UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO

“Identificación del diagnóstico de patología crítica en los informes radiológicos mediante procesamiento de lenguaje natural: Aplicación en Chile.”

Guillermo Javier Ortiz Calvo

TESIS PARA OPTAR AL GRADO DE MAGISTER EN INFORMÁTICA MÉDICA.

Director de Tesis: Prof. Dr. Mauricio Cerda

2016

UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO

INFORME DE APROBACION TESIS DE MAGISTER

Se informa a la Comisión de Grados Académicos de la Facultad de Medicina, que la Tesis de Magister presentada por el candidato

Guillermo Javier Ortiz Calvo

ha sido aprobada por la Comisión Informante de Tesis como requisito para optar al Grado de Magister en Informática Médica en el Examen de Defensa de Tesis rendido el día 20 de julio de 2016.

Prof. Dr. Mauricio Cerda
Director de Tesis
Santiago de Chile

COMISION INFORMANTE DE TESIS

Prof. Dr. ..... **Prof. Dr. .....**

Prof. Dr. .....
Presidente Comisión

Dedicado a la gente que me apoyó y que soportó la convivencia conmigo durante el extenso desarrollo de esta tesis. En especial a mi Nati por las largas horas de trasnoche.

Gracias a mis padres por el financiamiento. A mi papá por su ayuda en la redacción.

A mi equipo de trabajo por darme las facilidades para poder dedicarle tiempo a esta tesis.

Al departamento de Imagenología de Clínica Alemana por facilitar los datos utilizados para el desarrollo de esta herramienta.

ÍNDICE

1.- RESUMEN

Actualmente los informes radiológicos se redactan en texto libre sin un campo específico que los categorice según diagnóstico. Por este motivo, la identificación de los diagnósticos clasificados como patología crítica debe hacerse de forma manual, acarreando consigo problemas como el submuestreo y gran tiempo invertido. Este trabajo propone como solución desarrollar una herramienta utilizando métodos de procesamiento de lenguaje natural para analizar los texto de forma masiva.

En esta tesis se plantea como hipótesis que es posible identificar más del 80% de los diagnósticos existentes en SNOMED-CT (una terminología médica) presentes en las impresiones de los informes radiológicos, identificando la patología crítica con más de un 90% de sensibilidad mediante algoritmos de procesamiento de lenguaje natural (NLP).

Para clasificar los informes se utilizó SNOMED-CT por su amplio manejo de conceptos médicos y sinónimos. La tarea se realizó con 3 algoritmos: 1) un motor de búsqueda para encontrar los términos de SNOMED-CT contenidos en los informes utilizando indexación reversa, 2) un detector de negación basado en expresiones regulares y 3) se combinó ambas herramientas para identificar patología crítica. Los algoritmos propuestos fueron evaluados en muestra representativa (n=219) de 1973 informes de Angiografía Pulmonar por Tomografía Computada, etiquetada por 2 médicos.

Como resultados se obtuvo un valor *kappa* de acuerdo entre etiquetadores de 85.5%, $IC^{95\%}$[80.8-90.3%], $p < 0.001$. Por otra parte el motor de búsqueda presentó un rendimiento con medida F (F) de 0.94, sensibilidad (S) de 91.2% y valor predictivo positivo (VPP) de 98%. El detector de negación obtuvo una F de 0.99, S de 98.7% y VPP de 99.3%. Para medir el rendimiento en la detección de patología crítica se utilizó como referencia el diagnóstico de tromboembolismo pulmonar (TEP), obteniendo valores F de 0.94, S de 96.3% y VPP de 92.86%

Como conclusión, el presente trabajo de tesis muestra que es posible construir una herramienta para identificar la patología crítica basada en NLP utilizando la regularidad de los patrones de expresión en el texto, lo que permitirá en futuros trabajos crear herramientas de soporte para la toma de decisiones.

2.- ABSTRACT

Currently radiology reports are written in free text without a specific field to categorize according to diagnosis. Therefore, identification of diagnostics listed as critical result, group characterized by having a high risk of harm to the patient, must be done manually. As a solution is proposed the use of natural language processing tools to analyze big volume of texts.

This thesis pose the hypothesis that it is possible to identify more than 80% of existing diagnostics from impressions of radiology reports on SNOMED-CT, a clinical terminology, identifying critical results with more than 90% sensitivity, using natural language processing (NLP) algorithms.

To identify reports, SNOMED was used because of its wide management of medical terms and synonyms. Identification was built as a 3 steps algorithm: 1) A search engine was built to find terms of SNOMED contained in reports using reverse indexing, 2) a negation detector based on regular expressions, and 3) both tools were combined to identify critical results. The proposed algorithms were tested against a representative sample (n = 219) of 1973 Computed Tomography Pulmonary Angiography (CTPA) reports, which were tagged by 2 medical doctors.

The obtained results were an inter-rater reliability *kappa* value of 85.5% for taggers, was obtained $IC^{95\%}$ [80.8-90.3%]. Moreover, search engine had a performance of measure F (F) of 0.94, sensitivity (S) of 91.2% and positive predictive value (PPV) of 98%. The negation detector had a F of 0.99, S of 98.7% and VPP of 99.3%. The measurement of performance for critical results detection was made using pulmonary embolism as reference, obtaining values; F of 0.94, S of 96.3% and VPP of 92.86%

In conclusion, this thesis shows that it is possible to build a tool to identify critical results using NLP by making use of the specific regularity of text expressions in the case of radiology reports, allowing in future researchs to create decision support tools.

3.- INTRODUCCIÓN

Actualmente los informes radiológicos se redactan en texto libre sin un campo específico que los categorice según diagnóstico. Esto conlleva a que para identificar la información relevante se debe leer e interpretar el texto completo, consumiendo gran cantidad de tiempo, produciendo fatiga, e induciendo a errores humanos, entre otras desventajas. [1]

Una solución a este problema se conoce como procesamiento de lenguaje natural (NLP, *Natural Language Processing*). El NLP consiste en extraer información desde un texto no estructurado, permitiendo guardarlo en lenguaje computacionalmente comprensible y analizable.

Las aplicaciones médicas plantean nuevos desafíos al NLP debido a la ambigüedad intrínseca de los textos[2] y a la abundante negación[3, 4], producto de la necesidad de comunicar posibilidades diagnósticas descartadas, como es posible observar en la sección 6.2.2. Adicionalmente para la extracción de información se utilizan ontologías que corresponden a vocabularios especializados y estructurados. Por otra parte, cada vez existen más vocabularios diseñados para diversos objetivos, dificultando la tarea del investigador a la hora de elegir cuál utilizar.

Han existido múltiples aplicaciones de NLP en medicina [5-42], en su mayoría en idioma inglés. La falta de desarrollo de esta área en el idioma español, plantea un desafío y a la vez una motivación a desarrollarla.

La cantidad de aplicaciones de NLP en el área médica es enorme. Dado a su aplicación y al beneficio que obtendrían los pacientes, se eligió trabajar en identificar los reportes radiológicos con diagnóstico de patología crítica. Son definidos como un grupo de diagnósticos caracterizados por poseer un gran riesgo de daño para el paciente y que a su vez, puede ser evitado con un oportuno aviso al clínico para el tratamiento del paciente. [43]

Se plantea la presente tesis como un paso importante para desarrollar futuras aplicaciones de gestión. Esto permitiría realizar la detección en tiempo real de patologías críticas, así como un seguimiento cronometrado del tiempo que demora en dar aviso al clínico responsable del paciente. Se podría crear un catálogo de

patologías críticas de forma centralizada por la institución y que pueda ser modificable en el futuro, dejando de depender de la subjetividad y la memoria del radiólogo que informa

3.1.- Estado del Arte

El procesamiento de lenguaje natural posee principalmente 3 fases. 1) Reconocer la estructura del texto mediante el uso de reglas gramaticales e identificar patrones semánticos. 2) Regularizar y estandarizar los términos a través de mapeos de conceptos que posean múltiples palabras. 3) Codificar y mapear dichos términos a un vocabulario controlado [44], conocido como ontología. Estos pasos permiten a un computador extraer la información de forma comprensible para una máquina, estructurada y codificada, lo que es en sí la definición del concepto “extracción de información”. [45, 46]

3.1.1.- Procesamiento de lenguaje natural en medicina

En medicina hay numerosos avances en el tema, pero en su mayoría en idioma inglés. A continuación se presenta una breve descripción de estos métodos.

En docencia se ha aplicado NLP. Denny J.C. et al. (2015) realiza la monitorización de competencias médicas adquiridas por médicos en formación, desde las anotaciones que realizan en la ficha clínica de los pacientes [10]. Esto pretende realizar una evaluación de forma puntual en el tiempo como sería en una prueba tradicional, además pretende evaluar todos los registros del médico en formación y dar retroalimentación con respecto a la presencia o ausencia de observaciones o evaluaciones clínicas que debieron ser aplicadas, habiéndose logrado el objetivo en el área evaluada. Se dio retroalimentación a los alumnos, enviando material de apoyo en relación a 2 competencias geriátricas; directrices anticipadas y estado mental. Se logró una sensibilidad del 100% y de un 93% respectivamente, considerándose la herramienta exitosa.

Por otra parte, con datos provenientes de los registros clínicos, se han realizado estudios en temas como; la vigilancia en tiempo real de infecciones asociadas a catéter urinario[5], la extracción de información de alergias en las historias

clínicas[6], la extracción de antecedentes familiares en notas médicas[7]. Los resultados obtenidos para la vigilancia de infección de catéteres fue pobre, 65% de sensibilidad y 45.2% de valor predictivo positivo. Se hace referencia a que la herramienta sí sería útil para la extracción de características del registro, pero que requiere mayor trabajo para identificar infecciones asociadas a catéter urinario. La identificación de alergias, en cambio, fue exitosa, con una sensibilidad global de 91% y un valor predictivo positivo de 84.4%. La extracción de antecedentes familiares tuvo como resultados rendimientos mayores al 80% en todos los puntos evaluados pero con falencias en la detección de negación que sólo llegó a una sensibilidad de 68% y un valor predictivo positivo de 49.5%.

Con el fin de optimizar el tiempo de los clínicos invertido en revisión de fichas, se ha realizado la detección de información nueva en las notas de evolución médica.[8] En el ámbito de la gestión se ha trabajado en la identificación de inmigrantes en registros clínicos.[9] Éstas son las principales evidencias del intento que se realiza por obtener información estructurada a partir de los textos libres con el fin de generar mejor gestión, vigilancia y estadística.

Estos ejemplos muestran el número creciente de intentos de utilización de la información que se encuentra en los campos de texto libre de los registros clínicos, demostrando por un lado que sería posible crear herramientas con este fin y por otro dejando en evidencia las dificultades que se podrían presentar durante su elaboración.

Otro documento objeto de explotación mediante NLP son los informes médicos. En dicho campo el área radiológica presenta un importante número de trabajos. A continuación se agrupa por objetivos y utilidad de las herramientas debido a su interés para el presente trabajo de tesis. No fue posible agrupar según método o herramienta, ya que gran parte de los trabajos citados a continuación utilizan softwares comerciales, o describen vagamente el método haciendo énfasis en la aplicación de la herramienta.

3.1.1.1.- Uso en estructuración de la información

Existen estudios sobre la priorización de información clínica relevante, siendo ésta una de las formas a través de las cuales, se facilita la comprensión del informe por el clínico no radiólogo [14], así como en otro estudio se utiliza el reconocimiento de localización anatómica de patologías críticas para facilitar la comunicación hacia el clínico [31].

