

Recuperación de Información y Recomendaciones en la Web

Curso 2015

Grupo 17

Lourdes Cairelli 4.639.504-5

Francisco Paroli 4.753.388-4

Adrian Gomez 4.190.901-3

Indice

- [Indice](#)
- [Introducción](#)
- [Propuesta de la Solución](#)
 - [Diseño y Arquitectura](#)
 - [Herramientas utilizadas](#)
- [Estructura de la base de datos](#)
- [Implementación](#)
 - [Herramientas](#)
 - [fuzzy_match](#)
 - [ruby-stemmer](#)
 - [stopwords](#)
 - [Algoritmo](#)
- [Interfaz de usuario](#)
- [Pruebas realizadas](#)
 - [Medidas de Efectividad](#)
 - [Comentarios sobre las pruebas](#)
- [Problemas encontrados](#)
- [Conclusión](#)
- [Trabajo a futuro](#)
- [Referencias](#)

Introducción

Pediasis es un sistema de historias clínicas electrónica, brinda información sobre cada internación del paciente y un registro de dicha internación.

Nuestro objetivo de aplicar recuperación de datos en la web se basa en que nuestra aplicación pueda sugerir tres diagnósticos cuando el médico está ingresando el resumen de ingreso del paciente, basándonos en los diagnósticos ingresados en resúmenes de ingreso similares. Esto podría ayudar al médico a tener en cuenta algún diagnóstico que no se le ocurrió en el momento o simplemente le sirve de guía de lo que pusieron otros colegas o él mismo en otra oportunidad.

Propuesta de la Solución

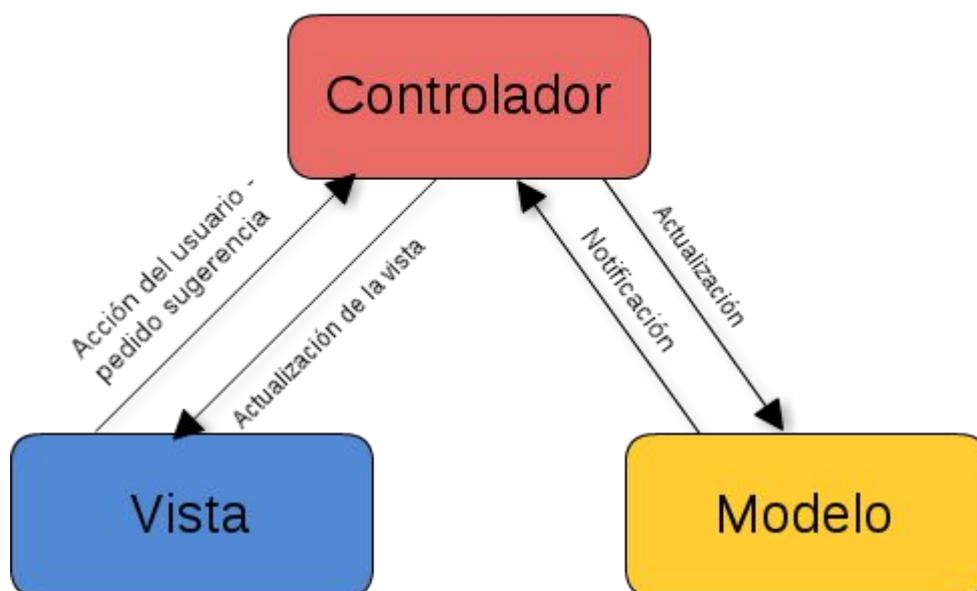
El problema planteado tal como comentamos arriba es sugerir al médico, cuando está ingresando el resumen de ingreso del paciente posibles diagnósticos.

Nuestra solución plantea una interfaz de usuario donde el médico ingresa la nueva internación y cuando éste ingrese el resumen de ingreso para la dicha internación el sistema le sugiere posibles diagnósticos. Los diagnósticos sugeridos corresponden a internaciones cuyos resúmenes de ingreso tienen coincidencias con el resumen de ingreso que se está ingresando en la nueva internación.

