



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

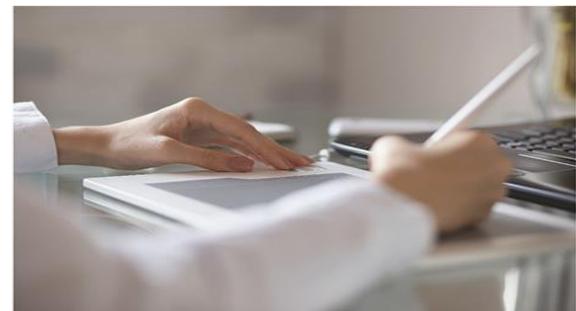
EDUCACIÓN
PROFESIONAL

Diplomado Big Data y Analítica de Datos 2021

Curso: Fundamentos Machine & Deep Learning

Educación Profesional Escuela de Ingeniería UC

Profesor: Rodrigo Sandoval



Contenidos en temas de NLP, NLU y NLG

Clase 6 NLP

- Pre-procesamiento de Texto
- Clasificación de texto
- Extracción de entidades

Clase 7 NLU & NLG

- Word Embedding
- Contextualización y Secuencia
- Transformers

Y seguimos hablando de Redes Neuronales Profundas para lenguaje



Definir Uds, en sus propias palabras:

¿En qué consiste el procesamiento de lenguaje natural (NLP)?

¿Qué ejemplos del día a día se les ocurre en NLP?



NLP – Procesamiento de Lenguaje Natural

La manipulación automática del lenguaje natural, como texto o habla, por software.

La extracción de información o datos útiles y significativos, desde texto natural, sin una estructura pre-definida.

1 min Me gusta Responder

Damian Vega
De que color y que marca es tu computadora?

Osbaldo Encinas
Negra hp

Hace un momento Me gusta Responder

Escribe una respuesta... 

5:09    ..  84% 

!

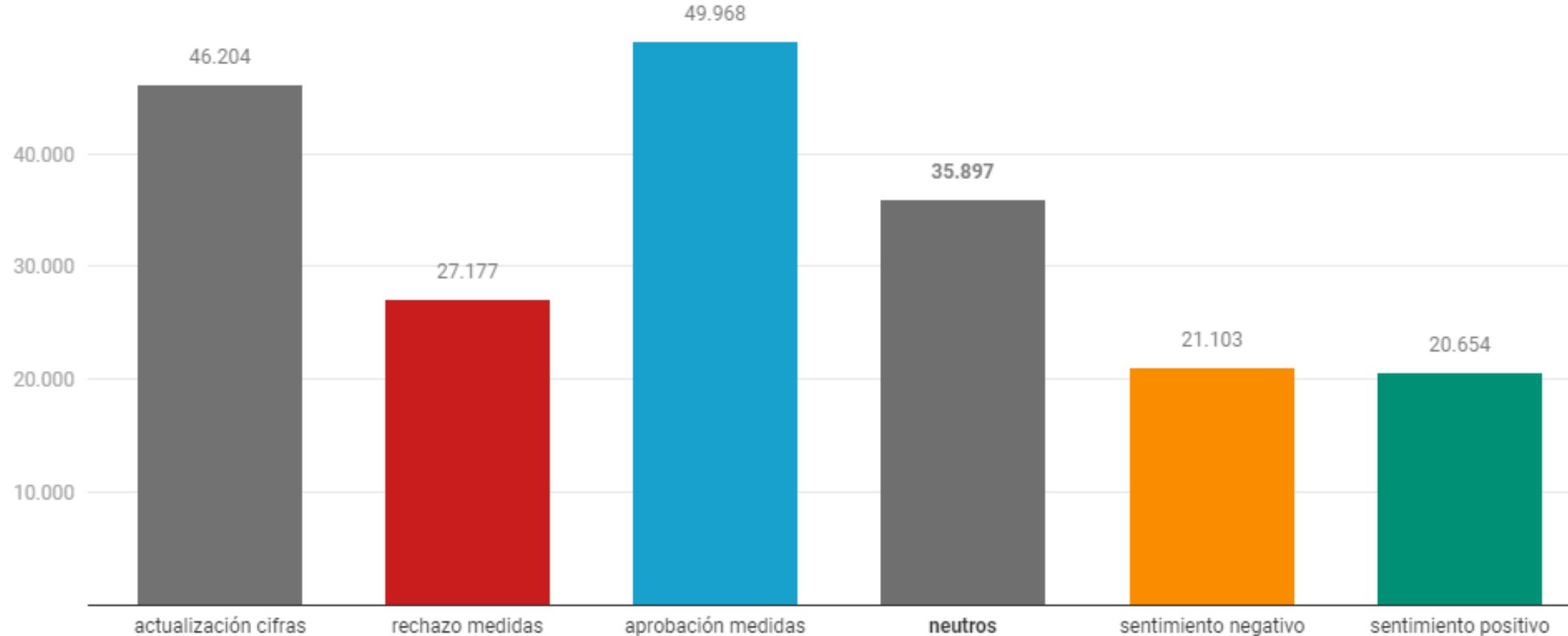
No puedes publicar ni comentar durante 3 días

Esto se debe a que ya publicaste contenido que no cumplía nuestras Normas comunitarias.

Esta publicación infringe nuestras normas sobre lenguaje que incita al odio, por lo que solo tú puedes verla.