Con el fin de almacenar estructuradamente la información se realizó la extracción del estado de conectividad neuronal en estudios de neurorradiología.[15] Dado que la negación abunda, debido a la necesidad de transmitir información descartada, la identificación de ésta es un paso crítico en el NLP. Un estudio trabajó en mejorar la detección de negación en reportes radiológicos para identificar lo descartado con excelentes resultados (sensibilidad de 92.6%, valor predictivo positivo de 98.6%).[32]

Si quisiéramos buscar informes con un concepto en particular, lo ideal sería que los informes estén guardados con el concepto ya identificado, se obtendría un una retribución rápida de los resultados. Por dicho motivo Zhang et al. (2012) trabajaron en la anotación automatizada de informes que permite una mejor gestión de las imágenes una vez informadas.[38] Las mediciones en radiología son un dato frecuente en los informes y a la vez clave, dado que el tamaño de lo observado muchas veces define la normalidad o la patología. Viendo este potencial se realizó la extracción de mediciones en informes de tomografías computadas (TC) de abdomen[25], permitiendo guardar este tipo de datos de forma estructurada.

Los estudios de investigación en oncología se realizan con un gran volumen de casos. Habitualmente se utilizan documentos estructurados a ser llenados por los investigadores, con el fin de homogeneizar la información contenida. A través de NLP se realizó la extracción y normalización de hallazgos en reportes imagenológicos de cáncer[34], para facilitar a las investigadores el registro de los datos de interés. Los japoneses Imai et al. (2003) realizaron en su idioma, la extracción de diagnósticos de reportes radiológicos[35] y la indexación automatizada de hallazgos en estos reportes[36] para una posterior búsqueda. Lamentablemente, la herramienta es inutilizable en

español. Otro grupo, ahora en inglés, realizó la extracción de información en informes de niños con neumonía[33] con el mismo fin.

Lo avances actuales en estructuración de la información demuestran que es posible extraer información de los textos no estructurados para realizar análisis a mayor nivel, agregando información, permitiendo búsquedas eficaces en librerías extensas de documentos.

3.1.1.2.- Uso en facilitar la redacción

Se avanzó en la construcción de un motor de anotación sensible al contexto[16] de forma tal que el radiólogo sea capaz de ir etiquetando conceptos mientras informa. Otro grupo de investigadores avanzó en la creación de un sistema de soporte en la decisión diagnóstica[17], para poner a disposición del radiólogo los datos estadísticos históricos con respecto a lo que describe, sugiriendo posibles diagnósticos al analizar lo que el radiólogo ya ha descrito.

En cuanto a los informes de mama, se desarrolló una herramienta que permite la detección automatizada de BIRADS[18, 19], sugiriendo la clasificación final y facilitando el análisis al médico. No sólo con BIRADS se ha trabajado, otro grupo construyó una herramienta similar para la identificación del estado tumoral en Resonancia Magnética.[41] Con lo descrito en la primera sección del informe, facilita la conclusión por parte del radiólogo en la impresión diagnóstica.

La búsqueda de imágenes antiguas durante redacción de un informe se utiliza para tener un punto de comparación con el estudio actual. Muchas veces las imágenes encontradas son de la misma sección del cuerpo pero obtenidas con otro fin, por lo que se tomaron con otra técnica o con foco en otro órgano haciendo que la imagen no sea adecuada para comparar. Un grupo de investigadores trabajó en la identificación de parte del cuerpo en las descripción del estudio en DICOM [26] de forma de identificar el órgano objeto del estudio y no sólo la sección corporal, lo que ayuda al radiólogo a encontrar imágenes anteriores para comparar de forma correcta con el estudio actual.

Con el fin de estructurar informes se trabajó en la extracción de la descripción de fracturas[27], esto parece ser bastante especializado, pero finalmente se obtiene también el beneficio de poder buscar informes por las características de las fracturas.

En esta sección se han expuesto herramientas que pretenden ayudar al radiólogo a realizar la redacción de su informe. Se destaca la idea de incluir sistemas de soporte a la toma de decisiones durante la redacción de los informes. Esto propone futuro usos herramientas que analicen la redacción en tiempo real.

3.1.1.3.- Uso en la detección de errores

El grupo de investigadores que trabajó en BIRADS, escala de riesgo de malignidad en lesiones mamarias, siguió ahondando en el tema. Desarrolló una herramienta capaz de detectar la ambigüedad entre las características informadas y la clasificación BIRADS[20]. Esto permite identificar errores en la clasificación realizada por el radiólogo. A su vez, la extracción de recomendaciones de complementación de estudio[21-24] permiten evitar la negligencia por omisión de la información entregada al paciente.

Se ha desarrollado también una herramienta capaz de identificar los nódulos descritos en la sección de hallazgos en informes de TC y evaluar si éstos se encuentran o no descritos en la impresión diagnóstica[40], permitiendo realizar un monitoreo de errores de omisión en los informes.

La revisión retrospectiva de errores es crucial en cuanto ofrecer un servicio de calidad en radiología. La monitorización de errores es una tarea permanente en el que ya se cuenta con herramientas computacionales para detectar dichos errores.

3.1.1.4.- Usos en el área financiera y administrativa

En el área administrativo-financiera de la radiología se describe la creación de un sistema de codificación en CIE 9-MD[37] que facilita la cobranza en Estados Unidos.

Se trabajó en el reconocimiento de la entidad y su seguimiento en el tiempo [29]. Esto permite gestionar de forma masiva los informes, evitando que se pase por alto alguna patología que ameritara seguimiento. También se ha trabajado en una herramienta dirigida a detectar de forma automática reportes con hallazgos adrenales[28], permitiendo gestionar estos de forma masiva.

En relación a la recuperación de la información, se ha desarrollado sistemas que buscan inteligentemente imágenes[39], explotando la información contenida en los informes para devolver resultados. También se ha trabajado en establecer correlación automática de partes del cuerpo y hallazgos[30] para estructurar la información con mayor detalle ante una eventual búsqueda.

Por último, existe una herramienta que utiliza el NLP para buscar imágenes radiológicas de publicaciones, a partir de elementos contenidos en el texto de éstas[42], siendo una excelente herramienta de docencia y autoaprendizaje desde fuentes confiables.

Los desarrollos aquí planteados motivan a la construcción de una herramienta versátil, en el que podamos conocer el contenido de los informes de forma estructurada con el mayor detalle posible, para luego poder sacar provecho de dicha información.

3.1.2.- Ontologías

Para estructurar la información se necesita tener los elementos o unidades previamente establecidas y codificadas. En el caso de las nomenclaturas estos corresponde a los términos. Cuando a este vocabulario de referencia o terminología, le agregamos propiedades y/o relaciones entre sus términos podemos hablar de ontología.

Existen múltiples ontologías en el ámbito de las ciencias biológicas y un número creciente en el área médica.

Entre las nomenclaturas más conocidas se encuentran:

3.1.2.1.- CIE-10 – Clasificación Internacional de Enfermedades

La Clasificación Internacional de Enfermedades se creó con el objetivo de clasificar las causas de mortalidad de pacientes hospitalizados para estudios epidemiológicos. Los primeros intentos de clasificar enfermedades datan de 1592 en registros del documento “London’s Bills of Mortality” escrito por John Graunt. En 1700 se publica “Nosología Methodica” de François Bossier de Lacroix como uno de las primeras clasificaciones de mortalidad. Hoy en día se utiliza la Clasificación Internacional de Enfermedades en su versión 10. Lamentablemente se ha hecho mal

uso de ésta, utilizándose con otros fines como el pago de prestaciones médicas como ocurre en Estados Unidos, país en el que para que un médico pueda cobrar su atención, debe clasificar previamente el diagnóstico según la versión 9-CM. En Chile es utilizada para codificación de enfermedades en licencias médicas, en altas de hospitalización, en el diagnóstico del registro clínico ambulatorio, entre otros usos. [47]

3.1.2.2.- UMLS – Unified Medical Language System

Fundado y publicado por la *US National Library of Medicine*, posee más de un millón de conceptos, estructura mono-jerárquica y poco manejo de sinónimos. Sólo 65% de sus conceptos no tienen restricción de propiedad intelectual[48], no pudiendo utilizarse la nomenclatura completa de forma libre sin requerir una licencia comercial.

3.1.2.3.- CIAP2 – Clasificación Internacional de Atención Primaria

Nomenclatura que recoge principalmente motivos de consulta de atención primaria. Es mantenida por la Organización Mundial de Médicos de Familia (WONCA). [49]

3.1.2.4.- RADLEX – Radiology Lexicon

Desarrollada en el año 2005 por la Sociedad de Radiología de América del Norte (RSNA) con el fin de desarrollar un vocabulario útil para la radiología. Actualmente posee sobre 30.000 términos con una estructura mono-jerárquica. Sólo está disponible en inglés, idioma en el que se utiliza para clasificación de informes, asociación de hallazgos y diagnósticos, sistemas de soporte para la toma de decisiones, entre otros.[50]

3.1.2.5.- Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT)

Es una terminología estándar que facilita la interoperabilidad en los registros clínicos electrónicos. Posee más de 311.000 conceptos únicos organizados en jerarquías múltiples de complejidad (granularidad) creciente. Al considerar los términos correspondientes a sinónimos de los conceptos descritos, el número asciende a más de 800.000 términos activos en idioma español. Las relaciones existentes en la multi-jerarquía ascienden a más de 1.360.000. Fue creado con los términos clínicos habituales en la jerga médica, lo que lo hace un poderoso lenguaje

de salida para la codificación de textos médicos, sin perder el nivel de detalle de lo codificado.

3.1.3.- Métodos de Búsqueda y clasificación

3.1.3.1.- Métodos de indexación

La indexación corresponde al proceso previo a la consulta, por el cual ordenamos los datos de una colección de documentos de manera de poder realizar la búsqueda y obtener la información de manera rápida.

Matriz de términos y documentos

Corresponde a una matriz en la que en un eje tenemos cada documento y en el otro cada palabra del vocabulario usado en la colección. En el cruce de cada palabra, va el número de veces en que se encuentra la palabra en el documento. Esto permite hacer combinatoria de consultas del tipo, que posea a y b, pero no c.

Tener cada documento asociado a una lista de identificadores de palabras que se conoce como Indexación. Si asociamos además la posición de cada aparición se llama indexación posicional.[51] El proceso de indexación es lento, pero una vez construido pueden realizarse consultas de manera rápida.

Indexación invertida o reversa

Cada palabra del vocabulario posee una lista de identificadores de documentos en los que está presente. Al poseer además almacenada la posición de la palabra en cada documento se conoce como indexación posicional invertida.[51]

Los métodos de indexación nos permiten mantener la información de posición de la palabra facilitando la búsqueda de términos de múltiples palabras. Permitiendo hacer consultas del tipo: “que contenga el término A seguido del término B”. Este tipo de indexación es aún más eficiente, optimizando el número de tareas a la hora de buscar, debido a que habitualmente el vocabulario utilizado es menos numeroso que la cantidad de documentos pertenecientes a una colección.

3.1.3.2.- Búsqueda aproximada por texto

Además de realizar la búsqueda clásica por calce perfecto de palabras, podemos realizar búsqueda aproximada por texto. Para ésta se utilizan mediciones de la distancia entre las palabras como métricas de comparación, habitualmente usados en los correctores ortográficos.

