

Plan TL

Plan de Impulso de las
Tecnologías del Lenguaje



GOBIERNO
DE ESPAÑA

MINISTERIO
DE ECONOMÍA
Y EMPRESA

Actuaciones 2019

Plan de impulso de las Tecnologías del Lenguaje

BSC 2019

Eje 0: Gobernanza

- OTG Sanidad BSC:
 - Alcance: 6M € / 4 años
 - OTG Sanidad / Plataforma HPC PlanTL
 - Comienzo del encargo BSC (enero 2019)
- OTG:
 - Sistemas conversacionales: David Griol
 - Inteligencia Artificial: Jerónimo Arenas
- Informes:
 - Sistemas conversacionales
 - Datos Abiertos en AAPP
 - Análisis de campañas evaluación español y lenguas cooficiales

Eje I: Apoyo desarrollo infraestructuras lingüísticas



- RAE Corpus Capitel (Corpus Anotado del Plan de Impulso de las Tecnologías del Lenguaje): Textos periodísticos de distintas fuentes y dominios (anotación morfosintáctica, entidades nombradas, anotación sintáctica)
- BNE Crawling web española 10 años
- Corpus Legal Jurix2019

Eje II: Impulso de la Industria TL

- Ayudas SEAD, SGFSI TDH 2019 (9M €), incluyendo TL (Abril 2019)
- INAP Formación funcionariado AGE + CCAA + EELL (MOOCs + capsulas formativas)
- Comunicación: Página web. LinkedIn
- Infodays:
 - Vigilancia Sectorial Mayo
 - Infoday TL Salud (BSC, Barcelona), diciembre 2019.
 - Infoday Justicia (Jurix2019, Madrid), diciembre 2019
- Catálogo productos/svc. Censo empresas y grupos investigación
- Estrategia formativa (UPV, SEPLN)
- Congresos. Participación en congresos y encuentros nacionales e internacionales sobre TL e Inteligencia Artificial

- Plataforma Traducción Automática
 - Proyectos EU CEF: MTHub / ELRI / NTEU / NecData
 - Taller ELRI en España (marzo 2019) en la sede de la Comisión Europea en Madrid: Infraestructura de Recursos y Servicios de Traducción Automática para las Instituciones Públicas
- Plataforma Sistemas Conversacionales
 - Informe Sistemas Conversacionales
 - Incorporación David Griol OTG
 - SGAD
- Plataforma Procesamiento de Lenguaje Natural
 - Resultados Proyecto eInfra OpeMinTed y adaptación al entorno HPC (ej. procesado crawling BNE)

Eje IV: Proyectos Faro: Salud

- CNIO: publicación de desarrollos en Github y Zenodo
 - Recursos y componentes sw
- Corpus de casos clínicos en español: Anotación morfosintáctica
- Base de datos de abreviaturas médicas en español
- Glosarios médicos bilingües
- Archivos para traducción automática neuronal
- Crawling webs medicina (600Gb texto)
- Campañas de evaluación en Sanidad
 - Reconocimiento de sustancias en textos médicos en español
 - Anonimización
 - Indización semántica médica en español

Eje IV: Proyectos Faro: Vigilancia sectorial

- Convenio SEUIDI + FECYT
 - Dirección de políticas públicas I+D
 - Análisis sector TL, IA
- Incibe: Análisis contratación y empleo sector ciberseguridad
- Apertura código CorpusViewer 1.0
- Integración desarrollos experimentales (perfilado, evolución temporal ... CorpusViewer 1.5)
- Infoday: Jornada sobre Inteligencia Artificial aplicada a las políticas públicas de ciencia, innovación y emprendimiento
- Participación en las Jornadas Técnicas RedIRIS (Abril)
- Corpus Viewer finalista en los premios CNIS 2019 (marzo 2019)

Eje IV: Proyectos Faro: Otros

- Legislación y Justicia
 - Jurix2019
 - Corpus Legal 1.250M palabras. Modelos embeddings, Transformers, tópicos, grafos relaciones interdocumentales
 - BOE
 - MPR
- Cultura
 - BSC-BNE: Procesado crawling web españolas 10 años, 5Tb
 - EFE: Anotación corpus noticias 20M
 - Vanguardia

- Presidenta Comisión anunció prioridades Digital Single Market + Energía
- 2500M € / año, 8% ES
- IA europea -> IA ciudadano
 - Orientación servicio público (sectores estratégicos: salud, justiciar, educación, I+D, ...)
 - XAI: orientación a evaluación, generación de evidencias
 - Cuestiones éticas: evaluación regulador
- Sector TL dentro de IA: 45% sector IA nacional TL
- Colaboración Planes CCAA: Agenda Digital Vasca, PlanIA Cat



Gracias

www.plantl.gob.es

plantecnologiaslenguaje@mineco.es

A decorative abstract graphic in the bottom left corner, featuring blue and green wavy lines and a grid pattern.