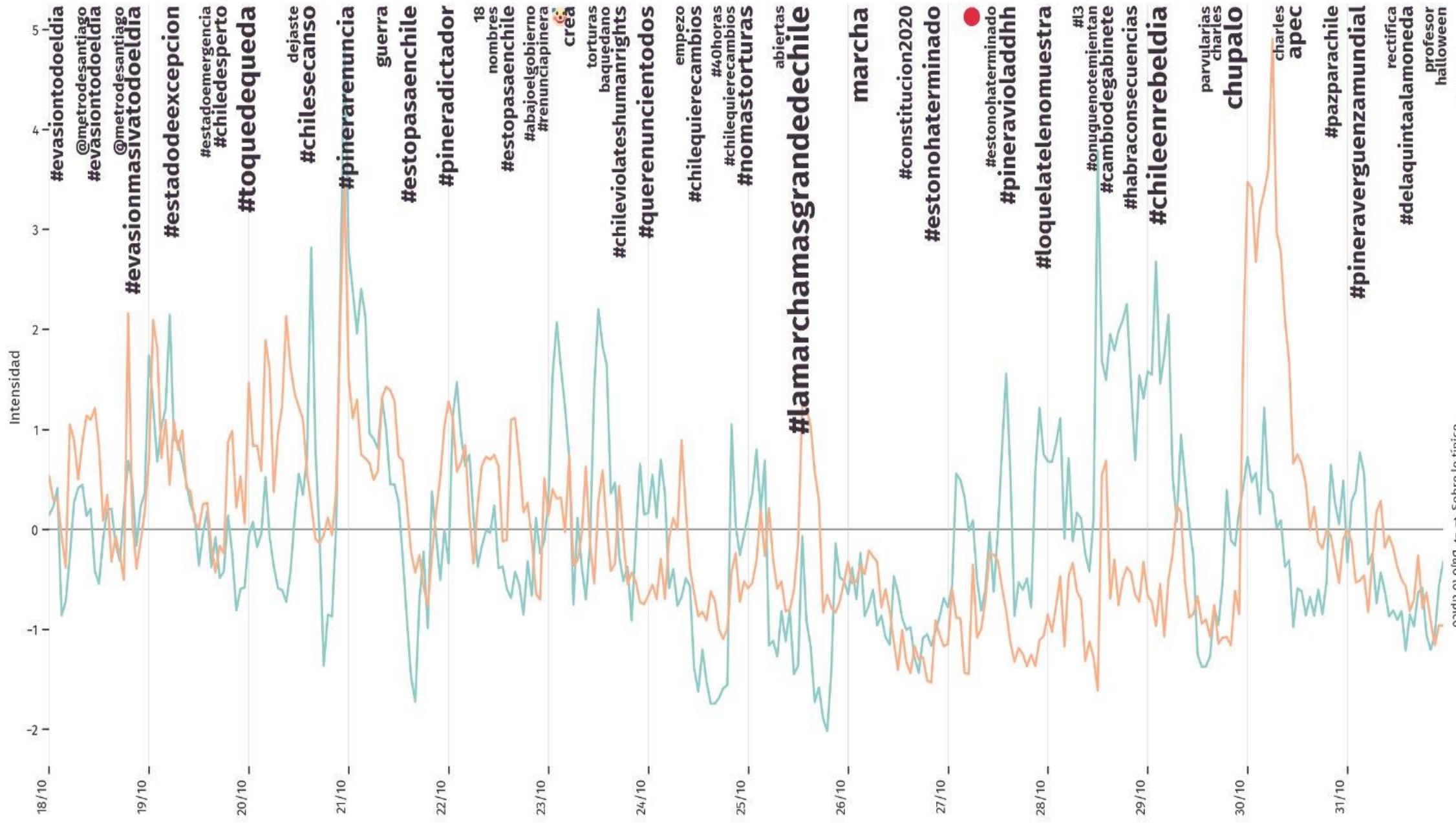
Osbaldo Encinas
Negra hp

Análisis de Tweets acerca de la Pandemia (Uruguay 2020)



<https://www.elpais.com.uy/multimedia/informes-especiales/analisis-mensajes-pandemia-reaccionaron-redes-coronavirus-uruguay.html>

Rabia y Ansiedad en Estallido Social (18 al 31 de Octubre, 2019, según Twitter)





Motorola



¡Compreza accesorios para obtener más ayuda!

18:05

Obtenga la solución

Hola. No me dice nada el mensaje.

18:06 ✓✓

Voy a tratar de ayudarte con eso. ¿Cuál de los siguientes se parece más a su problema? Escriba un número para continuar, por ejemplo, "1"

18:06

- [1] Problemas con mensajes
- [2] Solicitar reporte de envíos
- [3] Quiero revisar el estado de mi orden.
- [4] Ninguno de los anteriores

18:06

4

18:06 ✓✓

¿Podría parafrasear su problema?

18:06

Quiero un lápiz de reemplazo para mi moto g pro.

18:06 ✓✓

Necesitamos la información de su país que donde está ahora. ¿Es chile correcto?

18:06

- [1] Si
- [2] No

18:06

1

18:07 ✓✓

¡Se ha encontrado una solución! Consulte este artículo "reparar o reemplazar accesorios" para obtener más ayuda.

18:07

Obtenga la solución

77

18:07 ✓✓

¡Encantada de escuchar eso!

18:07



+56 9 3710 0099



Este chat es con una cuenta de 18/6/2020 -az clic para más información.

Hola Rodrigo Sandoval Urrich, soy Daniela Asistente Virtual de Redsalud, le escribo de parte del Dr ALEJANDRA BARBARA CARRILLO PUGA para confirmar su asistencia a la cita del 19/06/2020 a las 09:00, en Clínica Dental ARAUCO-PARQUE ARAUCO. ¿Asistirá a su cita?

10:02

SI. Confirmo asistir

Excelentel, recuerde llegar **solo 15 minutos antes y sin acompañantes** (a excepción de pacientes menores de edad o estricta necesidad) con el fin de evitar aglomeraciones. En caso de no poder asistir, agradecemos avisar para anular su hora y dejar la disponibilidad para otro paciente. En tiempo de Coronavirus: ¡Todos nos cuidamos!

De: Daniela Asistente Virtual de Redsalud <daniela@....cl>
Enviado el: lunes, 20 de julio de 2020 17:07
Para: rodrigo@....com
Asunto: RE: Rodrigo, confirme su asistencia.

Queda confirmada su hora con DRA NNNN AAAAA, para el 27/07/2020 a las 09:45, en Clínica XXXXXXXXX.

Muchas gracias,
Daniela. Ejecutiva Clínica Dental ARAUCO-PARQUE ARAUCO

De: Rodrigo Sandoval <rodrigo@...com>
Si. Confirmo que asistiré el lunes 27

De: Daniela Asistente Virtual de Redsalud
<daniela@confirmaciones.redsalud.cl>
Enviado el: lunes, 20 de julio de 2020 17:05

Pero en la práctica, los Bots Conversacionales sólo resuelven algunas tareas muy simples, basando su operación en una interacción muy acotada de lenguaje natural, siempre enfrentando el diálogo con temáticas específicas. Los Bots NUNCA ENTIENDEN las inquietudes reales de sus interlocutores humanos.

respuesta.

AI – Comprensión de Lenguaje

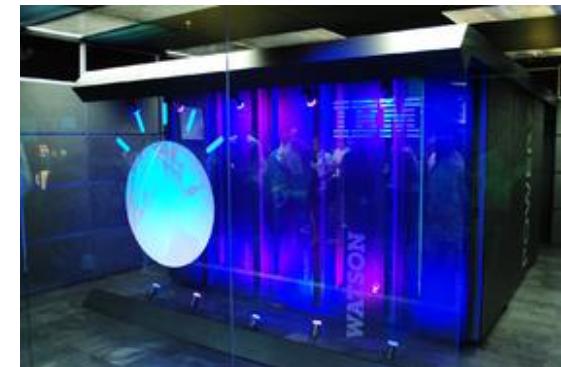
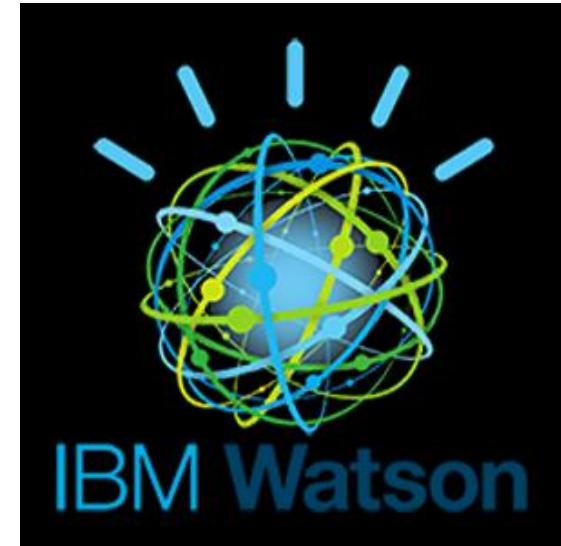
Primeras versiones ...

¿comprensión de lenguaje?



Natural Language Processing with Watson

- IBM Watson: combina hardware y modelos altamente sofisticados, que analizan un texto no-estructurado, reconociendo elementos relevantes.
- Saltó a la fama cuando pudo ganarle a expertos en el juego de conocimientos Jeopardy



IBM Watson



IBM Watson: How it Works

IBM Watson

Hace 1 año. • 439,246 vistas

Learn how IBM Watson works and has similar thought processes to a human.

<http://www.ibm.com/watson>.



IBM's Watson Supercomputer Destroys Humans in Jeopardy | Engadget

Engadget

Hace 5 años. • 1,519,611 vistas

IBM's Watson supercomputer destroys all humans in Jeopardy. » Subscribe To Engadget Today: <http://bit.ly/YA7pDT> » Watch ...



IBM Watson: Smartest Machine ever built - Documentary

Pro Documentary

Hace 1 año. • 305,720 vistas

Açıklama For more documentaries of any kind, you can subscribe to my channel Pro Documentary.

Entonces ... ¿basta con procesar el lenguaje natural?

NLP

*Natural Language
Processing*
(Qué se dijo)

NLU

*Natural Language
Understanding*
(Qué significa)

NLG

*Natural Language
Generation*
(Qué se responde)



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

7.1. Procesamiento Lenguaje Natural

Procesamiento Lenguaje Natural

Objetivo:
responder
preguntas
como éstas

- ¿A qué se refiere un texto? ¿De qué clase es un texto en particular?
- ¿Qué cosas están mencionadas en un texto?
- ¿Cómo poder atender y reaccionar ante las necesidades de un usuario o cliente desde sus comentarios?
- ¿Cuál es la opinión o percepción de personas en sus comunicaciones?