Nuestro sistema es una aplicación web realizada con ruby on rails, basada en arquitectura de sistema MVC. Planteamos mantener nuestra arquitectura agregando la funcionalidad de sugerencia, para implementar dicha funcionalidad en el sistema vamos a utilizar gemas de rails y también vamos a extender nuestra base de datos agregando nuevas tablas que necesitamos para esto.

Diseño y Arquitectura

La arquitectura de nuestro sistema es arquitectura MVC (modelo, vista controlador) como ya mencionamos anteriormente



Herramientas utilizadas

Ruby
Ruby on Rails
FuzzyMatch
Ruby-Stemmer
Stopwords

Estructura de la base de datos

La estructura de la base de datos para soportar este algoritmo de búsqueda involucra dos tablas:

- **diagnostic_keyboards**: guarda la relación entre diagnóstico y la keyboard con la que está relacionado.
- **keyboards**: guarda todas las keyboard relevantes que se encontraron en los resúmenes de ingreso.

Implementación

Herramientas

Usamos 3 herramientas como base para el algoritmo de reconocimiento: fuzzy_match, ruby-stemmer y stopwords.

fuzzy_match

1. Utiliza una combinación de par distancia (2-gram) y distancia de Levenshtein. Al obtener la lista de resultados, en caso de coincidencia en par distancia, desempata utilizando Levenshtein.
2. Dada una lista de objetos asociada a cadenas de caracteres y una cadena de caracteres, retorna el objeto que más se aproxima.
3. Sin configuración se puede obtener un 90% de precisión, pero si se quiere más se pueden agregar expresiones regulares, agrupaciones y stop words.

ruby-stemmer

Esta librería se basa en determinadas reglas y algoritmos que llevan cada palabra a su expresión mínima.

stopwords

A esta librería se le pasa una lista de palabras y retorna la lista sin las palabras elegidas por nosotros para que se filtren. En nuestro caso en vez de seleccionar de a una palabra utilizamos la opción que nos da la librería de seleccionar el idioma español.

Algoritmo

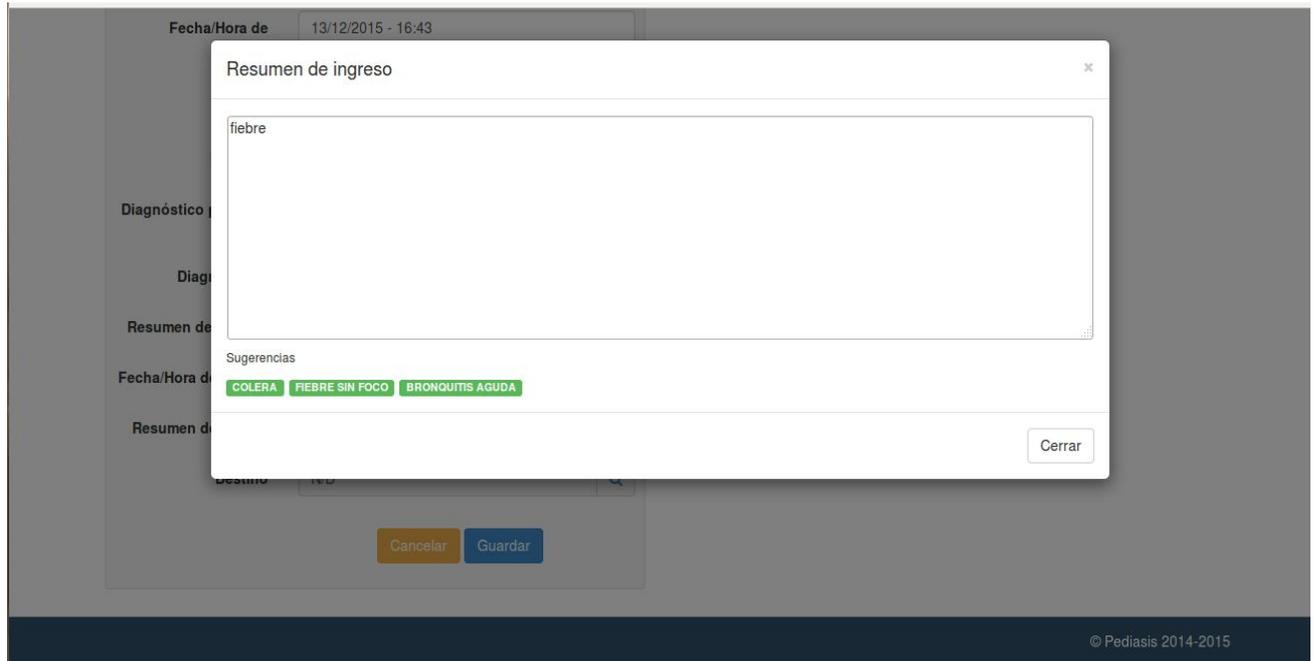
El algoritmo de recomendación se puede dividir en 2 partes, primero se extraen las palabras asociadas a un diagnóstico cuando el médico hace el ingreso de una internación. Vamos a ir guardando a medida que se ingresan diagnósticos las palabras asociadas a este diagnóstico en la tabla `diagnostic_keyboards` para hacer más eficiente el algoritmo y no buscar todo en memoria. Esto es impracticable en un sistema grande.

