Universidad del País Vasco Euskal Herriko Unibertsitatea
BILBOKO INGENIARITZA ESKOLA
ESCUELA DE INGENIERÍA DE BILBAO

GRADO EN INFORMÁTICA DE GESTIÓN Y SISTEMAS DE INFORMACIÓN

TRABAJO FIN DE GRADO

IDENTIFICADOR AUTOMÁTICO DE RELACIONES TEMPORALES EN TEXTOS CLÍNICOS BASADO EN REDES NEURONALES

Alumno/Alumna: Andrés, Santamaría, Edgar
Director/Directora (1): Atutxa, Salazar, Aitziber
Director/Directora (2): Casillas, Rubio, Arantza

Curso: 2018-2019

Fecha: Bilbao, 19, Julio, 2019

Índice

Índice de figuras

Índice de tablas

Directoras: Aitziber Atutxa y Arantza Casillas

Grupo de investigación: IXA (http://ixa.si.ehu.eus)

1. Introducción

Los datos son ancestrales, hace aproximadamente 5000 años se escribieron los primeros registros contables en las ciudades Ur y Uruk, estas pertenecían a la antigua Mesopotamia, el lenguaje utilizado era simple y pretendía que la información de las cosechas persistiese. Según la teoría de la evolución, existe una necesidad de competir y colaborar entre especies para garantizar la supervivencia, la humanidad, por tanto, escaló puestos con la aparición de la escritura, esto es debido a la adquirida capacidad de memoria persistente, a partir de entonces se concibe la historia como motor de la Historia. Posteriormente se lograron avances del lenguaje escrito, aparecieron idiomas complejos y ricos, estos eran capaces de sintetizar las opiniones de los eruditos, a partir de entonces, se podría registrar información e incluso conocimiento puesto que se registrarían procesos cognitivos complejos. Llegados a este punto, el legado de conocimiento a las sucesivas generaciones se convierte en el principal recurso para el progreso, " El conocimiento es poder. "(**Francis Bacon** 1597: Meditationes Sacrae). [1], pero, el aprendizaje necesario para descifrar el conocimiento en muchos casos conlleva años de estudio, en el presente estudio afrontamos el reto de extraer conocimiento de información clínica mediante una máquina automáticamente.

La información clínica relevante acerca de cada paciente se compendia en su historia clínica, esta información trata: antecedentes, enfermedad actual, exploración física, pruebas y tratamientos. Constituye, por tanto, el histórico de información relevante y conocida sobre el paciente para el pronóstico clínico. Dado que la mayoría de procesos clínicos se encuentran estandarizados, es crucial el apropiado tratamiento de esta información para conocer : evolución del paciente, eficacia del tratamiento y posibles ensayos a realizar. El presente estudio pretende aportar soluciones Software enfocadas al aprendizaje máquina capaz de sintetizar la evolución de los pacientes, para esto, investigamos sobre el conjunto de historias clínicas provistas en el reto internacional i2b2 [25] [13].

[1] filósofo, político, abogado y escritor 1561 - 1626

Por otro lado se pretende acercar al alumno al mundo de la investigación, en este caso trabajando dentro del grupo IXA, un grupo con una larga trayectoria en aplicación de técnicas de procesamiento de lenguaje natural ' Natural language proccesing' (NLP), http://ixa.si.ehu.eus, el gobierno vasco lo clasificó como equipo de investigación tipo B (IT935-16, 2016). En este TFG se aplicarán técnicas de NLP sobre historias clínicas, para diseñar un sistema de extracción de conocimiento clínico, posteriormente se generan representaciones para comprensión humana de los resultados, este proceso se basa en la teoría del conocimiento [12, Cap 1.2, pág 7] . Las técnicas de extracción consistirán en la aplicación de técnicas NLP y de aprendizaje máquina 'machine learning' (ML) proporcionadas por las librerías: NLTK [5], Python3 [7, Part 3], Tensoflow [1], keras [9] y Gemsim [20]. En el presente trabajo se explicará como se han utilizado dichas librerías con el fin de apoyar futuros trabajos técnicos en el campo NLP.

A continuación se explica la aplicación de técnicas NLP a través de un ejemplo sencillo 1, en este caso se pretende modelizar matemáticamente tres oraciones simples y sus respectivas categorías ($texto \rightarrow clase$), consiste en una aplicación sencilla de análisis de sentimientos 'Sentiment analysis'. Primeramente se genera un espacio de representación común a todas las palabras del conjunto para establecer una correlación numérica ($palabra \rightarrow A(word)$), además se define el tipo de dato objetivo (clase) en este caso binario, este proceso permite la representación ($A(word) \rightarrow Bin(clase)$), por último se simplifica la representación para permitir su comprensión humana ($(x, y) \rightarrow Bin(clase)$). La representación matemática modelada en el presente estudio se logrará mediante la aplicación de redes neuronales 'Neural network' (NN), los llamados vectores de palabra ' word embbedings ', estas generarán un espacio de representación común para la interpretación de la información relativa al lenguaje, concretamente trataremos las técnicas: TFIDF, Word2vec y Doc2vec.

Una vez se ha modelado matemáticamente el problema, en la figura 2 podemos entender de manera sencilla en que consiste el entrenamiento y predicción siguiendo el ejemplo de aplicación NLP, primeramente advertir que el conjunto de datos provisto es ínfimo y no necesariamente representativo, exponemos la información preprocesada ($(x, y) \rightarrow Bin(clase)$) en un eje cartesiano, posteriormente aplicamos el algoritmo 'OneR' que consiste en predecir la clase mayoritaria, por tanto aplicamos un enfoque frecuentista simple, en este caso de clasificación binaria $P(a) = Nfavorables/Ntotales$ y $P(b) = 1 - P(a)$, esto es

debido a que ambas clases son mutuamente excluyentes $P(a \cap b) = 0$, es decir, no es ninguna oración positiva y negativa al mismo tiempo. Aplicados los cálculos obtenemos $P(a) = 2/3$ y $P(b) = 1/3$, por tanto la clase estimada es $\hat{y} = 0$, finalmente se estima la calidad del modelo ML, en este caso se utilizará la totalidad del conjunto de entrenamiento, esto se denomina evaluación no honesta, y según la cual se acierta el 60 % de los casos, este porcentaje representa la cota superior de acierto, a efectos prácticos se espera un desempeño mucho menor en ejemplos todavía desconocidos. El presente artículo tratará modelos de predicción más complejos entrando en detalles de su realización y posterior afinamiento para permitir al lector realizar sus propios acercamientos, concretamente trataremos el modelo BiLSTM y CRF.

Los sistemas propuestos buscarán seguir conceptos de arquitecturas unificadas de NNs [11], es decir, se pretende que tanto la aplicación PLN como ML se realicen mediante NNs especializadas, en este estudio la información clínica relevante es: evento clínico, expresión temporal y sección temporal (Event, Timex3 y SecTime), además se trata su relación temporal con respecto a la fecha de creación de Documento 'DocTime 0'. Los resultados se visualizan para que el conocimiento logrado por el proceso lo interprete un experto humano, en este caso un médico, el sistema unificado descrito coincide con el prototipo 3.

Se profundizará en las siguientes áreas:

- Aprendizaje profundo 'deep learning' (DL), para la aplicación de técnicas ML .
- Álgebra, métodos numéricos y cálculo.
- Evaluación de la información, extracción de información y secuenciación de información.
- Aplicación de técnicas NLP para el preprocesado de datos.
- Desarrollo de módulos en Python y utilización de librerías NLP y ML.
- XML (eXtended Markup Language) y XLSX (hojas de cálculo), se trata el almacenamiento de la información a través de estos estándares.
- Visualización de información mediante interfaces de usuario GUI, así como visualización de datos.

Conjunto de datos en formato (texto → clase)

el perro esta libre → positivo

el perro esta enfermo → negativo

el perro esta triste → negativo

Diccionario en formato (palabra → A(word))

el → 0

perro → 1

esta → 2

libre → 3

enfermo → 4

triste → 5

Conjunto de datos en formato (A(word) → Bin(clase))

0 1 2 3 → 1

0 1 2 4 → 0

0 1 2 5 → 0

Simplificación de la representación ((x,y) → Bin(clase)):

1 3 → 1

1 4 → 0

1 5 → 0

Figura 1: Ejemplo preprocesado (NLP)

Tabla con la información obtenida ((x,y) → Bin(clase)):

0				
0				
1				

OneR (ML)

datos para enfoque frecuentista

a = 0

b = 1

P(a) = N.º casos favorables / N.º casos totales ↔ P(a ∩ b) = 0

P(b) = 1 – P(a)

P(a) = ⅔

P(b) = ⅓

Resultado y comprensión:

y = 0 y según los datos acertará el 60% de las veces según casos conocidos.

Figura 2: Ejemplo entrenamiento y predicción (ML)

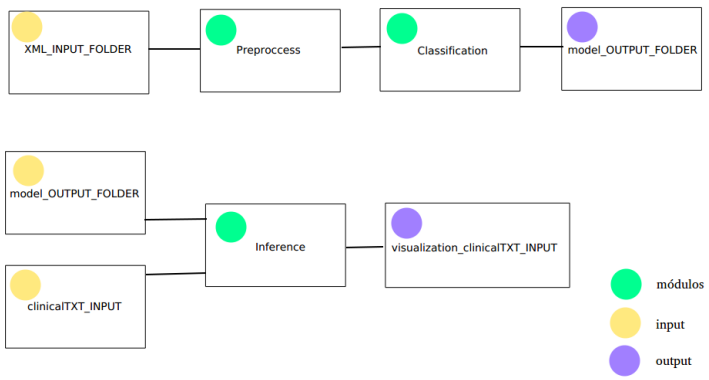

Figura 3: Prototipo para aplicación NLP + ML y predicción.

1.1. Origen del proyecto

El proyecto fue propuesto por el grupo de investigación IXA, asimismo Aitziber Atutxa y Arantza Casillas fueron designadas para guiarlo hacia los mejores resultados posibles, se fundamenta según los siguientes principios:

- Innovación, el análisis de datos 'data minning' constituye una pseudo ciencia en desarrollo por lo tanto los acercamientos a este área podrían conllevar nuevas tecnologías o nuevas aplicaciones de las existentes, esto es debido al constante avance tecnológico y la alta capacidad de cómputo de las máquinas actuales, por otro lado como mencionaba previamente al no ser un área completamente estudiada quedan aún muchos dominios pioneros de aplicación, entre ellos el procesamiento de lenguaje natural (PLN).

  Además, existe una necesidad generalizada de sistemas expertos capaces de interpretar lenguaje natural de manera automática, estos intrínsecamente son complejos, laboriosos e innovadores por ello actualmente se conciben como campo de investigación, el grupo IXA concretamente los desarrolla desde las técnicas PLN.

- Técnicas multidisciplinares, el procesamiento de lenguaje natural es aplicable de manera indiferente a todos aquellos sectores que hayan alcanzado la era digital, y también permite establecer las bases informacionales para aquellos en proceso de digitalización.

- Salidas Profesionales, la salida profesional que defiende este estudio es analista de datos 'data scientist', ya que se espera que el constante desarrollo tecnológico demande este perfil profesional en los próximos años, esta previsión se fundamenta en el fenómeno 'Google' sucedido en EE.UU el cual marco un antes y un después en la aplicación de estas tecnologías mostrando a nivel global sus beneficios y repercusiones.

- El sector clínico, es uno de los prioritarios en cualquier país moderno por tanto los esfuerzos en este área se justifican y se estiman beneficiosos a nivel social y desde el punto de vista económico como potencial nicho de mercado

  Existen numerosos acercamientos a este sector mediante técnicas PLN, estas son aplicadas sobre datos clínicos provistos por retos internacionales como 'i2b2' [25] [13], posteriormente en el estado del arte se detallarán varios de estos acercamientos para entender la repercusión del constante cambio tecnológico en las técnicas PLN y los hallazgos logrados, además se mostrarán las ideas relevantes utilizadas para este estudio.

Estos principios resultan en un trabajo de 'calidad' puesto que se estima que satisface sus necesidades sociales de avance tecnológico conociendo métodos de PLN aplicables, asimismo actualiza el estado del arte sobre aplicación de técnicas PLN así como resultados empíricos de su aplicación ayudando a concretar aún más esta pseudo ciencia del análisis de datos 'data minning'.

1.2. Motivaciones para la elección del proyecto

La elección de un proyecto de esta índole y categoría es debido a los Valores personales del investigador, y además al reconocimiento que el mismo aspira a lograr.

En cuanto a Valores, destacaremos el 'socialismo' como valor, por lo que se pretende devolver a la sociedad lo que en el investigador invirtió, y además lograr su emancipación como 'hombre ilustrado'. Por otra parte citamos el 'derecho a la vida' como moral, por lo que se pretende facilitar el trabajo al conjunto de médicos para garantizar su mejor, y más rápido desempeño. Citamos la 'perseverancia' y el 'esfuerzo' como medios para lograr una vida sostenible, y en busca de la felicidad, estos son estratégicos puesto que gracias a ellos mi madre logró aceptar su enfermedad degenerativa (síndrome post-polio), y remontó su vida dado que su paladín nunca desfallecía. Por último le daremos un puesto honorífico al 'orgullo', gracias a el hemos podido proseguir con nuestra labor en entorno hostil, hemos podido sobreponernos a la incertidumbre y el riesgo, hemos podido llegar más lejos con menos medios, y lo más importante podemos hoy postular este escrito.

En cuanto al reconocimiento, se busca lograr una aceptación como especialista en inteligencia artificial (IA), concretamente en el campo de NLP, y por ello se aplican las técnicas punteras en el estado del arte DL, todo en busca de lograr una aplicación hasta ahora desconocida (INNOVACIÓN). Por otro lado, el hecho de ayudar a la sociedad mediante avances tecnológicos en el ámbito de la sanidad, es un reto que aporta valor añadido en lo profesional y personal, además de garantizar una categoría de proyecto (COMPLEJIDAD). Por último, reconocer que el ámbito clínico es una disciplina difícil, en cuanto a cantidad de información necesaria y existente para el diagnóstico de un paciente, y requiere apoyo para que el médico pueda ofrecer servicios de calidad y eficaces, por tanto la colaboración entre la informática y la medicina puede ser sustancialmente beneficiosa para ambos sectores (UTILIDAD).

Edgar Andrés Santamaría

1.3. Preámbulo

Antes de comenzar con la planificación del proyecto y con su despliegue nos centraremos en entender conceptos relacionados con el área de investigación 'Text Mining' [27, Cap 9.5], veremos primeramente ciertas definiciones y conceptos relacionados, posteriormente se expone el contexto de uso de estas tecnologías, y por último se tratan otras posibles áreas de aplicación aparte de la referida en este estudio 'clinical domain'.

Definición 1. inteligencia artificial

La inteligencia artificial (IA) según [18, Cap 0, Pág 1-8], es una ciencia que busca por un lado teorizar la inteligencia en base al método informático y por otro guiar el diseño de máquinas capaces de replicar actividades mentales del hombre como: Tratamiento de lenguaje natural, recuperación inteligente de datos, Sistemas expertos, demostración de teoremas, realización de acciones físicas (robótica), programación, resolución de problemas (combinatorios, de planificación y de percepción). Estos sistemas denominados de producción [18, Cap 1.1] se componen de datos o espacios de estados (posibles y objetivo), un conjunto de reglas de transformación y una lógica de control para aplicar las reglas, con esta arquitectura es posible resolver problemas complejos como el 8-puzzle [18, Cap 1.1.1].

Definición 2. reconocimiento de formas

El reconocimiento de formas o 'pattern recognition' según [3, Cap 1.1, Pág 3] , es una técnica utilizada para solucionar problemáticas de: visión por computador, reconocimiento de voz y robótica, los algoritmos de aprendizaje máquina, en este caso, buscan patrones en: imágenes, sonidos y características biométricas. Un ejemplo de aplicación descrito en [6, Cap 1, Pág 1-3] , y [3, Cap 1.2.2, Pág 6-8], es el reconocimiento de dígitos escritos 'handwritten digit recognition' que consiste en clasificar, en diez clases distintas (0- 9), símbolos escritos a mano y digitalizados.

Definición 3. Extracción de conocimiento

La extracción de conocimiento o 'knowledge discovery from data' consiste en que al aprender reglas sobre ciertos datos logramos una explicación del proceso subyacente, y por tanto un experto en dicho proceso sería capaz de interpretarlas, extrayendo por tanto su conocimiento [3, Cap 1.2.2, Pág 8], el proceso se lleva a cabo en 7 pasos: preprocesado, integración, selección, transformación, minería, evaluación y presentación [12, Cap 1.2, Pág 7].

Definición 4. Aprendizaje máquina

El aprendizaje máquina o 'machine learning' tiene diferentes definiciones según en punto de vista de su aplicación, algunos autores lo conciben según la teoría probabilística como [17, Cap 1.1 , Pág 1], o [6, Cap 1.2, Pág 12], y otros lo conciben según la teoría de conjuntos como [27, Cap 1.4 , Pág 28], o [6, Cap 2], ambos modelos están muy relacionados, ambos siguen la teoría de la probabilidad diferenciándose en el enfoque que toman hacia la resolución del problema 'a priori / a posteriori' y 'fronteras de decisión' respectivamente.

En cuanto a su objeto [3, Cap 1], consiste en que la máquina extraiga un algoritmo para una problemática de la que solo tenemos información de ejemplo y se rige por la teoría de la información para obtener dicho conocimiento, otros autores relacionan estas técnicas directamente con la inteligencia artificial [15, Cap 1.2] propone un sistema capaz de jugar a las damas.

Definición 5. Minería de datos

Minería de datos o 'data minning' consiste en la aplicación de métodos de 'machine learning' a grandes bases de datos según [3, Cap 1, Pág 2], y [12, Cap 1.2, Pág 5], las bases de datos o 'data warehouses' son las estructuras modernas que alimentan los distintos sistemas con datos [12, Cap 1.3.2, Pág 10], la aplicación de las distintas técnicas busca el 'reconocimiento de formas' para su posterior 'extracción de conocimiento' a favor del usuario del sistema.

Contexto de uso 1. Necesidad de la minería de datos

La necesidad de la minería de datos surge de las crecientes cantidades de información disponibles para su procesado y la consecuente búsqueda de métodos escalables y eficientes de análisis [3, Cap 1, Pág 4], a esto se le suma un contexto empresarial competitivo con nuevas problemáticas y necesidad de soluciones prácticas [4, Cap 4], estas necesidades requieren del perfil 'analista de datos' el cual se pretende demostrar con este trabajo.

Contexto de uso 2. Utilidad de la minería de datos

Según [12, Cap 1.4, Pág 15] existen distintas funcionalidades en minería de datos divididas en dos grandes grupos descripción y predicción, el primer grupo busca la caracterización de los datos disponibles 'clustering' lo que incluye: caracterización y discriminación [12, Cap 1.4.1] , minería de patrones frecuentes, asociación y correlación [12, Cap 1.4.2] , análisis de agrupaciones [12, Cap 1.4.4], y análisis de datos atípicos [12, Cap 1.4.5], por otro lado el segundo grupo 'classification' busca realizar inducciones sobre los datos provistos para la realización de predicciones lo que incluye: clasificación y regresión [12, Cap 1.4.3].

Estas dos grandes problemáticas tienen sus propias técnicas resolutivas y modelos de análisis complementándose en los diferentes estadios del proceso práctico como veremos posteriormente en el desarrollo.

Aplicación real 1. Web

según [12, Cap 1.3] existen diferentes sistemas estructurados/des-estructurados de información interesantes de analizar (Bases de datos, almacenes de datos e información transaccional), concretamente en [12, Cap 1.3.4] advierten de la importancia de las técnicas de extracción de datos web 'web scrapping' para el análisis de información masiva que involucra distintos formatos (imagen, audio, texto ...) y en constante crecimiento.

Concretamente en [4, Cap 4] se expone la metodología para generar marketing dirigido en base a estudios estadísticos de la información de potenciales clientes, en ese caso se basan en el nivel de cualificación y la distribución espacial (zona de residencia) de los perfiles estudiados.

Aplicación real 2. Negocio

según [4, Cap 1, Pág 7] la minería de datos puede utilizarse con fines de mejora de: marketing, ventas o procesos de apoyo, mediante una profunda comprensión de sus clientes gracias a la información que disponen sobre ellos.

