vocali

Tecnologías del lenguaje

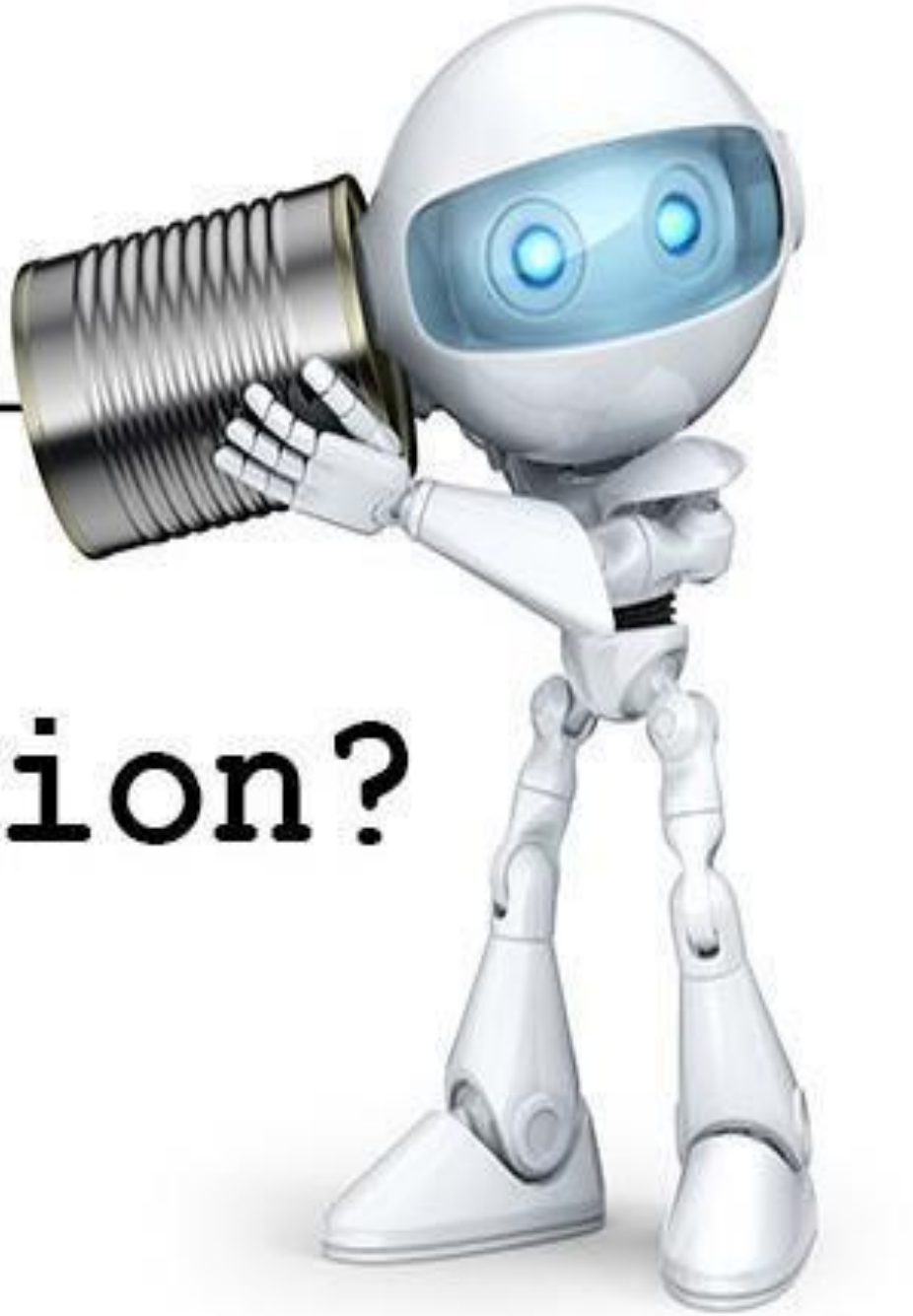
Saturdays.AI
Murcia

María José Cano - Fernando Molina

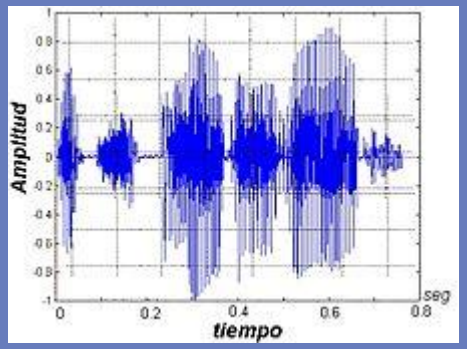

vocali

vocali

Speech Recognition?