Distancia de edición de Levenshtein

La distancia de Levenshtein corresponde a la distancia medida en el número de operaciones necesarias para llegar de la palabra original a la palabra deseada. Las inserciones, borrados y sustituciones de letras, cuentan como una operación.[52]

Algoritmo Needleman-Wunsch

El algoritmo de Needlemann-Wunsch proviene de la bioinformática para comparar secuencias de letras que representan aminoácidos de proteínas o bases nucleotídicas de ADN y ARN.

En este algoritmo se utiliza una matriz de similitud desde la cual, se obtiene una ponderación del valor de cada operación. A diferencia de la distancia de Levenshtein, ésta comienza por realizar una alineación global de la palabra, permitiendo la existencia de espacios vacíos.[53]

Algoritmos fonéticos

En los algoritmos fonéticos se calcula la distancia entre las palabras por su similitud al pronunciarlas. Para realizar esta tarea se traduce la palabra a una representación de su fonética. La más conocida es llamada Metaphone.[54]

Técnicas de procesamiento de lenguaje natural

Dentro de las herramientas del procesamiento de lenguaje natural para realizar búsqueda aproximada, encontramos el *stemming* y la lematización. En ambas técnicas se intenta dejar la parte invariable de la palabra. En la primera técnica se realiza remoción del final de la palabra de forma heurística, a través de reglas, esperando obtener un resultado correcto. La lematización, en cambio, intenta realizar esta tarea utilizando una base de conocimientos del idioma para obtener las raíces.[55]

En general, todos los métodos de búsqueda aproximada consumen más recursos computacionales al tener que realizar un mayor número de tareas y cálculos. En estos casos se prioriza la plasticidad de la búsqueda por sobre la rapidez.

3.1.3.3.- Clasificador bayesiano

Uno de los modelos matemáticos más importantes para clasificar documentos es el clasificador bayesiano.

Para clasificar un documento (d) en una clase (c) calculamos las diferentes probabilidades que el documento tiene de pertenecer a cada una de las clases, siendo la clase correspondiente la con probabilidad más alta. Esto se basa en la siguiente fórmula:

$$\mathrm{C_{MAP}} = \max_{c \in C} \frac{\mathrm{P(d|c)} \times \mathrm{P(c)}}{\mathrm{P(d)}}$$

Donde MAP es la máxima probabilidad.

Dado que la probabilidad del documento es constante (comparamos el mismo documento contra distintas clases) podemos simplificarla obteniendo como fórmula final la siguiente:

$$\mathrm{C_{MAP}} = \max_{c \in C} \mathrm{P(d|c)} \times \mathrm{P(c)}$$

Esto corresponde en estadística bayesiana a la multiplicación del *likelihood* por el *prior*.

La probabilidad del documento, dado la clase, corresponde a la multiplicación de las probabilidades de ocurrencia de cada característica del documento dada la clase, es decir:

$$\mathrm{C_{NB}} = \max_{c \in C} \mathrm{P}(c_j) \times \prod_{i \in palabras} \mathrm{P}\left(x_i \middle| c_j\right)$$

Donde NB corresponde a la probabilidad del clasificador de bayes.

Podemos considerar características en un documento la aparición de las palabras. Para esto realizamos 2 suposiciones:

1. Consideramos el documento como un conjunto de palabras.
2. La probabilidad de aparición de cada palabra es independiente.

Si bien ambas suposiciones son falsas, esto permite simplificar el modelo a las fórmulas antes mencionadas.[51]

En cuanto a velocidad es rápido, aunque su performance disminuye con el aumento del número de clases, ya que el número de tareas aumenta de forma exponencial.

3.1.3.4.- Modelo de Espacio Vectorial

El método más aceptado en la actualidad de realizar búsqueda en grandes colecciones de documentos se conoce como modelo de espacio vectorial (VSM, *Vector Space Model*).

El VSM comienza convirtiendo los documentos en vectores, donde cada una de las coordenadas del vector corresponde a la frecuencia de las distintas palabras. Como resultado de esta operación se obtiene un espacio altamente multidimensional (una dimensión por cada palabra utilizada).

Para poder comparar los documentos entre sí, no basta con la frecuencia. Se deben normalizar estos valores para hacerlos comparables. Para esto se consideran dos ponderaciones:

1.- Ponderación de la frecuencia de la palabra (*TF*, Term Frequency)

Dado a que la relevancia del documento no aumenta de forma lineal con la frecuencia de aparición de una palabra, se utiliza el logaritmo para modelarlo de la siguiente forma:

$$W_{ft} = \begin{cases} 1 - \log_{10} ft_{t,d}\,, & ft_{t,d} > 0 \\ 0\,, & ft_{t,d} = 0 \end{cases}$$

Para:

W_{ft}= Peso de la frecuencia del término.

$ft_{t,d}$= frecuencia del término t en el documento d.

2.- Ponderación inversa a la frecuencia en la colección de documentos (IDF)

Esto plantea que las palabras menos comunes son más informativas que las más comunes. Justamente en esto se basa que en ciertos flujos de NLP retiremos las palabras que catalogamos como *stopwords,* dado a que su frecuencia es tan alta que no entregan mayor información. Esto puede ser cierto o no, dependiendo de nuestro objetivo. En nuestro caso en particular, al interpretar el contexto en el que se encuentra el término, los *stopwords* son necesarios, ya que por ejemplo, la palabra “sin” puede cambiar completamente el sentido de la frase.

La ponderación inversa se calcula como:

$$ITF_t = \log_{10}\left(\frac{N}{df_t}\right)$$

Para:

N= Número de documentos de la colección.

df_t= Número de documentos con término buscado.

En el VSM utilizamos la combinación de ambas ponderaciones, conocida como *tf-idf,* que consiste en la multiplicación de estos pesos para cada término.

$$W_{t,d} = \left(1 - \log_{10} ft_{t,d}\right) \times \log_{10}\left(\frac{N}{df_t}\right)$$

Hasta el momento tenemos nuestra matriz de documentos y vectores de frecuencias de las palabras ponderadas por $W_{t,d}$, esto se conoce como matriz de pesos.

Cuando queremos realizar una búsqueda, tratamos nuestra consulta de la misma forma, la convertimos en un vector de frecuencias de las palabras y la ponderamos por $W_{t,d}$.

Teniendo nuestros vectores, ya podríamos comparar y entregar resultados por cercanía, sin embargo, estaríamos calculando la distancia de vectores de forma euclidiana, en la cual el peso de la frecuencia de cada palabra dentro del documento pesa mucho. De esta forma una consulta corta quedará siempre distante de términos en que las palabras mencionadas sean frecuentes. Para solucionar esto, normalizamos por el largo del documento y medimos la distancia representada por el coseno del ángulo de la siguiente forma:

$$\cos(\vec{q}, \vec{d}\,) = \frac{\vec{q}}{\|\vec{q}\|} \cdot \frac{\vec{d}}{\|\vec{d}\|}$$

Donde,

$\frac{\vec{q}}{\|\vec{q}\|}$= vector de la consulta normalizado por su largo.

$\frac{\vec{d}}{\|\vec{d}\|}$= vector del documento normalizado por su largo.

Con esto todos los vectores que deseemos comparar se encontrarán a la misma distancia del origen, en el límite de una híper esfera. Representándose la similitud como el coseno de la distancia angular.[51, 56]

Este método, al basarse en puro cálculo numérico, es el más rápido de los descritos. Sus limitaciones serán planteadas en la discusión.

3.2.- Contexto

En la atención clínica habitual de un paciente, al requerirse apoyo por imágenes durante el estudio de un cuadro, su médico solicita un examen a través de una orden de examen. Usualmente, el paciente acude físicamente al Departamento de Imagenología donde se adquiere la imagen. Alternativamente, se puede obtener una imagen con algún dispositivo portátil en el lugar donde se encuentra el paciente. Esta imagen será analizada posteriormente por un médico especialista en Imagenología, quien redactará un informe como producto de su análisis. Este informe podría no ser entregado al tratante hasta que éste consulte explícitamente su estado y lo solicite.

3.1.1- Definición de patología crítica

En 1972, del Dr. George Lundberg introduce el término “valores críticos” para resultados de pruebas de laboratorio. Los define como indicadores de un estado fisiopatológico alejado de la normalidad que puede poner en peligro la vida del paciente si no se actúa rápidamente, y para el que existe tratamiento[57].

Rápidamente el uso de dicho término se extendió a estudios de Anatomía Patológica e Imagenología. La definición de hallazgos críticos en Imagenología según las recomendaciones chilenas[58] es:

> *“Aquellos hallazgos nuevos o inesperados que indican que el paciente tiene un elevado riesgo de morbimortalidad si no se toman las adecuadas medidas de diagnóstico o tratamiento en forma oportuna.*
>
> *También podemos incluir en esta categoría a aquellos hallazgos cuya interpretación difiere de manera significativa de una interpretación e informe preliminar que ya ha sido entregado (discrepancias).”*

Luego de aceptada esta definición se modificó el flujo de trabajo, y ahora, cuando un radiólogo encuentra una patología que corresponda a la definición planteada, dará aviso al médico que solicitó el examen o al paciente de manera directa, con el fin de iniciar tratamiento a la brevedad.

Si bien el uso del concepto se extendió rápidamente, hay pocos estudios que muestren la importancia de notificar a la brevedad las patologías consideradas valores críticos. En estudios con resultados de laboratorio, se pudo medir que los paciente con exámenes de sodio, potasio o glucosa plasmáticas en rangos críticos, presentaban una tasa de mortalidad 13% más alta al no ser notificado el médico tratante[59, 60]. A pesar de la falta de conocimiento, existen múltiples estudios que intentan mejorar el tiempo de comunicación entre el radiólogo y el médico tratante para poder prevenir la morbimortalidad del paciente, entendiéndose esta rápida comunicación como beneficio implícito para el paciente[43, 59-61].

Lo que comenzó como una buena práctica fue luego avalado por importantes entidades como la "*Joint Commission*", el "*American College of Radiology*", y la "Coalición para la Prevención de Errores Médico de *Massachusetts*" quienes han creado recomendaciones respecto a que patologías considerar críticas para disminuir la morbimortalidad de este grupo de pacientes y de qué forma se debería notificar.[43, 59, 62] En Chile la Superintendencia de Salud y la Sociedad Chilena de Radiología proponen también sus recomendaciones.[58, 63] La acreditación de las instituciones de Salud, realizada por la Superintendencia de Salud en Chile, dentro del ámbito de Acceso Oportunidad y Continuidad de la Atención, mide la aplicación de "procedimientos para asegurar la notificación oportuna de situaciones de riesgo, detectadas a través de exámenes diagnósticos en las áreas de Anatomía Patológica, Laboratorio e Imagenología".[64]

Es esperable una falta de estudios de tipo caso control debido a los componentes éticos que implicarían el no avisar patologías graves premeditadamente. Cuando se detecta un neumotórax es requerida de forma urgente la instalación de un tubo pleural para que el paciente no fallezca ahogado por problemas ventilatorios de índole mecánicos. Cuando se detecta un

tromboembolismo pulmonar (TEP), se debe administrar a la brevedad tratamiento anticoagulante con el fin de recuperar la circulación pulmonar. Patologías como los pseudoaneurisma, los cuerpos extraños bronquiales y disección aortica requerirán intervenciones quirúrgicas inmediatas con el fin de evitar las complicaciones que ocurrirán sin tratamiento. [65]

Un estudio de 1960, no reproducible con los estándares éticos actuales, realizó un estudio tipo caso control randomizado en el que siguieron a pacientes con diagnóstico de TEP con y sin tratamiento anticoagulante . El estudio comenzó a tratar a todos sus pacientes tras acumular 19 pacientes con TEP del grupo sin administración de tratamiento, observando una mortalidad mayor al 26% (5). Para el grupo con tratamiento sólo se presentaron 2 muertes en 54 pacientes.[66]

3.1.2.- Dificultades en la medición de Patología Crítica

Una dificultad al momento de establecer políticas de declaración, es que la recomendación del Ministerio de Salud chileno[58] y de la Sociedad Chilena de Radiología (SOCHRADI)[63] es poco conocida. La exigencia de la acreditación en Salud de la Superintendencia exige superar un umbral autoimpuesto por la institución, indicador que se calcula de la siguiente forma:

$$\frac{\text{Nº de patologías críticas declaradas comunicadas a tiempo según protocolo}}{\text{Total de patologías críticas declaradas}}$$

En esta acreditación no se exige un listado de patologías, quedando las recomendaciones sólo como una declaración de buenas intenciones que cada médico puede o no poner en práctica. Las instituciones no utilizan el listado recomendado por la SOCHRADI, dado la dificultad que presentaría cumplir con un índice de mayor cobertura. No se considera en el indicador las patologías críticas no declaradas (patologías que el radiólogo olvidó catalogar como crítica y debieron serlo según protocolo), incluso estando definidas en los protocolos institucionales.