Luego se utilizan estas palabras para hacer las comparaciones y sugerencias, las mismas no son case-sensitive.

Para hacer las recomendaciones nos vamos a quedar con las coincidencias obtenidas con el algoritmo de par distancia con una similitud mayor a 0,2 (la similitud tiene valores entre 0 y 1). Decidimos filtrar por este valor porque vimos que los que tenían peso mayor a 0,2 eran los que más se acercaban a lo buscado. En caso de que queden más de 5 resultados con la misma similitud, mostramos solo los primeros 5.

Tanto para extraer las palabras de una internación ingresada como para obtener las posibles sugerencias al texto ingresado por el médico se le aplica el mismo tratamiento. Primero utilizamos stopwords para eliminar las palabras que no vale la pena considerar. Luego utilizamos ruby-stemmer para llevar las palabras a su mínima expresión. Y por generamos una cadena de caracteres con todas las palabras resultantes. Esta cadena de caracteres es la que indica a los diagnósticos o la que utilizamos a la hora de realizar una búsqueda.

Interfaz de usuario



Pruebas realizadas

Para realizar estas pruebas tomamos resúmenes de ingreso existentes y comparamos el diagnóstico realizado por el médico con los diagnósticos sugeridos por el sistema.

Prueba 1

Resumen de ingreso: Bronquiolitis con insuficiencia

Diagnóstico sugerido:

1. BRONQUITIS AGUDA
2. BRONQUIOLITIS AGUDA DEBIDA A VIRUS SINCICIAL RESPIRATORIO
3. CELULITIS DE OTRAS PARTES DE LOS MIEMBROS
4. INFECCION INTESTINAL VIRAL, SIN OTRA ESPECIFICACION
5. CELULITIS DE LA CARA

Diagnóstico del médico: BRONQUIOLITIS VIRUS RESP.SINCICIAL

Prueba 2

Resumen de ingreso: Crisis asmatica moderada prolongada con IR.