Objetivos en NLP

Clasificación

- Procesar una porción de texto y asignarle una clase o categoría.

Extracción

- Desde una porción de texto, determinar (extraer) un dato de un tipo definido

Comparación

- Determinar la semejanza o similitud entre dos porciones de texto.



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

7.1.1. Clasificación de Texto No-Estructurado

ALBERTO MOZO AGUILAR
TEATINOS 332 SANTIAGO
Fono 698-4264
e-mail: ama@notariamozo.cl



REPERTORIO N° 2.884/2014.-

TRANSFORMACION DE SOCIEDAD

"RSOLVER SOLUCIONES TECNOLOGICAS LIMITADA"

A

"RSOLVER SpA"

-A-A-A-

En Santiago, República de Chile, a **ocho de Abril de dos mil catorce**, ante mí, **ALBERTO MOZO AGUILAR**, abogado, Notario Público, Titular de la Cuadragésima Notaría de Santiago, con Oficio ubicado en calle Teatinos número trescientos treinta y dos, comuna de Santiago, comparecen: don **RODRIGO ANDRES SANDOVAL URRICH**, chileno, casado y separado totalmente de bienes, ingeniero civil, cédula de identidad número ocho millones ocho mil quinientos ochenta y seis guión nueve y doña **ISABEL MARGARITA ANAYA FERNANDEZ**, chilena, casada y separada totalmente de bienes, según se acreditará, administradora hotelera, cédula nacional de identidad número ocho millones quinientos cincuenta y dos mil quinientos setenta guión cero, ambos domiciliados para estos efectos en calle Félix de Amesti, número novecientos setenta, departamento ochenta y tres, Comuna de Las Condes, ciudad de Santiago; **UNO: Antecedentes. /Uno/** Por escritura pública de fecha doce de abril de dos mil doce, otorgada en la notaría de Santiago de don **RAUL IVAN PERRY PEFAUR**, se constituyó la sociedad de responsabilidad limitada llamada "**Rsolver Soluciones Tecnológicas Limitadas**", en adelante la "**Sociedad**", cuyo extracto se inscribió a fojas veinticinco mil trescientos treinta y dos, número diecisiete mil ochocientos trece en el Registro de Comercio del Conservador de Bienes Raíces de Santiago del año dos mil doce, y se publicó en el Diario Oficial con fecha diecinueve de abril de dos mil doce. **/Dos/** De

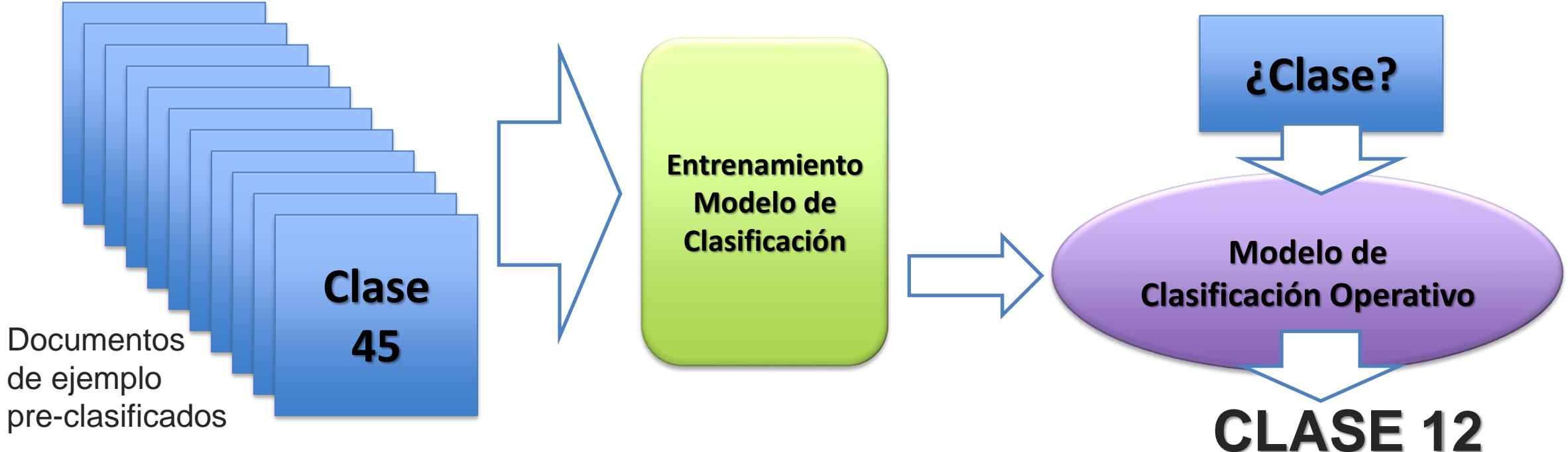
TRAN362979.0414 RSOLVER SOLUCIONES TECNOLOGICAS LIMITADA ahora RSOLVER SpA

Un robot debe poder leer el contenido de un documento jurídico y determinar de qué tipo de documento legal se trata

Tipos
(Constitución sociedad, contrato servicios, contrato compraventa, ...)

Multitipos
(Documentos que describen más de un tipo de doc jurídico)

Modelos de Clasificación Supervisada



- Los modelos de clasificación supervisada en Machine Learning reciben una cantidad grande de ejemplos etiquetados (referencias pre-clasificadas) con la cual se entrena, apuntando a buenos desempeños (~90% exactitud).
- Luego, al recibir un nuevo ejemplo no-etiquetado, producen una etiqueta adecuada, con la exactitud lograda en entrenamiento.

Modelos de *Machine Learning* Supervisado

- Árbol de Decisión
- Random Forest
- Naive Bayes
- Support Vector Machine
- Regresión Logística
- K-Nearest Neighbors
- Redes Neuronales (simples)
- Otros...

REDES NEURONALES PROFUNDAS

- FFN (Feed Forward Network)
- CNN (Convolucional)
- RNN (Recurrente)
- GRU, LSTM
- Transformers
- Otros...

Todos estos modelos responden a problemas de clasificación categórica, pero dependiendo de la naturaleza de los datos, puede convenir un modelo sobre otros.

Multi-Clase y Multi-Clasificación

ALBERTO MOZO AGUILAR
TEATINOS 303 SANTIAGO
Fono 699-4264
e-mail: amm@notariamozo.cl



REPERTORIO N° 2.884/2014.-

TRANSFORMACION DE SOCIEDAD

"RSOLVER SOLUCIONES TECNOLOGICAS LIMITADA"

A

"RSOLVER SpA"

-A-A-A-

En Santiago, República de Chile, a ocho de Abril de dos mil catorce, ante mí, ALBERTO MOZO AGUILAR, abogado, Notario Público, Titular de la Cuadragésima Notaría de Santiago, con Oficio ubicado en calle Teatinos número treinta y tres, comuna de Santiago, comparecen: don RODRIGO ANDRES SANDOVAL URRICH, chileno, casado y separado totalmente de bienes,¹ Ingeniero civil, cédula de identidad número ocho millones ocho mil quinientos ochenta y seis guión nueve y doña ISABEL MARGARITA ANAYA FERNANDEZ, chilena, casada y separada totalmente de bienes, según se acreditará, administradora hotelera, cédula nacional de identidad número ocho millones quinientos cincuenta y dos mil quinientos setenta guión cero, ambos domiciliados para estos efectos en calle Félix de Amesti, número novecientos setenta, departamento ochenta y tres, Comuna de Las Condes, ciudad de Santiago; UNO: Antecedentes. /Uno/ Por escritura pública de fecha doce de abril de dos mil doce, otorgada en la notaría de Santiago de don RAUL IVAN PERRY PEFAUR, se constituyó la sociedad de responsabilidad limitada llamada "Rsolver Soluciones Tecnológicas Limitadas", en adelante la "Sociedad", cuyo extracto se inscribió a fojas veinticinco mil trescientos treinta y dos, número diecisiete mil ochientos trece en el Registro de Comercio del Conservador de Bienes Raíces de Santiago del año dos mil doce, y se publicó en el Diario Oficial con fecha diecinueve de abril de dos mil doce. /Dos/ De

TRAN362979.0414 RSOLVER SOLUCIONES TECNOLOGICAS LIMITADA ahora RSOLVER SpA

1

Red Neuronal

SVM
por cada tipo



Naive
Bayes

Tipo 43: Modificación de Sociedad
Tipo 47: Declaración Poderes

Preguntas esenciales

¿Cuál es la dimensión del X de entrada si el texto es de largo variable?