Reconocimiento Automático de Habla (ASR)



Extracción de características

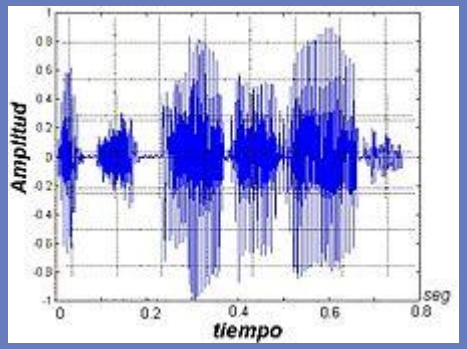
Modelo de lenguaje o gramática

Modelo Acústico (“fonemas”)

Léxico (G2P)

MOTOR DE RECONOCIMIENTO

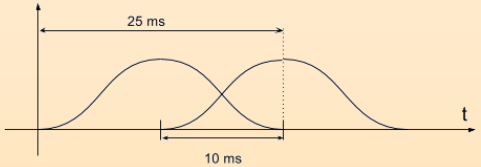
Reconocimiento Automático de Habla (ASR)



Extracción de características

Extracción de características. MFCC:

- Enventanado: ventanas de 25 ms con 10 ms de solapamiento. Se aplican ventanas de Hamming para evitar discontinuidades



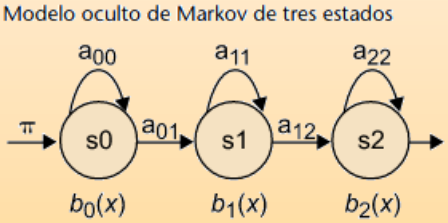
- Transformada Discreta de Fourier: pasar al dominio de las frecuencias
- Escala de Mel: adaptación al oído humano
- Coeficientes Cepstrales: Logaritmo y Transformada Discreta del coseno
- 12 componentes + E = 13 coef
- Añadimos derivada y segunda derivada --> 39 coeficientes

Reconocimiento Automático de Habla (ASR)

Modelo Acústico

Trifonemas: fonemas con contexto

HMM



PDFs: GMM o DNNs (TDNN, LSTMs …)

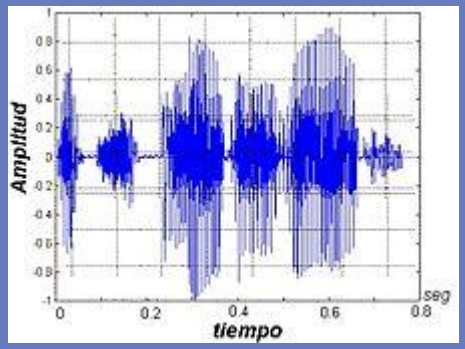
Modelo de lenguaje o gramática

Modelo Acústico (“fonemas”)

Léxico (G2P)

MOTOR DE RECONOCIMIENTO

Reconocimiento Automático de Habla (ASR)



Grapheme 2 phoneme

casa k a s a
saturdays s a t u r d e i s

- Basados en reglas
- Sequitur, Phonetisaurus...

Modelo de lenguaje o gramática

Modelo Acústico (“fonemas”)

Léxico (G2P)