ANTECEDENTES: Preescolar de 2 años

Diagnóstico sugerido:

1. ASMA
2. NEUMONIA VIRAL, NO ESPECIFICADA
3. CELULITIS DE OTRAS PARTES DE LOS MIEMBROS
4. NEUMONIA BACTERIANA, NO ESPECIFICADA
5. NEFRITIS TUBULOINTERSTICIAL AGUDA

Diagnóstico del médico: ASMA

Prueba 3

Resumen de ingreso: CBO moderada con insuficiencia respiratoria

Diagnóstico sugerido:

1. OTRAS ENFERMEDADES PULMONARES OBSTRUCTIVAS CRONICAS ESPECIFICADAS
2. BRONQUIOLITIS AGUDA DEBIDA A VIRUS SINCITAL RESPIRATORIO
3. NEUMONIA VIRAL, NO ESPECIFICADA
4. NEUMONIA DEBIDA A OTROS VIRUS
5. INFECCION DE VIAS URINARIAS, SITIO NO ESPECIFICADO

Diagnóstico del médico: CRISIS BRONCOOBSTRUCTIVAS REITERADA

Prueba 4

Resumen de ingreso: DAI grado III

Diagnóstico sugerido: Como ninguna sugerencia tuvo similitud mayor a 0.3 no se mostraron resultados

Diagnóstico del médico: DIARREA AGUDA INFECCIOSA

Prueba 5

Resumen de ingreso: diarrea aguda infantil grado I

Diagnóstico sugerido:

1. ENTERITIS DEBIDA A ROTAVIRUS
2. ABSCESO CUTANEO, FURUNCULO Y CARBUNCO
3. INFECCION DE VIAS URINARIAS, SITIO NO ESPECIFICADO
4. NEUMONIA VIRAL, NO ESPECIFICADA
5. CONVULSIONES FEBRILES

Diagnóstico del médico: DIARREA A ROTAVIRUS

Prueba 6

Resumen de ingreso: Neumonía viral con coinfección

Diagnóstico sugerido:

1. NEUMONIA DEBIDA A OTROS VIRUS
2. INFECCION DE VIAS URINARIAS, SITIO NO ESPECIFICADO
3. NEUMONIA VIRAL, NO ESPECIFICADA
4. CONVULSIONES FEBRILES

5. NEFRITIS TUBULOINTERSTICIAL AGUDA

Diagnóstico del médico: NEUMONIA A OTROS

Prueba 7

Resumen de ingreso: NAC basal izquierda

Diagnóstico sugerido: Como ninguna sugerencia tuvo similitud mayor a 0.3 no se mostraron resultados

Diagnóstico del médico: NEUMONIA BACTERIANA NO ESPECIFICADA

Prueba 8

Resumen de ingreso: Gastroenterocolitis sin deshidratación. Alta tasa de diarrea

Diagnóstico sugerido:

1. INFECCION INTESTINAL VIRAL, SIN OTRA ESPECIFICACION
2. CONVULSIONES FEBRILES
3. NEFRITIS TUBULOINTERSTICIAL AGUDA
4. GASTRITIS, NO ESPECIFICADA
5. CELULITIS DEL OIDO EXTERNO

Diagnóstico del médico: DIARREA VIRAL NO ESPECIFICADA

Prueba 9

Resumen de ingreso:

ANTECEDENTES PERSONALES

Primera Infancia 1a 8m.

Sin APP a destacar. Crecimiento en P3-15%. CEV Vigente

Herniorrafia bilateral a los 2 meses de vida

Hiperreactividad Bronquial sin tto preventivo

AF: Asma de primera línea. Tabaquistas intradomiciliarios

ASEC: Deficitario

MOTIVO DE INGRESO: EGRESO DE UCIN CBO CON IR

ENFERMEDAD ACTUAL

Ingresa hace 48hrs por cuadro de corta evolución de tos perruna con fatiga, sin fiebre TAL 9 mantenido enviado desde Hospital.

No requirió AVM persiste en CIPAP por 30 Hrs con buena evolución, no presentó complicaciones respiratorias ni hemodinámicas.

Aminofilina por 30hrs.

