan las necesidades de información de los usuarios, y una serie de juicios de relevancia que indican qué documentos son relevantes para cada tópico.

Esta colecciones normalmente se construyen en foros de evaluación como TREC. TREC es una iniciativa del NIST, el cual depende del Departamento de Comercio de los Estados Unidos. Esta iniciativa dispone de recursos para contratar asesores, que usualmente son agentes retirados de la CIA, para juzgar la relevancia de los documentos que forman la colección. El coste de contratar a estos asesores es elevado, además de que es necesario organizar competiciones en las que los grupos participantes proporcionan sus resultados para las tareas de búsqueda, que se utilizan para elegir los documentos que los asesores deben juzgar. Todo esto hace que construir colecciones nuevas, y aún más para tareas y escenarios distintos, sea un proceso muy costoso. Esto hace imposible que personas e investigadores independientes puedan construir sus propias colecciones de prueba cuando abordan nuevas tareas, ya que organizar una competición requiere muchos recursos. Existen muchas tareas distintas dentro de la recuperación de información, y muchas de ellas dependen de la construcción de nuevos conjuntos de datos de prueba. Un ejemplo de estas tareas es la elaboración de métodos para la detección temprana de riesgos en Internet, especialmente aquellos relacionados con la salud, como la esquizofrenia o el daño autoinfligido. Hoy en día no existe una manera fácil de poder crear una nueva colección para ser empleada en un ámbito específico.

El principal objetivo de este proyecto fue diseñar y construir una plataforma que permita la etiquetación de documentos eficiente por parte de los asesores asociados a casos de trastornos psicológicos y mentales, como psicólogos. Esta plataforma se usó para construir la colección de prueba en la competición de CLEF (<http://www.clef-initiative.eu>) eRisk (<https://early.irlab.org>) de 2020, que se celebra con el objetivo de evaluar la efectividad de metodologías y métricas para la detección temprana de casos de riesgo en Internet, como la anorexia o la depresión. El sistema está pensado para ser utilizado por dos tipos de usuarios: administrador y etiquetador. El primero es el encargado de emplear la aplicación para descargar y organizar los textos. En este caso se utiliza la red social Reddit para obtener los textos que forman la colección. El segundo es el encargado de etiquetar los documentos de la colección.

La aplicación se desarrolló empleando la metodología Scrum con algunas adaptaciones, como por ejemplo, que el equipo de desarrollo estaba compuesto por una sola persona, lo que lleva a relajar algunas exigencias como los daily meetings. Para el desarrollo de la aplicación se emplearon tecnologías como Spring (<https://spring.io>) para la capa



de lógica de negocio y de acceso a datos, Lucene (<https://lucene.apache.org>) para la indexación y búsqueda de documents, Thymeleaf (<https://www.thymeleaf.org>) para la capa web y D3.js (<https://d3js.org/>) para la elaboración de gráficas.

Posibles aplicaciones / Posibles aplicaciones:

La principal aplicación de este proyecto es proveer de colecciones grandes y de calidad a tareas donde puedan ser necesarias, para así mejorar la calidad de los resultados obtenidos en dichas tareas. Por ejemplo, se puede emplear la aplicación para elaborar colecciones sobre diversas dolencias relacionadas con la salud: anorexia, daño autoinfligido, depresión, esquizofrenia, entre otros. Así, se pueden construir colecciones que contienen textos de personas que escriben sobre estos temas y su etiquetación según si estas personas sufren dicha enfermedad o no. De esta manera se puede apoyar la investigación en la detección temprana de estos riesgos.