MOTOR DE RECONOCIMIENTO

Recor omático de Habla

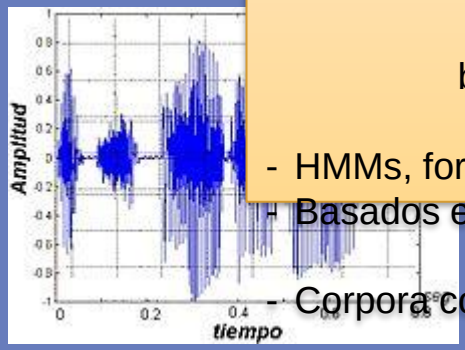
Gramáticas y LMs

casa
en
restaurante

comer
una pizza
ensalada

un
bocadillo

- HMMs, formato ARPA
- Basados en redes neuronales
- Corpora con errores: NLP



Extracción de características

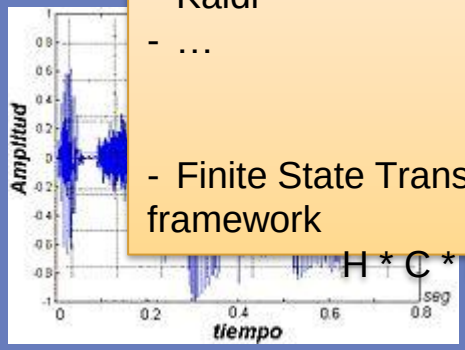
MOTOR DE RECONOCIMIENTO

Modelo de lenguaje o gramática

Modelo Acústico (“fonemas”)

Léxico (G2P)

Reconocimiento Automático de Habla (ASR)



Toolkits y Frameworks

- HTK
- Sphinx
- Kaldi
- …

- Finite State Transducer framework

H * C * L * G

Extracción de características

Modelo de lenguaje o gramática

Modelo Acústico (“fonemas”)

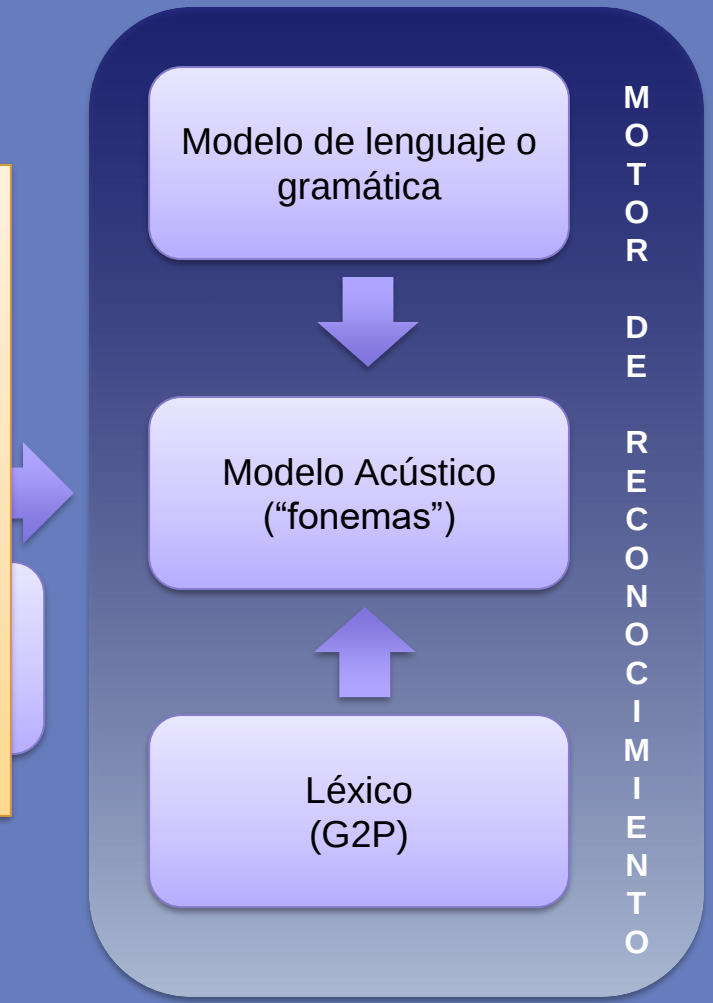
Léxico (G2P)

MOTOR DE RECONOCIMIENTO

Métricas

Reconocimiento

- Word Error Rate
  WER = 100* (S + D + I) / nº de palabras ref.
- Sentence Error Rate
  SER = frases incorrectas / frases totales
- Real time factor
  RTF = tiempo de transcripción / tiempo audio



Modelos de lenguaje

- Asignan probabilidades a secuencias de palabras a partir de un corpus de entrada
- Ayudan al modelo acústico a “organizar” los fonemas prediciendo la siguiente palabra a partir de *n* anteriores
  - ¿ m e d i s t e l a c a j a ? → ¿Me diste la caja? vs ¿Mediste la caja?
- No se usan solo en reconocimiento de voz: OCRs, corrección automática de textos, question answering, traducción automática...