RxTX torax hiperinsuflado. sin foco de consolidación

Hemograma: Hb 10.3 Htc 31.2% GB: 9.100 Neu: 65% Plq: 266.000 PCR: 18.2 PCT: 0.15

Hemocultivo Pendiente

Al ingreso a Sala: paciente VEA sin SFR sin requerimientos de O2, en apirexia con buen estado general.

Tolerando vo. TU y TD normal.

PP: No SDR FR 40cpm MAV + bilateral subcrepitantes difusos

CV: RR 100cpm RBG pulsos positivos bilateral

ABD: Blando depresible indoloro sin defensa

SNM: no sg meningeos no sg focal neurologico

Vista en Neumologia indica inicio de tto Fluticasona 125 c/12hrs, Montelukast 5mg dia

Solicita Hemograma IgE, alfa 1 antitripsina test sudor, dosificacion de Inmunoglobulinas subclase IgG

En SUMA: 1a 8m

Asma del Lactante

CBO C/IR

Diagnóstico sugerido:

1. BRONQUIOLITIS AGUDA, NO ESPECIFICADA
2. CEFALEA
3. ASMA
4. BRONQUIOLITIS AGUDA DEBIDA A VIRUS SINCITIAL RESPIRATORIO

Diagnóstico del médico: ASMA

Prueba 10

Resumen de ingreso:

DATOS ANTROPOMETRICOS:

Peso ING.: 12,700

Talla ING.: 95cm

MOTIVO DE INGRESO: diarrea aguda infantil grado I

ANTECEDENTES: Preescolar 2A 10M procedente de Montevideo. AP: CBO reiterados en el primer año

Diagnóstico sugerido:

1. BRONQUIOLITIS AGUDA, NO ESPECIFICADA
2. ASMA
3. OTRAS ENFERMEDADES PULMONARES OBSTRUCTIVAS CRONICAS ESPECIFICADAS
4. CELULITIS DE OTRAS PARTES DE LOS MIEMBROS
5. NEFRITIS TUBULOINTERSTICIAL AGUDA

Diagnóstico del médico: DIARREA A ROTAVIRUS

Para las siguientes pruebas ingresamos 2 pacientes a los que les inventamos un resumen de ingreso y un diagnóstico principal.

Resumen 1: Paciente ingresa con dolor abdominal , fiebre alta , náuseas y vómitos

Diagnóstico 1: DIARREA Y GASTROENTERITIS DE PRESUNTO ORIGEN INFECCIOSO

Resumen 2: Dolor de cabeza y fiebre. Presenta la garganta roja con una manchas amarillas

Diagnóstico 2: DOLOR DE GARGANTA

Prueba 11

Resumen de ingreso: Paciente ingresa con dolor abdominal , fiebre alta , náuseas y vómitos

Diagnóstico sugerido:

1. DIARREA Y GASTROENTERITIS DE PRESUNTO ORIGEN INFECCIOSO
2. GASTRITIS CRONICA, NO ESPECIFICADA
3. DOLOR DE GARGANTA
4. CONVULSIONES FEBRILES
5. NEFRITIS TUBULOINTERSTICIAL AGUDA

Prueba 12

Resumen de ingreso: Dolor de cabeza y fiebre. Presenta la garganta roja

Diagnóstico sugerido:

1. DOLOR DE GARGANTA
2. DIARREA Y GASTROENTERITIS DE PRESUNTO ORIGEN INFECCIOSO
3. BRONQUIOLITIS AGUDA DEBIDA A VIRUS SINCICIAL RESPIRATORIO
4. CELULITIS DE OTRAS PARTES DE LOS MIEMBROS
5. GASTRITIS CRONICA, NO ESPECIFICADA

Prueba 13

Resumen de ingreso: Malestar estomacal y vista nublada

Diagnóstico sugerido: Sin sugerencias

Prueba 14

Resumen de ingreso: fiebre slts con dolor de farganta

Diagnóstico sugerido:

1. DOLOR DE GARGANTA
2. DIARREA Y GASTROENTERITIS DE PRESUNTO ORIGEN INFECCIOSO

Medidas de Efectividad

Precisión

$P = \text{Documentos relevantes recuperados} / \text{Documentos recuperados}$

$P = 26/51$

Recall

$R = \text{Documentos relevantes recuperados} / \text{Documentos relevantes}$

Para realizar estos cálculos tuvimos que decidir nosotros cuáles de las sugerencias eran correctas y cuáles no. En el caso del recall, cómo estamos utilizando toda la base de diagnósticos, necesitaríamos a un médico que nos indique cuales de estos serían relevantes dependiendo del resumen, por lo tanto no podemos realizar esta medida.

Comentarios sobre las pruebas

Como se puede ver en las pruebas, dependiendo de cómo se escriba el resumen pueden obtenerse resultados diferentes, por ejemplo en las pruebas 4 y 5 los resúmenes eran casi iguales, pero como el primer médico escribió diarrea aguda infantil como DAI y no teníamos otro resumen de ingreso con esa sigla, el sistema no le dio ninguna sugerencia.

El resumen de la prueba 9 fue tomado de otro hospital que los de las anteriores y se puede ver que por más que pongan muchos más datos, las sugerencias son correctas.

En la prueba 14 realizada con un resumen ingresado por nosotros se puede ver cómo por más que se ingresen mal las palabras el resultado de la búsqueda es correcto.

Problemas encontrados

No todos los médicos escriben de la misma forma los resúmenes de ingreso, por lo tanto para un mismo diagnóstico y los mismos síntomas del paciente podemos tener 2 resúmenes de ingreso completamente distintos.

Estas diferencias también se dan dependiendo del hospital. Hay en algunos que solo ingresan los síntomas del paciente y datos como peso y talla. Mientras que en otros también se ingresa una descripción de la familia del paciente, barrio en el que vive y otros datos.

Otro de los problemas encontrados fue que al trabajar con el CTI pediátrico muchos de los diagnósticos principales son "BIEN NUTRIDO", por lo tanto tuvimos que filtrar estos casos y no tomarlos en cuenta.

Un problema que tuvimos por tratarse de un ámbito académico fue la demora en procesar los resúmenes de ingreso para sacar los datos de los médicos y de los pacientes para conservar el anonimato.

Conclusión

No es problema fácil de resolver ya que depende mucho de la forma en que los usuarios interactúan con el sistema, pero creemos que la solución lograda se acerca mucho y es una base sólida para continuar expandiendo.

Consideramos que este sistema puede ser muy útil para el médico ya que por lo que pudimos observar en las pruebas, aunque los resúmenes fueran distintos muchas veces nos sugirió el diagnóstico elegido por el médico o uno similar.

Para lograr tener sugerencias más exactas tendría que haber un protocolo de ingreso de los resúmenes, así no habría que lidiar con las distintas formas de escribir de los médicos ni con los datos elegidos por cada hospital.

De todas formas nos parece que con una base de datos con una cantidad suficientemente grande de internaciones el sistema va a funcionar mejor aunque en ese caso podríamos encontrarnos con limitaciones desde el punto de vista de performance, este problema podría ser mitigado utilizando otras herramientas para el almacenamiento y procesamiento de las internaciones.

Trabajo a futuro

La base de datos que usamos de prueba fue obtenida con datos de cti, como trabajo a futuro también nos planteamos conseguir otros datos de otras internaciones y ver cómo se comporta nuestro sistema en ese caso.

Tener una visión de un médico para mejorar nuestro algoritmo de sugerencias sería sumamente productivo.

Diseñar con la ayuda de un médico una forma de ingresar los resúmenes para que no sea texto libre.

Referencias

https://github.com/seamusabshere/fuzzy_match

<https://github.com/brenes/stopwords-filter>

<https://github.com/aurelian/ruby-stemmer>