No existen en Chile sistemas automatizados de apoyo a la decisión con respecto a la patología crítica, que recuerden al radiólogo que la patología descrita se considera crítica.

3.3.- Problema

Existiría una sub declaración de patología crítica que iría en desmedro del resultado en el tratamiento del paciente. Ésta está favorecida por la naturaleza del informe radiológico, el cual, si bien en muchos centros es de naturaleza digital, es en texto libre. No existe una nomenclatura de salida formal como intento de clasificar los reportes según resultado.

4.- HIPÓTESIS Y OBJETIVOS

4.1.- Hipótesis

Mediante algoritmos de procesamiento de lenguaje natural pueden identificarse más del 80% de los diagnósticos existentes en SNOMED-CT de las impresiones de los informes radiológicos, identificando patología critica con más de un 90% de valor productivo positivo.

4.2.- Objetivos

4.2.1.- Objetivo General

Identificar más del 80% de los diagnósticos de los informes radiológicos de manera correcta y clasificar los diagnósticos identificados en crítico o no, con al menos un 90% de precisión (o valor predictivo positivo).

4.2.2.- Objetivos específicos

1.- Elaborar un diseño experimental y construir la base de datos. En particular construir un corpus lingüístico y procesar terminología SNOMED-CT.

2.- Identificar los diagnósticos de los informes radiológicos en más de un 80% de los casos, considerando como caso cada diagnóstico.

3.- Identificar con precisión mayor al 90% los diagnósticos de patología crítica informados en el texto libre de los informes radiológicos.

5.- METODOLOGÍA

5.1.- Elaborar un diseño experimental y construir la base de datos. En particular construir un corpus lingüístico y procesar terminología SNOMED

5.1.1.- Diseño experimental

Para lograr construir un diseño experimental se obtuvo los informes de Angiotomografía solicitados en Clínica Alemana durante los años 2013 y 2014. La autorización del comité de ética de la institución fue obtenida mediante addendum a proyecto de investigación en curso[67], en el que se agregó al autor de la presente tesis como co-investigador. El proyecto citado tiene como objetivo identificar mediante procesamiento de lenguaje natural los informes con diagnóstico de tromboembolismo pulmonar.

Se comenzó con un total de 2330 mensajes de HL7. Se inició la depuración de dichos mensajes por la eliminación de los mensajes duplicados, con lo que 245 mensajes fueron eliminados.

En caso de existir correcciones o mensajes con informes parciales, se consideró sólo el mensaje con fecha más reciente. Tras esto 94 mensajes fueron descartados considerándose antiguos.

Se identificó mediante detección de patrones en la estructura de los informes el segmento correspondiente a la impresión diagnóstica. Se iteró para mejorar la detección de estos patrones. Después de esto, en 18 informes no fue posible identificar un segmento de impresión, corroborándose la no existencia de ésta manualmente por lo que también fueron descartados.

Finalmente se obtuvo, tras la depuración de casos, 1973 impresiones diagnósticas, las cuales fueron identificadas asignando un número correlativo a los mensajes. Se guardó la impresión y su número asociado en otra base de datos, con la cual se procedió a realizar la investigación.

Ya obtenidos los informes a analizar, se diseñó un motor de búsqueda de conceptos existentes en la base de datos de SNOMED-CT. Se eligió esta terminología por ser amplia y cubrir gran cantidad de sinónimos. Además presenta

estructura multi-jerárquica, con la cual se podría agrupar la información obtenida por diversas jerarquías según su propósito.

Para seguir con el diseño de las herramientas, se decidió utilizar indexación reversa de la terminología SNOMED-CT con el fin de poder buscar los términos de múltiples palabras para cada una de ellas de forma independiente.

Se decidió no realizar búsquedas aproximadas por texto, con el propósito de priorizar la velocidad de la herramienta.

Si bien el objetivo final tiene como dato de salida una clasificación dicotómica, se favoreció mantener el nivel de detalle de los conceptos SNOMED-CT para ampliar el propósito de la herramienta a futuro, pudiendo hacer consultas complejas si se mantenían los índices de SNOMED-CT. Por este motivo se diseñó una herramienta de búsqueda de términos en lugar de utilizar un clasificador bayesiano.

Como el buscador de términos sólo nos informaría la mención de un término y su ubicación, fue necesario desarrollar un detector de negación para conocer la real presencia del término. Esta última tarea se realizó utilizando expresiones regulares por el gran éxito de herramientas como NegEx[68] en el área. De esta forma podemos obtener como dato de salida un índice de un término SNOMED-CT asociado al contexto de la frase. Utilizando un modelo de espacio vectorial, este nivel de información no hubiese sido simple de obtener, por lo que se prefirió la utilización de la combinación de herramientas descritas.

5.1.2.- Preparación de la base de datos de SNOMED-CT para la identificación de términos

Se obtuvo la terminología de SNOMED-CT, edición nacional chilena vigente a la actualidad, con la licencia correspondiente para uso en investigación otorgada por el representante nacional.

Se consideraron para la identificación de diagnósticos sólo los términos de conceptos pertenecientes a la jerarquía de hallazgos correspondientes a 90245 términos.

Se procedió a realizar una indexación reversa de los términos SNOMED. En base a la tabla de descriptores se identificaron los *tokens* (una palabra como una unidad) que componían los términos, omitiendo los *tokens* considerados *stopwords*.

Como resultado, se obtuvo una nueva tabla en la que se tenía un *token*, el *id* al que pertenecía y el largo del término. 330.828 *tokens* fueron identificados (23.684 diferentes).

Se contó además el largo en *tokens* de cada término sin *Stopwords*.

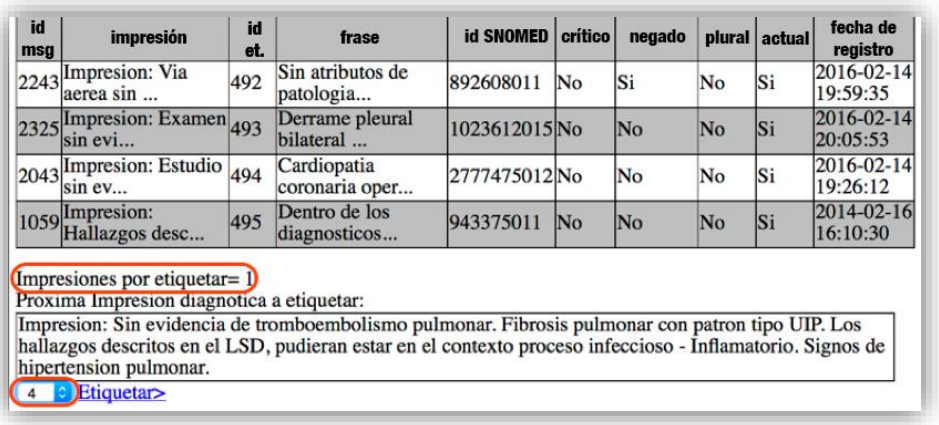
Se agregó de forma símil a las extensiones de descripciones de SNOMED términos correspondiente a acrónimos asignando un *descriptor id* de SNOMED que correspondiera al acrónimo expandido. Para esto se identificó en los informes expresiones regulares de 2 a 4 letras mayúsculas consecutivas, separadas o no por puntos. Se identificaron 12 acrónimos utilizados en las impresiones diagnósticas, sólo uno correspondiente a diagnóstico. Se analizó ante la presencia del acrónimo y su contexto de forma manual sin identificar en ningún caso ambigüedad, por lo que la tarea de desambiguación fue omitida.

5.1.3.- Anotación de las Impresiones

Se calculó el número de informes que representaba adecuadamente la muestra con un error alfa de 5% y una potencia estadística de 95% (error beta de 5%), corregido por tamaño del universo. Se obtuvo un tamaño muestral de 219 impresiones.

Se generó un formulario *web* en PHP para realizar el etiquetado, solicitando a dos médicos con el *curso SNOMED CT Foundation* aprobado, identificar la mención del término indicando el *descriptor id* de SNOMED-CT; si el informe completo debía clasificarse como crítico por el diagnóstico etiquetado, si estaba negado, si era nuevo, si estaba en plural, y posteriormente adjuntar la frase en la que se encontró el término (Figuras 1 y 2). En caso de no haber diagnóstico en la impresión, se guardó el *id* del informe con una línea en blanco. Finalmente para el análisis realizado en el presente estudio, se utilizó el id del descriptor SNOMED-CT y las variables booleanas de informe critico, negación y actualidad.

Figura 1. Primera ventana de pre visualización de impresión a etiquetar.

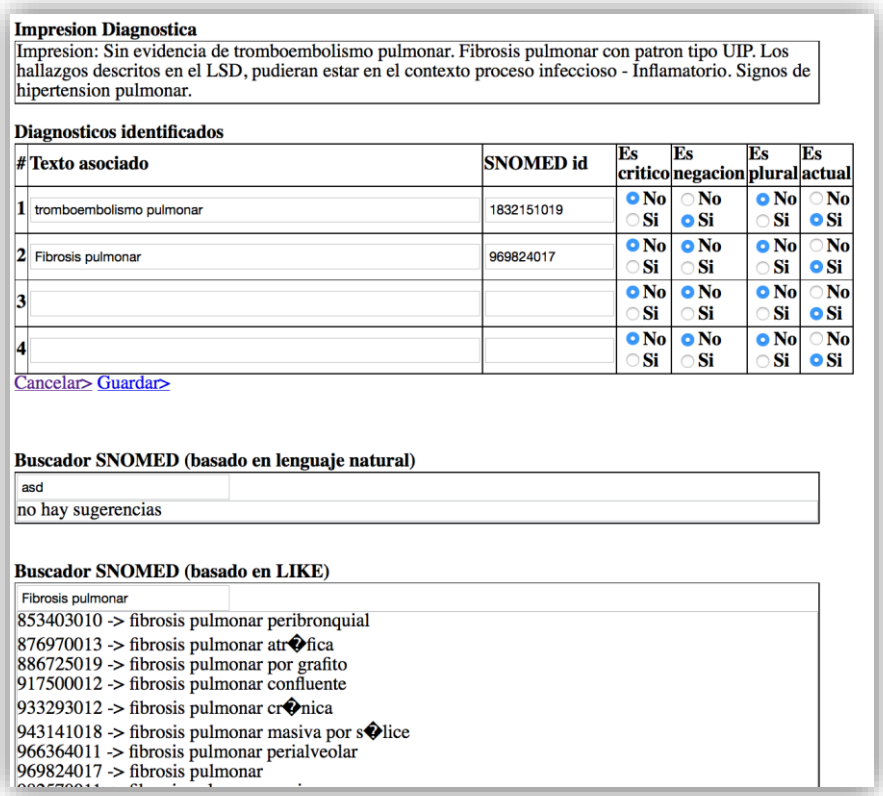


id msg	impresión	id et.	frase	id SNOMED	crítico	negado	plural	actual	fecha de registro
2243	Impresion: Via aerea sin ...	492	Sin atributos de patologia...	892608011	No	Si	No	Si	2016-02-14 19:59:35
2325	Impresion: Examen sin evi...	493	Derrame pleural bilateral ...	1023612015	No	No	No	Si	2016-02-14 20:05:53
2043	Impresion: Estudio sin ev...	494	Cardiopatia coronaria oper...	2777475012	No	No	No	Si	2016-02-14 19:26:12
1059	Impresion: Hallazgos desc...	495	Dentro de los diagnosticos...	943375011	No	No	No	Si	2014-02-16 16:10:30