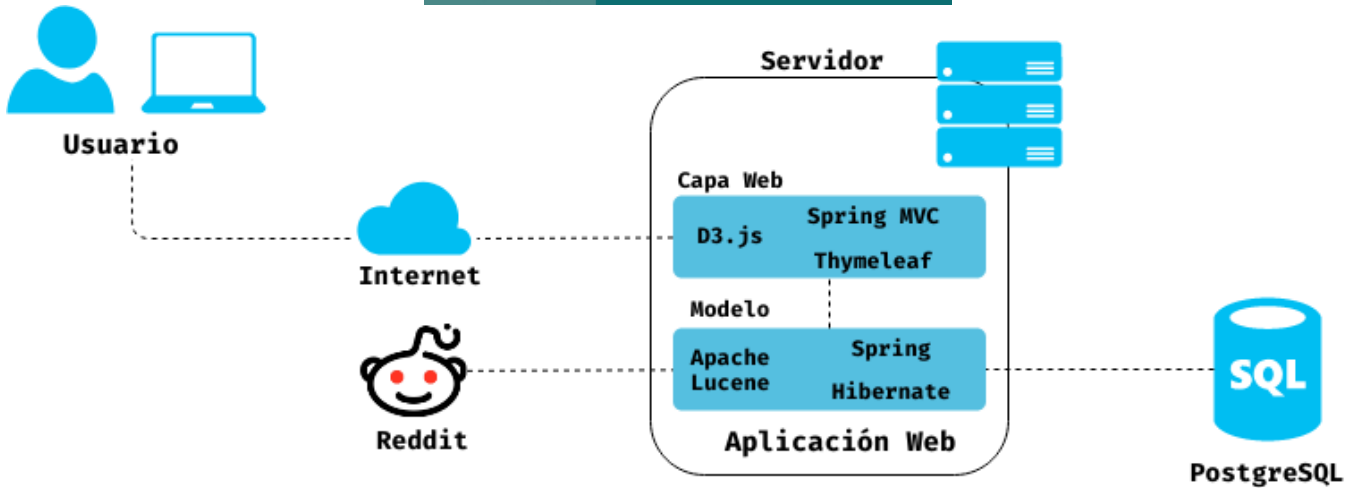
Normalmente estas enfermedades se tratan de manera reactiva, es decir, se aplica el tratamiento una vez la enfermedad ha sido diagnosticada. Sin embargo, es sabido que algunos indicadores, como el lenguaje, están bastante relacionados con distintos trastornos mentales. Utilizar estos indicadores para la detección temprana de dichos riesgos ha demostrado ser una línea de investigación muy fructífera. Por lo tanto, esta aplicación es útil para apoyar las investigaciones en este ámbito.

A mayores de todo esto, la aplicación puede ser utilizada en otros ámbitos muy diversos. Actualmente la aplicación solamente descarga los textos de la red social Reddit, pero no existe limitación en el contenido de los escritos. Se pueden obtener colecciones sobre otros temas que no estén relacionados con los riesgos de salud.

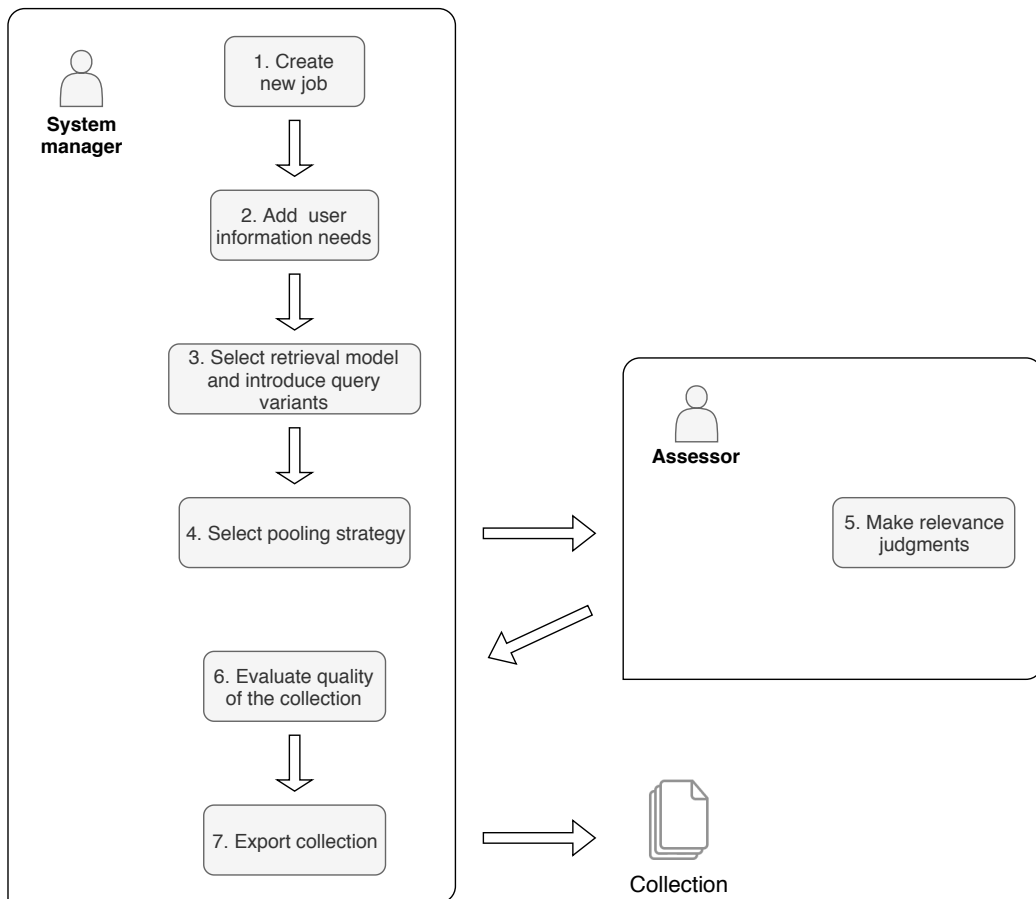
Etapas para o seu desenvolvemento futuro / Etapas para su desarrollo futuro:

- Extender la plataforma a otros dominios o a otras fuentes. Se empleó Reddit para obtener los textos porque es una fuente apropiada ya que contiene foros específicos sobre enfermedades mentales. Sin embargo, la aplicación podría extenderse fácilmente para trabajar con otras fuentes de datos y otros dominios.
- Extender la aplicación para ser empleada en otro tipo de entornos. Ahora mismo la aplicación está pensada para ser utilizada solamente por un equipo, pero podría extenderse para que utilizaran diversos equipos en colaboración para construir nuevas colecciones.
- Explotación de las colecciones obtenidas mediante el desarrollo de funcionalidades de acceso y búsqueda de información, clustering y clasificación de textos, incluyendo el estudio de la evolución temporal de los tópicos producidos por los usuarios.

Imaxes representativas / Imágenes representativas:



Arquitectura de la aplicación



Roles de la aplicación y tareas asignadas a cada uno de los roles. El resultado final es una colección con los juicios realizados.



cut myself

Search

Judging user:
3197748

Post ID	Date	Body	
3197756	09-10-2019	That was me the other night, kept thinking about cutting myself but was tired to care. I cut myself the next day though...	www.reddit.com/r/selfharm/comments/dfiblz/that_funny_moment_when/f3n3re/
3197760	27-09-2019	I hooked up with my friend and he saw the cuts on my legs. I didn't want anyone to know but now he does and I'm still seeing him so I know he'll know if I do it again but I want to cut myself again	www.reddit.com/r/selfharm/comments/d9ynbw/how_did_people_find_out_about_your_self_harm/f1mw1mz/
3197758	08-10-2019	I'm 20 and still waiting to feel happy, I tell myself it will bet better but it hasn't.	www.reddit.com/r/depression/comments/df57ej/how_many_people_here_are_1420_years_old/f31i030/
3197759	02-10-2019	I used to pinch my arms and then that wasn't satisfying and I started cutting my thighs	www.reddit.com/r/selfharm/comments/dc9a72/my_friends_took_my_scissors/f28d50z/
3197757	09-10-2019	Update: I have been cutting my thighs pretty consistently since a couple days after posting this	www.reddit.com/r/selfharm/comments/d3ywbq/pinching/f31yc75/

1

Relevant

Non relevant

Etiquetador analizando los textos publicados por un usuario para decidir si sufre algún riesgo o no

X	Autorizo a consulta por parte dos membros da comisión evaluadora da memoria do meu proxecto / Autorizo la consulta por parte de los miembros del tribunal de la memoria de mi proyecto.
---	--

Instruccions para o depósito da memoria / Instrucciones para el depósito de la memoria:

Débase depositar no OneDrive da UDC, dentro da carpeta co seu nome de usuario incluída en:

4 edición - Premio TFG Aplicado

https://udcg-my.sharepoint.com/:f:/r/personal/nieves_pedreira_udc_es/Documents/4%20edici%C3%B3n%20-%20Premio%20TFG%20Aplicado?csf=1&e=y6SVha

Se debe depositar en el OneDrive de la UDC, dentro de la carpeta con su nombre de usuario incluída en:

4 edición - Premio TFG Aplicado

https://udcg-my.sharepoint.com/:f:/r/personal/nieves_pedreira_udc_es/Documents/4%20edici%C3%B3n%20-%20Premio%20TFG%20Aplicado?csf=1&e=y6SVha