¿Es factible/necesario diferenciar versiones de la misma palabra (p. ej: “Contrato”, “CONTRATOS”, “Contratación”, “contratante”)?

¿Cómo se pueden tratar los errores ortográficos y gramáticos?

Se necesita ...

Pre-Procesamiento de Texto No-Estructurado

Necesidad 1: reducción de diversidad lingüística

Se sabe que en lenguaje natural hay diferentes posibles verbalizaciones para decir lo mismo. El analizar texto en forma sintáctica estricta no permite reconocer la equivalencia entre estas posibles alternativas, por lo que se hace necesario encontrar formas “generales” del texto.

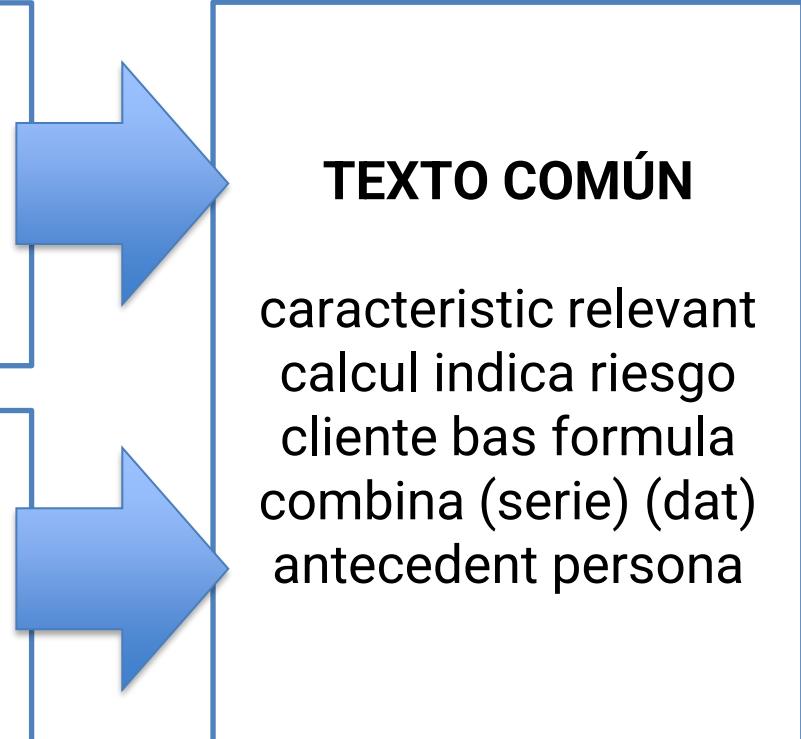
El pre-procesamiento de texto se centra en diferentes técnicas de “normalización”. En el fondo, se busca generalizar las expresiones, de modo de poder compararlas con más facilidad.

TEXTO 1

Dentro de las características relevantes, está el “Cálculo de Indicador de Riesgo por Cliente”, que se basa en una fórmula que combina una serie de datos de los antecedentes personales.

TEXTO 2

La característica más relevante es el “Cálculo de Indicador de Riesgo por Cliente”, basado de acuerdo a una fórmula que combina antecedentes de las personas.



Normalización de texto (reducción diversidad)

Tokenización

- Grandes porciones de texto en segmentos más cortos (segmentación: párrafos, frases; tokenización: palabras), según la necesidad
- Más que una operación mecánica, requiere incorporar cierta lógica propia del lenguaje: qué palabras se podrían separar, etc.

Simplificación

- Todo a minúsculas.
- Eliminación de tildes y ~ (si es factible, dado el contexto).
- Eliminación signos puntuación.
- Eliminación palabras innecesarias (stop words – dependiendo del contexto)

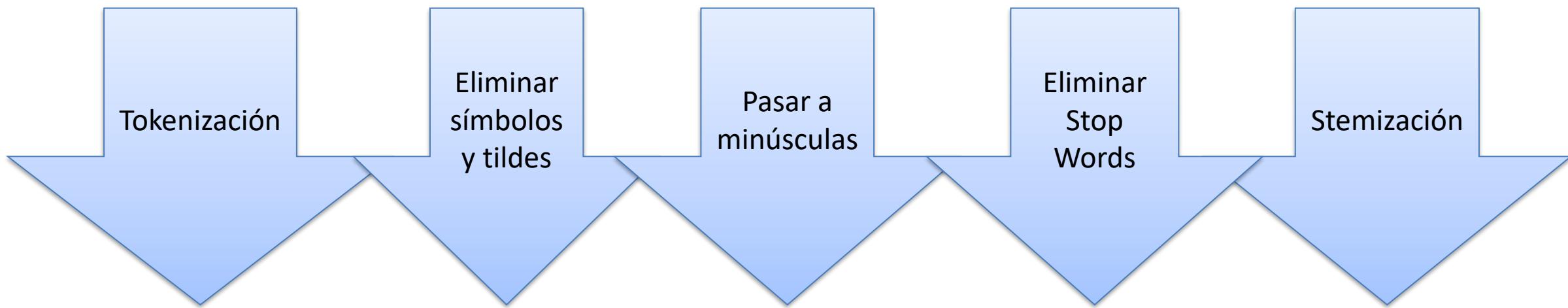
Semántica básica

- Stemming: base de la palabra. Ej: corriendo → corre
- Lemmatization: similar, unificando conceptos similares en una única palabra, conservando la idea de la palabra original. Ej. Mejor → Bueno

Referencia: <https://www.kdnuggets.com/2017/12/general-approach-preprocessing-text-data.html>

Ejemplo de Normalización

“Este es un TEXTO de ejemplo, que sirve para ver qué términos quedan después de la limpieza de palabras, símbolos, y otros elementos.”



[un texto ejemplo sirv ve termin qued despues limpieza palabr simbol demás otr element]

Necesidad 2: Uso de un Vocabulario para Clasificación

Todos los modelos de clasificación reciben un X de dimensión fija, pero los textos son de largo variable. Por ello, se vuelve necesario cambiar la representación de los vectores (“documentos” = porciones de texto) que reciben los modelos.

La mejor estrategia es convertir los textos en vectores que refieran palabras existentes en un Vocabulario.

Luego, cada nuevo texto a analizar se mapea a la ocurrencia de palabras en el vocabulario, simplificando su representación general.

Texto = colección de palabras o términos

El modelo tiene un desempeño adecuado para la necesidad

Normalización

model desempeñ adecuad necesi

Vocabulario = términos relevantes en un corpus

model dimens text estrateg vector palabra simpl adecuad ...