Impresiones por etiquetar= 1
Proxima Impresion diagnotica a etiquetar:
Impresion: Sin evidencia de tromboembolismo pulmonar. Fibrosis pulmonar con patron tipo UIP. Los hallazgos descritos en el LSD, pudieran estar en el contexto proceso infeccioso - Inflamatorio. Signos de hipertension pulmonar.
4 Etiquetar>

Se indica cuantas impresiones restan por etiquetar (círculo rojo superior), y cuantos diagnósticos se etiquetarán en el siguiente informe (4 en la imagen, circulo rojo inferior). En la tabla superior se puede ver todos los diagnósticos ya etiquetados. Los datos almacenados de izquierda a derecha corresponden a; identificador del mensaje, impresión etiquetada, identificador del etiquetado, frase en la que se encuentra el término, identificador SNOMED-CT del término etiquetado, si el informe es crítico debido a ese término, si el termino se encuentra negado, si el término se encuentra en plural, si el término es actual y la fecha y hora del etiquetado.

Figura 2. Ventana de etiquetado



Impresion Diagnostica

Impresion: Sin evidencia de tromboembolismo pulmonar. Fibrosis pulmonar con patron tipo UIP. Los hallazgos descritos en el LSD, pudieran estar en el contexto proceso infeccioso - Inflamatorio. Signos de hipertension pulmonar.

Diagnosticos identificados

#	Texto asociado	SNOMED id	Es critico	Es negacion	Es plural	Es actual
1	tromboembolismo pulmonar	1832151019	No / Si	No / Si	No / Si	No / Si
2	Fibrosis pulmonar	969824017	No / Si	No / Si	No / Si	No / Si
3			No / Si	No / Si	No / Si	No / Si
4			No / Si	No / Si	No / Si	No / Si

Cancelar> Guardar>

Buscador SNOMED (basado en lenguaje natural)

asd
no hay sugerencias

Buscador SNOMED (basado en LIKE)

Fibrosis pulmonar
853403010 -> fibrosis pulmonar peribronquial
876970013 -> fibrosis pulmonar atr[illegible]fica
886725019 -> fibrosis pulmonar por grafito
917500012 -> fibrosis pulmonar confluente
933293012 -> fibrosis pulmonar cr[illegible]nica
943141018 -> fibrosis pulmonar masiva por s[illegible]lice
966364011 -> fibrosis pulmonar perialveolar
969824017 -> fibrosis pulmonar

El experto indica cuantos diagnósticos hay en la impresión (4 en el ejemplo) y el *id* SNOMED asociado, entre otros atributos.

La validación del acuerdo entre etiquetadores se realizó considerando la combinación de diagnóstico y estado de negación. Las principales diferencias estuvieron dadas en el etiquetado de sinónimos no existentes en la nomenclatura, en el que uno de los médicos asignó identificadores de otros términos similares. Se consideró como *gold standard* el etiquetado del médico que no homologó sinónimos, debido a que una correcta utilización de SNOMED debiese contemplar la incorporación de nuevos descriptores de conceptos con un nuevo id no siendo correcto homologar.

Se anotaron 445 líneas, correspondiendo 13 de ellas a informes sin diagnóstico. Los diagnósticos más frecuentes encontrados fueron: tromboembolismo pulmonar (159), derrame pleural (26), enfisema centrolobulillar (13), atelectasia (12), hipertensión pulmonar (12), derrame pleural bilateral (11).

5.2.- Identificar los diagnósticos de los informes radiológicos en más de un 80% de los casos, considerando como caso cada diagnóstico

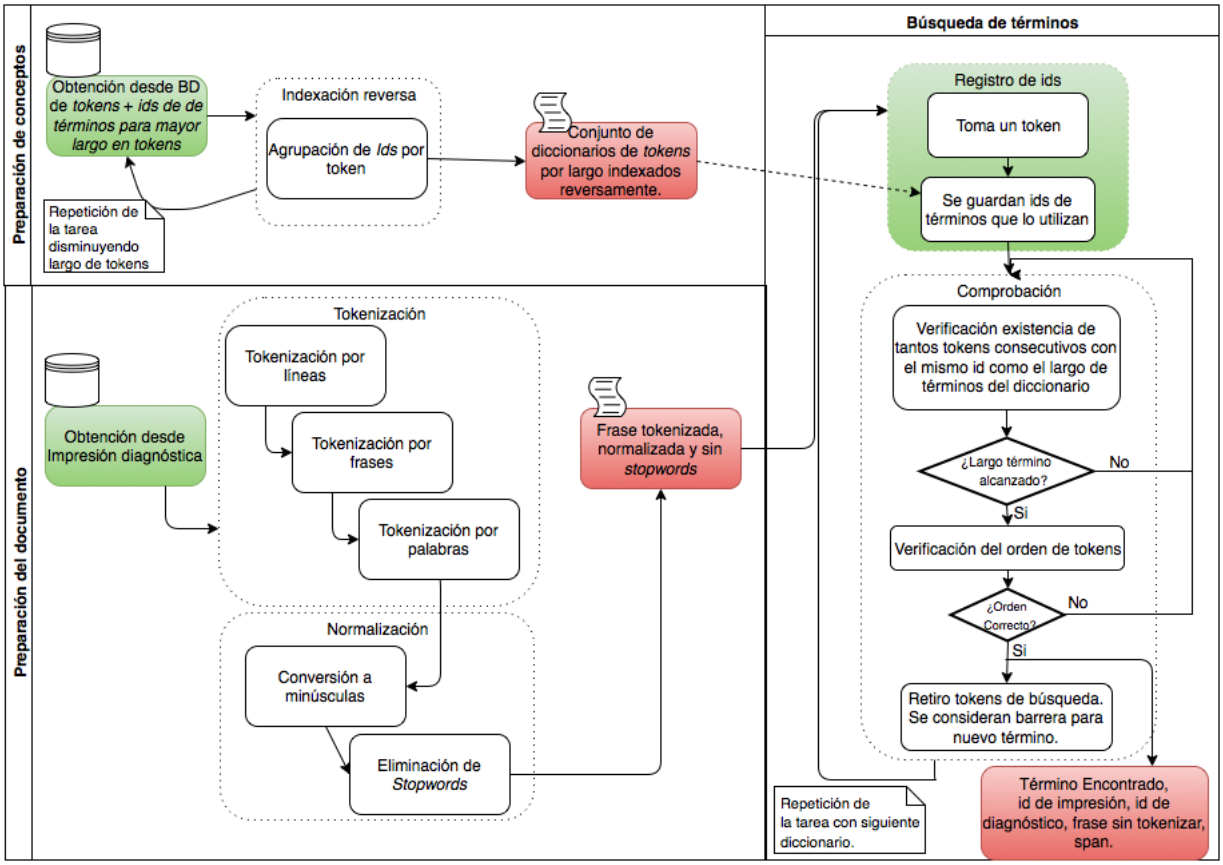
5.2.1.- Identificación del diagnóstico: Detección de los términos, *search engine* con herramientas de procesamiento de lenguaje natural

El proceso general, tanto de indexación como de búsqueda, se ilustra en la Figura 3. Para cada informe se realizó una *tokenización* en 3 niveles. Primero se dividió la impresión en líneas, luego en frases y por último en palabras. A continuación se removieron los *stopwords* del informe y se normalizó a minúsculas, siempre manteniendo la relación con la frase original.

Luego, en la *tokenización* por palabras y considerando el nivel frase como una barrera, se buscaron los términos existentes en SNOMED. Para esto se utilizó la técnica de indexación reversa aplicada sobre los términos SNOMED. En vez de utilizar el *id* de los informes para la indexación se consideró el *id* del descriptor. Finalmente, se obtuvo de la base de datos un vocabulario segmentado por largo del término en el que cada *token* poseía una lista de *id* de descriptor SNOMED en el que se utilizaba.

Este cruce permitió identificar el comienzo de un término por la identificación de un *token*, realizando iteraciones de búsquedas por los vocabularios de los descriptores de mayor largo hacia los de menor largo. Al encontrar una coincidencia de un *token* se procedía con el *token* siguiente. De haber ocurrencia de *tokens* de forma consecutiva, con el mismo descriptor, por una largo equivalente al largo del descriptor medido en *tokens,* se realizaba la verificación del orden en el descriptor completo. De encontrarse en el mismo orden, estos *token* eran marcados como encontrados, para no buscar términos contenidos en ellos en una próxima iteración, en búsqueda de términos de menor largo. De esta forma se aseguraba no detectar términos cortos contenidos dentro de términos de mayor longitud.

Figura 3. Esquema de procesos del motor de búsqueda.



Se detalla los 3 procesos individuales con los cuales se trató los informes. Es posible observar como confluyen los 2 primeros (“Preparación de conceptos” y “Preparación de documentos”) en el tercero (“Búsqueda de términos”). La preparación de conceptos tiene como fin cargar en memoria los

conceptos a buscar y agruparlos por palabras de forma de poder acelerar la búsqueda de éstos. La preparación de los documentos tiene como objetivo identificar las palabras y frases como unidades manejables por un computador. El tercer flujo en si corresponde a la búsqueda propiamente tal.

Al identificar un término SNOMED en la frase se obtenía como salida el *id* del la impresión, el *id* del término, la frase sin *tokenizar* y el alcance (*span*) del término dentro de la frase.

5.2.2.- Construcción del detector de negación

Se utilizó el algoritmo de la herramienta NegEx ampliamente validada por la comunidad científica como detector de negación por patrones en idioma inglés[68]. La herramienta ha evolucionado al considerar la dirección de la expresión de negación cambiando su nombre a ConText[69]. Dado que en la literatura estaban descritos resultados pobres al traducir las reglas de ConText al español[70] se decidió utilizar el algoritmo de ConText construyendo expresiones regulares del español desde cero. Para entregarle a la herramienta los términos encontrados por nuestro *search engine*, se decidió programar nuevamente la herramienta por lo que, finalmente, sólo fue utilizada la idea base de su algoritmo.

Se implementó además 2 nuevos pasos para mejorar los resultados de ConText:

1.- Se asignó un nuevo atributo correspondiente al tipo de la categoría de la expresión regular con el fin de evitar la detección simultánea de 2 reglas pertenecientes al mismo tipo. Por ejemplo: el tipo polaridad tiene reglas cuyas características son negación o aseveración. Existe una expresión regular de negación que detecta el *substring* “sin evidencias de” sin embargo la expresión de aseveración “evidencias de” esta contenida en su interior. Al asignar tipo a la categoría podemos considerar que sólo la regla más amplia de la misma categoría permanezca, evitando ambigüedades.

