

Obtención automática de palabras claves en textos clínicos: aplicación en sospechas diagnósticas chilenas

Problema

La falta de estructura en los textos clínicos dificulta la extracción e integración de la información, especialmente en el caso de datos masivos. A pesar de que se puede realizar estadística descriptiva sobre textos carentes de estructura, las técnicas actuales de recuperación de información no se encuentran fácilmente disponibles para investigadores en medicina por lo que la información no estructurada a menudo se descarta en proyectos que intentan apoyar la decisión clínica (1).

Solución

Aplicar técnicas de procesamiento de lenguaje natural (PLN) a las sospechas diagnósticas de la lista de espera (LE) chilena para la recuperación de las palabras claves que definen cada especialidad contenidas en ella, junto con la generación de una visualización intuitiva (2) de las mismas en una plataforma disponible en línea para los tomadores de decisiones y el público general. Además se trata de un solución escalable para su utilización masiva en repositorios de textos clínicos no estructurados.

Método

Se obtuvo la LE para nueva consulta de especialidad de 17 servicios de salud en Chile y un total de 2.592.925 sospechas diagnósticas fueron identificadas. La obtención de las palabras clave de especialidades se realizó utilizando la técnica de PLN frecuencia de término - frecuencia inversa de documentos (FT-FID) (3).

Variables y Métricas

Se utilizará la frecuencia de cada palabra dentro de cada una de las especialidades, junto con la frecuencia de documentos que contienen cada una de las palabras para calcular el puntaje FT-FID y detectar las palabras que más información aportan a cada una de las especialidades de la LE chilena.

Hipótesis

Con la aplicación de técnicas de PLN para la ponderación de frecuencia de palabras sobre un tópico en específico, se pueden extraer las palabras que resultan de mayor importancia en cada especialidad contenida en la LE chilena. Además, esta importancia de términos se pueden comunicar usando nubes de palabras.

Objetivos

- 1) Extraer las 5 palabras clave de cada especialidad de la LE chilena
 - a) Preprocesar las listas de espera para normalizar su contenido
 - b) Aplicar el método de FT-FID sobre el texto libre de la LE chilena
- 2) Generar nubes de palabras para cada una de las especialidades de la LE Chilena.

Resultados

Palabras claves para cada especialidad fueron obtenidas. Se tabularon 5 palabras clave, y para una representación visual, se crearon nubes que muestran palabras con frecuencia ponderada codificada por tamaño, las cuales están disponibles en <https://cimt.cl/lechile/>. Cirugía maxilofacial, cirugía y traumatología maxilofacial y cirugía bucal, comparten 4 de 5 palabras claves.

Perspectiva

Este trabajo es una prueba de conceptos en referato (4) para el análisis automático de grandes volúmenes de textos con el fin de describir su contenido fácilmente.

Referencias

1. Dalianis H. Clinical Text Mining: Secondary Use of Electronic Patient Records. Clinical Text Mining. Springer; 2018.
2. Felix C, Franconeri S, Bertini E. Taking Word Clouds Apart : An Empirical Investigation of the Design Space for Keyword Summaries. 2018;24(1):657–66.
3. Manning CD, Raghavan P, Schutze H. An Introduction To Informational Retrieval. Inf Retr Boston. 2009;(c):1–18.
4. Villena F, Dunstan J. Obtención automática de palabras claves en textos clínicos: aplicación en sospechas diagnósticas chilenas [trabajo no publicado]. Rev Med Chil.