Esto denota una cercana relación entre los negocios y la minería de datos, en [4, Cap 12] se expone la metodología llevada a cabo para el análisis de riesgos en Marketing, concepto común a toda empresa de excelencia.

en [27, Cap 1.3] Se comentan otras potenciales áreas de negocio en las que aplicar estas tecnologías como pueden ser: Banca, Seguros, Energía, Meteorología, Medicina (como el presente estudio) , y Producción.

1.4. Otra información relevante para la comprensión del estudio

En este apartado ofrecemos una serie de conocimientos y explicaciones técnicas generales para poder entender la nomenclatura técnica del estudio, no se entra en detalles pero se ofrece una visión general para apoyar la lectura del estudio.

Primeramente hablaremos de 'neural network' NN [27, Cap 6, pág 232], que consiste en una red con múltiples capa compuestas de unidades 'perceptron' (Neurona), esta arquitectura tiene la peculiaridad de permitir la modelización de funciones no lineales, esto es indispensable para tareas de predicción complejas. A grandes rasgos su entrenamiento consiste en el ajuste de los pesos internos que conectan la red, mediante el paso de instancias de entrenamiento por lotes hacia delante, y la propagación de errores hacia atrás.

Una vez introducidas NN, explicaremos sucintamente 'recurrent neural network' RNN [3, Cap 11.12.2], esta arquitectura está basada en NN con la peculiaridad de que sus capas ocultas, es decir aquellas capas entre la de 'Input' y la de 'Output', son NNs con la capacidad de clonarse y procesar secuencialmente los valores de entrada, esta característica es crítica en la resolución de tareas 'seq2seq', es decir, en las que se requiere del contexto de cierto Input (en este caso palabra), concretamente la arquitectura BiLSTM ofrece excelentes resultados teniendo en cuenta tanto el contexto cercano como el lejano a la palabra, en este caso son la base de la representación de oraciones para su etiquetado IOB.

Etiquetado 'inside outside begining' IOB, es una tarea 'name entity recognition' NER que consiste en predecir secuencias de etiquetas para oraciones de entrada, las predicciones del sistema se basan en la predicción de estas cadenas. Podemos apreciar un etiquetado IOB en el siguiente fragmento.

Admission I-FACTUAL
Date O
: O
2012-01-20 O
Discharge I-FACTUAL
Date O
: O
2012-01-23 O
Service O

La parte izquierda representa la palabra y la derecha la entidad a la que se refiere (I-entidad), cuando se presenta O nos referimos a que no pertenece a ninguna entidad. Cuando dos entidades se encuentran contiguas usamos (B-entidad) para diferenciar la primera de la segunda, podemos apreciar unejemplo en el siguiente fragmento.

Admission I-FACTUAL
Discharge B-FACTUAL
Date O
: O
2012-01-23 O
Service O

De esta manera se generan las secuencias de entrenamiento para la RNN, y esta se encarga de la representación de las secuencias de palabras (oraciones).

Hablamos de representación, y ahora explicaremos sucintamente en que consiste, hasta ahora conocemos los 'word embbedings' [3, Cap 6.8] como un diccionario de palabras, pero existen maneras más complejas de representación, en este caso la RNN genera un 'sequence embedding' ,con la información resultante de una secuencia de palabras de entrada, esta tarea requiere de dos pasos 'encoding' 4 y 'decoding' 19, en este caso la decodificación la llevará a cabo una capa CRF especialista en 'IOB tagging' de 'sequence embeddings'.

A partir de este momento, para referirnos al sistema encargado de la representación vectorial de cierto Input, nos referiremos como sistema WE. Entre los sistemas que comentamos se encuentran 'Doc2vec' [14], red neuronal encargada de la representación de textos, 'TFIDF'[24], modelo para el cálculo de 'word embeddings' en base a la frecuencia de aparición de palabras en los textos de entrenamiento , y 'Word2vec' [2], red neuronal encargada de la representación de 'word embeddings', esta es la aproximación utilizada para realizar los vectores pre entrenados utilizados en este estudio, concretamente su variante 'continous bag of words' CBOW [21], que consiste en representar la palabra según las palabras más representativas para su predicción

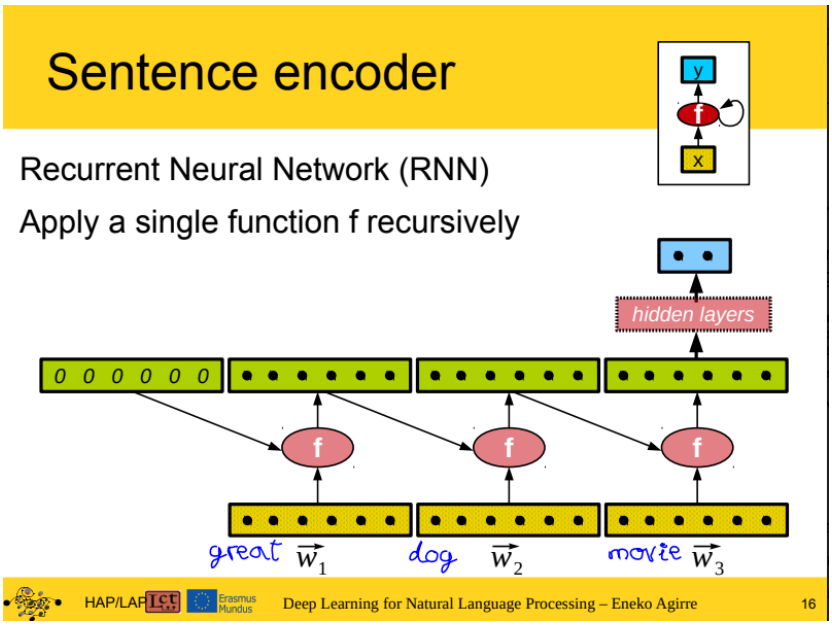

Figura 4: Representación de secuencia mediante RNN.

Por último, la capacidad más importante de estas representaciones, y es que son capaces de transferir conocimiento de una tarea a otra, es decir, unos 'word embbedings' pre entrenados en una tarea similar pueden reutilizarse ayudando enormemente a modelar un problema dado que se basan en más información de la disponible, concretamente la transferida del inicial problema en el que se generaron. Esta noción es clave para entender el alto rendimiento del sistema, ya que los 'word embeddings' pre entrenados utilizados eran de gran calidad.

2. Documento Objetivos de Proyecto

A continuación expondremos la planificación asociada a este proyecto de investigación, dicha planificación consta de: Descripción, Objetivos, Herramientas, Alcance, Planificación temporal, Riesgos y Evaluación económica.

La Descripción del alcance ofrece la división de tareas realizada para el proyecto de investigación proveyendo tiempos reales y estimados.

Los Objetivos ofrecen los distintos hitos que componen el proyecto de investigación, y constituyen los medidores últimos para conocer el estado del proyecto.

La Elección de Herramientas ofrece la comparativa de herramientas realizada para la selección de : IDE, lenguajes, librerías, modelos de representación y modelos de aprendizaje máquina.

La planificación temporal muestra la disposición temporal de las tareas descritas de manera estimada y la contrasta con la seguida realmente.

Los Riesgos muestran los eventos inciertos relacionados con el proyecto y que podrían generar un impacto negativo, asimismo se detalla como se minimizará su impacto en el denominado plan de sostenibilidad.

La Evaluación Económica muestra el desglose económico de las tareas descritas de manera estimada y lo contrasta con el real, asimismo se dispone como retornar la inversión realizada en este proyecto.

2.1. Descripción del alcance

A continuación se describen las tareas que componen el proyecto de investigación:

1. Documentación, esta tarea consiste en generar la documentación relativa al proyecto, dicha documentación consistirá en:

   a) Introducción
   b) Documento Objetivos del Proyecto (DOP):
   c) Antecedentes
   d) Análisis de requisitos
   e) Prototipo
   f) Diseño
   g) Desarrollo
   h) Pruebas
   i) Conclusiones
   j) Referencias

   tiempo estimado : 75h

   tiempo real: 30h

2. Revisión y corrección, esta tarea consiste en revisar la documentación y corregir los fallos: léxicas, semánticas y sintácticas.

   tiempo estimado : 15h

   tiempo real: 9h

3. Análisis, esta tarea constituye el objetivo estratégico de la investigación, y consiste en analizar la problemática expuesta (i2b2 challenge) hasta lograr un modelo conceptual que cumpla las expectativas del reto. Se divide en los siguientes apartados:

   a) Prototipo

b) Análisis de requisitos

c) Diseño

tiempo estimado : 15h

tiempo real: 70h

4. Implementación, esta tarea consiste en implementar los 3 componentes de los que se compondrá el sistema:

   a) Preproccess

   b) Clasification

   c) Inference

   tiempo estimado : 110h

   tiempo real: 120h

5. Pruebas, esta tarea consiste en validar el sistema mediante las métricas de evaluación provistas en ï2b2 challenge".

   tiempo estimado : 25h

   tiempo real: 13h

6. Seminarios, esta tarea consiste en aprender el conocimiento necesario par realizar el acercamiento al reto 'i2b2 challenge'.

   tiempo estimado : 112h

   tiempo real: 120h

7. Investigación, esta tarea consiste en la revisión y comprensión de los artículos relacionados con el reto 'i2b2 challenge'.

   tiempo estimado : 50h

   tiempo real: 80h

8. Reunión, esta tarea consiste en la revisión y supervisión del trabajo de investigación por parte de Aitziber Atutxa y Arantza Casillas.

   tiempo estimado : 6h

   tiempo real: 3h

9. Visualizar, esta tarea consiste en la generación de artefactos visuales sobre los distintos apartados generados durante el proyecto, deben ser útiles y fáciles de entender en previsión a la presentación de resultados.

   tiempo estimado : 60h

   tiempo real: 35h

10. Presentación, esta tarea consiste en la generación de la presentación asociada al proyecto así como los preparativos necesarios para su consecución.

    tiempo estimado : 5h

    tiempo real: Aun no realizada

tiempo total estimado : 472h

tiempo total real: 480h

2.2. Objetivos

En esta sección establecemos las metas y objetivos a cubrir en el proyecto, estos han de ser concretos y cuantificables, y además deben poder controlarse en el tiempo para garantizar su consecución, y por tanto la calidad del presente proyecto.

El objetivo principal, consiste construir un sistema de aprendizaje automático que fuese capaz de participar en la competición **The 2012 i2b2 challenge is on temporal relations** que tuvo lugar en en 2012 aplicando técnicas modernas de aprendizaje automático. Esta competición consistía en, dado un informe médico de ingreso hospitalario, identificar automáticamente los eventos mdicos y fechas que en el aparecen, para posteriormente establecer relaciones temporales entre ellos. Por eventos clínicos nos referimos a:

- conceptos clínicos (problemas, ensayos, o tratamientos)
- departamentos (como quirófano, o entrada principal)
- evidencias clínicas (resultados de ensayos clínicos)
- sucesos clínicos (por ejemplo admisiones, transferencia, etc).

La calidad del sistema construido se medirá a través de figuras de mrito como son la precisin, cobertura y FB1 (precision, accuracy y f-measure).

Como objetivo secundario, sería ideal obtener mejores resultados que los participantes del 2012.

Con respecto a las expresiones temporales, estas serían horas, duraciones, y frecuencias temporales.

Pasemos ahora a detallar los requisitos previos a la consecución de los objetivos para defender que en efecto el proyecto cuenta con un plan de ruta. Como metas contamos con:

1. Investigación: leer y comprender el estado del arte en aplicación de técnicas NLP para el reto 'i2b2'.
2. Diseño: Aprender y comprender arquitecturas DL.
3. Implementación: Aprender programación en Python.
4. Documentación: Aprender y comprender como gestionar un proyecto.

5. Revisión y corrección: Entender conceptos generales de Filosofía , Historia y Medicina.
6. Análisis: Entender las necesidades de los médicos en cuanto al uso de historias clínicas.
7. Pruebas: Aprender y comprender diferentes modelos de evaluación de sistemas ML.
8. Seminarios: Introducirse en la aplicación de técnicas DL.
9. Reunión: Estrechar lazos con el departamento de informática, y conocer el grado de satisfacción de profesionales expertos con respecto a mi trabajo.
10. Visualizar. Aprender y comprender diferentes técnicas para mostrar información, y tecnologías 'Frontend'.
11. Presentación: Desarrollar un prototipo comprensible para apoyar la explicación.

Con estas metas pretendemos garantizar capacidad de maniobra en cualquier estadio del proyecto para minimizar riesgos y lograr el mayor rendimiento posible en cada fase del proyecto, este modelo de gestión coincide con la metodología SCRUM dentro del repertorio de metodologías ágiles.

Finalmente introducimos los objetivos del proyecto, es decir los requisitos que debe satisfacer el proyecto para determinar su Calidad final. Como objetivos contamos con:

1. Investigación: Compendiar el estado del arte para facilitar futuros acercamientos.
2. Diseño: Postular una arquitectura viable para cubrir las necesidades del Análisis.
3. Implementación: Generar un prototipo de alta fidelidad.
4. Documentación: Transmitir la mitad del conocimiento adquirido en el ámbito investigado.
5. Revisión y corrección: Evitar imprecisiones ortográficas y semánticas.
6. Análisis: Proponer las cualidades que debe poseer el prototipo de alta fidelidad.
7. Pruebas: superar las evaluaciones obtenidas en la pasada revisión del reto 'i2b2' [13] .
8. Seminarios: Conocer a los profesores del Master NLP y lograr un curriculum que nos permita aspirar a dicho master.
9. Reunión: Validar el trabajo realizado y garantizar que la beca IKASIKER se aprovecho convenientemente.
10. Visualizar: Apoyar la presentación y garantizar que el tribunal del TFG comprende el trabajo realizado.
11. Presentación: Lograr una Matrícula de Honor.

Creemos que con los objetivos propuestos se puede valorar objetivamente el grado de consecución del trabajo, y estimar adecuadamente la Calidad del mismo, además cubrimos en rigor los valores : competitividad, innovación y rendimiento.Los cuales consideramos eje principal y por supuesto motor de la investigación.

2.3. Elección de Herramientas

En esta sección explicamos las herramientas utilizadas en el presente estudio y el porque de su selección frente a otras alternativas analizadas:

1. Lenguaje de programación: El seleccionado fue Python dada su alta versatilidad en el procesamiento de documentos y su gran desempeño en operaciones de cálculo, frente a JAVA con gran capacidad de visualización GUI.

2. Librerías NLP: La seleccionada fue 'NLTK' dada su facilidad de uso y la gran cantidad de documentación disponible, frente a 'Textacy' con gran capacidad de cálculo y mejor desempeño en despliegues reales.

3. Librerías ML: La seleccionada fue Glample Tagger dada su facilidad de uso y que cuenta con la tecnología idónea con capacidad de afinamiento, frente a 'Keras' con gran capacidad de diseño de arquitecturas NN que requieren afinarse, y 'SkLearn' que cuenta con modelos desfasados con respecto al estado del arte DL.

4. Soporte documental: El seleccionado fue 'Overleaf' dada su alta prestación gracias al motor 'LaTex' y su garantía de confidencialidad, frente a 'Google Docs' que viola el principio de confidencialidad del presente trabajo, y Libre 'Office' que carece del soporte suficiente para la realización de textos científicos.

5. Versión de Interprete: La seleccionada fue 'Python 3.6' dadas las mejoras en las librerías internas y de rendimiento que acarrea, además soporta el uso interno 'Python 2.7', frente a 'Python 2.7' que cuenta con menor desarrollo de sus librerías internas aunque 'Glample Tagger' fue desarrollada en este soporte.

6. Entorno de desarrollo: El seleccionado fue 'PyCharm' dada su alta prestación y ayuda a la programación, frente a 'PyDev' basado en el proyecto Eclipse con menor capacidad de configuración interna de la configuración de proyecto.

7. Entorno de Edición de texto: El seleccionado fue Visual 'Studio Code' dada su capacidad de visualización y ayuda a la programación en diferentes nomenclaturas relevantes como XML y Html, frente a 'Notepad ++' que carece de las ayudas a la visualización.

2.4. Planificación temporal

En esta sección abordaremos la planificación del proyecto, esta es dinámica existiendo dos estimaciones, una planificación inicial antes de iniciar el proyecto y otra una vez realizada la fase de investigación, es decir, una vez entendida la problemática en profundidad. por último añadimos una tercera planificación con la distribución temporal final del proyecto.

A continuación se muestra la distribución relativa a las 'a priori' contempladas tareas implicadas en el proyecto:

1. Seminarios
2. Investigación
3. Reuniones
4. Análisis
5. Diseño
6. Implementación
7. Evaluación
8. Guía de uso
9. Memoria del proyecto

La primera fase fue asistir y superar las asignaturas de 'inteligencia artificial' y 'minería de datos con duración de 100h, esta fase me permitió entender de manera más profunda los puntos de vista sobre el aprendizaje máquina y las diferentes técnicas conocidas para su aplicación, tras finalizar la segunda fase 'investigación' se entendió la problemática desde un punto de vista diferente, dicha fase duró al rededor de 80 h iniciándose en el verano de 2018 (Mayo) terminando en invierno de 2019 (Enero), en dicho periodo se leyeron bastantes artículos relacionados con la problemática expuesta, se aprendió a programar en el lenguaje Python y a aplicar técnicas PLN mediante aprendizaje profundo 'deep learning'. Con este conocimiento fresco se procedió a la implementación de un modelo SVM, para replicar el estado del arte en sus fases más tempranas, esto fue costoso debido al preprocesamiento requerido para el posterior entrenamiento ML, este sistema primitivo, como era de esperar, contaba con evaluaciones parecidas a las descritas en el estado del arte. En este momento además, se

estima que dicha planificación inicial era inexacta en su descripción del alcance de las tareas involucradas. Esto es debido a que la arquitectura necesaria del sistema es compleja, y requiere mucho conocimiento.

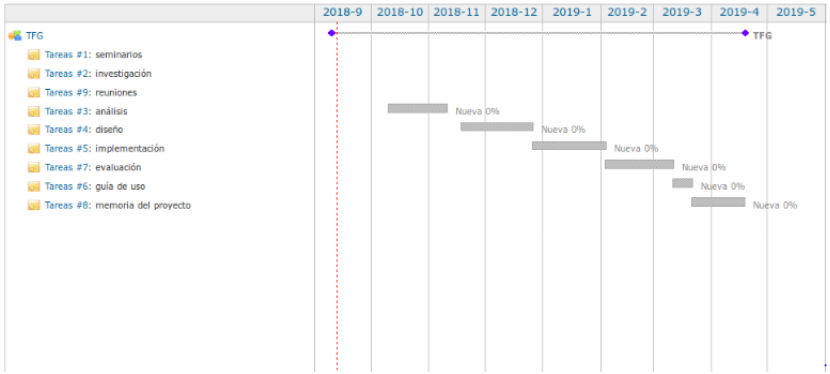

Figura 5: Distribución estimada a priori de tareas relativas al proyecto.

Por tanto en la Tabla general de tareas relativas al proyecto 'a posteriori' 6 pueden verse las tareas involucradas realmente en el desarrollo del proyecto además de una visión detallada del progreso hasta el momento, una de las decisiones tomadas en las reuniones de seguimiento consiste en la aplicación de un modelo de desarrollo (Análisis - implementación - Pruebas y Visualización) cíclico para poder generar correctamente el sistema. En este punto se estima que con tres iteraciones sobre el problema de aplicación es suficiente para terminar el desarrollo del prototipo de alta fidelidad.

A continuación mostramos la nueva distribución temporal del alcance actualizado del proyecto, cuenta con la información temporal estimada y real compendiada, así como las fechas de inicio / final de cada tarea para ofrecer una visión completa, y el tiempo de la consecución del proyecto. Esta aproximación constituye la planificación temporal hasta Mayo, ya que por entonces (Febrero-Marzo) se estimaba la finalización del sistema por aquella fecha. Acabamos de terminar la primera fase de análisis e implementación del modelo basado en NN ('Multilayer perceptron'), fue relativamente sencillo entender su arquitectura dado que era materia estudiada en las asignaturas Sistemas de apoyo a la decisión y Minería de datos. El desempeño se este sistema era ligeramente superior al 'Baseline' SVM mencionado previamente.