2.- Si en una misma frase existía más de un *substring* de la misma categoría, y uno de éstos se encontraba dentro del alcance del otro, el contenido al interior del alcance, limitaba el alcance del otro considerando la regla de dirección. Por ejemplo: En la frase “Se observa fractura costal sin

evidencias de derrame pleural". Tenemos 2 *substrings* de polaridad; "Se observa" y "sin evidencias de", ambos con regla de negación hacia adelante. En el algoritmo ConText el primero aseveraría hasta el final de la frase. En el algoritmo propuesto su alcance se ve limitado por la presencia del segundo *substring* por lo que se asevera sólo fractura costal y no derrame pleural.

Identificación de los patrones de negación

Para la construcción del detector de negación se analizó la literatura para encontrar la herramienta más adecuada. Se decidió realizar la tarea identificando patrones dentro del nivel de la frase.

Para la extracción de estos patrones:

1.- Se consideró sólo las impresiones diagnósticas no etiquetadas.

2.- De estas impresiones se obtuvo todas las frases en que, con el *search engine,* se logró identificar algún término en ellas.

3.- El término encontrado fue reemplazado por un término común, en este caso por la palabra SNOMED CT.

4.- Se eliminó la puntuación y los espacios finales en la frase.

5.- Se ordenó las frases por frecuencia de aparición.

6.- Se procedió a identificar *substrings* de las frase que indicaran la negación. Se consideró que se debía marcar la sección de la frase más larga posible que no contuviera el término.

7.- Se agrupó y ordenó por frecuencia los trozos de frases de negación.

8.- De forma manual se analizó estos *substrings* para identificar patrones.

9.- Con estos patrones se construyeron expresiones regulares que identifican negación.

10.- Se asignó una regla a estas expresiones para identificar si el término negado se encontraría anterior o posterior a la expresión en el texto. Se incluyó la regla de bi-direccionalidad.

5.3.- Identificar con precisión mayor al 90% los diagnósticos de patología crítica informados en el texto libre de los informes radiológicos

5.3.1.- Definición de patología crítica

Se revisó la norma técnica número 9 de la Sociedad Chilena de Radiología (SOCHRADI), que entrega recomendaciones acerca de qué considerar patología crítica. Muchas de las recomendaciones de la guía no dependen sólo de la presencia o no de cierto diagnóstico, por lo que se decidió evaluar sólo con la patología categorizable de esta forma. De las 10 condiciones catalogadas por la norma como patología crítica, únicamente 2 se pueden evaluar por sólo la presencia de ésta. “Tromboembolismo pulmonar” y “Neumotórax a tensión”.

5.3.2.- Creación de los subconjuntos de diagnósticos críticos

Para la creación de los subconjuntos se utilizó el *browser* de la IHTSDO para buscar términos correspondientes a los conceptos de “Tromboembolismo Pulmonar” (TEP) y “Neumotórax a Tensión”. Se crearon 2 sets de *id* de descriptores correspondiendo a sinónimos de cada uno de estos conceptos, los cuales se detallan en la sección 6.3.

5.3.3.- Identificación de patología crítica

Los resultados obtenidos se evaluaron por informe, requiriéndose un TEP no negado, sin presencia de un TEP histórico dentro de éste, para considerar todo el informe como positivo.

6.- RESULTADOS

6.1.- Elaborar un diseño experimental y construir la base de datos. En particular construir un corpus lingüístico y procesar terminología SNOMED

Se realizó satisfactoriamente el diseño experimental y la construcción de la base de datos. Se obtuvo un corpus de 219 informes etiquetados con un kappa de acuerdo entre etiquetadores de 85.5%, $IC^{95\%}$[80.8-90.3%], $p < 0.001$.

Se procesó la terminología realzándole una indexación reversa.

Finalmente se obtuvo 2 bases de datos, por un lado la base de conceptos del SNOMED-CT, utilizando 2 tablas. La tabla de descriptores original de SNOMED-CT y una tabla que guarda información de la palabra, id del término y largo en palabras del término. La otra base de datos corresponde a la que aloja los documentos. En ésta se utilizó una tabla en la que se encuentró el Mensaje HL7 originales, y la sección de la impresión radiológica ya identificada entre otros datos.

6.2.- Identificar los diagnósticos de los informes radiológicos en más de un 80% de los casos, considerando como caso cada diagnóstico

6.2.1.- Identificación del diagnóstico

Se realizó un test diagnóstico para evaluar el *search engine* sólo evaluando la detección de la mención del término. Los resultados se muestran en la tabla 1.

Tabla 1. Test diagnóstico para evaluar *search engine*.

	Término Detectado	Término no detectado.	Total
Término Presente	394	38	432
Término Ausente	8	13	21
Total	402	51	453

Sensibilidad / *Recall*:	98.01%	$IC^{95\%}$[96.1 - 99.1%]
Especificidad:	25.49%	$IC^{95\%}$[14.3 - 39.6%]
Prevalencia de Enfermedad:	95.36%	$IC^{95\%}$[85.4 - 91.5%]
Valor Predictivo Positivo / *Precision*:	91.20%	$IC^{95\%}$[88.1 - 93.7%]
Valor Predictivo Negativo:	61.90%	$IC^{95\%}$[38.4 - 81.8%]
Medida F balanceada:	0.9448	

χ^2 de independencia: *p-value* < 0,001.

Fuente: Elaboración propia.

Podemos ver un excelente rendimiento para buscar diagnósticos habiendo presentado tan sólo 8 falsos positivos y 38 falsos negativos de un total de 453 términos. El *p-value* obtenido nos indica la dependencia existente entre el *gold standard* y la herramienta de búsqueda.

6.2.2.- Identificación de los patrones de negación

Tras el procedimiento de extracción de patrones se obtuvo 14 reglas de negación. En la tabla 2 se pueden observar algunos ejemplos.

Tabla 2. Ejemplos de patrones de negación.

Referencia	Tipo	Característica	Expresión regular	Dirección
Sin evidencias	Polaridad	NEG	((([Aa]ngio(( )?(T(A)?C|t(a)?c))?((de torax|de aorta(toracoabdominal)?))?|[Ee]studio|[Ee]xamen))?s|S)in evidencias((evidentes|sugerentes|categoricas|actuales)) ?(de)?	Adelante
Sin evidencia	Polaridad	NEG	((([Aa]ngio(( )?(T(A)?C|t(a)?c))?((de torax|de aorta(toracoabdominal)?))?|[Ee]studio|[Ee]xamen))?s|S)in evidencia((evidente|sugerente|categorica|actual))?(de)?	Adelante
Negativo/a	Polaridad	NEG	((([Aa]ngio(( )?(T(A)?C|t(a)?c))?((de torax|de aorta(toracoabdominal)?))?|[Ee]studio|[Ee]xamen))?n|N)egativ[oa](para)?	Adelante
Sin signos	Polaridad	NEG	((([Aa]ngio(( )?(T(A)?C|t(a)?c))?((de torax|de aorta(toracoabdominal)?))?|[Ee]studio|[Ee]xamen))?s|S)in signos((evidentes|sugerentes|categoricos|actuales)) ?(de)?	Adelante
Sin elementos	Polaridad	NEG	((([Aa]ngio(( )?(T(A)?C|t(a)?c))?((de torax|de aorta(toracoabdominal)?))?|[Ee]studio|[Ee]xamen))?s|S)in elementos((evidentes|sugerentes|categoricos|actuales)) ?(de)?	Adelante
Sin hallazgos	Polaridad	NEG	((([Aa]ngio(( )?(T(A)?C|t(a)?c))?((de torax|de aorta(toracoabdominal)?))?|[Ee]studio|[Ee]xamen))?s|S)in hallazgos((evidentes|sugerentes|categoricos|actuales)) ?(de)?	Adelante

Fuente: Elaboración propia.

A partir de las expresiones regulares listadas en la tabla, las cuales son todas finitas, se obtienen coincidencias con múltiples frases. Tan sólo la primera expresión “Sin evidencias” se expande a 780 posibles cadenas de texto, tan diferentes como por ejemplo “AngioTAC de tórax sin evidencias sugerentes de” o simplemente “Sin evidencias de”, manteniendo en común el texto de referencia “Sin evidencias”. La segunda regla, “Sin evidencia” se separó de la primera ya que las palabras que acompañan a la expresión cambian al ser singulares. Dicha frase coincide con cadenas como “Angio TC de aorta toracoabdominal sin evidencia categórica de”. Las reglas siguientes poseen los siguientes ejemplos:

Negativo:	“Estudio negativo para”
Sin signos:	“Examen sin signos sugerentes de”
Sin elementos:	“TAC de tórax sin elementos actuales de”
Sin hallazgos:	“Angiotac de aorta sin hallazgos de ”

6.2.3.- Detección de la negación

Para testear la herramienta de detección de negación, se ingresaron al algoritmo los diagnósticos etiquetados. Sólo se le solicitó a la herramienta que identificara la negación, de esta forma el resultado obtenido al detectar negación es independiente del funcionamiento del motor de búsqueda en el paso previo. Se le entregaron a la herramienta 433 términos y sus frases. En la tabla 3 se observa la tabla de contingencia de etiquetado de negación versus detección de negación.

Tabla 3. Tabla de contingencia de la detección de negaciones.

	Detectado Negado	No detectado Negado	Total
Etiquetado Negado	148	2	150
Etiquetado Aseverado	1	281	282
Total	149	283	432

Sensibilidad / *Recall*:	99.33%	$IC^{95\%}$[96.3 - 99.9%]
Especificidad:	99.29%	$IC^{95\%}$[97.4 - 99.9%]
Prevalencia de Enfermedad:	34.49%	$IC^{95\%}$[30.0 - 39.1%]
Valor Predictivo Positivo / *Precision*:	98.67%	$IC^{95\%}$[95.2 - 99.8%]
Valor Predictivo Negativo:	99.65 %	$IC^{95\%}$[98.0 - 99.9%]
Medida F balanceada:	0.9899	

χ^2 de independencia: *p-value* < 0,001.

Fuente: Elaboración propia.

6.3.- Identificar con precisión mayor al 90% los diagnósticos de patología crítica informados en el texto libre de los informes radiológicos

En la tabla 4 se puede observar la relación entre los términos SNOMED y la recomendación de la norma técnica número 9. El set de TEP contenía 7 *ids* de descriptores SNOMED y el set de neumotórax a tensión 4 *ids*.

Tabla 4. Diagnósticos críticos y su relación con SNOMED

Diagnóstico Crítico según norma 9	Termino (Ids) SNOMED	Requiere análisis adicional.
Masas colecciones que afecten vía aérea		Requiere herramienta adicional de localización en vía aérea.
Neumotórax a tensión	neumotórax a tensión (1153211017) neumotórax espontáneo a tensión (1149734010) neumatocele a tensión (1288817017) neumomediastino a tensión (2682512019)	No
Derrame pleural masivo		El termino puede estar codificado con otra expresión volviéndose subjetivo.
Derrame pericardio con riesgo de taponamiento cardiaco		Se requeriría herramienta que identifique posibles riesgos.
Tromboembolismo Pulmonar.	embolia pulmonar (988479017) tromboembolismo (1423242011) tromboembolismo pulmonar (1832151019) embolismo pulmonar (1838453018) trombosis (2799133019) tromboembolia pulmonar (2853413013) tromboembolia pulmonar aguda (3069648013)	No
Síndrome aórtico agudo		Conjunto grande y ambiguo requeriría herramienta que identifique condiciones como, aneurisma actualmente roto.
Neumonía extensa o multifocal		Requeriría identificar extensión. Limite de clasificación ambiguo.
Signos de sangrado activo de tórax		No es un diagnóstico en particular. Se identificar dirigidamente cualquier tipo de sangrado y su estado.
Sonda o tubo endotraqueal en bronquios		Habría que identificar localización del tubo.
Sospecha de aneurisma aórtico o visceral complicado		Término complicado es un grupo no acotado de diagnósticos además en estado de sospecha.