Modelos de lenguaje

- Los LMs más sencillos son los de *n-grams*: asignan probabilidades a secuencias de palabras de longitud *n*
- Conjetura de Markov: es posible predecir la probabilidad de la palabra *n* a partir de las *n-1* palabras anteriores
- Suelen ser modelos de trigramas: incrementar *n* aumenta exponencialmente la complejidad sin mejora significativa
- ¿Cómo son de buenos los LMs? → “perplexity”
- Problema de los n-grams: su “historia de predicción” es corta → aproximaciones basadas en NN que son capaces de modelar mucho mejor el contexto

Modelos de lenguaje

- Generar LMs es sencillo si la cantidad y calidad del corpus de entrada es adecuada
  - SRILM - The SRI Language Modeling Toolkit
  - The CMU-Cambridge Statistical Language Modeling Toolkit
  - KenLM Language Model Toolkit
  - Tensorflow
  - Fast.ai
  - ….

NLP y lingüística de corpus

- Calidad de los corpus de entrada
- Necesidad de replicar el trabajo para cada idioma
- Recursos específicos para cada dominio
- Dificultad para acceder a *datasets*
- Privacidad de los datos
- Ambigüedad y características propias del lenguaje natural (uso de siglas y abreviaturas, sinonimia, negación, ironía,...)

Detección y corrección automática

- Dos grandes grupos de errores:
  - Non real-words: palabras que no existen (“lenguage”, “vlanca”, “radiológicoas”)
  - Real words: palabras correctas pero no en ese contexto (“intestino hueso”, “contrate intravenoso”)
- Embeddings: aprenden los contextos en los que aparecen las palabras, así que nos pueden ayudar a detectar palabras que aparecen en contextos sospechosos (Word2Vec, Glove…)

Anonimización de textos

- Especialmente importante en dominios como el médico o el legal
  - REGEX + gazzetters + reglas (GATE)
  - NER en Spacy
  - CRFs: conditional Ramdom Fields (sklearn-crfsuite)
  - XGBoost
  - NCRF++ (CNNs, LSTMs, CRFs en Pytorch)
  - ...

Complejidad del lenguaje

- Ejemplos:
  - “La historia clínica indica que la tía de de la paciente sufrió un cáncer de pulmón que se descarta para ella tras el informe radiológico”
  - “No se identifican en los estudios de 2020 hallazgos sugestivos de cáncer de pulmón”
  - “Se identifican nódulos en LSD”
  - “Paciente con CA pulmón”
- PoS-tagging, embeddings, CRFs, LSTMs….

Tareas / Competiciones

IberLEF 2020: Iberian Languages Evaluation Forum

Accepted Tasks

ALexS 2020: Lexicon Analysis Task at SEPLN

Cantemist 2020: CANcer TExt Mining Shared Task

CAPITEL 2020: Named Entity Recognition and Classification, and Universal Dependency Parsing of Spanish News

eHealth-KD 2020: eHealth Knowledge Discovery

FACT 2020: Factuality Analysis and Classification Task

IbeRDI 2020: NLP for R&D&I domain in Spanish: multi-task evaluation of document similarity metrics

IberLegal 2020: NLP for Legal Domain in Spanish: Named Entity Recognition and Classification in Legislative Texts