Cabe destacar que poco después se comenzó una formación en modelos DL basados en RNN (Abril), lo cual aparentemente sería tarea fácil pero como veremos en la distribución temporal final 9, supuso muchos días de análisis (Analisis2) e implementación (Implementación2) para entender en profundidad tanto la arquitectura como su programación. Esta materia era totalmente desconocida y no relacionada con ningún estudio realizado durante el grado por lo que se decidió asistir a un seminario en Julio para entender el tema en cuestión, durante el lapso desde Abril hasta Julio dedicamos esfuerzos al aprendizaje autodidacta de la tecnología en cuestión logrando el primer modelo a finales de mayo 19, fue entonces cuando se descubrió un problema añadido ¿Cómo afinamos el sistema?, y entonces se recomendó el uso de la librería Glample Tagger que ya contaba con sistemas internos de afinamiento y concretamente la arquitectura mencionada.

GANTT Planificación general del proyecto

Tarea	Fecha inicio	dia inicio	Fecha fin	Duración Real (Días)	Días Completados	Días Restantes	Porcentaje de proceso	Real		Estimado	
								Duración (horas / día)	Duración (horas)	Duración (horas / día)	Duración (horas)
Planificación general del proyecto											
Revisión de la documentación 0	9/20	129	9/25	5	5	0	100%	1	5	1	5
Análisis 0	8/25	103	9/25	31	31	0	100%	0.1612903226	5	0.1612903226	5
implementación 0	11/20	190	4/11	142	142	0	100%	0.2816901408	40	0.2536211268	36
Pruebas 0	3/13	303	3/21	8	8	0	100%	0.5	4	0.625	5
Visualizar 0	1/1	232	2/5	35	35	0	100%	0.4285714286	15	0.5714285714	20
Reunión 0	8/29	107	8/30	1	1	0	100%	0.5	0.5	1	1
Reunión 1	10/4	143	10/5	1	1	0	100%	0.5	0.5	1	1
Reunión 2	2/21	283	2/22	1	1	0	100%	0.5	0.5	1	1
Reunión 3	4/18	339	4/19	1	0	1	0%	0	0	1	1
Reunión 4	5/10	361	5/11	1	0	1	0%	0	0	1	1
Documentación	1/1	232	5/9	128	89.6	38.4	70%	0.234375	30	0.5859375	75
Revisión de la documentación 1	5/6	357	5/10	4	0	4	0%	0	0	2.5	10
Análisis 1	1/1	232	1/18	17	17	0	100%	0.2941176471	5	0.2941176471	5
implementación 1	1/1	232	3/29	87	87	0	100%	0.4597701149	40	0.4137931034	36
Pruebas 1	3/7	297	3/14	7	7	0	100%	0.1428571429	1	1.428571429	10
Visualizar 1	3/11	294	4/11	31	31	0	100%	0.4838709677	15	0.6451612903	20
Presentación	5/11	363	5/15	4	0	4	0%	0	0	0	0
Análisis 2	4/11	332	4/15	4	0	4	0%	0	0	1.25	5
implementación 2	4/16	337	4/29	13	0	13	0%	0	0	2.923076923	38
Pruebas 2	5/1	352	5/5	4	0	4	0%	0	0	2.5	10
Visualizar 2	4/25	346	4/29	4	0	4	0%	0	0	5	20
Seminarios	9/15	124	1/1	108	108	0	100%	0.9259259259	100	1.037037037	112
Investigación	5/15	1	1/1	231	231	0	100%	0.3463203463	80	0.2164502165	50

Figura 6: Tabla general de tareas relativas al proyecto.

En este punto se añadió el reto de superar las evaluaciones obtenidas en la revisión de 2017 12, por tanto para poder realizar dicha evaluación se debía lograr un sistema funcional por cada tarea involucrada, para esto, se generaron funciones adicionales en el preprocesador, para obtener conjuntos de entrenamiento en formato IOB, y se generaron conjuntos para 6 tareas (Event-Type, Event-Polarity, Event-Modality, Timex3-Type, Timex3-Modifier, Tlink-Type), debido a la complejidad del cálculo de Timex3-value se decidió no invertir esfuerzos en dicha tarea. Tras la generación de los conjuntos de entrenamiento se entrenaron los 6 modelos utilizando la librería Glample Tagger, estos a su vez forman parte del módulo Inference aportando predicciones sobre un texto dado, y además se evaluaron para contrastar el desempeño del sistema con los resultados mencionados anteriormente.

Por otro lado se ofrece el diagrama 7 de los datos anteriores 6 , basado en los días de proyecto desde su inicio en mayo 2018 hasta su final estimado en mayo de 2019, las barras de las tareas coinciden con el estado de proceso descrito en la Tabla general de tareas relativas al proyecto, y a su vez con el estado del proyecto en Abril.

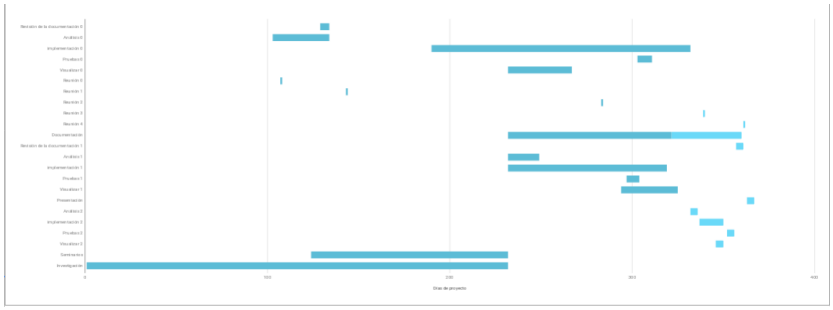

Figura 7: Distribución general de tareas relativas al proyecto hasta mayo.

En la siguiente tabla 8 podemos apreciar el estado del proyecto en su fase final, todavía no se había planteado concretamente el modelo de presentación para meditar tranquilamente que aspectos resaltar del proyecto, por otro lado destacar el incremento de horas invertidas por día en las últimas fases de análisis y diseño, esto es debido al aprendizaje de las tecnologías DL. Asimismo, en las horas invertidas para formación (seminarios) vemos una gran actividad por día, esto es debido a la cantidad de conocimiento necesario para realizar el prototipo de alta fidelidad. Por último en el diagrama 9 podemos apreciar l planificación Gantt anual del proyecto ya finalizado.

						Real		Estimado	
Tarea	Fecha inicio	dia inicio	Fecha fin	Duración Real (Días)	Porcentaje de proceso	Duración (horas / día)	Duración (horas)	Duración (horas / día)	Duración (horas)
Planificación general del proyecto									
Revisión de la documentación 0	9/20/2018	129	9/25/2018	5	100%	1	5	1	5
Análisis 0	8/25/2018	103	9/25/2018	31	100%	0.1612903226	5	0.1612903226	5
implementación 0	11/20/2018	190	4/11/2019	142	100%	0.2816901408	40	#¡REF!	36
Pruebas 0	3/13/2019	303	3/21/2019	8	100%	0.5	4	0.0352126761	5
Visualizar 0	1/1/2019	232	2/5/2019	35	100%	0.4285714286	15	0.5714285714	20
Reunión 0	8/29/2018	107	8/30/2018	1	100%	0.5	0.5	1	1
Reunión 1	10/4/2018	143	10/5/2018	1	100%	0.5	0.5	1	1
Reunión 2	2/21/2019	283	2/22/2019	1	100%	0.5	0.5	1	1
Reunión 3	4/18/2019	339	4/19/2019	1	100%	0.5	0.5	1	1
Reunión 4	6/10/2019	361	6/10/2019	1	100%	0.5	0.5	1	1
Reunión 5	5/10/2019	361	5/11/2019	1	100%	0.5	0.5	1	1
Documentación	1/1/2019	232	7/19/2019	128	100%	0.234375	30	0.5859375	75
Revisión de la documentación 1	7/19/2019	357	5/22/2019	4	100%	1	4	2.5	10
Análisis 1	1/1/2019	232	1/18/2019	17	100%	0.2941176471	5	0.2941176471	5
implementación 1	1/1/2019	232	3/29/2019	87	100%	0.459770114	40	0.4137931034	36
Pruebas 1	3/7/2019	297	3/14/2019	7	100%	0.1428571429	1	1.428571429	10
Visualizar 1	3/11/2019	294	4/11/2019	31	100%	0.4838709677	15	0.6451612903	20
Presentación	5/11/2019	363	5/15/2019	4	0%	0	0	1.25	5
Análisis 2	4/11/2019	332	7/5/2019	62	100%	0.9677419355	60	0.0806451613	5
implementación 2	4/16/2019	337	6/7/2019	40	100%	1	40	0.95	38
Pruebas 2	6/10/2019	352	7/2/2019	17	100%	0.4705882353	8	0.5882352941	10
Visualizar 2	7/10/2019	346	7/17/2019	6	100%	0.8333333333	5	3.333333333	20
Seminarios	9/15/2018	124	7/5/2019	111	100%	1.081081081	120	1.009009009	112
Investigación	5/15/2018	1	1/1/2019	231	100%	0.3463203463	80	0.2164502165	50

Figura 8: Distribución general final de tareas relativas al proyecto hasta mayo.

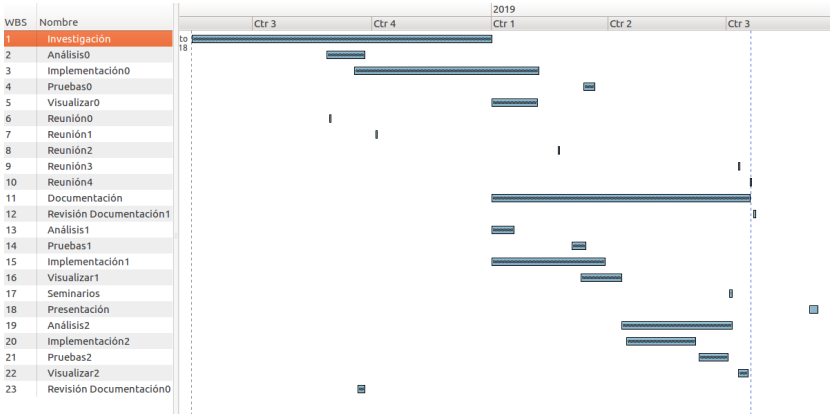

Figura 9: Distribución Temporal final de tareas relativas al proyecto (por fechas).

2.5. Riesgos

En esta sección se identifican los riesgos relativos al proyecto, se valora su impacto y se establece un plan de contingencia con tres planes aplicados a cada riesgo (prevención, emergencia y restauración), este proceso busca minimizar el impacto de la materialización de dichos riesgos, y asimismo garantizar la consecución de un trabajo de calidad.

Dada la importancia de la gestión del riesgo en cualquier proyecto, hemos realizado un especial énfasis en aquellos riesgos que nos han parecido de especial relevancia, es decir, aquellos con una valoración superior a 15: Obtener un artefacto disconforme, Incapacidad para desarrollar una tarea y Recursos insuficientes, como veremos posteriormente en la fase de desarrollo se realizaron rigurosas medidas para evitar un artefacto disconforme, y durante la planificación temporal hablaremos de la estricta formación llevada a cabo para evitar la incapacidad para desarrollar una tarea, por último en cuanto a recursos insuficientes el investigador solicitó junto a la beca IKASIKER un ordenador de resguardo en la universidad por si su ordenador fallase, además se compró un tercero portátil para minimizar la posibilidad de problemas.

2.5.1. Identificación de riesgos

Riesgos generales

1. Pérdida de datos de repositorios, consiste en perder información del proyecto durante su desarrollo.
2. Filtración de datos de repositorios, consiste en la revelación de información del proyecto durante su desarrollo.
3. Fallo de los repositorios, consiste en que la información del proyecto no se encuentre disponible.
4. Obtener un artefacto disconforme, consiste en que los resultados Software del trabajo no sean satisfactorios.
5. Incapacidad para desarrollar una tarea, consiste en que alguna de las tareas propuestas en el alcance sea irresoluble para el investigador.
6. Fallo de Hardware, consiste en que las herramientas de trabajo fallen.
7. Recursos insuficientes, consiste en la falta de herramientas para la consecución de los objetivos.

Riesgos específicos

1. Falta de información, consiste en que falte información para lograr unos resultados óptimos del estudio.
2. Preproceso ineficaz, consiste en que el módulo Preproceso sea insuficiente, ineficaz o lento en la tarea de procesado de datos
3. Clasificación ineficaz, consiste en que el módulo Clasificación sea insuficiente, ineficaz o lento en la tarea de aprendizaje sobre la información provista.
4. Métricas de evaluación inconsistentes, consiste en no cumplir las métricas de evaluación comunes impuestas por ï2b2 challenge"[25] [13].
5. Inferencia ineficaz, consiste en que el módulo Inferencia sea insuficiente, ineficaz o lento en la tarea de predicción sobre la información provista.
6. Visualización ineficaz, consiste en que la visualización propuesta del proceso sea insuficiente, ineficaz o lenta en la tarea de proveer conocimiento al experto humano usuario del Software 'médico'.

2.5.2. Valoración de riesgos

	probabilidad estimada frente a impacto esperado				
	insignificante	pequeño	moderado	grande	catástrofe
Extremadamente probable	medio(5)	alto(10)	alto(15)	muy alto(20)	muy alto(25)
Muy probable	medio(4)	medio(6)	alto(12)	alto(16)	muy alto(20)
Probable	bajo(3)	medio(5)	medio(9)	alto(12)	alto(15)
Poco probable	bajo(2)	bajo(4)	medio(6)	medio(8)	alto(10)
Excepcional	bajo(1)	bajo(2)	bajo(3)	bajo(4)	medio(5)

Tabla 1: Matriz de riesgo

Riesgos generales

1. Pérdida de datos de repositorios es, probable y catástrofe por tanto 15 de valoración.
2. Filtración de datos de repositorios es, probable y catástrofe por tanto 15 de valoración.
3. Fallo de los repositorios es, probable y grande por tanto 12 de valoración.
4. Obtener un artefacto disconforme es, muy probable y grande por tanto 16 de valoración.
5. Incapacidad para desarrollar una tarea, es extremadamente probable y grande por tanto 20 de valoración.
6. Fallo de Hardware, es probable y catástrofe por tanto 15 de valoración.
7. Recursos insuficientes, es extremadamente probable y catástrofe por tanto 25 de valoración

Riesgos específicos

1. Falta de información, es probable y catástrofe por tanto 15 de valoración.
2. Preproceso ineficaz, es probable y grande por tanto 12 de valoración.
3. Clasificación ineficaz, es probable y moderado por tanto 9 de valoración.
4. Métricas de evaluación inconsistentes, es probable y catástrofe por tanto 15 de valoración .
5. Inferencia ineficaz, es probable y pequeño por tanto 5 de valoración.
6. Visualización ineficaz, es muy probable e insignificante por tanto 4 de valoración.

2.5.3. Plan de contingencia

A continuación se exponen los 3 planes dispuestos para cada riesgo (Prevención, emergencia y restauración), así se pretende minimizar el impacto de su posible materialización

Prevención:

1. Para evitar pérdida de datos de repositorios, se establece un sistema de copias de seguridad para el código, y un sistema de control de versiones para la documentación.
2. Para minimizar filtración de datos de repositorios, se establece que solo se comparta información del proyecto con Aitziber Atutxa y Arantza Casillas.
3. Para evitar fallo de los repositorios , se establecerá un sistema de permisos concreto para el acceso y manipulación información del proyecto.
4. Para evitar obtener un artefacto disconforme, se establecen reuniones periódicas con Aitziber Atutxa y Arantza Casillas en las que se orientará la investigación.
5. Para evitar incapacidad para desarrollar una tarea,, Edgar Andrés es responsable de conseguir el material lectivo necesario bajo la supervisión de Aitziber Atutxa y Arantza Casillas.
6. Para evitar fallo de Hardware, se solicito un ordenador de investigación y además Edgar Andrés aporta sus propias herramientas de respaldo.
7. Para minimizar recursos insuficientes, se solicitó una Beca de Colaboración y recursos por parte de la institución UPV/EHU, y además Edgar Andrés aporta sus propios recursos de respaldo.
8. Para evitar falta de información, se utilizan técnicas de normalización y estratificación sobre los datos estudiados.
9. Para minimizar preproceso ineficaz, se estudian diversas técnicas para la representación de los textos clínicos, además se establece la orientación de Aitziber Atutxa y Arantza Casillas.
10. Para minimizar clasificación ineficaz, se estudian diversas técnicas para el aprendizaje máquina sobre los textos clínicos, además se establece la orientación de Aitziber Atutxa y Arantza Casillas.

11. Para minimizar métricas de evaluación inconsistente, se estudian diversas métricas de evaluación sobre los modelos de clasificación de textos clínicos, además se establece la orientación de Aitziber Atutxa y Arantza Casillas.

12. Para minimizar Inferencia ineficaz, se estudia exhaustivamente el estado del arte conocido para el reto i2b2, además se establece la orientación de Aitziber Atutxa y Arantza Casillas.

13. Para minimizar visualización ineficaz, se estudian diversas técnicas para la visualización de datos y diverso Software de matemática aplicada, además se establece la orientación de Aitziber Atutxa y Arantza Casillas.

Emergencia:

1. En caso de pérdida de datos de repositorios, se recuperará la copia de seguridad más reciente.

2. En caso de filtración de datos de repositorios, se comunicará a Aitziber Atutxa y Arantza Casillas.

3. En caso fallo de los repositorios, se tratará de arreglar con permiso de administrador corriendo el riesgo (1).

4. En caso obtener un artefacto disconforme, se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas en las que se orientará la investigación.

5. En caso de incapacidad para desarrollar una tarea, se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

6. En caso de fallo de Hardware, se recurrirá a alguna de las herramientas de respaldo existentes.

7. En caso recursos insuficientes, se recurrirá a alguno de los recursos de respaldo existentes.

8. En caso falta de información, se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

9. En caso preproceso ineficaz, se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

10. En caso de clasificación ineficaz,se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

11. En caso de métricas de evaluación inconsistente, se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

12. En caso de Inferencia ineficaz,se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

13. En caso de visualización ineficaz, se seguirán las recomendaciones de Aitziber Atutxa y Arantza Casillas.

Restauración:

1. Para recuperarse de pérdida de datos de repositorios, se volverá a generar la información no recuperada.

2. Para recuperarse de filtración de datos de repositorios, se planteará de nuevo el plan de prevención de este riesgo y se harán efectivas las recomendaciones.

3. Para recuperarse de fallo de los repositorios , se establecerá el sistema de permisos nuevamente.

4. Para recuperarse de obtener un artefacto disconforme, se harán efectivas las recomendaciones.

5. Para recuperarse de incapacidad para desarrollar una tarea, se harán efectivas las recomendaciones.

6. Para recuperarse de fallo de Hardware, se restaurarán todos los sistemas necesarios en la máquina de respaldo.

7. Para recuperarse de recursos insuficientes, se destinarán los recurso de respaldo a aquella tarea que los requiera.

8. Para recuperarse de falta de información, se harán efectivas las recomendaciones.

9. Para recuperarse de preproceso ineficaz, se harán efectivas las recomendaciones.

10. Para recuperarse de clasificación ineficaz, se harán efectivas las recomendaciones.

11. Para recuperarse de métricas de evaluación inconsistente, se harán efectivas las recomendaciones.

12. Para recuperarse de Inferencia ineficaz, se harán efectivas las recomendaciones.

13. Para recuperarse de visualización ineficaz, se harán efectivas las recomendaciones.

En este punto comentamos que por suerte y buena predisposición ante el riesgo no llegaron a darse ninguno de estos riesgos, concretamente gracias a la flexibilidad de gestión que ofrecen las metodologías ágiles. El único inconveniente digno de destacarse fue la falta de conocimiento para el desarrollo de un modelo DL, el cual se adquirió sin demasiadas complicaciones invirtiendo el tiempo necesario en formarse practica y teóricamente.

2.6. Evaluación económica

En este apartado tratamos las implicaciones monetarias en €, que suponen la realización del proyecto, primeramente se definirán los gastos contemplados para su posterior desglose mensual, y finalmente se tratará el retorno de la inversión.