Origen: Patologías críticas según recomendación de la norma técnica Nº 9 de la Sociedad Chilena de Radiología. Índice SNOMED-CT obtenido de la edición nacional 2016. Relación entre ambos y análisis de factibilidad corresponde a elaboración propia.

6.3.1- Identificación de patología crítica

En los informes etiquetados no se anotó ningún diagnóstico de Neumotórax a tensión por lo que no fue posible realizar un test diagnóstico. Se revisó la detección de la herramienta la cual tampoco fue capaz de detectar un Neumotórax a tensión en el universo completo de informes (etiquetados y no). Esto se explica por la naturaleza del diagnóstico en cuestión, ante el cual su detección en imágenes es considerado negligencia médica, ya que debe ser diagnosticado por clínica y tratado de inmediato. Por este motivo se realizó la medición del rendimiento sólo con el set de TEP.

Para la identificación de TEP, el resultado para identificarlo como crítico en base a la mención, no negación y no histórico, fue el siguiente.

Tabla 5. Tabla de contingencia identificación TEP.

	Detección de TEP actual	Sin detección de TEP actual	Total
Informe con TEP actual	26	1	27
Informe sin TEP actual	2	190	192
Total	28	191	219

Sensibilidad / *Recall*:	92.86%	$IC^{95\%}$[76.5 - 99.1%]
Especificidad:	99.48%	$IC^{95\%}$[97.1 - 99.9%]
Prevalencia de Enfermedad:	12.79%	$IC^{95\%}$[8.6 - 17.9%]
Valor Predictivo Positivo / *Precision*:	96.30%	$IC^{95\%}$[81.0 - 99.9%]
Valor Predictivo Negativo:	98.96%	$IC^{95\%}$[96.2 - 99.8%]
Medida F balanceada:	0.9455	

χ^2 de independencia: *p-value* < 0,001.

Fuente: Elaboración propia.

7.- DISCUSIÓN

Los resultados fueron satisfactorios al obtener una sensibilidad de detección de diagnósticos de 98.01%, $IC^{95\%}$[96.1 - 99.1%] y un valor predictivo positivo para identificación de patología crítica de 96.30%, $IC^{95\%}$[81.0 - 99.9%], ambos por sobre lo propuesto. En general se pudo identificar la gran mayoría de los diagnósticos existentes en SNOMED a pesar de utilizar un método de búsqueda por términos exactos.

A continuación se discutirá el método utilizado en la presente tesis en cuanto a: complejidad computacional, comparación del método utilizado con otros, y por último, una discusión acerca de la presencia de *tailoring* versus *overfiting*.

7.1.- Complejidad computacional del algoritmo propuesto

7.1.1.- Orden computacional del cálculo, velocidad versus calidad

Para estimar una complejidad computacional simplificaremos su función dividiendo el tratamiento de los documentos en 3 tareas tomando N como el término SNOMED más largo (N=26):

1.- Tokenización a 3 niveles y búsqueda. Asignación de los id correspondientes a cada palabra.

Esta tarea es dependiente del número de párrafos que tenga el informe (recordar que el primer nivel de tokenización es por "*salto de línea*"). Luego del número de frases de cada párrafo y finalmente del número de palabras de cada frase.

Podemos representar esto de la forma:

$$Tareas = Párrafos \times Frases \times Palabras$$

En la tabla 6 es posible observar que en el peor caso teórico el número de tareas (T) correspondería a 50 x 5 x 66 = 16.500. Cabe destacar que en el mejor caso es 1, y en el caso promedio 31 tareas por informe.

Tabla 6. Descripción detallada del universo de impresiones.

	Rango (Min-Max)	Promedio	Total Neto
Párrafos	(1-50)	3,83 párrafos/impresión	8903
Frases	(1-5)	1,09 frases/párrafos	9684
Palabras	(1-66)	7,4 palabras/frase	71705

Fuente: Elaboración propia.

Esta tarea se realiza 26 veces, una vez por cada vocabulario de términos al segmentar el vocabulario global por largo en *tokens*. Es decir la complejidad es O(TN) dado que los vocabularios están en una tabla de hash.

2.- Lectura de los id palabra por palabra verificando si el largo corresponde al largo del diccionario.

Luego de la asignación se realizó la verificación del largo. Para esto se buscó por cada palabra coincidencias con el set de *ids* de la siguiente palabra en la frase hasta el largo del vocabulario, es decir O(TN). Por otra parte esta operación se debe repetir para todos los vocabularios en otras palabras $O(TN^2)$.

3.- Verificación de orden

La verificación del orden también depende del largo del término a verificar. Siendo esta otra tarea de largo N.

Sintetizando podemos describir la complejidad del algoritmo propuesto en el peor caso como $O(TN^2)$, dado que N es un término constante (N=26), por lo que puede simplificar a una complejidad de O(T).

A pesar de la complejidad computacional y la abundancia de iteraciones condicionales la duración total de la búsqueda de términos en las 1973 impresiones tomó sólo *2' 30"* lo que corresponde a un promedio de 76 *ms* por documento.

Evidenciando que el algoritmo podría ser utilizado a futuro para identificación en tiempo real.

7.2.- Comparación entre métodos presentados

7.2.1.- Búsqueda aproximada por texto

El sistema de escritura donde se redactan los informes radiológicos posee corrector ortográfico. Por otro lado, al estar el radiólogo escribiendo un documento legal, revisa minuciosamente la ortografía por lo que no se utilizó corrector por no encontrarse provechoso. De la misma forma no se aproximó tampoco por distancia fonética. SNOMED posee gran cantidad de sinónimos por lo que tampoco se utilizó stemmer o lematizador. Por ejemplo, la diferencia entre “infarto del miocardio” e “infarto miocárdico” ya está considerada en SNOMED con 2 *id* distintos, siendo estos sinónimos entre sí.

7.2.2.- Desventajas del clasificador bayesiano

Una desventaja del modelo bayesiano para lograr el objetivo planteado, es que tenemos múltiples textos cortos en los que el orden de las palabras es crucial, por lo que no podemos perderlos. Es así como un clasificador bayesiano clasificaría de igual forma los informes;

“Sin evidencia de tromboembolismo pulmonar.

Signos de neumonía derecha.”

y

“Signos de tromboemolismo pulmonar.

Sin evidencia de neumonía derecha.”

Existen variaciones del clasificador bayesiano que sí pueden manejar negaciones utilizando búsqueda de n-gramas, por lo que el orden de palabras estaría implícitamente considerado. Esta área no fue investigada en la presente tesis, pero podría ser una aproximación a explorar en el futuro.[71]

A pesar de ser un buen modelo, motivo por el cual ha sido ampliamente utilizado en diversos escenarios, el modelo *naive* posee debilidades evidentes por lo

que se prefirió utilizar otro método que mantuviera la información de la posición de las palabras y su relación.

7.2.3- Desventajas del VSM

Al agrupar las palabras por sus frecuencias, perdemos la posición y la relación entre éstas, volviendo al modelo de conjunto de palabras. Con este modelo, tampoco podremos distinguir entre las dos frases mostradas en la sección anterior.

Como podemos concluir tras la explicación, el VSM se utiliza para la retribución de información de forma renqueada (*Ranked Information Retrival*).

Las ventajas de la utilización de VSM sería poder traer con gran rapidez documentos relevantes a la consulta realizada y ordenados por relevancia. Sin embargo, este sistema de retribución de información no nos permite conocer el lugar ni el orden de las palabras como si lo hacen los sistemas de retribución de información dicotómicos como el utilizado.

Se ha utilizado el VSM con otros fines además de la retribución de información ranqueada como lo es la desambiguación. Al obtener una puntuación (similitud del coseno del ángulo de los espacios vectoriales) podemos discernir entre contextos. Este modelo fue utilizado en la herramienta polyfind[72] para evaluar el contexto en el que se encontraban los acrónimos.

El VSM si bien podría utilizarse adaptándolo para el uso en la herramienta desarrollada no plantea grandes beneficios, pues sólo nos devolvería el documento, pero no el lugar en donde se encuentra el término. Igualmente tendríamos que realizar la búsqueda del término dentro del documento una vez devueltos estos últimos.

7.3.- *Overfiting* versus *tailoring*

El problema del *overfiting* está descrito para los métodos de aprendizaje no supervisado. Al entregarse un conjunto de documentos previamente clasificados, el sistema podrá aprender de ellos. Se inferirán tanto reglas correctas como reglas erróneas, ajustándose éste al máximo a los datos entregados. Al medir la eficacia de la herramienta obtendremos un valor falsamente óptimo, dando la impresión de que

la herramienta clasifica muy bien, sin embargo esto está dado por clasificar los errores de la misma forma que se realizó en los documentos manualmente clasificados.

Cuando se realizan búsquedas ingresando reglas manuales, ocurre el fenómeno de *tailoring*, que podríamos traducir como “entallaje” (hecho a medida). Este fenómeno se caracteriza por el ajuste de la herramienta al contexto de los documentos de los cuales se obtuvieron los datos. Al entregarse las reglas de manera manual se evitan las inferencias erróneas. Al medir la eficacia de la herramienta presenta resultados óptimos, reales para el contexto de los documentos. Hay que tener presente que los resultados no pueden ser extrapolados a otros tipos de documentos. En el caso del presente trabajo, el detector de negación funciona de forma excelente, ya que los radiólogos de Clínica Alemana niegan con gramática y vocabulario muy regular, estableciéndose patrones fáciles de identificar. Si con la misma herramienta analizamos otros documentos médicos, por ejemplo, resúmenes de alta, podemos predecir que el resultado de la herramienta no será similar al obtenido en los informes radiológicos, ya que los médicos que redactan los resúmenes de alta utilizan patrones de expresión distintos a los radiólogos para negar.

Con el fin de evitar el *overfiting* y explotar el *tailoring* para la herramienta desarrollada, se generaron las reglas con los patrones de negación analizando informes del *pool* de 1973 informes, siendo ciego el investigador a los 219 utilizados para medir la herramienta. Hay que tener presente también que tras la obtención de patrones de los documentos, éstos fueron analizados para generalizar reglas a partir de la combinación de patrones, siendo posible encontrar formas de negación gramaticalmente correctas que no estaban presentes de manera textual en los documentos analizados.

8.- CONCLUSIONES

La creación de una herramienta de NLP para la identificación del diagnóstico crítico radiológico es posible, como quedó demostrado en esta tesis.

Existen diversos elementos a aprovechar, como lo son la semi estructuración de los informes (texto libre dividido en secciones fijas) y la utilización de patrones regulares por parte de los radiólogos para expresarse. La búsqueda de patrones regulares permitió un gran manejo del contexto en que se encontraban los términos, con muy pocos errores de clasificación. Se propone ésta como la forma que debe ser explotada a futuro, para generar herramientas de extracción de la información contenida en informes radiológicos, de forma de poder tener previamente analizado el texto y estructurada la información de interés, tanto para la retribución de informes como para la agregación de la información de éstos.