Mapeo al
Vocabulario

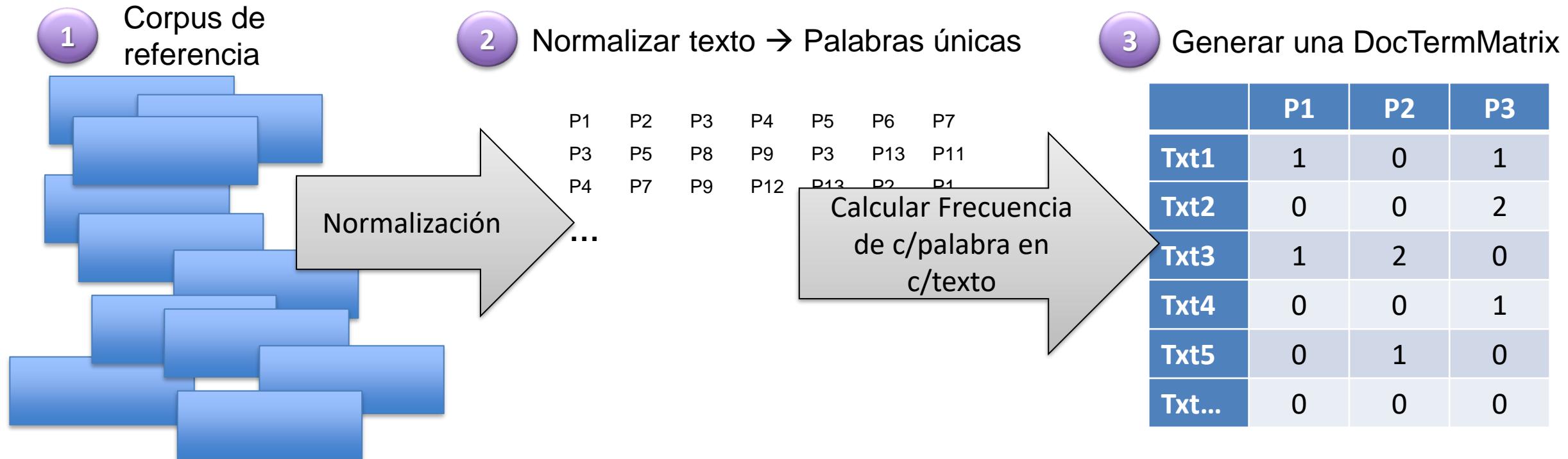
0 1 0 0 0 0 1 0 1 0 0 1 0 1 0 0 1 0 1 0 0 0 ...

Tamaño = cantidad de términos relevantes distintos en el corpus

Dimensión = Tamaño del Vocab

Construcción del Vocabulario (parte 1)

Para construir un vocabulario se necesita un “corpus” o colección de “documentos” (ejemplos de texto del lenguaje de referencia). Luego de normalizar y armar la lista de términos diferentes, se determina la ocurrencia de éstos en los documentos de origen (Document Term Matrix).



Construcción del Vocabulario (parte 2)

Contando con la DocTermMatrix, que típicamente tiene una alta dimensionalidad (cantidad de términos diferentes en el corpus), se puede **reducir su dimensionalidad** al eliminar los términos “escasos” (*sparse*).

DocTermMatrix

	P1	P2	P3	P4	P5	P6	P7	...
Txt1	1	0	1	0	0	1	0	
Txt2	0	0	2	0	1	0	0	
Txt3	1	2	0	0	0	0	0	
Txt4	0	0	1	0	0	0	0	
Txt5	0	1	0	0	0	0	2	
Txt..	0	0	0	0	0	0	0	

Dimensión = Tamaño original del Vocab

4

Se eliminan términos con un factor “sparse” menor que un umbral.

DocTermMatrix
Dimensión reducida

	P1	P5	P6	...
Txt1	1	0	1	
Txt2	0	1	0	
Txt3	1	0	0	
Txt4	0	0	0	
Txt5	0	0	0	
Txt..	0	0	0	

Dimensión = Tamaño reducido del Vocab

Vocabulario
Reducido

P1
P5
P6
P10
P17
P22
...

Arma
Vocab

Algunas consideraciones

Diccionario ≠ Vocabulario

- Diccionario: lista de palabras de un lenguaje (definición de diseño)
- **Vocabulario**: lista de palabras encontradas en un texto (empírica)
- Por tanto, **Vocabulario** se usa en la práctica (que puede contener errores de tipeo y ortográficos, pero que son empíricos)

Representación vectorial mapeando el Vocabulario no considera:

- Orden de las palabras en el texto
- Cantidad ocurrencias de la palabra (aunque depende de la implementación)

Aplicando la Clasificación de Texto con Pre-procesamiento





PANDEMIA DE CORONAVIRUS

PORADA NOTICIAS CASOS EN CHILE CUARENTENAS CASOS EN EL MUNDO MIRADA GLOBAL

Presidente de la Cámara dice que hay "voluntad" para discutir este jueves proyecto de 10%

Asimismo el Senado adelantó la sesión de este miércoles y programó que a partir de las 13:00 horas se votará en general y en particular la iniciativa.

21 de Julio de 2020 | 17:54 | Por Verónica Marín, Emol



190



El presidente de la Cámara anunció que hoy hay reunión de comité a las 19:00 horas donde se discutirá la tabla de los próximos días.

Aton

El presidente de la Cámara de Diputados, **Diego Paulsen (RN)**, aseguró que existe voluntad para discutir este jueves el proyecto que permite el retiro del 10% de los fondos de pensiones en el contexto de la pandemia, esto considerando que mañana será votado en el Senado y pasaría a un tercer trámite.

Noticias y Comentarios

Francisco Orrego ★
Frente a esta situación al gobierno solo le queda asumir e informar en forma transparente, para que las personas decidan en forma informada si retirar sus fondos. Obviamente hay personas que necesitan en forma urgente fondos y otros que podrían pensarlo definitivamente no hacerlo.
Responder Denunciar 11 1 Hace una hora

Diego Letorre
Francisco Orrego Ojalá Piñera se atreva a vetarlo o demore la promulgación. Así la calle cura rápido al país de este inútil de presidente y se hace cargo un interino hasta pronta elección...
Responder Denunciar 8 21 Hace una hora

Eduardo Torres
Diego Letorre O sea, tú estás anhelando el caos, la destrucción de Chile... ¿eres de izquierda por casualidad?
Denunciar 14 2 Hace 50 minutos

Gabriel Berenguela ★
Millones de chilenos hoy están al borde del colapso económico y ven como su esperanza más cercana es el retiro del 10% de sus fondos de capitalización individuales para la vejez y así solucionar en parte sus necesidades básicas y las deudas adquiridas por culpa de la pandemia.
Personas sencillas, la mayoría de sectores humildes y de clase media. Todos los cuales han esperado por cinco meses una ayuda que no ha llegado.
Ahora para muchas personas de poder, de gobierno y del mundo parlamentario, esto se ha transformado en una verdadera cruzada para mantener a salvo el santo grail, no importando las consecuencias colaterales que esto conlleve. Pero para miles de hogares y millones de compatriotas, este proyecto es una esperanza real y concreta de poder sortear esta dantesca crisis económica, que los tiene con el agua al cuello. Es importante señalar con claridad, que aquí están en juego probablemente el futuro de miles de familias chilenas y no un trofeo político, ni menos el salvaguardar los intereses de grupos económicos.
Responder Denunciar 15 4 Hace una hora

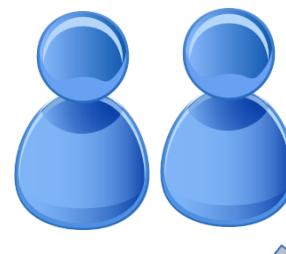
Julio Rosales
Gabriel Berenguela el problema es que esos millones de chilenos lo único que harán es entregar su dinero a aquellos grupos económicos a los cuales le deben. El dinero seguirá en manos de los que aborenecen y quedarán sin uno al mes siguiente. Todos piensan que recibirán una millonada sin embargo la realidad es que recibirán su dinero para entregarlo a otros, si es que no van al super por el 4K de 55".
Responder Denunciar 2 8 Hace una hora

Fernando Urzúa
Julio Rosales Cada uno hará lo que desee con su dinero. Que malo hay en comprarse un 4K????
Denunciar 0 0 Hace 3 minutos

Aaron Aguirre
Obvio, no hay que darle muchas vueltas al proyecto, existe la voluntad transversal de ayudar a la mayoría de la gente, hay casi unanimidad en la cámara baja como la alta, incluso con los