2.6.1. Gastos Directos

La tareas se han dividido en cuatro grupos:A(Gestor), B(Diseñador), C(Desarrollador) y D(Publicidad).Cada grupo tendrá asignado un sueldo diferente según el Boletín oficial del estado [22] :

1. Grupo A:9,7/hora: Documentación, revisión y corrección , Presentación y Reunión.
2. Grupo B:8,3/hora: Análisis, Seminarios e Investigación.
3. Grupos C:6,7/hora: Implementación, Pruebas y Visualización

Estos sueldos vienen dados por Convenio colectivo estatal de empresas de consultoría y estudios de mercado y de la opinión pública emitido en el Boletín Oficial de Estado "BOE" y constituyen la cota inferior permitida por ley. Los gastos indirectos son calculados parcialmente sobre los gastos directos según las fórmulas descritas en este apartado.

2.6.2. Gastos Indirectos

Se han considerado los siguientes gastos indirectos, y se han sumado a los directos de cada mes:

1. Luz, agua y costes telefónicos (10 % del total)
2. Ordenador: $15 = CosteOrdenador/5 * 12$, donde el ordenador utilizado costo 900 €, y se amortiza a 5 años.

2.6.3. Desglose

En esta sección se desglosan los ingresos-gastos del desarrollo del proyecto, estos valores coinciden con la distribución de horas reales finales vistas previamente. Los ingresos contemplados en las tablas coinciden con los importes recibidos por la Beca IKASIKER de colaboración.

Concepto	Mayo	Junio	Julio	Agosto	Septiembre	Octubre
Ingresos	0	0	0	0	0	0
Gastos directos	85.8	85.8	85.8	111.4	245.52	181.12
Gastos indirectos	23.58	23.58	23.58	26.14	39.55	33,11
Ingresos-Gastos	-109.38	-109.38	-109.38	-137.54	-285.07	-214.23
Ingresos-Gastos ACC	-109.38	-218.76	-328.14	-465.68	-750.75	-964.98

Tabla 2: Desglose 2018

En la tabla anterior 2 podemos observar el desglose correspondiente a 2018, en mayo, junio y julio se tuvieron en cuenta 2h mensuales correspondientes a documentación y 8h mensuales correspondientes a investigación. Por otro lado, en Agosto se sumaron media hora de reunión y 2.5h mensuales de análisis. En septiembre, se sumaron a la base de Mayo, junio y Julio, 10.9h mensuales de seminarios y 5h para revisión de documentación. Por último en octubre, se sumaron a la base de Julio (85.8), 10.9h mensuales de seminarios y media hora de reunión. Cabe destacar que hasta Enero nos encontramos en el primer Sprint del proyecto.

Concepto	Noviembre	Diciembre	Enero	Febrero	Marzo	Abril	Mayo
Ingresos	0	0	0	1650	0	0	0
Gastos directos	225.34	225.34	406.2	298.3	326.95	422.8	328.33
Gastos indirectos	37.53	37.53	55.62	44.83	47.7	57.3	47.83
Ingresos-Gastos	-262.87	-262.87	-461.82	1304	-374.65	-480.1	-376.16
Ingresos-Gastos ACC	-1227.85	-1490.72	-1952.54	-648.54	-1023.19	-1503.29	-1879.45

Tabla 3: Desglose 2018 - 2019

En la tabla anterior 3 podemos observar el desglose correspondiente al final de 2018 (Sprint 0) y la transición hasta Mayo 2019 (Sprint 1), en Noviembre y Diciembre, se sumaron a la base de Octubre (181.12), 6.6h mensuales de Implementación, esta corresponde al Sprint 0 y se prolongará hasta Abril solapándose con la implementación del Sprint 1, dado que para el correcto desempeño del sistema NN se requería un sistema WE de calidad, y este último pertenecía al

primer Sprint. En Enero , se sumaron a la base de Diciembre (225.34), 13.3h mensuales de Implementación y 5h mensuales de Análisis, es mes con mayor trabajo realizado durante el proyecto, dado que se entendió que las planificaciones iniciales eran erróneas, esto suponía un cambio de ruta y se tradujo en tiempo de trabajo. En Febrero , se sumaron a la base de Diciembre (225.34), 13.3h mensuales de Implementación y media hora de reunión, esto es por la rápida finalización del diseño del sistema NN WE, además se recibe el 67 % de la Beca IKASIKER de colaboración. En Marzo , se sumaron a la base de Febrero (298.3), 4h mensuales de pruebas y 7.5 h de visualizar correspondientes al final del Sprint 0 de la Implementación. En Abril , se sumaron a la base de Marzo (326.95), 15h mensuales de Análisis y media hora de reunión, en este caso recién terminada la Implementación del Sprint 1, se inició la del Sprint 2. En Mayo , se sumó a la base de Marzo (326.95), media hora de reunión para validar los resultados correspondientes al final del Sprint 1.

Concepto	Junio	Julio	Agosto	Septiembre
Ingresos	0	0	812	0
Gastos directos	355.13	333.47	0	48.5
Gastos indirectos	50.51	48,35	0	19.85
Ingresos-Gastos	-405.64	-381.82	812	0
Ingresos-Gastos ACC	-2285.09	-2666.91	-1854.91	-1923.26

Tabla 4: Desglose 2019

Finalmente desglosamos las horas del Sprint 2, En Junio , se sumaron: 10.9h mensuales de Seminarios, 2h mensuales de Documentación, 4h mensuales de Pruebas, 13.3h mensuales de Implementación, 15h mensuales de análisis y media hora de Reunión para validar los resultados correspondientes al final del Sprint 1, es el segundo mes con mayor actividad dentro del proyecto dado que para la generación del prototipo de alta fidelidad basado en RNN se requería un esfuerzo extra. En Julio , se sumaron a la base de Julio (326.95) habiendo finalizado la implementación (261.17), 5h de Visualización para terminar de ensamblar el prototipo de alta fidelidad y 4h de Revisión de documentación para finalizar la memoria, en este plazo damos por concluido el proyecto pero añadimos Agosto dado que se espera recibir el 33 % restante de la Beca IKASIKER y el Septiembre se estiman 5h para la preparación de la Presentación.

2.6.4. Retorno de la inversión

En este apartado trataremos el beneficio generado y la inversión necesaria para el desarrollo del proyecto.

En cuanto a inversión se requirieron 2666.91 €, es decir el mayor valor acumulado (Julio), y el proyecto a supuesto un beneficio negativo final de 1923.26 €, esto es debido a la cantidad de horas necesarias para el desarrollo del proyecto, que temporalmente se traducen en más o menos 150h de sobre coste con respecto a las 300h en las que se estimaba el proyecto inicialmente.

A pesar de los esfuerzos realizados, asumimos positivamente el resultado dado que se ha generado un sistema funcional sobre el que basar posteriores estudios, además se ha actualizado el estado del arte, y se considera meritorio de menciones curriculares dado que se ha basado en estudios introductorios de Master, y profundizado suficientemente como para garantizar el acceso y superación del master Language Analysis and Processing (HAP/LAP) ofertado por el equipo IXA, y en el que se pretende mejorar el presente trabajo.

Por otro lado, se considera incalculable el valor del conocimiento adquirido, dado que el sistema creado es genérico y versátil, pudiendo aplicarse en diferentes áreas, además garantiza y demuestra la experiencia que Edgar Andrés cuenta en el desarrollo de sistemas DL y ML. Por otro lado demuestra conocimientos transversales en diversas tecnologías y técnicas lo que supone el posible acceso a puestos de trabajo de alta cualificación.

Por último, se considera saldada la deuda contraída por el estudiante con el sistema educativo del País Vasco, dado que este proyecto es suficientemente extenso como para, garantizar la perpetuación y calidad de la investigación en el ámbito NLP, técnica cuya aplicación y desarrollo beneficiará enormemente a la sociedad. En este trabajo consideramos haber contribuido enormemente a la concreción de la tecnología DL, y a fundamentar el potencial de su aplicación a modo de carta de presentación para Edgar Andrés.

3. Antecedentes

En esta sección abordaremos los distintos trabajos realizados a nivel internacional para abordar la tarea en cuestión 'i2b2 challege' [25] [13], entre los acercamientos se encuentran tanto sistemas completos, soluciones parciales a distintos problemas que surgidos como ontologías de trabajos relacionados, para obtener la mayor información posible sobre la problemática y los distintos pasos realizados hasta ahora por investigadores internacionales.

Los artículos mencionados en esta sección se consideran inspiradores para la idea expuesta en el presente trabajo, y mediante la aproximación de dichos experimentos se ha logrado el desarrollo del prototipo de alta fidelidad expuesto posteriormente. En cualquier caso, dada las notables diferencias entre los acercamientos y el presente estudio, se considera invención propia de Edgar Andrés, y es que el desarrollo integro del sistema lo realizó por sus propios medios uniendo diversas fuentes de información disponibles a su alcance, y en colaboración con Aitziber Atutxa y Arantza Casillas.

Se espera que el presente trabajo infunda inspiración en posteriores trabajos, y que la aproximación al presente sistema genere posteriores invenciones en la linea de investigación expuesta en este apartado, recomendándose encarecidamente la utilización de la arquitectura presentada, para la aplicación de tareas NER en otras áreas de conocimiento como periodismo, web scrapping etc, ya que únicamente se deben desarrollar los conjuntos de datos etiquetado IOB para reutilizar esta arquitectura.

3.1. Neural Network Methods for Natural Language Proccessing

En el articulo [11, Cap 2] se expone que el problema del procesamiento del lenguaje natural (PLN) abarca muchas tareas.

En el caso del presente estudio se tratarán de aplicar técnicas PLN óptimas y unificadas mediante redes neuronales para la extracción de etiquetas y sus categorías, finalmente mediante técnicas de representación se buscará resumir y sintetizar la información clínica para facilitar trabajo al personal médico.

Las redes neuronales, son la tecnología más novedosa para la aplicación de técnicas de aprendizaje máquina, en el caso de PLN son mecanismos que dada una sentencia como entrada, devuelven una serie de predicciones referentes a la(s) tarea(s) PLN para la que fue entrenada, estas arquitecturas se componen de varias capas de transformación "Lookup tables"para el aprendizaje de roles semánticos y análisis sintáctico, estas capas se entrenan mediante el algoritmo 'backPropagation', cada vector de palabra obtenido se normaliza Çonvolution", y el resto de la red establece fronteras de decisión para posteriores predicciones [11, Cap 3], la arquitectura descrita es (Ver Fig.10).

Para entrenar la red, se entrena una capa neuronal para cada categoría de pares (expresión, tipo) [11, Cap 4], donde con expresión nos referimos al atributo etiqueta del conjunto , y con tipo a las categorías de etiqueta del conjunto, para ello se dispone de los conjuntos de datos propuestos por la competición SemEval(2018) [13], una vez entrenada la red se logrará una modelización unificada de todos los elementos del problema PLN como veremos más adelante en el desarrollo, sobre esta representación común se realizarán los estudios de 'minería de datos'.

En este caso la aplicación de la red neuronal no seguirá este esquema intrínsecamente puesto que existirán dos, la primera encargada de vectorizar las etiquetas y la segunda encargada de aprender los conceptos, posteriormente entraremos en profundidad teórica en este respecto.

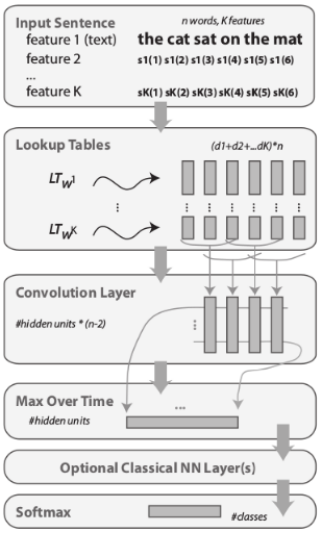

Figura 10: Arquitectura de red neuronal [11, Figura I]

3.2. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text

En el artículo [25] se introducen los tres objetivos propuestos para el reto internacional 'i2b2 challenge', primeramente la extracción de conceptos clínicos (Event, Timex3 y Sectime) sobre los informes clínicos propuestos 'i2b2 datasheet' , posteriormente se propone la asignación de tipos a dichos conceptos , y finalmente se propone extraer relaciones entre los conceptos de tipo (problem, test, y treatment).

A estos retos se propusieron 59 sistemas totales de los que se determinó que los sistemas basados en reglas muestran mejores resultados para la extracción de conceptos, relaciones y tipos, por otro lado se postuló que los sistemas ensamblados de clasificadores muestran mejores capacidades para este tipo de datos y por último que las fuentes de información externas pueden ayudar al entrenamiento.

Como introducción al trabajo se exponen los datos provistos, el conjunto en cuestión es 'i2b2 partnered with VA Salt Lake City Health Care System' (i2b2/VA) generado manualmente para apoyar el desarrollo de técnicas PLN y

permitir a la comunidad científica la comparación individual de sistemas, cabe destacar que estos retos son de carácter incremental, existiendo posteriores como semEval (2018) [13] y anteriores retos.

En este caso se dispusieron: 394 informes clínicos de entrenamiento, 477 informes para test y 877 textos sin anotación para la prueba de los sistemas. Estos están disponibles en https://i2b2.org/NLP/DataSets.

Los métodos sugeridos para el acercamiento al reto consisten en la generación de sistemas capaces de clasificar las referencias a los conceptos estándar anotados en los textos de entrenamiento así como sus respectivos tipos (e.g posible , condicional o hipotético para conceptos etiquetados como Event). en cuanto a la extracción de relaciones se busca la detección de pares de conceptos con interrelación, existen tres tipos (Event - Event, Event - Timex3 y timex3 - Timex3).

Antes de continuar con el artículo, en la siguientes tablas se desglosa y explica la información provista para la extracción con el fin de poder explicar en profundidad los objetivos anteriores:

	Event	
Atributo	Tipo	Descripción
Tipo	nominal	asigna una categoría al evento
Id	cadena caracteres	distingue unívocamente el evento en el texto
Etiqueta	cadena caracteres	contiene el fragmento de texto que representa el evento
Modalidad	nominal	contiene la modalidad verbal del evento
Polaridad	binario	contiene la polaridad objetiva del evento
Inicio	entero	contiene la posición de inicio del fragmento Etiqueta
Fin	entero	contiene la posición final del fragmento Etiqueta

Tabla 5: Información Event

Los eventos por tanto pueden considerarse conceptos relacionados con el ámbito clínico en general y se identifican unívocamente dentro del texto mediante el ID (e.g 'Admission' [E0] o 'elevated PSA' [E2]) categorizados según su atributo tipo en: ocurrencia, problema, prueba, evidencia, departamento clínico o tratamiento. Continuando con el ejemplo, 'Admission' su categoría es ocurrencia y 'elevated PSA' es de tipo problema.

En cuanto al atributo polarity, esta puede ser POS o NEG otorgando una valoración objetiva del evento. Siguiendo el ejemplo tanto 'Admission' como 'elevated PSA' son de polaridad POS.

El atributo etiqueta será el que se utilice para aplicar las técnicas PLN tratando de generar por una parte un sistema automático que las detecte y además las categorice, en este caso de ejemplo las etiquetas son las propias cadenas 'Admission' y 'elevated PSA', este atributo se complementa con los enteros Inicio y fin.

	Timex3	
Atributo	Tipo	Descripción
Tipo	nominal	asigna una categoría a la expresión temporal
Id	cadena caracteres	distingue unívocamente la expresión temporal en el texto
Etiqueta	cadena caracteres	fragmento de texto que representa la expresión temporal
Modificador	nominal	contiene información temporal adicional
Valor	cadena de caracteres	contiene la representación ISO: 8601
Inicio	entero	contiene la posición de inicio del fragmento Etiqueta
Fin	entero	contiene la posición final del fragmento Etiqueta

Tabla 6: Información Timex3

Las expresiones temporales por tanto pueden considerarse conceptos relacionados con el tiempo en general y se identifican unívocamente dentro del texto mediante el ID (e.g 'one to two weeks' [T9] o '2012-01-23' [T1]) categorizados según su atributo tipo en: fecha, duración, frecuencia o tiempo . Continuando con el ejemplo, 'one to two weeks' y '2012-01-23' son de tipo fecha.

En cuanto al modificador, es un atributo encargado de proveer información adicional: Aproximado, NA, inicio, fin, por encima o intermedio ,con el fin de permitir concretar secuencias temporales entre las expresiones temporales.

El valor provee la representación estándar ISO:8601 para concretar el significado de todas las expresiones temporales.

El atributo etiqueta será el que se utilice para aplicar las técnicas PLN tratando de generar por una parte un sistema automático que las detecte y además las categorice, en este caso de ejemplo las etiquetas son las propias

cadenas 'one to two weeks' y '2012-01-23', este atributo se complementa con los enteros Inicio y fin. En este caso se añade la dificultad de tratar caracteres especiales (e.g '-') y estructuras acortadas no estandarizadas (e.g ' b.i.d.' = repetir cada 12 h).

	Sectime	
Atributo	Tipo	Descripción
Tipo	nominal	asigna una categoría a la sección
Id	cadena caracteres	distingue unívocamente la sección temporal en el texto
Etiqueta	cadena caracteres	fragmento de texto que representa la sección temporal
Valor	cadena de caracteres	contiene la representación ISO: 8601
Inicio	entero	contiene la posición de inicio del fragmento Etiqueta
Fin	entero	contiene la posición final del fragmento Etiqueta

Tabla 7: Información Sectime

Las secciones temporales por tanto pueden considerarse conceptos relacionados con el tiempo en general y se identifican unívocamente dentro del texto mediante el ID (e.g '2012-01-20' [S0] o '2012-01-23' [S1]) categorizados según su atributo tipo en: admisión, o alta . Continuando con el ejemplo, '2012-01-20'es de tipo admisión y '2012-01-23' es de tipo alta.

El valor provee la representación estándar ISO:8601 para concretar el significado de todas las secciones temporales.

El atributo etiqueta será el que se utilice para aplicar las técnicas PLN tratando de generar por una parte un sistema automático que las detecte y además las categorice, en este caso de ejemplo las etiquetas son las propias cadenas 'one to two weeks' y '2012-01-23', este atributo se complementa con los enteros Inicio y fin. En este caso se añade la dificultad de tratar caracteres especiales (e.g '-').

Una vez conocemos los datos y los retos propuestos, de los cuales, como se comenta previamente, se pretende generar un sistema automático para la extracción de conceptos y sus categorías. Nos disponemos a presentar las métricas de evaluación propuestas, en este caso : precision, recall y f1-score.

$precision(P) = TP/(TP + FP)$.

$$recall(R) = TP/(TP + FN).$$

$$f1 - score = 2 * P * R/(P + R).$$

Estas serán las utilizadas para la evaluación del sistema de extracción de conceptos en textos clínicos, y poder así compararse con los sistemas evaluados en este artículo.

Table 2 Exact and inexact evaluation on the concept extraction task

Concept extraction					
System by	**Medical experts**	**Method**	**External?**	**Exact F measure**	**Inexact F measure**
deBruijn *et al*[25]	N	Semi-supervised	N	0.852	0.924
Jiang *et al*[16]	Y	Hybrid	Y	0.839	0.913
Kang *et al*[17]	N	Hybrid	Y	0.821	0.904
Gurulingappa *et al*[18]	N	Supervised	Y	0.818	0.905
Patrick *et al*[19]	N	Supervised	Y	0.818	0.898
Torii and Liu[20]	N	Supervised	N	0.813	0.898
Jonnalagadda and Gonzalez[21]	N	Semi-supervised	N	0.809	0.901
Sasaki *et al*[22]	N	Supervised	N	0.802	0.887
Roberts *et al*[23]	N	Supervised	N	0.796	0.893
Pai *et al*[24]	Y	Hybrid	N	0.788	0.884

Figura 11: Evaluación de sistemas de extracción de conceptos.

3.3. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge

El artículo [23] nos propone un sistema capaz de extraer los conceptos previamente introducidos y además se centra en la extracción de conceptos TLINK o relaciones temporales, estos son pares de conceptos (Event - Event, Event - Timex3 o timex3 - Timex3) entre los cuales existe una relación temporal.