MEX-A3T 2020: Fake news and offensive language detection in Mexican Spanish

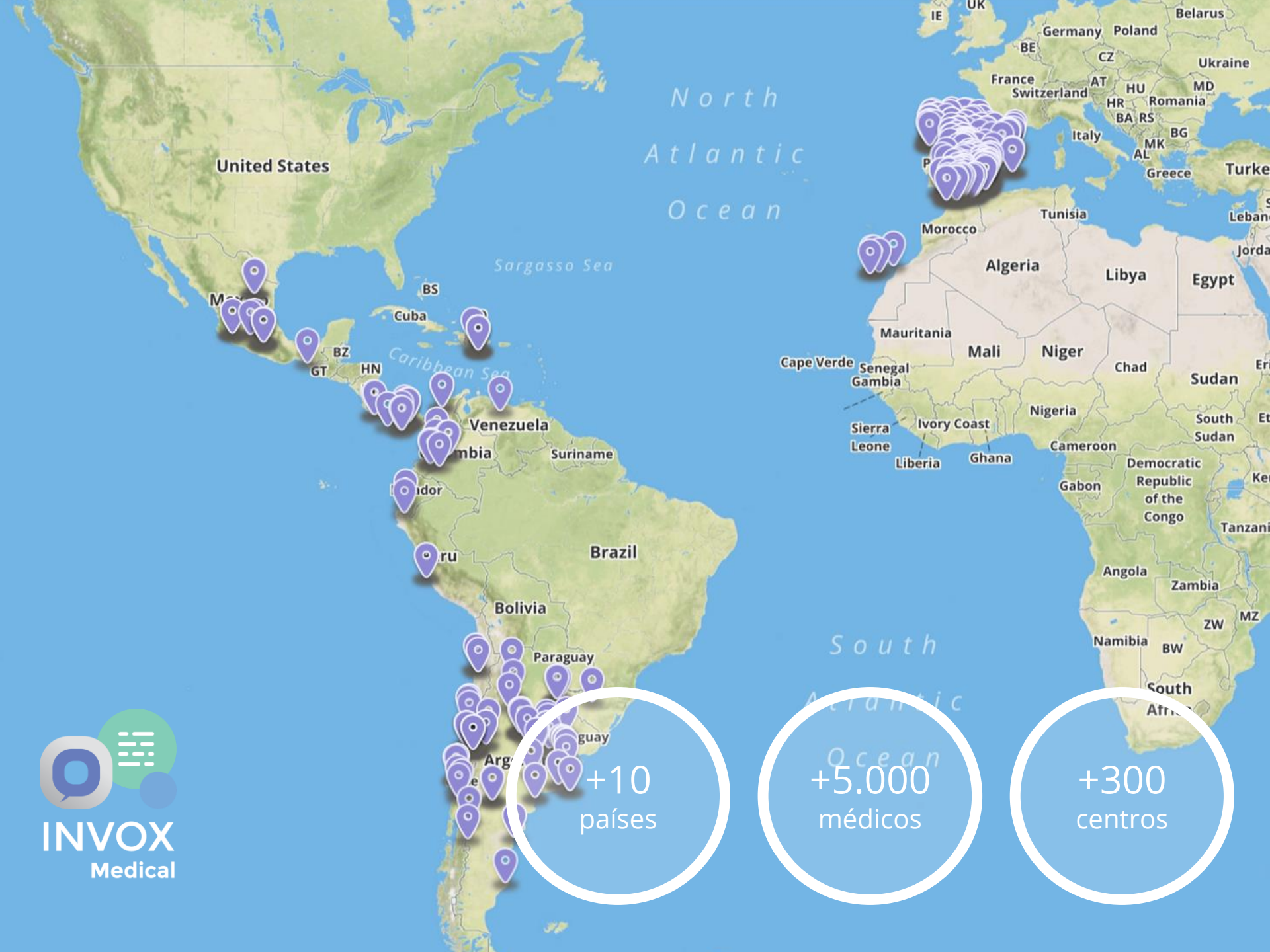
TASS 2020: Sentiment Analysis in Spanish

Tareas / Competiciones

Tasks

- eHealth-KD 2019: eHealth Knowledge Discovery
- FACT: Factuality Analysis and Classification Task
- HAHA 2019: Humor Analysis based on Human Annotation
- IroSvA: Irony Detection in Spanish Variants
- MEDDOCAN (Medical Document Anonymization) task
- MEX-A3T: Authorship and aggressiveness analysis in Twitter: case study in Mexican Spanish
- Named Entity Recognition and Relation Extraction for Portuguese
- NEGES: Negation in Spanish
- TASS: Sentiment Analysis Task at SEPLN

Algunos
PROYECTOS



INVOX Medical

+10 países

+5.000 médicos

+300 centros

Justicia implanta tres nuevas tecnologías que agilizarán el trabajo de jueces y fiscales

Publicado 19/11/2019 12:53:07 CET

Se trata de una calculadora de acumulación de condenas, la firma digital manuscrita y la conversión automática en texto de las grabaciones de las vistas judiciales

MADRID, 19 Nov. (EUROPA PRESS) -

El Ministerio de Justicia ha presentado este martes tres nuevas aplicaciones tecnológicas que se irán implantando progresivamente en todos los juzgados españoles que permitirán agilizar el trabajo de jueces y fiscales.