Líderes Políticos

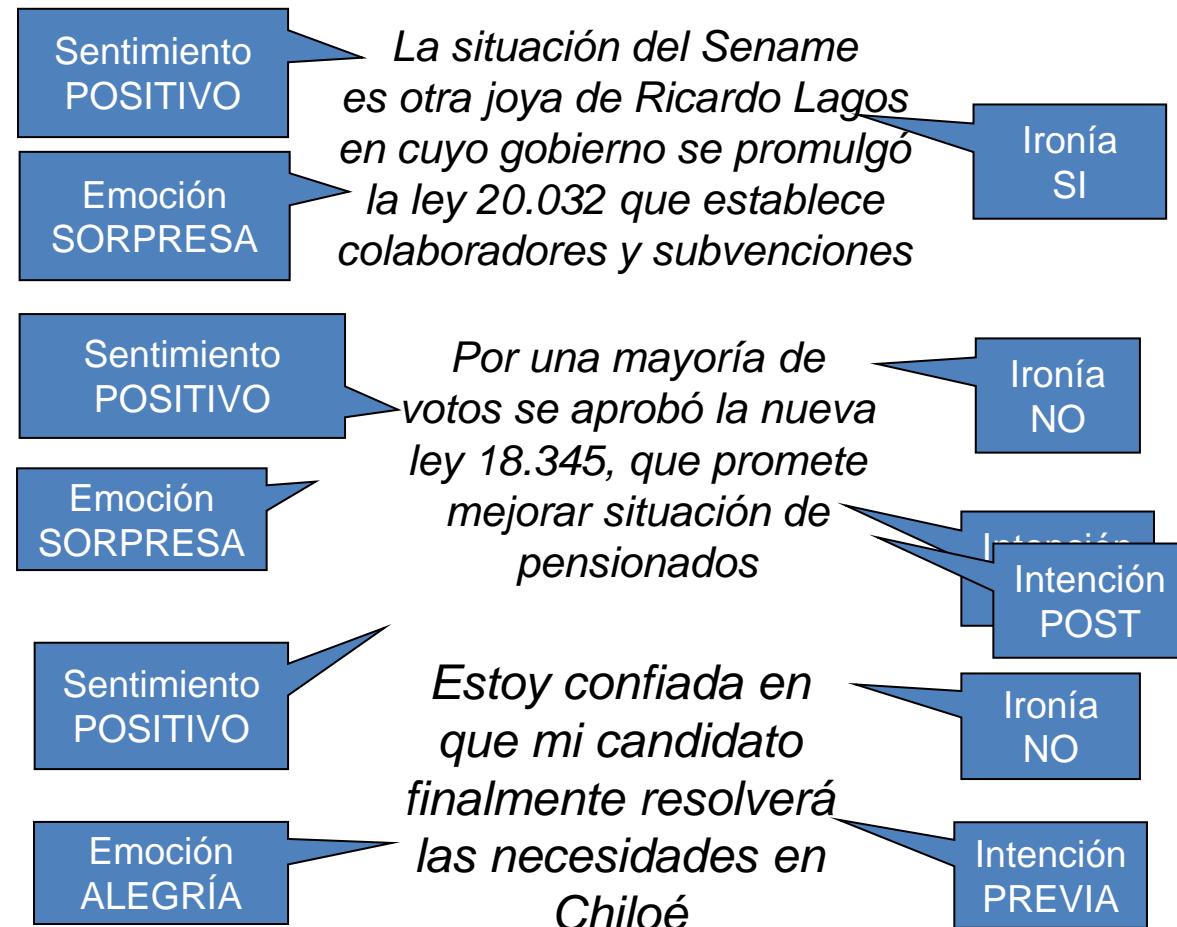


Asesores comunicacionales

Clasificación multi-dimensión

Dimensiones:

- **Sentimiento:** positivo, negativo, neutro. Se considera una opción de no mostrar neutros → en política.
- **Emoción:** enojo, sorpresa, alegría, etc.
- **Intención:** previo, posterior (por ej, diferenciar una intención de compra de una evaluación de compra).
- **Ironía:** es o no es. En este caso, se pretende usar la ironía de un comentario como “negador” del sentimiento y emoción.



Entrenamiento Contextual

- Los humanos que clasifican manualmente comentarios de referencia, tienen disponible un contexto y potencialmente links a las noticias de origen.
- Esta clasificación contextual permite mejorar sustancialmente la calidad de clasificación.
- Áreas contextuales: Política-Economía, Artista, Deportes, Empresa Servicios, Empresa Producto.

Deportes

"Buonanotte y su arribo a la UC: ""Hay que pelear en todos los frentes"""

<http://www.latercera.com/noticia/deportes/2016/07/656-689507-9-buonanotte-y-su-arribo-a-la-uc-hay-que-pelear-en-todos-los-frentes.shtml>

Comment:

Bienvenido Diego vamos mi uc

Polarity

Positive Neutral Negative Undefined

Emotion

Happiness Sadness Fear Anger Surprise Disgust Undefined

Intention

After Before Undefined

Irony

Yes No Undefined

Classify

Resultados Procesamiento Mensajería Clientes

ANTES (MANUAL)

CAPACIDAD

Un ejecutivo de centro de atención puede revisar hasta 30 mensajes por hora

PRECISIÓN

85% en ejecutivos humanos

Tasa de error de 15%

ESTANDARIZACIÓN

Ciertos criterios dependen de la persona

DESPUÉS (AUTOMÁTICO)

CAPACIDAD

El robot revisa 60 mensajes por minuto

PRECISIÓN

92% en robot

Tasa de error de 8% en el robot

ESTANDARIZACIÓN

Criterios dependen sólo de los ejemplos de entrenamiento

Recibí la chaqueta y me pareció de una excelente calidad y terminaciones

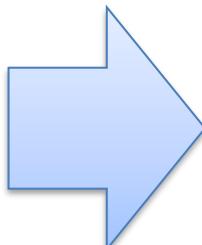


Área: Post-venta
Tópico: Felicitaciones
Producto: Chaqueta
Sentimiento: Positivo
Emoción: Alegría
Post-evento

Muchos de estas dimensiones de interpretación siguen siendo modelos de Machine Learning (clasificación supervisada) y otros son NER.

Tengo serios problemas con el servicio de Internet. Se corta repetidamente, a veces por varios minutos

He llamado y no contestan.