En este caso se decidió realizar el acercamiento a las relaciones temporales para lo cual se tuvo que plantear un sistema de extracción de expresiones temporales y eventos clínicos, así como de sus categorizaciones para poder nutrir de información al sistema de extracción de relaciones temporales, en la tarea de extracción de conceptos concluyeron que los acercamientos basados en reglas mostraban mejor desempeño para la extracción de eventos y que no existía diferencia entre técnicas de aprendizaje máquina y sistemas basados en reglas en cuanto a la extracción de expresiones temporales.

Los sistemas estudiados hasta el momento consistían en clasificadores SVM o CRF para la tarea de extracción de conceptos clínicos siendo la mayor puntuación de f-score 0.92 hasta este momento por parte de 'Beihang University; Microsoft Research Asia, Beijing; Tsinghua University'

De esta manera los acercamientos de extracción evolucionaron a métodos 'tagging' o de etiquetado que consistían en aprender las palabras contenidas en el texto para predecir etiquetas, de esta manera se introduce el potencial del aprendizaje máquina al paradigma de extracción de conceptos clínicos.

En el presente estudio se aplica el etiquetado 'Inside–outside–beginning tagging ' (IBO), formato adecuado para la aplicación de técnicas de aprendizaje máquina, posteriormente explicaremos este concepto en profundidad.

Table 2 System results for EVENT, TIMEX, and TLINK tracks

Organization	Span F measure	Type accuracy	Polarity accuracy	Modality accuracy	Method
EVENT					
Beihang University; Microsoft Research Asia, Beijing; Tsinghua University	0.92	0.86	0.86	0.86	CRF
Vanderbilt University	0.9	0.84	0.85	0.83	CRF + SVM
The University of Texas, Dallas	0.89	0.8	0.85	0.84	CRF+SVM
The University of Texas, Dallas—deSouza	0.88	0.71	0.85	0.05	CRF
University of Arizona, Tucson	0.88	0.73	0.79	0.8	CRF+SVM+NegEx
University of Novi Sad, Novi Sad, Serbia; University of Manchester	0.87	0.82	0.79	0.82	CRF+dictionary based
Siemens Medical Solutions	0.86	0.71	0.78	0.77	CRF+MaxEnt
MAYO Clinic	0.85	0.76	0.75	0.76	CRF
LIMSI–CNRS; INSERM; STL CNRS; LIM&BIO	0.83	0.8	0.84	0.85	CRF+SVM
University of Illinois at Urbana-Champaign	0.83	0.74	0.75	0.77	Integer Quadratic Program

Organization	Primary score–Value F-measure	Span F measure	Type accuracy	Value accuracy	Modifier accuracy	Method
TIMEX3						
MAYO Clinic	0.66	0.9	0.86	0.73	0.86	Regular Exp
Beihang University; Microsoft Research Asia, Beijing; Tsinghua University	0.66	0.91	0.89	0.72	0.89	CRF+SVM+rule based
University of Novi Sad, Novi Sad, Serbia; University of Manchester	0.63	0.9	0.85	0.7	0.83	Rule based
Vanderbilt University	0.61	0.87	0.85	0.7	0.85	Rule based +HeidelTime
University of Arizona, Tucson	0.61	0.88	0.81	0.69	0.8	HeidelTime+CRF
The University of Texas, Dallas	0.55	0.89	0.78	0.62	0.79	CRF+SVM+rule based
Siemens Medical Solutions	0.53	0.89	0.86	0.6	0.8	SUTime
The University of Texas, Dallas—deSouza	0.53	0.89	0.78	0.59	0.79	GUTime+CRF +rule base
Bulgarian Academy of Sciences; American University in Bulgaria; University of Colorado School of Medicine	0.49	0.8	0.72	0.61	0.71	Regular Exp
LIMSI–CNRS; INSERM; STL CNRS; LIM&BIO	0.45	0.84	0.75	0.54	0.72	HeidelTime

Figura 12: Evaluación de sistemas de extracción de conceptos (Actualizada).

3.4. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge

Este artículo [19] consiste en un acercamiento de un sistema híbrido compuesto por un lado de un sistema CRF (conditional random fields) y otro SVM (support vector machine). El primero era encargado de la extracción de conceptos y la técnica utilizada consiste en la modelización del problema estructurando un grafo basado en cierta función de pérdida [16], este asimismo cuenta con un desempeño similar a las redes neuronales con el impedimento de la escalabilidad y configuración posibles, es decir, tiene la capacidad de generar fronteras de decisión para problemas no lineales.

Para poder alimentar el sistema CRF mencionado previamente se enfatizó en el pre procesamiento palabra a palabra de los textos generando un denominado

'Lexicon' para mapear cada palabra con la respectiva etiqueta que el sistema debería aprender, este acercamiento constituyó la aplicación de reconocimiento de entidades (name entity recognition).

Llegados a este punto, el sistema constituye fuente de inspiración dedo que se pretende generar un análogo basado en redes neuronales recurrentes (RNN) BiLSTM para la aplicación de reconocimiento de entidades IBO introducidas previamente.

El segundo sistema SVM se utilizó para el reconocimiento de relaciones entre los conceptos extraídos, su naturaleza lineal lo convierte en ideal para la clasificación de entidades.

En el presente estudio se postula un sistema SVM a modo de sistema 'Baseline', este como es de esperar no ofrece competencia alguna a las demás tecnologías mencionadas hasta ahora (RNN BiLSTM y CRF) pero es suficientemente simple para realizar los primeros estudios sobre el conjunto de datos, y así determinar que el 'qui' de los datos provistos consiste en la des estructuración de los datos provistos, posteriormente se profundizará en estos datos y en la introducida 'primera impresión' del conjunto i2b2.

Por último hablaré de el sistema SNOMED CT encargado de realizar el preprocesado previamente comentado, los autores del artículo [19] enfatizan en la alta dependencia de los resultados del sistema integrado (SNOMED CT + CRF + SVM) con respecto al desempeño de la fase inicial del proceso (preprocesado).

Por tanto la mayoría de los esfuerzos realizados en este presente trabajo se enfocan a una exacta detección de las cadenas IOB para garantizar un correcto desempeño de pasos posteriores gracias a las experiencias previas registradas en tareas de carácter similar.

3.5. À la Recherche du Temps Perdu: extracting temporal relations from medical text in the 2012 i2b2 NLP challenge

En el siguiente artículo [8] vemos el acercamiento propuesto por especialistas Canadienses a la extracción de relaciones temporales mediante sistemas estadísticos, este acercamiento consiste en dos pasos (1) la extracción de conceptos clínicos y (2) la extracción de relaciones temporales, y dadas las características del reto , enfocaron el reto generando cuatro sistemas expertos según conceptos 13, y es que los acercamiento probabilísticos logran mayor desempeño en tareas de clasificación para múltiples clases (multi-label classification) enfocando el sistema como un ensamblado de expertos binarios [17, Cap 1.2.2].

Table 3 Ablation test results

	Specialist tested alone			System without specialist			
Specialist	Precision	Recall	F1 score	Precision	Recall	F1 score	Delta F1
All	0.7273	0.6449	0.6837	0.7273	0.6449	0.6837	–
Sectime	0.9274	0.2661	0.4136	0.6541	0.3997	0.4962	−0.1875
Local	0.5960	0.3573	0.4468	0.8333	0.3005	0.4417	−0.2420
Non-local overlap	0.6166	0.0358	0.0676	0.7316	0.6142	0.6678	−0.0159
Non-local rules	0.7623	0.0030	0.0060	0.7499	0.6431	0.6924	0.0083

Figura 13: Errores acumulados del sistema en cascada.

Estas técnicas se utilizaron en el desarrollo del sistema 'Baseline' (SVM) dado que este es de carácter lineal y aunque no sea un acercamiento probabilístico intrínsecamente, sufre parecidas carencias en la clasificación 'multi label'. Por tanto se generaron tres clasificadores binarios (es o no es) para cada etiqueta de la tarea de extracción (Event, Timex3 y SecTime), en nuestro caso no implementamos la extracción de relaciones temporales.

El sistema propuesto en [8] ensamblaba la votación realizada por los clasificadores para resolver las predicciones en cuanto a la extracción de conceptos.

En el presente estudio no se ensamblan las predicciones de los expertos base puesto que se pretende realizar un estudio meramente informativo sabiendo de antemano su pobre y complejo desempeño en tareas de clasificación 'multi label' sobre conceptos clínicos. Esta experiencia de la que hablamos consiste en un sistema ensamblado mediante cascada (AdaBoost) de bosques aleatorios (Random forest), este sistema mostraba resultados menores a los del presente artículo

aunque ambos bajos relativamente , con la diferencia de que los expertos no se dividían por concepto sino por característica del texto recogida a modo de valor numérico de un vector (embbeding) TFIDF (Tern Frequency Inverse Document Frequency), aquel sistema fue enfocado a la extracción de las causas de muerte dadas las autopsias `https://github.com/EdgarAndresSantamaria/AdaBoost`, en este trabajo dada la des estructuración de los datos provistos (muy similar al conjunto de datos utilizado en el presente trabajo) la evaluación era baja(f1 score 0.19) para el modelo a pesar de mostrar poco error en las predicciones de test 14, este modelo está planteado pero sin ajustar sus parámetros dada la alta complejidad del mismo.

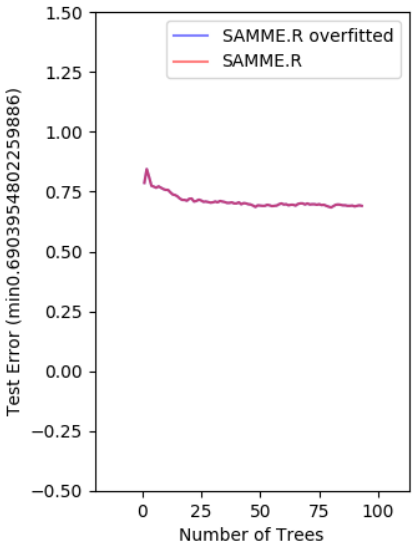

Figura 14: Errores acumulados del sistema en cascada.

Con esto se pretende concluir que los modelos probabilísticos principalmente caracterizados por la linealidad de las predicciones no son aptos para la realización de un sistema de extracción basado en información des estructurada.

3.6. Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives

En el presente artículo se propone un sistema de extracción de Eventos y Expresiones temporales híbrido, el cual utiliza tanto técnicas basadas en reglas como acercamientos basados en aprendizaje máquina.

En este caso también se enfatiza en el procesamiento de los datos aplicando estrategias basadas en 'Name Entity Recognition' (NER), en este punto entra en juego la parte del sistema compuesta de reglas puesto que aprovecha sistemas ya generados para el proceso de 'tagging' y posteriormente aplica el algoritmo 'Conditional Random Field' para el modelado del problema (f1 score 0.9) 15.

Table 4 Event recognition: per category performance on the test data (run 2, lenient matching)

Event type	Frequency	P (%)	R (%)	F (%)
Problem	4309	95.24	87.82	91.38
Treatment	3285	95.68	83.65	89.26
Occurrence	2499	63.43	66.91	65.12
Test	2173	95.05	87.48	91.11
Clinical department	732	76.02	83.61	79.64
Evidential	595	64.99	75.80	69.98

F, F score; P, precision; R, recall.

Figura 15: Resultado del sistema híbrido de extracción.

Por tanto el estado del arte nos sugiere que los acercamientos basados en técnicas NER para su posterior aplicación de técnicas de aprendizaje máquina es la tendencia en los últimos años, y UPV-EHU no será menos en su aportación a estos esfuerzos.

Por último cabe destacar la división realizada en este acercamiento según tipos de etiquetas lo que sugiere que estos modelos no lineales basados en los

principios de la neurona son aptos para clasificación de clases minoritarias, es decir, aquellas que cuentan con menor número de ejemplos sobre los que aprender y cobra en estos sistemas especial relevancia del proceso 'tagging' realizado sobre el conjunto de palabras de entrenamiento.

Posteriormente se enfatizará en el impacto que las clases minoritarias tienen sobre aquellos sistemas basados en la perspectiva probabilística (Naive Bayes, SVM ...), concretamente se tratará durante el estudio del sistema 'Baseline' en la fase 0 del desarrollo de la clasificación.

3.7. An end-to-end system to identify temporal relation in discharge summaries: 2012 i2b2 challenge

En el presente artículo [28] se expone un sistema integral para la extracción de conceptos clínicos y posteriormente la extracción entre estos conceptos, lograron el mayor desempeño hasta la fecha en la tarea de extracción de eventos (f1 score 0.9166), y un alto desempeño en extracción de expresiones temporales.

Lo relevante en este caso no es el desempeño del sistema sino que las relaciones se refieren a un 'Doctime 0', es decir que la fecha de admisión se convierte en el valor neutro de la linea temporal, y a partir de dicha premisa se establece la relación de tanto Eventos como Expresiones temporales, con respecto a la creación del documento. Esta idea es la llevada a cabo en el presente estudio, y por tanto la información mostrada en el valor 'historical' será 'OVERLAP' cuando el concepto se genere durante la presente historia, será 'BEFORE' cuando el concepto se de antes de la presente historia, y será 'AFTER' cuando el concepto se haya programado o se estime que sucederá posterior a la presente historia.

De esta manera logramos una manera sencilla y eficaz de extraer relaciones temporales entre los Eventos y Timex3 con respecto al 'DOCTIME 0', aunque no es la manera adecuada de tratar este tipo de información dado que se pierde información relevante como relaciones : Event- Event, Event -Timex3 y Timex3 - timex3. Se presume que en esta información no tratada reside la clave de extracción de relaciones causa - efecto, pero dada la envergadura y complejidad de su estudio no se ha realizado en el presente trabajo.

3.8. SemEval 2018 Task 6: Parsing Time Normalizations

En el articulo [13] se exponen distintos algoritmos base 'baselines' para el acercamiento a la competición 'i2b2 challenge', descomponiendo su funcionamiento en los distintos pasos a realizar:

1. modelo basado en caracteres (SCATE)
   - *a*) paso 1 : identificación de entidades
   - *b*) paso 2 : composición de entidades
   - *c*) paso 3 : composición semántica de entidades temporales
2. modelo basado en entidades temporales HeidelTime (TimeML)
   - *a*) paso 1 : etiquetado basado en reglas (muy preciso)
   - *b*) final : expresiones temporales normalizadas TIMEX3
3. modelo basado en entidades temporales CHRONO (SCATE)
   - *a*) paso 1 :identificar elementos temporales mediante 'regex'
   - *b*) paso 2 identificar frases temporales mediante Tokens consecutivos
   - *c*) paso 3 análisis sintáctico y normalización SCATE
   - *d*) paso 4 : composición semántica de entidades temporales (menos eficaz por lo general que el modelo basado en caracteres)

Según el citado artículo los avances se han dado en ámbito de análisis sintáctico y normalización temporal mediante la aparición del modelo SCATE que logra superar las limitaciones del anterior estándar TimeML (ISO :2012).

El modelo SCATE anota como entidades de tiempo composicionales capaces de ser interpretadas mediante una linea de tiempo por intervalos (ver Fig. 16) (interprete proporcionado por los organizadores).

Los redactores coinciden en que se debe enfatizar en la representación de expresiones temporales para lograr unas capaces de retener mayor información intrínsecamente, los modelos expuestos previamente logran resultados competitivos en este aspecto.

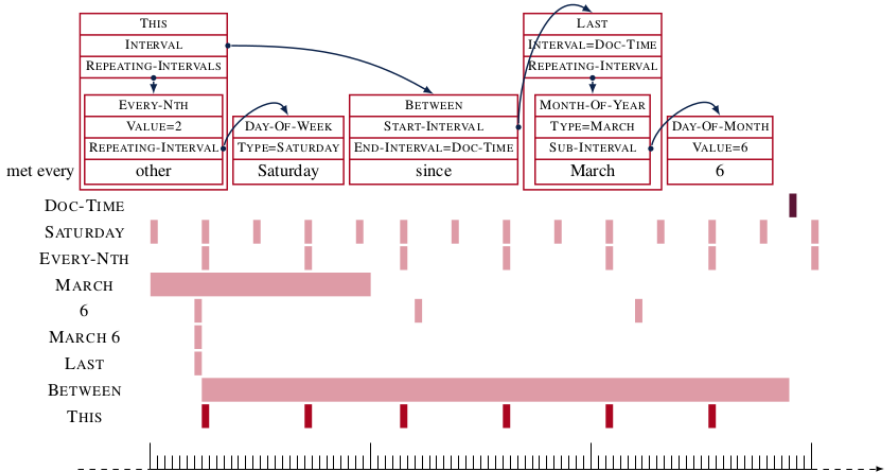

Figura 16: Ejemplo de anotaciones de tiempo semánticamente composicional y su interpretación. [13]

4. Análisis de requisitos

En este apartado se describen las necesidades de un sistema de extracción de información clínica, comprendidas a través del coloquio con un estudiante de Medicina UPV-EHU Leioa, en su 5º año de carrera, en las distintas conversaciones mantenidas se han determinado las siguientes premisas:

1. Las historias clínicas buscan determinar la evolución del paciente, la eficacia de un tratamiento y en ciertos casos decidir que ensayo clínico realizar.
2. La mayoría de procesos clínicos están estandarizados
3. La historia clínica compendia toda la información disponible relativa al paciente

Por tanto determinamos en base a las premisas previas las siguientes necesidades para el prototipo de alta fidelidad:

1. Debe reconocer conceptos clínicos, expresiones temporales y secciones temporales.
2. Debe ofrecer información adicional sobre eventos clínicos (modalidad, polaridad, tipo y relación histórica)

3. Debe ofrecer información adicional sobre expresiones temporales (modificador, valor, tipo y relación histórica)
4. Debe ofrecer información adicional sobre secciones temporales (valor y tipo)
5. Debe ofrecer favorecer la lectura del texto resaltando la información relevante, disponiendo el texto de manera sencilla y ofreciendo información adicional.

De esta manera se espera facilitar la comprensión de la historia clínica por parte del médico, y ofrecer así apoyo a su resolución de ensayos y tratamientos. Por otro lado se espera ayudar a diferenciar la información histórica e hipotética, teniendo en consideración que algunas historias son inmensas, y su redacción podría dar lugar a errores. Por último, se espera poder aportar información útil en cuanto a la polaridad de los eventos ,para facilitar la comprensión de la evolución y la eficacia de los tratamientos.

5. Diseño

En este apartado se proponen las tres arquitecturas principales sobre las que se basa la aplicación, no constituyen una superestructura puesto que las dos primeras únicamente interfieren en la generación de los modelos DL para su utilización como apoyo en la tercera, esta constituye el sistema final capaz de generar los prototipos de alta fidelidad.

Para poder comprender el segundo diseño introducimos una serie de conceptos y recomendaciones, primeramente mencionar que la parte inferior hasta 'Representation layer' constituye una arquitectura de 'recurrent neural neetworks' RNNs compuesta de capas intermedias 'Long Sort Tern Memory' LSTM [10], debido a su formación interna 17, son capaces de retener información del contexto ya procesado dada una secuencia de entrada, esto se realiza en cuatro pasos: decidir cuanta información olvidar del histórico, decidir cuanta información añadir al histórico, modificar el histórico y determinar una predicción para el estado actual. Para evitar que se pierda la información histórica inicial se propone la solución 'Bidirectional' LSTM que realiza el mismo proceso de ida y vuelta en la red, mediante esta solución se pretende lograr una representación numérica de las secuencias basándose en los 'word embbedings' de entrada.

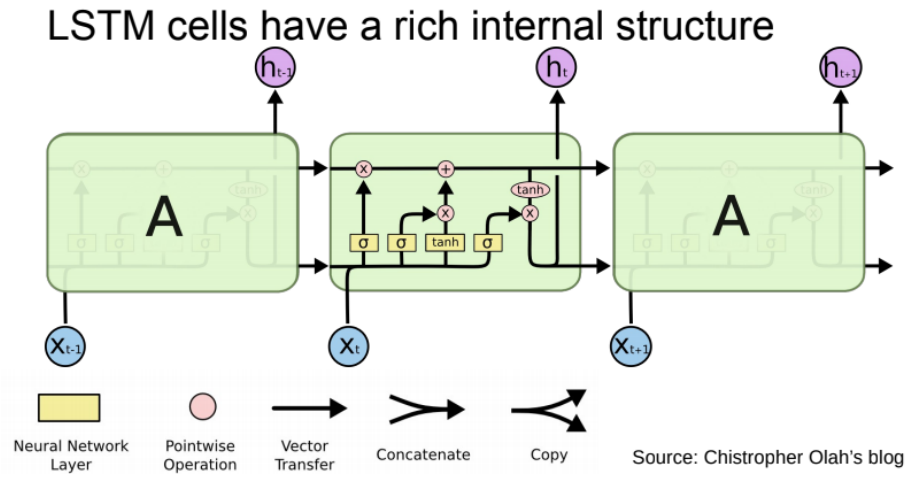

Figura 17: RNN LSTM.

