Plan TL

IBERLEGAL

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y EMPRESA

SECRETARÍA DE ESTADO PARA EL AVANCE DIGITAL

Hoja de ruta Desarrollo de la Tecnologías del Lenguaje en el dominio legal

David Pérez Fernández
Plan TL - SEAD

11 de diciembre de 2019

Algunos casos de uso

Plan TL
IBERLEGAL

- **Metadatos: metadatado automático, clasificación automática de documentos**
- **Reconocimiento de entidades nombradas (NERC): leyes, referencias sentencias, personas, lugares, organizaciones, referencias temporales ej. identificación de referencias en textos**
- **Extracción de información/Recuperación de información ej. recuperación de sentencias similares, indexado de textos a partir de las entidades referenciadas**
- **Simplificación de textos: expansión conceptual, ej. Enlace de conceptos con sus definiciones: BOE->RAE Diccionario panhispánico del español jurídico**
- **Análisis de grandes colecciones textuales: ej. visión de conjunto, temáticas principales sentencias, evolución temporal, áreas emergentes para planificar la formación de los jueces**

Algunos casos de uso

- Análisis de estructuras verbales: ej. búsqueda de hechos: factoides, storylines vida judicial individuos
- Visión de conjunto colecciones documentales, evolución temporal, visión multicorpus. ej. planificación formación áreas emergentes jueces
- Traducción automática
  ej. asistencia judicial multilingüe
- Análisis del habla y sistemas conversacionales
  ej. transcripción de vistas judiciales, consultas trámites judiciales similar 060

Sanidad vs Justicia

Corpus textuales:
Pubmed, patentes, Ensayos clínicos, publicaciones científicas, …
Metacorpus: ej. NIH ExPORTER

Corpus textuales:
PlanTL Corpus LegalES

Metadatos:
ISO/HL7 16527:2016 HL7 Personal Health Record System

Metadatos:
EJE, ELI, ECLI, e-Codex cross border justice

Modelos de lenguaje (w2v, BERT, ...), modelos de tópicos:
PlanTL BSC

Modelos de lenguaje (w2v, BERT, ...), modelos de tópicos:
No Existen -> PlanTL

Terminología y recursos semánticos:
extracción terminológica.
EN: Metatesauro UMLS + Terminologías ej. SnomedCT, LOINC …
ES: Completado MSSSI, PlanTL

Terminología y recursos semánticos:
Eurovoc, IATE
RAE Dic. Panhispánico jurídico
No Existe Metatesauro -> CGPJ, Mjusticia
Extracción terminolñogica -> Academia, PlanTL, TerminESP

Sanidad vs Justicia

Taxonomías:
ej. CIE-10, ICD-10 (enfermedades)

Corpus anotado:
EN: I2B2, ES: PlanTL informes médicos (segmentación, PoS, NERC(enfermedades, fármacos, anonimización), negación)

Campañas de evaluación:
EN: I2b2, CLEF eHealth, TREC CDS (Precision Medicine)
ES: PharmaCoNER, BARR1&2 (abrev), Meddocan (anonimiz), CIE-10, ChemProt, MESINESP, eHealthKD, WMT