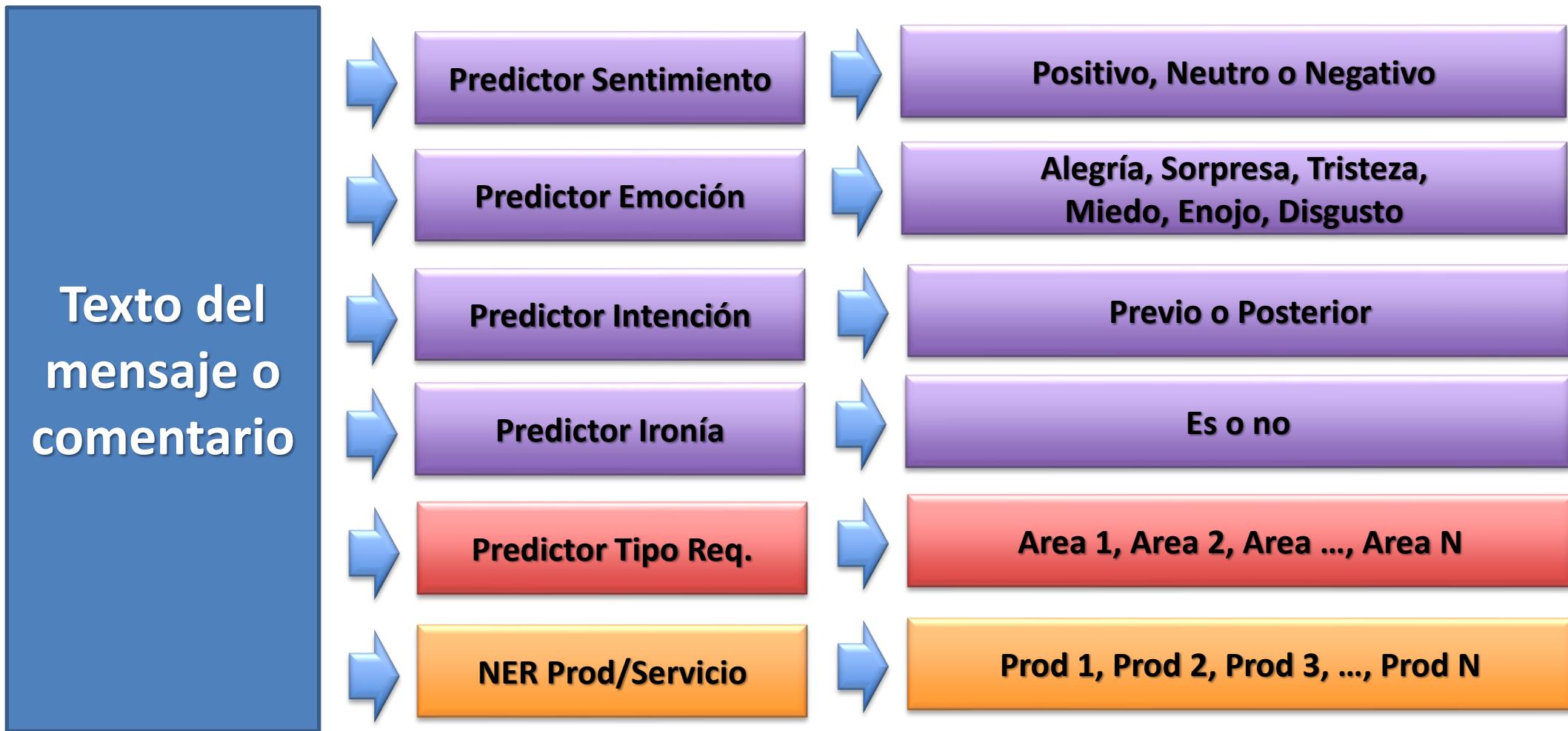


Área: Soporte técnico
Tópico: Reclamo
Producto: Internet
Sentimiento: Negativo
Emoción: Enojo
Post-evento

URGENCIA

Lo interesante es incorporar aspectos “humanos”, como es el caso de emociones e intenciones, dentro de los elementos a interpretar.

Clasificadores de texto



Caso atención servicios telefónicos

Análisis de Mensaje de Cliente

Estimados. Solicito bloquear el envío de SMS para las líneas adjuntadas en el excel. Saludos Cordiales

La petición del cliente es de u

Análisis de Mensaje de Cliente

por favor revisar las líneas 99999909999 y 99999989909, tienen problemas con la recepción de mensajes. se llamó a servicio tecnico y lo que indican es lo siguiente No tiene recepción de mensajería entrante solo voz

Este es claramente un **reclamo** respecto a una **línea**. Nótese que se siente disgusto en el mensaje.

Análisis de Mensaje de Cliente

Estimados buenas tardes. Junto con Saludar, agradecería gestionar Eliminación de Plan de Datos 9 Gb (Plan w89) a Línea N°989899998. Muchas Gracias.

de un **plan de datos**

Caso: clientes de banco

Análisis de Mensaje de Cliente

Hola Adrián, no puedo hacer transferencias, se queda pegado en la llamada de la clave dinámica.

La petición del cliente es de una **consulta sobre productos y servicios**, subcódigo 6, que pide que oc

Solicitud: Consulta sobre Productos y Servicios / otra
Subcódigo: 6

Emoción: **sorpresa**

Intención: **que pide que ocurra**

Análisis de Mensaje de Cliente

El 1 de marzo compro el SOAP con mi tarjeta de credito mastercard y a la fecha aun no recibo los creditos cabify ofrecidos por la compra.

Ojo, que el cliente está presentando un **reclamo** (problema con un producto) / cuenta corriente y servicios respecto a un código 7, que ya ocurrió.

Análisis de Mensaje de Cliente

Al revisar el día de hoy mi cuenta me percato que me han cobrado de forma automática el pago mínimo de la linea de crédito cargándomelo en la corriente, siendo que este fue pagado el 28 de febrero. Poseo el comprobante de pago realizado el 28 de febrero, y también el comprobante del cobro que me han realizado el día de hoy.

cto) / Tarjeta de crédito

Ojo, que el cliente está presentando un **reclamo** (problema con un producto) / cuenta corriente y servicios respecto a un código 1, que pide que ocurra. Nótese que se siente disgusto en el mensaje.

Solicitud: **Reclamo (problema con un producto) / Cuenta Corriente y Servicios**

Subcódigo: 1

Emoción: **disgusto**

Intención: **que pide que ocurra**



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

7.1.2. Extracción de Información desde Texto No-Estructurado

NER – Named-Entity Recognition

Técnica para la extracción de entidades desde un texto en lenguaje natural.

No se basa en reglas predefinidas (por ej, entregar una lista de nombres de personas para encontrarlos), sino que se “entrena” con ejemplos para reconocer Nombres, Lugares, Fechas, Montos, etc.

Poder Especial

Por el presente documento, en Santiago a 14 de abril del año 2018, declara don **Juan Pérez Soto**, en representación de la empresa ACME Ltda., para declarar que se autoriza a don **Miguel José Soto González** para realizar, en nombre de la empresa, todo tipo de trámites ante el servicio de impuestos internos, incluyendo, pero no limitándose a ...

Se busca encontrar, en el texto extenso, los nombres de personas mencionadas. NER permite reconocer estas porciones de texto, sin tener una lista de nombres previamente definida, sino que en base a ejemplos

En base a ejemplos, no a lista de alternativas

- Ejemplos de dónde aparece un nombre (NNN AAA):
 - ... se confiere el poder a don NNN AAA a partir de...
 - ... cuyo representante es NNN AAA, quien acepta ...
 - ... comparece NNN AAA, declarando ...
- Entonces, no se necesita una lista de nombres; basta con ver las palabras alrededor.
- Esto permite enfrentar desafíos como el siguiente:
 - ... confiere poder a don ARMANDO CABEZAS DEL RIO a ...

NER - Named-Entity Recognition

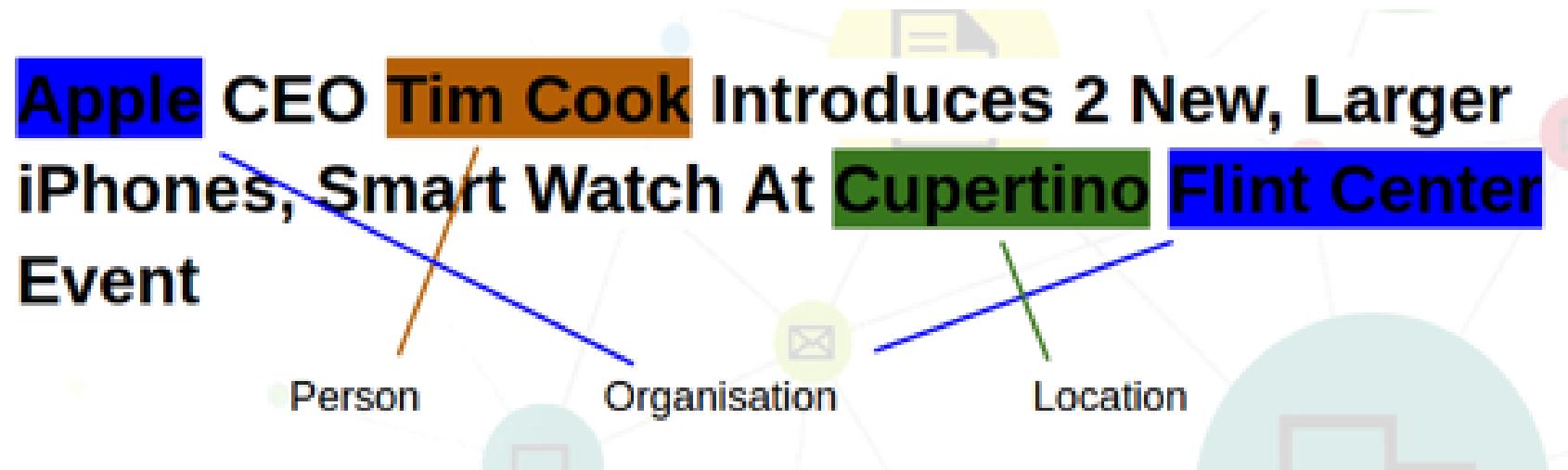


Figure 1: An example of NER application on an example text

- Por medio de ejemplos, es capaz de reconocer palabras que representan diferentes tipos: fechas, números o montos, nombres de personas, nombres de instituciones, lugares, ...
- Se entrena en forma similar a modelos de ML.