5.1. Diseño del Preprocesador

Primeramente contamos con un sistema de preprocesado para los datos 18, esta arquitectura es crucial para el correcto desempeño del sistema y cuenta con el mayor esfuerzo dedicado, fue diseñada con un output intermedio en archivo (.xlsx) para detectar los errores con mayor facilidad, es la encargada de procesar toda la información contenida en los archivos (.XML) dados como entrada, dicha información posteriormente se transforma a los archivos (.txt) de entrada para el sistema Glample tagger, cabe destacar que para el correcto entrenamiento debe proveerse en formato IOB lo que supone dificultad añadida al preproceso y requiere un exhaustivo control de errores, posteriormente en el desarrollo del preproceso. profundizaremos en ello.

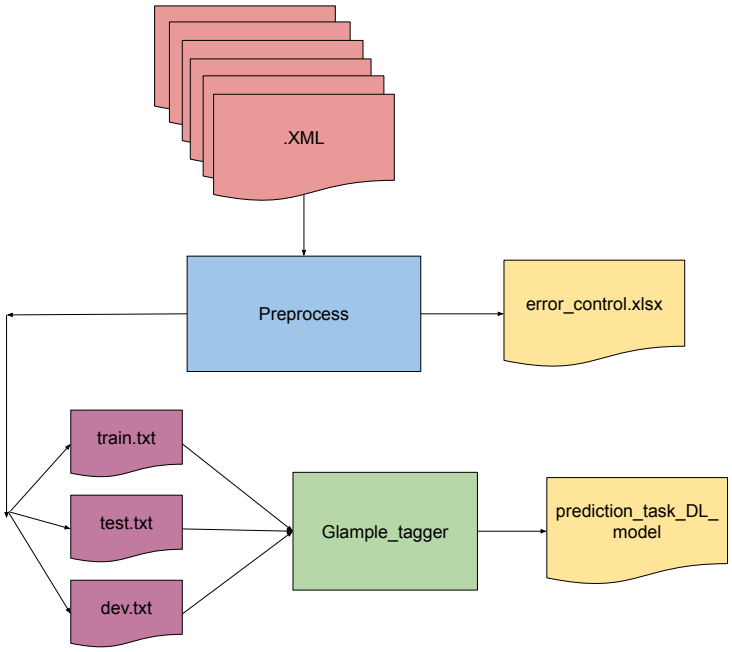

Figura 18: Arquitectura del preprocesado.

5.2. Diseño del Clasificador

A continuación se expone la arquitectura interna del modelo DL utilizado por el sistema Glample tagger 19, en nuestro caso utilizaremos 'word embeddings' pre entrenados para maximizar la eficacia del entrenamiento, asimismo utilizaremos la conexión bidireccional de las RNN LSTM, de esta manera en su última representación contendremos el compendio de la información contenida en la secuencia. Esta arquitectura constituye un gran avance en el estado del arte puesto que logra modelar secuencias con extremada eficacia.

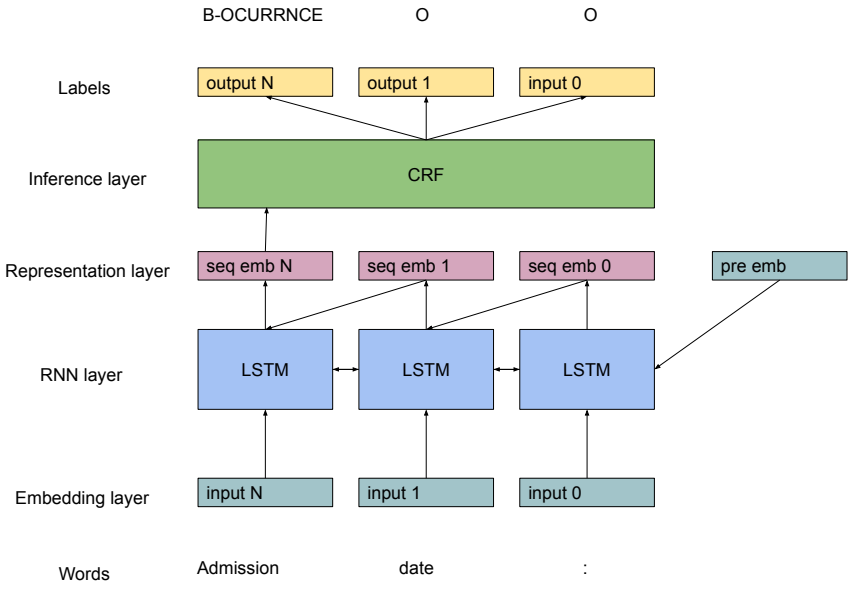

Figura 19: Arquitectura interna del sistema Glample Tagger.

Por último dicha representación de la secuencia dada se proveerá a un modelo 'conditional random fields' CRF [26], sistema especialista en etiquetado de secuencias y que cuenta con gran sinergia en la arquitectura debido a su capacidad de predicción no lineal.

5.3. Diseño del visualizador

Por último se expone la arquitectura utilizada para visualizar la información provista de los modelos DL pre entrenados para cierto archivo (.txt) de entrada 20. En esta arquitectura se realizan dos tareas, primeramente se solicitan predicciones a los modelos DL pre entrenados para obtener información del archivo de entrada, posteriormente se interpretan las predicciones de manera conjunta para generar dinámicamente un archivo (.html) de salida con toda la información compendiada, durante el desarrollo de la visualización no centraremos en ello.

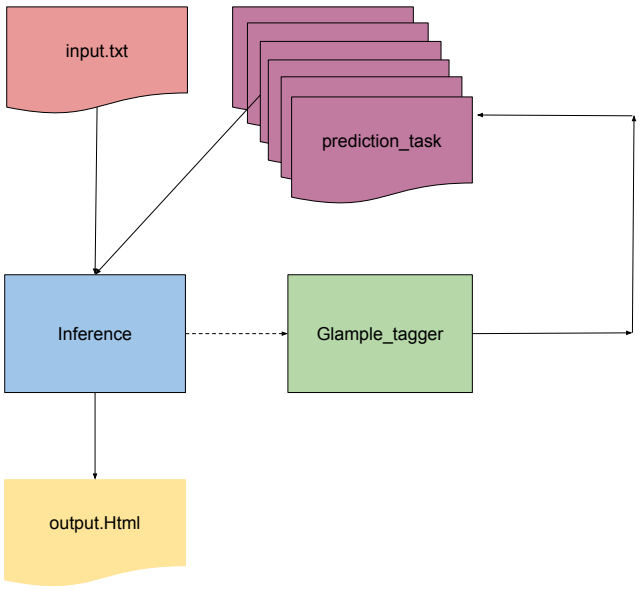

Figura 20: Arquitectura de la visualización.

6. Desarrollo

En este apartado se describe la labor de programación llevada a cabo para la generación del prototipo de alta fidelidad final.

6.1. Preprocesado

Constituye la tarea estratégica del sistema, puesto que la propagación de errores desde este nivel hasta los superiores es acumulativo, y por tanto los principales esfuerzos se realizaron en esta fase.

6.1.1. Sprint 0

En este paso se genero un sistema primitivo de procesamiento capaz de generar un archivo (.xlsx) con los datos contenidos en todo en directorio (.XML), de esta manera logramos centralizar la información y aplicar técnicas PLN a la información de las etiquetas almacenadas, además los datos relacionados (Tlink, Sectime 7, Timex3 6 y Event 5) con cada etiqueta se convertían a un espacio numérico. Esta idea era útil para su utilización en sistemas primitivos SVM [17, Cap 14.5] ya que su carácter es realizar fronteras de decisión, y esto es más fácil con información numérica. Aunque para su utilización en sistemas NN resulta una representación incompatible.

6.1.2. Sprint 1

Una vez realizada la primera aproximación del problema se procedió a la transición entre un modelo numérico tradicional y las arquitecturas NN, en esta fase se centraron los esfuerzos en la conversión de la información asociada a cada texto, la conversión se realizaría a formato IOB, por aquel entonces se formuló la idea de utilizar diccionarios para el etiquetado, en este caso , erróneamente, se decidió etiquetar a nivel de letra. De esta manera se lograba gran precisión a la hora de representar, pero a costa de la incapacidad de aprender las ristras IOB generadas, ya que estas eran demasiado grandes para su cómputo. En este momento se modificó la idea y se trato el etiquetado IOB a nivel de palabra.

6.1.3. Sprint 2

El modulo de preprocesado cuenta con tres tareas principales, primeramente procesar el directorio de archivos (.XML) para generar la representación intermedia (.XLSX), después debe generar diccionarios internos que relacionen la información procesada para generar el conjunto de entrenamiento que se hubiera especificado, finalmente devolverá el fichero intermedio y los ficheros train, test y dev a la ruta especificada para su utilización en Glample tagger.

Para cada tipo de etiqueta contenida en cada archivo (.XML), es decir: Tlink ('OVERLAP', 'BEFORE' o 'AFTER', respecto a 'Doctime 0'), Sectime 7, Timex3 6 y Event 5, se procesan todas las etiquetas en conjunto y se almacenan todos sus datos asociados 30.

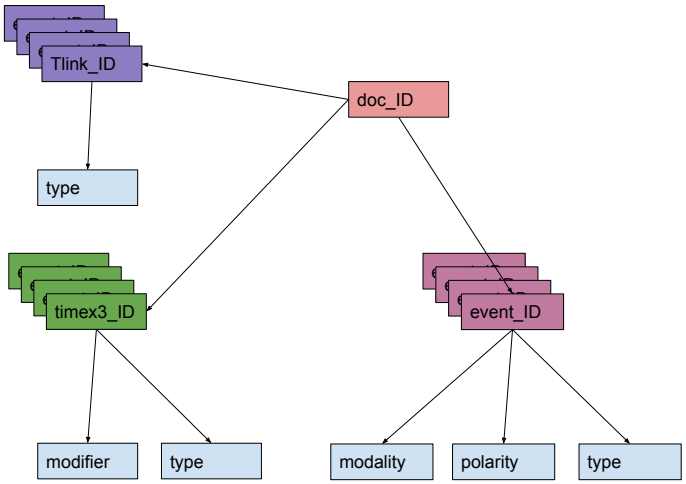

Figura 21: Estructura de información.

Una vez se ha generado la estructura de información convenientemente para el conjunto de textos, se procede a generar el conjunto de entrenamiento especificado en formato IOB, asociando cada etiqueta a una palabra por cada linea de un texto plano, el caso de etiquetado event-modality lo podemos apreciar en

el siguiente fragmento.

Admission I-FACTUAL
Date O
: O
2012-01-20 O
Discharge I-FACTUAL
Date O
: O
2012-01-23 O
Service O

Este etiquetado se lleva a cabo para cada tarea individual que se pretende aprender, en nuestro caso son 6 tareas: Event-type, Event-polarity, Event-modality, Timex3-type, Timex3-modifier y Tlink-type. Por último debemos dividir el conjunto en tres subconjuntos : train, test y dev. De esta manera estamos preparados para utilizar el sistema Glample tagger para entrenar modelos DL expertos en tareas de extracción de información clínica.

6.2. Clasificación

Esta tarea consiste en la aplicación de algoritmos ML a la información preprocesada anteriormente, al utilizar un sistema externo auto afinado logramos minimizar errores en esta fase.

6.2.1. Sprint 0

Este esfuerzo consistió en la generación del modelo SVM [17, Cap 14.5], para lo cual contábamos con una lista de etiquetas catalogadas en Event, Timex3 y Sectime, entonces surgió la necesidad de modelar matemáticamente dichas etiquetas, para ello se generó un sistema de WE Doc2Vec, en este caso primitivo se decidió modelizar los textos para generar un vocabulario interno y tomar de dicho vocabulario las representaciones de las etiquetas, esta aproximación mostraba grandes deficiencias dado que no se modelizaban todas las etiquetas existentes. En este punto se añadió un sistema WE de representación 'TFIDF' para las etiquetas desconocidas por 'Doc2Vec', el resultado de estas transformaciones alimentaba el modelo SVM logrando los primeros resultados. Este método era erróneo dado que no es recomendable la utilización de dos modelos de representación diferentes en una tarea de predicción, de ahí la baja precisión de los resultados, además se entendió que la representación 'Doc2Vec' era demasiado genérica y se requerían representaciones a nivel de palabra 'Word2Vec'.

6.2.2. Sprint 1

En esta fase se entendió que, la tarea crítica para la correcta predicción de las tareas era el sistema WE, y se programó la aproximación 'Word2vec', asimismo dadas las recomendaciones se inicializó dicho sistema con 'word embeddings' pre entrenados, esto fue un hito crucial para la mejora de desempeño en el sprint posterior, dado que se entendió el significado de la representación en un sistema unificado de NN. Por otro lado se genero una aproximación de la arquitectura 19 mediante las librerías 'Keras' y 'Tensorflow', era un sistema funcional pero con bajo desempeño debido a la carencia de conocimientos necesarios para su afinamiento.

6.2.3. Sprint 2

Entonces se sugirió la utilización del sistema Glample tagger, el cual contaba con la arquitectura deseada y además sistemas internos de afinamiento, en este momento se utilizaron los 'word embeddings' pre entrenados con los que se contaba para la investigación, este hecho fue un éxito dado que contaban con al rededor del 80 % del vocabulario a modelizar. Para realizar el entrenamiento del modelo que etiquete las palabras de los textos dados tendremos que hacer uso de un comando como:

python2.7 train.py –train path-train –dev path-dev –test path-test –pre_emb= path-emb –word_dim= dim-emb –lr_method= sgd –word_bidirect= 1

Internamente hemos modificado el número de 'epochs' (vueltas de entrenamiento) fijándolo en 10 para minimizar tiempos de ejecución, con el comando anterior especificamos el método de afinamiento 'sgd' (stochastic gradient descent) y el uso de 'Bidirectional' LSTM. Estos modelos entrenados son almacenados en la carpeta 'models/' de la librería y posteriormente se utilizan en el módulo de visualización.

Para realizar el etiquetado de un texto tendremos que hacer uso de un comando como:

python2.7 tagger.py –model path-modelo –input path-txt-entrada –output path-txt-salida

El texto de entrada no tiene porque seguir ningún formato como por ejemplo:

Admission Date :
2016-07-18
Discharge Date :
2016-07-21
Service :
MEDICINE

Aplicando el comando de etiquetado generamos un texto de salida similar al siguiente ejemplo de Event-modality:

Admission__B-FACTUAL Date__O :__O
2016-07-18 __O
Discharge__B-FACTUAL Date__O :__O
2016-07-21__O
Service__O :__O
MEDICINE__O

Llegados a este punto contamos con la capacidad de automáticamente etiquetar las palabras de un texto para las seis tareas expuestas previamente, y el etiquetado cuenta con la misma estructura inicial del texto y un cómodo separador de palabras y etiquetas (__).

Dado que los procedimientos expuestos consisten el la ejecución de lineas de comando, se tuvo que idear un mecanismo automático para ejecutarlas programáticamente, basándonos en nuestros conocimientos sobre hilos de ejecución indagamos sobre mecanismos de Python 3.6 capaces de generarlos y la respuesta residía en la librería nativa 'subprocess' la cual ofrecía una interfaz para la creación y control de hilos, parecida a la tecnología 'Web Workers' de JavaScript por lo que fue sencilla su implementación.

6.3. Inferencia

La inferencia consiste en la visualización de los resultados de aplicar un modelo ML pre entrenado sobre un conjunto de datos. El módulo mostrará un archivo (.html) que junto a un (.css) permite una fácil lectura clínica, además ofrece toda la información estimada por por los seis modelos DL expertos tratados anteriormente.

6.3.1. Sprint 0

En este caso primitivo la inferencia consistía en un apoyo a la interpretación de los resultados teóricos del sistema SVM mediante la librería 'Mathplotlib', se mostraban las distribuciones de datos Event 22, Timex3 23 y Sectime 24. Por otro lado se describían las predicciones realizadas, gracias a este acercamiento se pudieron detectar los fallos mencionados anteriormente en la representación, y que no era el sistema indicado para lograr el mejor desempeño en la presente tarea de extracción, esto lo trataremos en el apartado de pruebas.

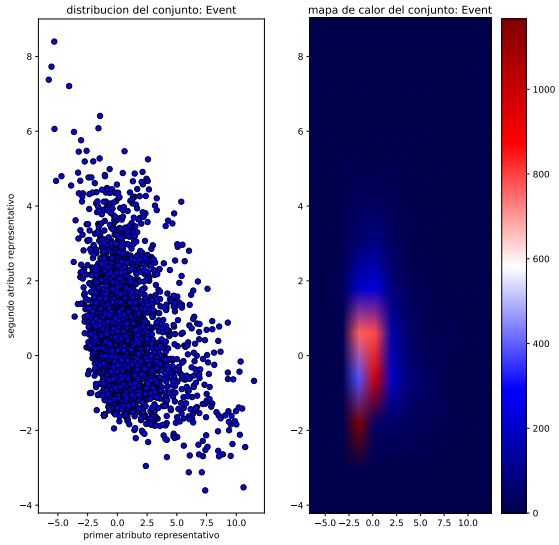

Figura 22: Distribución Event.

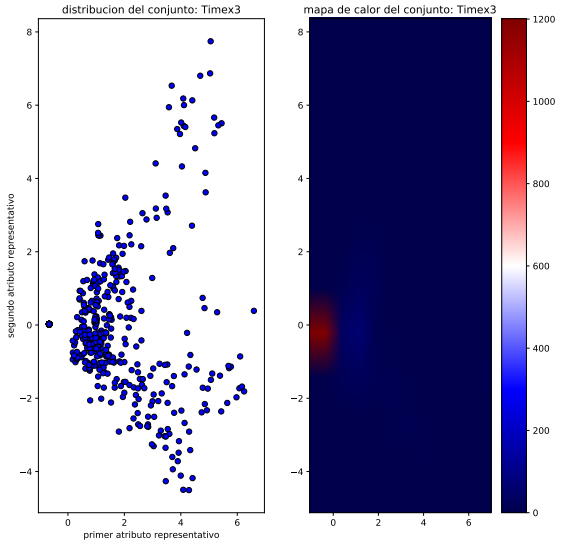

Figura 23: Distribución Timex3.

Las distribuciones que mostramos en este apartado se solaparán para ofrecer una mejor comprensión del espacio de representación, durante dicho solapamiento se entiende que no es posible el entrenamiento conjunto de las tres clases dado el gran des balance entre ellas, bajo esa premisa se decide entrenar diferentes expertos por tarea, además se empiezan los esfuerzos para ofrecer información comprensible al usuario, esto consistió en generar el primer archivo (.html) dinámico mediante Python.

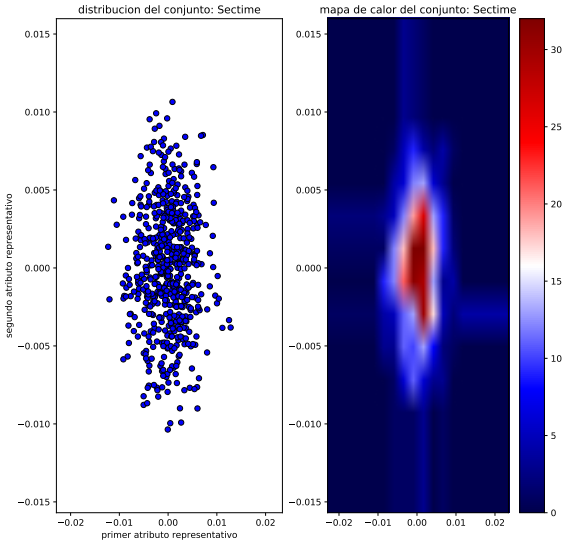

Figura 24: Distribución Sectime.