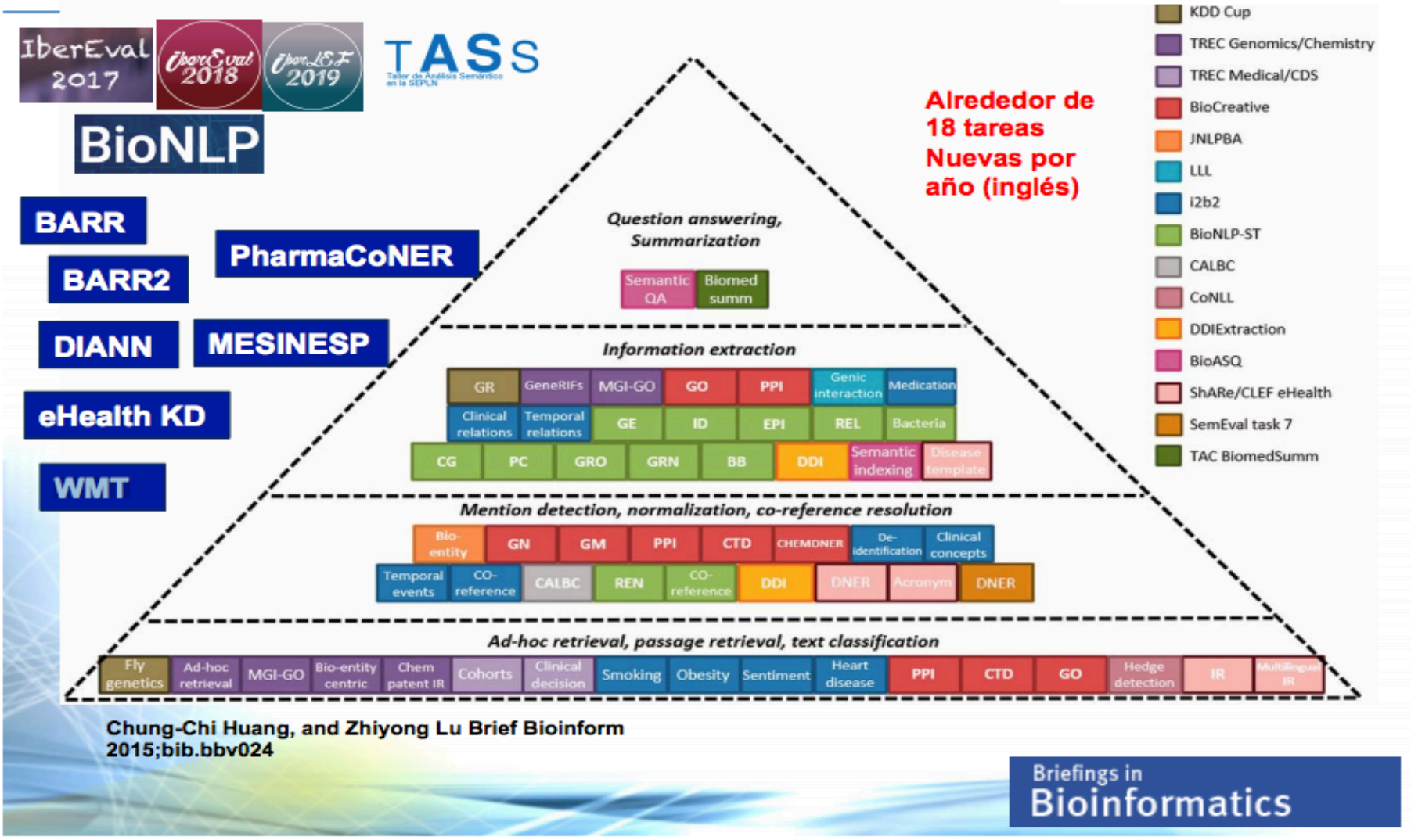
Taxonomías:

Corpus anotado:
EN: Cambridge Corpus Legal English, US Supreme Court
ES: No existe -> PlanTL

Campañas de evaluación:
EN: TREC legal
ES: No existe -> PlanTL Iberlef:Iberlegal

Plan TL
IBERLEGAL

Tareas en biomedicina/clínica desde perspectiva PLN



Chung-Chi Huang, and Zhiyong Lu Brief Bioinform 2015;bib.bbv024

Briefings in Bioinformatics

Sanidad vs Justicia

Plan TL
IBERLEGAL

Traducción:
Corpus paralelo, terminología multilingüe
PlanTL + CEF: eTranslate, ELRC

Componentes y plataformas NLP:
EN: ej cTakes
ES: PlanTL BSC

Plataformas escalables y HPC:
PlanTL BSC + RES

Traducción :
Eng: Eurlex, IATEEs: BOE + CCAAEU
Eurlex, DOUEUN Trib. Internacional

Componentes y plataformas NLP:
EN: LexNLP (eng + de +)
ES:

Plataformas escalables y HPC:
PlanTL BSC

Plan TL
IBERLEGAL

Gracias

https://www.plantl.gob.es

PlanTecnologiasLenguaje@mineco.es