NER - ¿Cómo funciona?

NER Involucra la identificación de nombres y palabras propias en texto y su clasificación en una categoría, de un set predeterminado de alternativas

El caso particular de nombres (“Named-Entity”) considera:

- Nombre de Persona
- Nombre de Organización
- Nombre de Lugar
- Nombre de Marca

Otros (Entity Recognition o Entity Identification) considera otros elementos:

- Fecha, monto, cantidad, ...

Desafíos de la ambigüedad

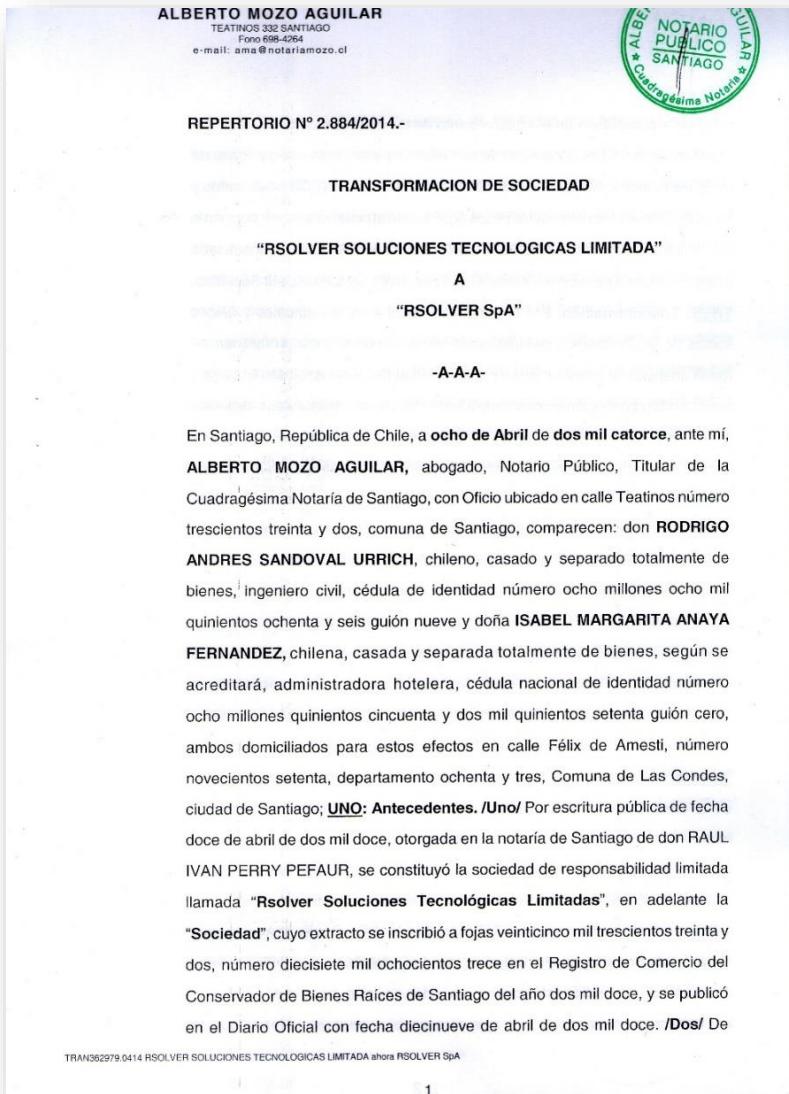
2010: ¿año o cantidad?

Washington: ¿persona o lugar?

Julio: ¿mes o persona?

Etc.

Extendiendo y Complementando el Modelo



Datos o Características complementarias

¿Cuál es la temática principal del documento?

¿Cuáles son las partes nombradas?

Si es un contrato de arriendo, ¿cuándo es el valor de arriendo definido?

Etc.

Caracterización Documental



Patente

Métodos y modelos son patentables, pero el software o el código que están escritos

Approved for use through 01/21/2014. GMB-0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number
PROVISIONAL APPLICATION FOR PATENT COVER SHEET – Page 1 of 2
This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).
Express Mail Label No. _____

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Rodrigo Andrés	SANDOVAL URRICH	Santiago, Chile
Juan Ignacio	SÁA HARGOUS	Santiago, Chile

Additional inventors are being named on the _____ separately numbered sheets attached hereto.

TITLE OF THE INVENTION (500 characters max):

DOCUMENT CHARACTERIZATION METHOD

Direct all correspondence to:

CORRESPONDENCE ADDRESS		
<input checked="" type="checkbox"/> The address corresponding to Customer Number:	30593	
OR		
<input type="checkbox"/> Firm or Individual Name	_____	
Address		
City	State	Zip
Country	Telephone	Email

ENCLOSED APPLICATION PARTS (check all that apply)

<input checked="" type="checkbox"/> Application Data Sheet. See 37 CFR 1.76.	<input type="checkbox"/> CD(s), Number of CDs _____
<input type="checkbox"/> Drawing(s) Number of Sheets _____	<input type="checkbox"/> Other (specify) _____
<input checked="" type="checkbox"/> Specification (e.g., description of the invention) Number of Pages _____	3

Fees Due: Filing Fee of \$260 (\$130 for small entity) (\$65 for micro entity). If the specification and drawings exceed 100 sheets of paper, an application size fee is also due, which is \$400 (\$200 for small entity) (\$100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.56(e).

METHOD OF PAYMENT OF THE FILING FEE AND APPLICATION SIZE FEE FOR THIS PROVISIONAL APPLICATION FOR PATENT

<input checked="" type="checkbox"/> Applicant asserts small entity status. See 37 CFR 1.27.	\$130.00
<input type="checkbox"/> Applicant certifies micro entity status. See 37 CFR 1.29.	
<input type="checkbox"/> Applicant must attach form PTO-268/15A or B or equivalent.	
<input type="checkbox"/> A check or money order made payable to the Director of the United States Patent and Trademark Office is enclosed to cover the filing fee and application size fee (if applicable).	
<input type="checkbox"/> Payment by credit card form PTO-268 is attached.	
<input checked="" type="checkbox"/> The Director is hereby authorized to charge the filing fee and application size fee (if applicable) or credit any overpayment to Deposit Account Number: 06-0750	

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by the Small Entity Compliance Guide for Patent Applications, which is to file (and by the USPTO to process an application. Confidentiality is guaranteed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 10 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Demostración de R:Docs

a Cognitiva Avanzada para Entendimiento de Contenido Documental

Análisis de Documento

chivo seleccionado

- Archivo: Contrato de Ingeniero Part-Time.pdf
- Tipo: Contrato de Prestación de Servicios
- Fecha emisión: 21 Jul 02020
- Notario: (no se menciona)
- Personas nombradas: Rodrigo Andrés Sandoval Urrich
- Personas jurídicas: SOCIEDAD RSOLVER SERVICIOS TECNOLOGICO



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

7.1.3. Comparación (simple) de Texto No-Estructurado

Comparación (simple) según mapeo a vocabulario

TEXTO 1

Dentro de las características relevantes, está el “Cálculo de Indicador de Riesgo por Cliente”, que se basa en una fórmula que combina una serie de datos de los antecedentes personales.

Mapeo al
Vocabulario

0 1 0 0 0 0 1 0 1 0 0 1 0 1 0 0 1 0 0 0 ...

TEXTO 2

La característica más relevante es el “Cálculo de Indicador de Riesgo por Cliente”, basado de acuerdo a una fórmula que combina antecedentes de las personas.

Mapeo al
Vocabulario

0 1 0 0 0 0 1 0 0 0 1 1 0 1 0 1 0 0 1 0 1 0 ...

Comparación de
conjuntos indexados

Coincidencia del 85%
→ Suficientemente iguales

Gracias

 **rsandova@ing.puc.cl**

rodrigo@RSolver.com

 **@RSandovalSolver**

 **/in/RodrigoSandoval**

www.RodrigoSandoval.net

www.RSolver.com