6.3.2. Sprint 1

En esta sección se trata la generación del primer archivo (.html) dinámico basado en el etiquetado a nivel de carácter, la visualización de las etiquetas era muy exacta dada la precisión con la que se estaba trabajando, llegados a este punto se determinó que el soporte (.html) era el indicado para mostrar la información debido a su facilidad de programación y su alta capacidad para contener información estructurada. De hecho esta parte se reutilizó posteriormente en el sistema final de inferencia como módulo de visualización, aunque recibió ciertas modificaciones para soportar la visualización por palabras y además añadir información adicional. Llegados a la introducción de la reutilización mencionamos que para metodologías basadas en iteraciones es conveniente la 'Programación orientada a objetos', la cual consiste en programar módulos que internamente cuentan con funciones, y así interactuar con los módulos directamente. La conveniencia se sostiene en que, posteriormente dichas funciones que componen los módulos, son fácilmente incluidas en otros módulos del sistema en caso de requerirse, y esto permite ahorrar tiempo en las pruebas de campo.

6.3.3. Sprint 2

En este apartado explicaremos como aprovechar de manera conjunta la clasificación manual expuesta previamente y además generar una visualización de la información provista por los diferentes modelos expertos.

la idea es generar visualizaciones dinámicamente sobre una entrada de texto plano especificada por lo que debemos solicitar programáticamente la generación de los seis modelos expertos en cada visualización, para ello utilizaremos subprocesos de python con instrucciones como:

```
args = shlex.split(comando manual de etiquetado)
process = subprocess.Popen(args)
try:
.       outs, errs = process.communicate(timeout=15)
except subprocess.TimeoutExpired:
.       process.kill()
.       outs, errs = process.communicate()
```

Debemos solicitar los seis modelos programáticamente y asegurarnos que sus archivos de salida se encuentren en la misma carpeta, posteriormente tomaremos dichos archivos con los diferentes etiquetados ya realizados y realizaremos una lista de séxtuplas de manera que para cada palabra contemos con toda la información estimada por los modelos DL expertos, como por ejemplo:

('follow_-B-OCCURRENCE', 'follow_-O', 'follow_-I-POS', 'follow_-O', 'follow_-B-AFTER', 'follow_-I-FACTUAL')

De esta manera en una única vuelta somos capaces de reconstruir toda la información conocida sobre el texto y podremos reconstruirla fijándonos en las coincidencias de los expertos de cada tipo y mostrar información adicional además de resaltar las palabras y oraciones relevantes.

Siguiendo con el ejemplo anterior y recordando la información de los eventos 30 , podemos decir que 'follow' es un evento clínico puesto que se le asigna:'type', 'polarity' y 'modality', además de esta información sobre la palabra, añadiremos adicionalmente que se refiere históricamente a un futuro evento debido a la estimación 'after', lo que coincide con el siguiente ejemplo 25 .

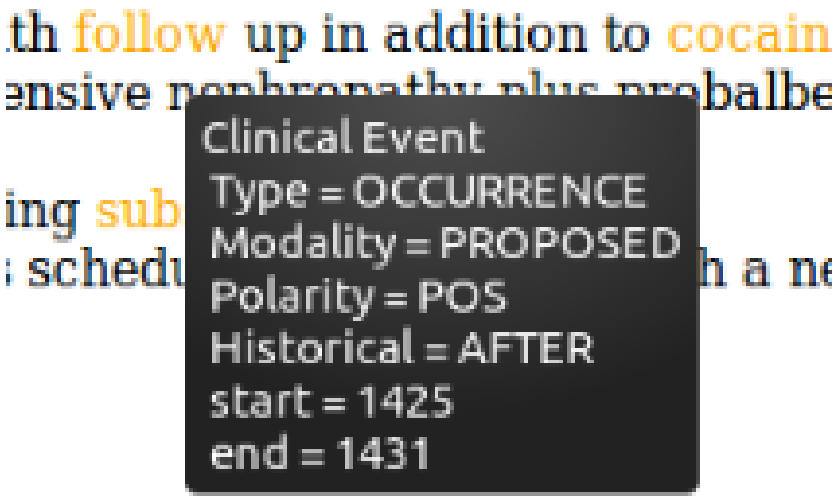

Figura 25: Ejemplo de ensamblado de predicciones.

Este sistema de ensamblado a pesar de lograr buenos resultados no es óptimo en cuanto a interpretación, puesto que, la decisión de que información tomar para oraciones siempre recae en la primera palabra, y tal vez no sea la más representativa. Y en cuanto a recursos, nos parece óptimo porque logra desglosar toda la información relevante con coste computacional lineal, lo que se traduce a un proceso por texto de 40s.

7. Pruebas

En esta sección se exponen los resultados de entrenamiento de los seis sistemas generados y se comparan con los compendiados en 12.

7.1. Sprint 0

A continuación se muestran los resultados del 'baseline' SVC realizado, en las figuras 26 27 y 28, el gráfico de la izquierda coincide con la representación unificada de las tres distribuciones Event(morado), Timex3 (naranja) y Sectime (azul), es decir, las distribuciones de 'word embbedings' según sus dos componentes principales, como podemos observar la mayor distribución es Event seguida de Timex3 y Sectime, respectivamente. Esto se traduce en el des balance entre las tres clases, lo que supone que la menor sea vagamente aprendida por el sistema 'baseline', como podemos observar en 28, concretamente en su parte derecha, no existe predicción alguna para sectime.

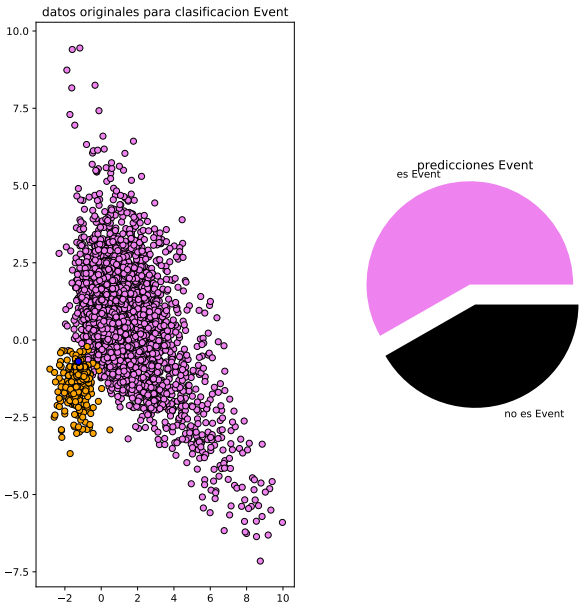

Figura 26: Predicciones SVC para Event.

Los resultados de evaluación de este sistema, por tanto, no pueden ser buenos, primeramente por la representación descompensada del problema, y además por el carácter lineal del clasificador SVC, en 29 queda ratificado, siendo poco mejor que 'tirar una moneda al aire'. Pero nos da una perspectiva diferente a la problemática, antes entendíamos que etiquetar dependía de la oración/palabra a etiquetar o en otras palabras su 'word embedding', ahora entendemos que depende del contexto de la oración, y para representar una oración requerimos resolver un problema añadido que consiste en modelar matemáticamente una palabra y su contexto, aquí se explica porque las RNN son necesarias para esta tarea, consiguen contener tanto la información de palabra, como la de las demás que la rodean en un único 'word embedding'.

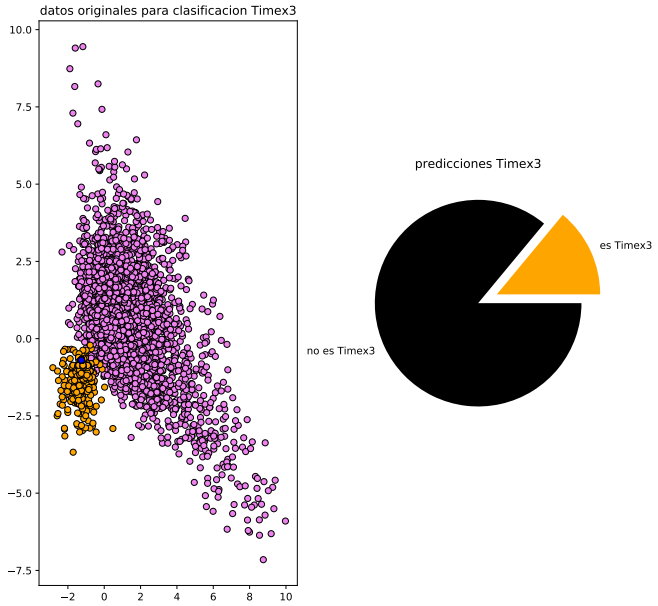

Figura 27: Predicciones SVC para Timex3.

Por otro lado, las distribuciones sugieren que las funciones lineales no son recomendables, ya que, los tres conjuntos se solapan, y por tanto requieren funciones no lineales para desambiguar en casos de decisión compleja, para minimizar este fenómeno decidimos dividir la tarea de decisión en distintos expertos capaces de dominar regiones del espacio de representación.

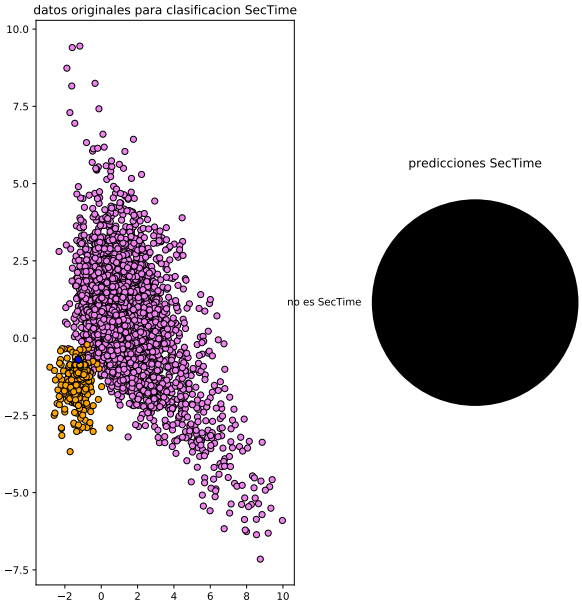

Figura 28: Predicciones SVC para Sectime.

En la siguiente Figura 29 podemos observar la estrategia de división de expertos, en este caso primitivo son 3, uno por cada etiqueta estudiada, la estrategia utilizada para su evaluación es '10 fold stratified cross validation', es decir del conjunto total se toman distintas distribuciones aleatorias para entrenar y validar, esta aleatoriedad a su vez busca favorecer las clases minoritarias para paliar el des balance. Asimismo cabe destacar que se ha aplicado una reducción dimensional 'principal component analysis' PCA, esto busca trabajar con un espacio bidimensional, ya que la dimensión de los 'word embeddings' es de 300 atributos.

```
Nota: estos predictores constituyen expertos para cada tipo de
etiqueta donde 0 coincide con pertenecer al tipo y 1 coincide con no
pertenecer al tipo.

-Event:

Log loss: 0.2615912902053098
Accuracy: 0.8508958777095602
f-score : 0.7848101265822786
Precision: 0.6514560192134494

-Timex3:

Log loss: 0.30780661219967953
Accuracy: 0.7819822077433906
f-score : 0.8685800604229608
Precision: 0.8375819373634378

-SecTime:

Log loss: 0.24062412631782654
Accuracy: 0.9256985340182935
f-score : 0.9614158370746307
Precision: 0.9256985340182935
```

Figura 29: Evaluación del sistema Baseline.

7.2. Sprint 1

En la sección actual se tratan los resultados obtenidos del sistema desarrollado mediante 'Keras' y 'Tensorflow', en este caso se entrenaron todas las etiquetas de 'type' de manera conjunta lo cual posteriormente se dividió en dos expertos (Event-type y Timex3-type), y una serie de reglas manuales para Sectime - type. Dado el erróneo enfoque, y además la carencia de capacidad para afinar el sistema haciendo uso del algoritmo 'sgd', propiciaron que los resultados no fueran demasiado precisos aunque se demuestra su gran capacidad para etiquetar secuencias.

Poco después de la consecución de este hito del proyecto, se recomendó el uso del sistema Glample tagger, el cual además de contar con la arquitectura necesaria, disponía de un sistema de auto afinado automático para garantizar su desempeño. Sus resultados se observan en la siguiente sección.

```
processed 25906 tokens with 4787 phrases; found: 4778 phrases; correct: 3332.
accuracy:  84.47%; precision:  69.74%; recall:  69.61%; FB1:  69.67
        ADMISSION: precision: 100.00%; recall:  88.46%; FB1:  93.88  46
    CLINICAL_DEPT: precision:  74.34%; recall:  80.38%; FB1:  77.24  226
             DATE: precision:  73.39%; recall:  79.68%; FB1:  76.41  342
       DISCHARGE: precision: 100.00%; recall:  86.79%; FB1:  92.93  46
        DURATION: precision:  60.87%; recall:  60.34%; FB1:  60.61  115
      EVIDENTIAL: precision:  62.56%; recall:  66.18%; FB1:  64.32  219
       FREQUENCY: precision:  55.56%; recall:  53.03%; FB1:  54.26  63
      OCCURRENCE: precision:  58.37%; recall:  57.56%; FB1:  57.96  783
         PROBLEM: precision:  71.59%; recall:  69.39%; FB1:  70.48  1232
            TEST: precision:  74.46%; recall:  73.27%; FB1:  73.86  740
            TIME: precision:  38.46%; recall:  45.45%; FB1:  41.67  13
       TREATMENT: precision:  71.77%; recall:  72.69%; FB1:  72.23  953

Score on dev: 68.98000
Score on test: 69.67000
```

Figura 30: Evaluación modelo DL extracción 'type'.

7.3. Sprint 2

En esta sección tratamos los resultados de los seis modelos DL utilizados por el prototipo de alta fidelidad, y lo contrastamos con el compendio de evaluaciones, realizado en la anterior revisión del reto internacional 'i2b2' 12, en la primera tabla 31 podemos observar un 'accuracy' sobre 'dev' de 0.8765 y sobre 'test' de 0.8851, ambos superiores al anterior registrado de 0.86 para 'Event Type accuracy'.

```
processed 15818 tokens with 2680 phrases; found: 2854 phrases; correct: 2055.
accuracy:  87.65%; precision:  72.00%; recall:  76.68%; FB1:  74.27
    CLINICAL_DEPT: precision:  65.45%; recall:  80.65%; FB1:  72.25  191
      EVIDENTIAL: precision:  75.81%; recall:  77.05%; FB1:  76.42  124
      OCCURRENCE: precision:  62.37%; recall:  62.37%; FB1:  62.37  558
         PROBLEM: precision:  74.16%; recall:  81.35%; FB1:  77.59  894
            TEST: precision:  80.49%; recall:  81.41%; FB1:  80.95  446
       TREATMENT: precision:  72.70%; recall:  79.12%; FB1:  75.77  641

processed 25906 tokens with 4174 phrases; found: 4223 phrases; correct: 3129.
accuracy:  88.51%; precision:  74.09%; recall:  74.96%; FB1:  74.53
    CLINICAL_DEPT: precision:  79.56%; recall:  85.65%; FB1:  82.49  225
      EVIDENTIAL: precision:  65.43%; recall:  76.81%; FB1:  70.67  243
      OCCURRENCE: precision:  64.31%; recall:  60.83%; FB1:  62.52  751
         PROBLEM: precision:  73.94%; recall:  75.69%; FB1:  74.81  1301
            TEST: precision:  79.59%; recall:  78.32%; FB1:  78.95  740
       TREATMENT: precision:  78.61%; recall:  80.45%; FB1:  79.52  963

Score on dev: 74.2700
Score on test: 74.5300
```

Figura 31: Evaluación extracción Event-type.

En la siguiente tabla 32 podemos observar un 'accuracy' sobre 'dev' de 0.8948

y sobre 'test' de 0.8938, ambos superiores al anterior registrado de 0.86 para 'Event Polarity accuracy'.

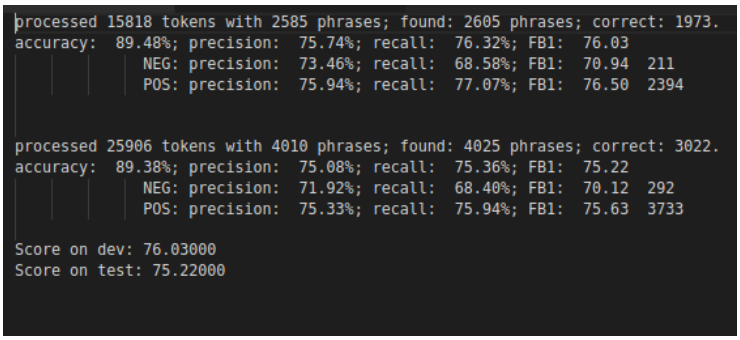

Figura 32: Evaluación extracción Event-polarity.

En la siguiente tabla 33 podemos observar un 'accuracy' sobre 'dev' de 0.8901 y sobre 'test' de 0.8885, ambos superiores al anterior registrado de 0.86 para 'Event Modality accuracy'.

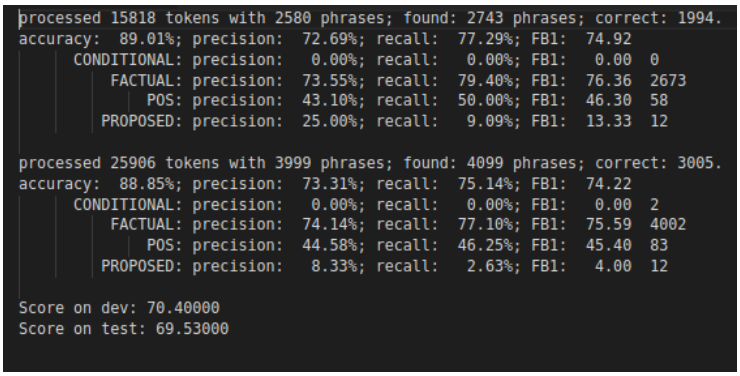

Figura 33: Evaluación extracción Event-modality.

En la siguiente tabla 34 podemos observar un 'accuracy' sobre 'dev' de 0.9855 y sobre 'test' de 0.9852, ambos superiores al anterior registrado de 0.89 para 'Timex3 Type accuracy'.

```
processed 15818 tokens with 395 phrases; found: 380 phrases; correct: 310.
accuracy:  98.55%; precision:  81.58%; recall:  78.48%; FB1:  80.00
              DATE: precision:  84.62%; recall:  83.02%; FB1:  83.81  260
          DURATION: precision:  69.49%; recall:  69.49%; FB1:  69.49  59
         FREQUENCY: precision:  78.95%; recall:  78.95%; FB1:  78.95  57
              TIME: precision: 100.00%; recall:  28.57%; FB1:  44.44  4


processed 25906 tokens with 621 phrases; found: 656 phrases; correct: 511.
accuracy:  98.52%; precision:  77.90%; recall:  82.29%; FB1:  80.03
              DATE: precision:  84.79%; recall:  86.38%; FB1:  85.58  434
          DURATION: precision:  64.66%; recall:  73.50%; FB1:  68.80  133
         FREQUENCY: precision:  62.82%; recall:  74.24%; FB1:  68.06  78
              TIME: precision:  72.73%; recall:  66.67%; FB1:  69.57  11


Score on dev: 80.00000
Score on test: 80.03000
```

Figura 34: Evaluación extracción Timex3-type.