La confidencialidad de datos (datos sensibles), plantea una barrera al desarrollo del área, en el que la presencia de la identidad del paciente en la información explotada es utilizada como el motivo por el cual se dificulta el acceso a dicha información. Cuesta hacer entender a los comités encargados de autorizar el acceso, que los datos serán anonimizados y que sólo se trabajará con el contenido de los informes, teniendo el investigador un total desconocimiento de a quien pertenecen. Es usual que estos comités soliciten un consentimiento informado de cada uno de los pacientes incluidos en el estudios, tarea que, por el volumen, es difícil de lograr, necesitando el investigador para cumplir dicha tarea acceder a las identidades y a datos personales que de lo contrarío no requeriría tener. El número de estudios que explotan grandes volúmenes de datos del registros clínicos electrónicos va en aumento, haciendo cada vez más familiar a los comités la aprobación de este tipo de estudios.

La creación de la herramienta descrita en este documento demandó un esfuerzo de desarrollo mayor al esperado. La aplicación de las herramientas de NLP puede ser tan amplia como podamos imaginar, sin embargo, requieren una gran inversión de tiempo y esfuerzo, el cual se ve acentuado al momento de tener que objetivar la eficacia de lo desarrollado.

9.- BIBLIOGRAFÍA

1. Reiner BI, Knight N and Siegel EL. Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol*. 2007; 4: 313-9.
2. Townsend H. Natural Language Processing and Clinical Outcomes. *Journal of AHIMA*. 2013; 84: 44.
3. Wu AS. Evaluation of Negation and Uncertainty Detection and its Impact on Precision and Recall in Search. *Journal of Digital Imaging*. 2011; 24: 234.
4. Matykiewicz P. Unambiguous concept mapping in radiology reports: graphs of consistent concepts. 2006: 1024.
5. Branch-Elliman W. Natural Language Processing for Real-Time Catheter-Associated Urinary Tract Infection Surveillance: Results of a Pilot Implementation Trial. *Infect Control Hosp Epidemiol*. 2015: 1.
6. Goss FR. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. 2014; 2014: 580-8.
7. Bill R. Automated extraction of family history information from clinical notes. 2014; 2014: 1709-17.
8. Zhang R. Navigating longitudinal clinical notes with an automated method for detecting new information. 2013; 192: 754-8.
9. Wieland MLMDMPH. Tracking Health Disparities Through Natural-Language Processing. *American Journal of Public Health*. 2013; 103: 448.
10. Denny JC. Using natural language processing to provide personalized learning opportunities from trainee clinical notes. *J Biomed Inform*. 2015.
11. Imler TD. Sa1400 Natural Language Processing for Detection of Quality Measures in Endoscopic Retrograde Cholangiopancreatography. *Gastrointestinal Endoscopy*. 2015; 81: AB198.
12. Allones JL. Automated mapping of clinical terms into SNOMED-CT. An application to codify procedures in pathology. 2014; 38: 134.
13. Hou JK. Automated Identification of Surveillance Colonoscopy in Inflammatory Bowel Disease Using Natural Language Processing. *Digestive Diseases and Sciences*. 2013; 58: 936.
14. Singh M. Prioritization of free-text clinical documents: a novel use of a bayesian classifier. 2015; 3: e17.
15. French L. Text mining for neuroanatomy using WhiteText with an updated corpus and a new web application. *Front Neuroinform*. 2015; 9: 13.
16. Mabotuwana T. A context-sensitive image annotation recommendation engine for radiology. 2014; 205: 1143-7.
17. Kahn CE, Jr. Ontology-based diagnostic decision support in radiology. 2014; 205: 78-82.
18. Sippo DA. Automated Extraction of BI-RADS Final Assessment Categories from Radiology Reports with Natural Language Processing. *Journal of Digital Imaging*. 2013; 26: 989.
19. Percha B. Automatic classification of mammography reports by BI-RADS breast tissue composition class. *Journal of the American Medical Informatics Association*. 2012; 19: 913-6.

20. Bozkurt S. Automated detection of ambiguity in BI-RADS assessment categories in mammography reports. 2014; 197: 35-9.
21. Yetisgen-Yildiz M. A text processing pipeline to extract recommendations from radiology reports. 2013; 46: 354-62.
22. Dutta S. Automated detection using natural language processing of radiologists recommendations for additional imaging of incidental findings. *Annals of Emergency Medicine*. 2013; 62: 162-9.
23. Dang PA. Natural language processing using online analytic processing for assessing recommendations in radiology reports. 2008; 5: 197-204.
24. Dang PA. Extraction of recommendation features in radiology with natural language processing: exploratory study. 2008; 191: 313-20.
25. Sevenster M. Automatically pairing measured findings across narrative abdomen CT reports. 2013; 2013: 1262-71.
26. Mabotuwana T. Determining scanned body part from DICOM study description for relevant prior study matching. 2013; 192: 67-71.
27. Do BH. Automatic Retrieval of Bone Fracture Knowledge Using Natural Language Processing. *Journal of Digital Imaging*. 2013; 26: 709.
28. Zopf J. Development of Automated Detection of Radiology Reports Citing Adrenal Findings. *Journal of Digital Imaging*. 2012; 25: 43-9.
29. Xu Y. Named entity recognition of follow-up and time information in 20 000 radiology reports. *Journal of the American Medical Informatics Association*. 2012; 19: 792-9.
30. Sevenster M. Automatically Correlating Clinical Findings and Body Locations in Radiology Reports Using MedLEE. *Journal of Digital Imaging*. 2012; 25: 240.
31. Roberts K. A machine learning approach for identifying anatomical locations of actionable findings in radiology reports. 2012; 2012: 779-88.
32. Huang Y. A novel hybrid approach to automated negation detection in clinical radiology reports. 2007; 14: 304-11.
33. Mendonça EA. Extracting information on pneumonia in infants using natural language processing of radiology reports. 2005; 38: 314-21.
34. Mamlin BW. Automated extraction and normalization of findings from cancer-related free-text radiology reports. 2003: 420-4.
35. Imai T. Extracting diagnosis from Japanese radiological report. 2003: 873.
36. Hersh W. Selective automated indexing of findings and diagnoses in radiology reports. 2001; 34: 262-73.
37. Farkas R. Automatic construction of rule-based ICD-9-CM coding systems. 2008; 9: S10.
38. Zhang S. Automatic image annotation and retrieval using group sparsity. 2012; 42: 838-49.
39. Gerstmair A. Intelligent image retrieval based on radiology reports. *European Radiology*. 2012; 22: 2750.
40. Gershanik EF. Critical finding capture in the impression section of radiology reports. 2011; 2011: 465-9.
41. Cheng LTE. Discerning Tumor Status from Unstructured MRI Reports--Completeness of Information in Existing Reports and Utility of Automated Natural Language Processing. *Journal of Digital Imaging*. 2010; 23: 119.
42. Kahn CE, Jr. GoldMiner: a radiology image search engine. 2007; 188: 1475-8.

43. Anthony SG, Prevedello LM, Damiano MM, et al. Impact of a 4-year quality improvement initiative to improve communication of critical imaging test results. *Radiology*. 2011; 259: 802-7.
44. Friedman C. A general natural-language text processor for clinical radiology. 1994; 1: 161-74.
45. Lacson R. Practical examples of natural language processing in radiology. 2011; 8: 872-4.
46. Lacson R. Natural language processing for radiology (part 2). 2011; 8: 583-4.
47. Organization WH. *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. World Health Organization, 1993.
48. Medicine NLo. *UMLS Knowledge Sources: Metathesaurus, Semantic Network, [and] SPECIALIST Lexicon*. U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine, 2003.
49. WONCA. CIdCdl. *Clasificación Internacional de la Atención Primaria segunda edición. CIAP - 2.* . 1999
50. Langlotz CP. RadLex: A New Method for Indexing Online Educational Materials. *RadioGraphics*. 2006; 26: 1595-7.
51. Manning CD, Raghavan P and Schèutze H. *An introduction to information retrieval*. Cambridge: Cambridge University Press, 2008, p.xxi, 482 p.
52. Wagner RA and Fischer MJ. The String-to-String Correction Problem. *J ACM*. 1974; 21: 168-73.
53. Needleman SB and Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970; 48: 443-53.
54. Philips L. Hanging on the Metaphone. *Computer Language*. 1990; 7: 39.
55. Bird S, Klein E and Loper E. *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009, p.512.
56. Jurafsky D and Martin JH. *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Pearson international ed. Upper Saddle River, N.J.
London: Pearson/Prentice Hall ;
Pearson Education, 2009, p.1024 p.
57. Lundberg GD. Critical (panic) value notification: an established laboratory practice policy (parameter). *JAMA*. 1990; 263: 709.
58. Álvarez ME, Valdés J and Jiménez L. Notificación de Valores o Resultados Críticos Recomendaciones Generales Laboratorio Clínico, Anatomía Patológica e Imagenología: (2013, accessed 29/05/2016).
59. Bates DW and Leape LL. Doing better with critical test results. *Jt Comm J Qual Patient Saf*. 2005; 31: 66-7, 1.
60. Kuperman GJ, Teich JM, Tanasijevic MJ, et al. Improving response to critical laboratory results with automation: results of a randomized controlled trial. *J Am Med Inform Assoc*. 1999; 6: 512-22.
61. Khorasani R. Optimizing communication of critical test results. *J Am Coll Radiol*. 2009; 6: 721-3.
62. Hanna D, Griswold P, Leape LL and Bates DW. Communicating critical test results: safe practice recommendations. *Jt Comm J Qual Patient Saf*. 2005; 31: 68-80.
63. Radiologia SCd. Nota Técnica N°9. Notificación de Resultados Críticos en Imagenología. 2015.

64. Salud Sd. Maual de Estándar general de Acreditación para prestadores Institucionales de Atención Cerrada. 2009.
65. Awais M, Hilal K, Waheed A, Khattak YJ, Rehman A and Ul-Ain Baloch N. Detection and Communication of Critical Findings Noted on Thoracic CT Scans by Radiology Residents. *J Am Coll Radiol*. 2015; 12: 1324-9.
66. Barritt DW and Jordan SC. Anticoagulant drugs in the treatment of pulmonary embolism. A controlled trial. *Lancet*. 1960; 1: 1309-12.
67. Silva C. Procesamiento de Lenguaje Natural del Informe Radiológico como test diagnóstico para pesquisa de Tromboembolismos Pulmonares en Clínica Alemana de Santiago durante el 2013-2014. Clinica Alemana de Santiago, 2016.
68. Chapman WW, Bridewell W, Hanbury P, Cooper GF and Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*. 2001; 34: 301-10.
69. Harkema H, Dowling JN, Thornblade T and Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *J Biomed Inform*. 2009; 42: 839-51.
70. Costumero R, Lopez F, Gonzalo-Martín C, Millan M and Menasalvas E. An Approach to Detect Negation on Medical Documents in Spanish. In: Ślęzak D, Tan A-H, Peters JF and Schwabe L, (eds.). *Brain Informatics and Health: International Conference, BIH 2014, Warsaw, Poland, August 11-14, 2014, Proceedings*. Cham: Springer International Publishing, 2014, p. 366-75.
71. Narayanan V, Arora I and Bhatia A. Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model. In: Yin H, Tang K, Gao Y, et al., (eds.). *Intelligent Data Engineering and Automated Learning – IDEAL 2013: 14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, p. 194-201.
72. Pustejovsky J, Castano J, Cochran B, Kotecki M, Morrell M and Rumshisky A. Extraction and Disambiguation of Acronym-Meaning Pairs in Medline. 2004.