Se trata de una calculadora que permite el cálculo automático de condenas en caso de acumulación de sentencias; otra encargada de convertir en textos las grabaciones de los juicios y comparecencias judiciales, así como una aplicación que posibilitará al ciudadanos la firma digital manuscrita de documentos en sus comparecencias personales en actos judiciales.

Las tres nuevas herramientas han sido presentadas por el secretario de Estado de Justicia, Manuel Jesús Dolz; y la directora general de Modernización de la Justicia, Sofía Duarte; quienes han destacado que se tratan de los primeros avances del Ministerio "hacia la Inteligencia Artificial".

"La Inteligencia Artificial significa ahorro de energías intelectuales y es un factor importante", ha subrayado el 'número dos' de Justicia porque permite mayor proyección de la actividad creativa del derecho, ya que ni jueces, ni fiscales ni ningún operador jurídico "tienen que preocuparse de cuestiones que pueden hacer las

Últimas noticias / España »

La Policía detiene a cuatro personas, tres detectives, por el 'caso Villarejo' que afecta al chantaje al juez Urquía

Ciudadanos ve la mesa de diálogo sobre Cataluña como "una muestra más de sumisión" de Sánchez a Torra

Los jueces del caso Rato y del espionaje en Madrid reciben las actuaciones del PP y Partido Laócrata por el 'Delcygate'

Lo más leído

1 Eugenia Martínez de Irujo: "Con Cayetano no tengo relación"

2 Nueva tecnología genera electricidad 'de la nada'

3 Billie Eilish abruma en los BRITs cantando a James Bond junto a


Quirófano
Artroscopia
Código: Int96 NHC: NHC96 Paciente: Laia Sanz Médico: Antonio López Cano
La paciente sufre una lesión deportiva en la rodilla derecha
Atrás
Finalizar
Registros realizados
08:51:59 Venda almohadillada 15 U
08:51:53 Electrodos
08:51:49 Paquete de gasas de 10 unidades
08:51:36 Venda almohadillada 10 U
08:51:18 Suero glucosado 100 ml.
08:51:09 Jeringa 5 ml.
08:51:06 Naropin
08:51:02 Midazolán
08:50:50 Fentanex
08:50:47 Atropina

Murcia Region to announce the 4 selected digital health companies

MAY 22, 2019

Good news! We are proud to announce the selected companies that will take part in the next co-creation phase (2nd inDemand iteration).

These are the solvers for the four challenges launched by Murcia Region in 2019.

Challenge 1: DEEP DIVER (Assistance in the search for diagnoses with suspicion of Professional Illness)

Selected company: **VÓCALI** (Spain)

inDemand

other **drug** use at all?

Patient

No, hu-huh.

Doctor

Just going to ask you a series of questions make sure we're not missing anything.

Patient

Sure.

Doctor

So in general, have you had any problems feeling really **tired** lately or anything like that?

Patient

I'm just a little **tired**.

Doctor

Okay.

Patient

From it.

Doctor

How about any problems with

Have you noticed a **fever** or just generally **feeling unwell**?

Para aprender más

- L. Rabiner; B. H. Juang (1993). Fundamentals of Speech Recognition. Engle-wood Cliffs, NJ: Prentice HallPTR.
- Reconocimiento Automático del habla con Adaptación al Género y al Locutor. PFC Ana B. Caballero
- Speech and Language Processing. Dan Jurafsky & James H. Martin (https://web.stanford.edu/~jurafsky/slp3/)
- Build your first speech-to-text model (Python)
- http://kaldi-asr.org/

vocali

¡Gracias!

María José Cano
mariajose.cano@vocali.net

Fernando Molina
fernando.molina@vocali.net