En la siguiente tabla 35 podemos observar un 'accuracy' sobre 'dev' de 0.9858 y sobre 'test' de 0.9845, ambos superiores al anterior registrado de 0.89 para 'Timex3 Modifier accuracy'.

```
processed 15818 tokens with 394 phrases; found: 404 phrases; correct: 318.
accuracy:  98.58%; precision:  78.71%; recall:  80.71%; FB1:  79.70
           APPROX: precision:  57.14%; recall:  41.67%; FB1:  48.19  35
              END: precision:  80.00%; recall:  72.73%; FB1:  76.19  10
             MORE: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
               NA: precision:  81.14%; recall:  86.59%; FB1:  83.78  350
            START: precision:  66.67%; recall: 100.00%; FB1:  80.00  9


processed 25906 tokens with 621 phrases; found: 651 phrases; correct: 496.
accuracy:  98.45%; precision:  76.19%; recall:  79.87%; FB1:  77.99
           APPROX: precision:  57.58%; recall:  32.20%; FB1:  41.30  33
              END: precision:  66.67%; recall:  57.14%; FB1:  61.54  6
           MIDDLE: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
             MORE: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
               NA: precision:  77.72%; recall:  86.89%; FB1:  82.05  597
            START: precision:  64.29%; recall:  56.25%; FB1:  60.00  14


Score on dev: 79.70000
Score on test: 77.99000
```

Figura 35: Evaluación extracción Timex3-modifier.

En la siguiente tabla 36 podemos observar un 'f-measure' sobre 'dev' de 0.6262 y sobre 'test' de 0.6319, ambos inferiores al anterior registrado de 0.69 para 'TLINK F measure'.

```
processed 9032 tokens with 1709 phrases; found: 1747 phrases; correct: 1082.
accuracy:  83.00%; precision:  61.93%; recall:  63.31%; FB1:  62.62
             AFTER: precision:  31.91%; recall:  17.44%; FB1:  22.56  47
            BEFORE: precision:  65.11%; recall:  72.38%; FB1:  68.55  1453
           OVERLAP: precision:  48.99%; recall:  38.29%; FB1:  42.98  247

processed 15176 tokens with 2740 phrases; found: 2973 phrases; correct: 1805.

accuracy:  82.53%; precision:  60.71%; recall:  65.88%; FB1:  63.19
             AFTER: precision:  25.00%; recall:   1.42%; FB1:   2.68  8
            BEFORE: precision:  62.44%; recall:  74.27%; FB1:  67.84  2556
           OVERLAP: precision:  50.61%; recall:  46.00%; FB1:  48.20  409


Score on dev: 62.62000
Score on test: 63.19000
```

Figura 36: Evaluación extracción Tlink-type.

Consideramos, por ende que la investigación fue exitosa y logró actualizar el estado del arte en cuanto a extracción de eventos y expresiones temporales clínicas, por otro lado entendemos que la extracción de relaciones temporales es mejorable, y junto con el atributo 'value' de las expresiones temporales, constituirán parte de los esfuerzos futuros del presente trabajo.

por otro lado, consideramos que los modelos no se encuentran debidamente entrenados, y esperamos que con los conocimientos necesarios, y herramientas adecuadas, los resultados del sistema actual puedan mejorar notablemente, esto se sostiene en el hecho de que cada modelo se ha entrenado únicamente con 10 'epochs', irrisorio para la envergadura del sistema.

8. Prototipo

En este apartado se describen los distintos prototipos realizados para la consecución de la investigación, se explica que cambia con respecto al anterior y porque. Los prototipos realizados son los siguientes:

8.1. Prototipo inicial

En el siguiente flujo se detalla como se planteó inicialmente la idea a realizar, se enfatizó en el resultado final y se genero un método intermedio abstracto para conseguirlo, esta lejos de la realidad.

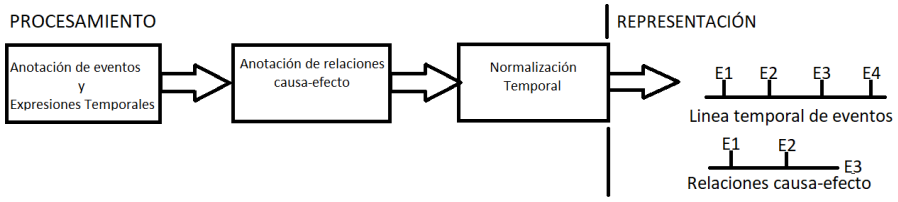

Figura 37: Sistema Prototipo inicial.

De esta idea 'en bruto' se basa el posterior desarrollo de prototipos hasta llegar al de alta fidelidad donde podemos apreciar que se mantiene la temporalidad en los eventos y expresiones temporales pero se desiste de los esfuerzos en conseguir relaciones causa-efecto debido a la no disponibilidad de información para realizar dicho estudio.

8.2. Prototipo baja fidelidad

En este prototipo se determinan los módulos necesarios para la arquitectura así como los ficheros de entrada y salida de cada módulo, esto es debido al desarrollo empírico realizado el cual sugiere la necesidad modular y que aparentemente se generaran los archivos iniciales, intermedios y finales mencionados, esta entre la realidad y las abstracciones del proyecto.

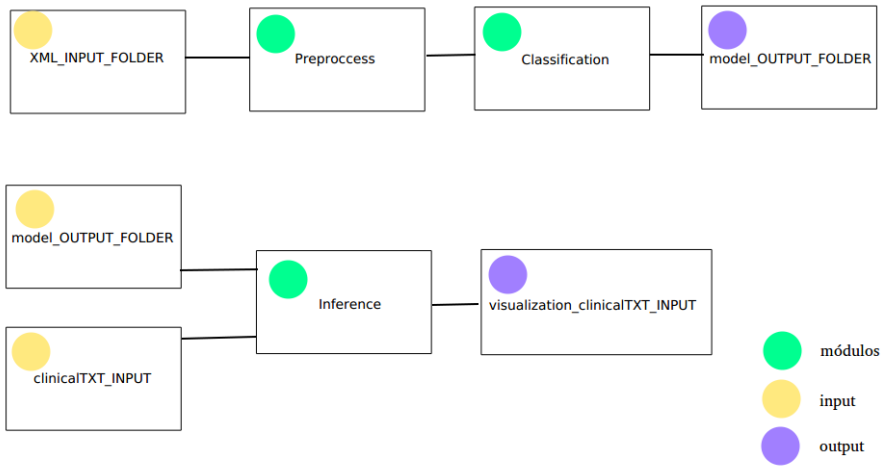

Figura 38: Sistema baja fidelidad.

Tras varias pruebas tratando de desarrollar un sistema para la predicción de valores en las Expresiones temporales y Secciones temporales se desiste en dichos esfuerzos debido a la alta complejidad que supone integrar dicho sistema con el estudiado en el presente artículo, por otro lado se presume que en dicho reto existen dos tareas, por un lado la predicción de fecha-hora (13/02/1997T15:00:00), y por otro lado lado la predicción de frecuencias y duraciones (q.b.i), lo cual se considera otro estudio aparte, por tanto en el prototipo de alta fidelidad los 'values' se definen por defecto como 'undefined'.

8.3. Prototipo alta fidelidad

Este prototipo coincide con la apariencia final del sistema, funcional aunque con errores debido al bajo afinamiento del sistema, muestra gran potencial en cuanto a la extracción de eventos (Naranja), secciones temporales (Azul) y expresiones temporales (Verde).

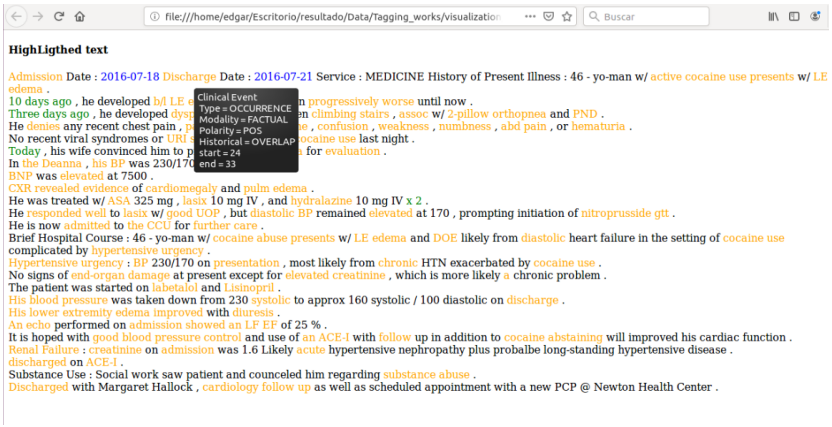

Figura 39: Prototipo alta fidelidad.

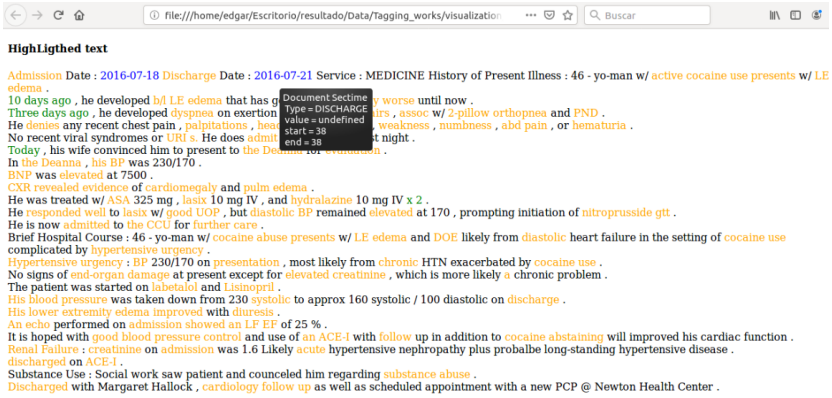

Figura 40: Prototipo alta fidelidad.

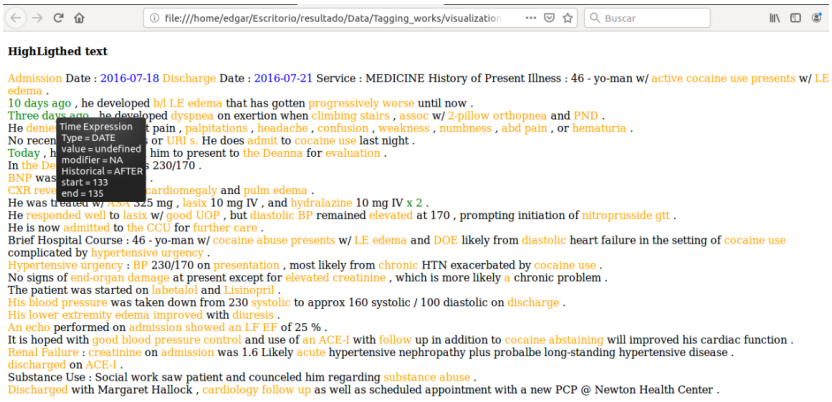

Figura 41: Prototipo alta fidelidad.

Como podemos apreciar en 39 el ratón se encuentra sobre la palabra 'Admission' de la cual se despliega un menú contextual para informarnos de que es un Evento clínico catalogado como ocurrencia de facto, positiva y que sucede durante la actual historia clínica. En 40 el ratón se encuentra sobre la palabra ' 2016-07-21' de la cual se despliega un menú contextual para informarnos de que es una sección temporal de tipo alta. Por último en 41 el ratón se encuentra sobre la frase ' 10 days ago' de la cual se despliega un menú contextual para informarnos de que es una expresión temporal de tipo fecha con valor concreto (sin modificador) e interpreta que sucederá en los días posteriores a la historia clínica.

Este problema nos resulta lógico debido a la manera de ensamblar las predicciones, solo se tiene en cuenta la primera palabra '10' valor que asocia con futuro cercano despreciando la última palabra 'ago' sobre la que recae realmente la decisión.

9. Conclusiones y trabajo futuro

Primeramente nos centraremos en las metas y objetivos del proyecto, de las cuales comentar que consideramos cumplidas todas las metas a excepción de Revisión y corrección puesto que hemos tenido finalmente que ceñirnos a la literatura técnica informática casi en exclusividad, ante la imposibilidad de introducir conceptos interdisciplinares por falta de conocimiento, asimismo la meta de Reunión, resultó ser demasiado subjetiva para una correcta valoración de su desempeño, aunque su segunda componente podemos darla por satisfecha, ya que el trabajo se encuentra validado. En cuanto a los objetivos, los consideramos cumplidos a excepción de la segunda componente de Revisión y corrección ya que no encontramos ninguna herramienta capaz de analizar la semántica del escrito, y por tanto lo consideramos ambicioso y difícil de medir, por otro lado el objetivo de la Presentación es medible pero 'a posteriori', por lo que lo consideramos impreciso pero con capacidad de demostrar nuestras aspiraciones ultimas del trabajo. En cualquier caso podemos recalcar el orgullo que supone haber alcanzado todos los demás objetivos y metas.

En cuanto a la elección de herramientas, tal vez se deberían tratar alternativas para la generación de 'word embeddings' a modo de apoyo para el uso de este texto como guía de ensayo NLP.

Nos sentimos orgullosos de la planificación temporal llevada a cabo a salvedad del sobre coste invertido, aunque lo consideramos lógico dada la complejidad del trabajo realizado. Asimismo la evaluación económica ha resultado satisfactoria, aunque en apartado de retorno de la inversión se considera anexo de valor para el trabajo más que un modelo económico capaz de devolver la inversión realizada, no consideramos la investigación con 'fines de lucro', sino con 'fines de desarrollo' y 'posibilidad de lucro'.

Por otro lado, en cuanto a riesgos podemos asegurar que en todo momento han sido contemplados, y se ha minimizado sus impactos exitosamente, la incapacidad para realizar una tarea fue el más acuciante, y se desvaneció gracias a la conciencia sobre el, y la capacidad de maniobra de las metodologías ágiles.

En cuanto al estado del arte nos sentimos seguros de ofrecer toda la información necesaria para el presente estudio, aunque no hemos compendiado la totalidad de artículos existentes, y es que preferimos condensar aquellos útiles y relacionados con el trabajo en su totalidad, frente a otros meramente útiles para tareas concretas del trabajo. Como ya hemos mencionado, son fuente de inspiración para nuevas creaciones, y ofrecemos aquellos más prometedores e inspiradores a nuestro entender.

En cuanto al desarrollo técnico y el análisis realizados nos sentimos enormemente plenos, dado que la calidad de enseñanza de la escuela es tal como dice ser, y poder realizar un trabajo práctico de esta envergadura nos permite sentirnos capaces de afrontar los retos futuros, además creemos que la labor anti escepticismo realizada tiene un gran valor añadido, ya que se concreta enormemente nuestro área de estudio permitiéndonos transmitir los conocimientos técnicos con mayor seguridad, y con apoyo de ejemplos.

El prototipo generado nos permite referirnos al concepto 'magia', y transformarlo en 'ciencia', ya que creemos cumple todos los métodos establecidos por la ciencia de la ingeniería, y los extrapola a la pseudo-ciencia de la minería de datos con espléndidos resultados.

Para concluir con la intervención, hablaremos sucintamente del trabajo futuro, en este caso nos decantamos por la mejora del prototipo en la extracción de relaciones temporales, dado que no se han optado por las mejores alternativas. Además, trataremos de entrenar un modelo más eficaz con los conocimientos y recursos del master Language Analysis and Processing (HAP/LAP), al que se aspira entrar con el presente trabajo, como último trabajo, y posiblemente, el más importante, tratare Edgar Andrés a titulo personal, de mejorar mis atrofiadas capacidades de trabajo en equipo, dado que como advertía acertada Aitziber Atutxa "si eres capaz de hacer esto solo, ¿ Que hará un grupo de diez como tu ?", y tras reflexión profunda, creo que debo respetar más el criterio ajeno para poder comenzar respetar el mio propio.

Edgar Andrés Santamaría

10. Agradecimientos

Me gustaría agradecer su participación a todos mis profesores de E.P.O y la E.S.O del colegio Madre de Dios Bilbao, su participación en cumplir mi sueño de ser 'ingeniero informático', desde chico me orientaron hacia las ciencias de la tecnología, y sobre todo, hacia conformarme como un hombre, puesto que una persona sin valores, es una persona vacía.

También agradecer a todos los profesores que he tenido el placer y honor, de conocer durante el grado de informática de gestión y sistemas de información, puesto que, me han dado la oportunidad de tener un futuro, de darle un sentido a mi lucha por sobrevivir.

Por supuesto, agradecer a todos mis amigos y conocidos, el poder compartir mi pasión por mi profesión con alguien, el poder charlar y sentirme cada vez más humano, el tener una mano cuando la oscuridad nos agobia, y el poder compartir lo que logro con el prójimo como el hace conmigo.

Finalmente me gustaría agradecer a mi familia, que por distancia que nos separe, algún día, estaremos lo más unidos posible, primero hay que ayudarse para poder ayudar. Y con especial deferencia, a mi madre Ana Santamaría Ruiz, sin la cual, hace mucho me hubiera perdido, hace mucho me hubiera roto, y hace mucho rendido. Gracias ama por existir, y darme ánimos cuando los demás me ven incapaz o poco cuerdo en mi manera de pensar, tu forma de ser me guié siempre.

Edgar Andrés Santamaría

Referencias

[1] Martín Abadi y col. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: `https://www.tensorflow.org/`.

[2] Jay Alammar. *word2vec*. `http://jalammar.github.io/illustrated-word2vec/`. 2019.

[3] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2014.

[4] Michael JA Berry y Gordon S Linoff. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2004.

[5] Steven Bird. *NLTK Documentation Release 3.2.5*. 2017.

[6] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[7] Vernon L Ceder, Kenneth McDonald y Daryl D Harms. *The quick Python book*. Manning, 2010.

[8] Colin Cherry y col. "A la Recherche du Temps Perdu: extracting temporal relations from medical text in the 2012 i2b2 NLP challenge". En: *Journal of the American Medical Informatics Association* 20.5 (2013), págs. 843-848.

[9] François Chollet y col. *Keras*. `https://keras.io`. 2015.

[10] Colah. *lstm*. `https://colah.github.io/posts/2015-08-Understanding-LSTMs/`. 2015.

[11] Weston J Collobert R. "A unified architecture for natural language processing: Deep neural networks with multitask learning". En: *Journal of the American Medical Informatics Association* (2008), págs. 160-167.

[12] Jian Pei Jiawei Han Micheline Kamber. *Data Mining Concepts and Techniques Third Edition*. Morgan Kaufmann Publishers is an imprint of Elsevier., 2013.

[13] Egoitz Laparra y col. "SemEval 2018 Task 6: Parsing Time Normalizations". En: *Proceedings of The 12th International Workshop on Semantic Evaluation*. 2018, págs. 88-96.

[14] Edward Ma. *doc2vec*. `https://towardsdatascience.com/understand-how-to-transfer-your-paragraph-to-vector-by-doc2vec-1e225ccf102`. 2018.

[15] Tom M. Mitchell. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.

[16] Andreas C. Müller y Sven Behnke. "pystruct - Learning Structured Prediction in Python". En: *Journal of Machine Learning Research* 15 (2014), págs. 2055-2060. URL: http://jmlr.org/papers/v15/mueller14a.html.

[17] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[18] Nils J Nilsson y Julio Fernández Biarge. *Principios de inteligencia artificial*. Díaz de Santos, 1987.

[19] Jon Patrick y Min Li. "High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge". En: *Journal of the American Medical Informatics Association* 17.5 (2010), págs. 524-527.

[20] Radim Řehůřek y Petr Sojka. "Software Framework for Topic Modelling with Large Corpora". English. En: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. http://is.muni.cz/publication/884893/en. Valletta, Malta: ELRA, mayo de 2010, págs. 45-50.

[21] Dipanjan Sarkar. *cbow*. https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-cbow.html. 2018.

[22] MINISTERIO DE EMPLEO Y SEGURIDAD SOCIAL. *boe*. https://www.boe.es/boe/dias/2018/03/06/pdfs/BOE-A-2018-3156.pdf. 2018.

[23] Weiyi Sun, Anna Rumshisky y Ozlem Uzuner. "Evaluating temporal relations in clinical text: 2012 i2b2 Challenge". En: *Journal of the American Medical Informatics Association* 20.5 (2013), págs. 806-813.

[24] Mayank Tripathi. *tfidf*. https://www.freecodecamp.org/news/how-to-process-textual-data-using-tf-idf-in-python-cd2bbc0a94a3/. 2018.

[25] Özlem Uzuner y col. "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text". En: *Journal of the American Medical Informatics Association* 18.5 (2011), págs. 552-556.

[26] hanna m. wallach. *crf*. http://www.inference.org.uk/hmw26/crf/. 2005.

[27] Ian H Witten y col. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[28] Yan Xu y col. "An end-to-end system to identify temporal relation in discharge summaries: 2012 i2b2 challenge". En: *Journal of the American Medical Informatics Association* 20.5 (2013), págs. 849-858.