

IA y tecnologías del lenguaje aplicado a datos clínicos: CARMEN-I recursos, sistemas y aplicaciones

Resumen

Los recientes avances en IA, técnicas de aprendizaje profundo, así como PLN, y en especial el acceso y uso de modelos del lenguaje, han facilitado mejoras significativas en la calidad de sistemas de procesamiento de texto.

Estas mejoras se están aprovechando en una diversidad de aplicaciones tecnológicas, tales como traducción automática, sistemas de pregunta-respuesta, asistentes virtuales, buscadores y herramientas de etiquetado semántico y analítica de datos.

El mayor obstáculo para el uso de estas tecnologías e incorporación de los consiguientes resultados en el sector de la salud con fines relacionados con la gestión y mejora de la calidad asistencial o investigación clínica es la dificultad de acceder a muestras de datos que sirvan para el desarrollo, validación y adaptación de sistemas de PLN clínicos.

El acceso a un conjunto de datos de historia clínica anonimizada, como es el caso del dataset MIMIC-III para textos clínicos en inglés, ha facilitado el desarrollo tecnológico de un número significativo de soluciones tanto académicas como comerciales para una diversidad de aplicaciones clínicas y constituye por lo tanto un motor en el desarrollo y avance tecnológico de IA adaptado al entorno clínico.

Por ello, y en el contexto del Plan de Impulso de las Tecnologías del Lenguaje (Plan TL) promovido por la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA) y del convenio entre el Barcelona Supercomputing Center y el Hospital Clínic de Barcelona, se ha generado el primer corpus anonimizado de distintos tipos de informes clínicos en español, llamado CARMEN-I (Corpus of Anonymized Records for Medical information Extraction).

El objetivo de CARMEN-I es servir como conjunto o base de datos de salud de libre acceso que permita la aplicación de la IA en salud, y que sirva como recurso con una adecuada estructura de información (modelo de extensiones, conformidad y versionado) para la creación de componentes de PLN clínicos debidamente documentados, evaluados y licenciados.

Este recurso se hará accesible públicamente, junto con el protocolo y guías de anonimización, con el fin de fomentar el desarrollo de tecnologías del lenguaje e IA aplicadas a datos clínicos y ofrecer unas guías y normas que sirvan para el proceso de anonimización de datos sensibles.

El objetivo de este recurso, no es solo impulsar el desarrollo tecnológico de sistemas de PLN clínico, sino también servir como base técnica para facilitar el proceso de creación de datos anonimizados a nivel nacional e internacional, en especial para latinoamérica y países con datos en lenguas romances.

Con el fin de presentar CARMEN-I, junto con los protocolos de generación de datos anotados y herramientas de alto impacto para el procesado de HCE en español, este Infoday contará con dos sesiones, una relacionada con la creación de este conjunto de datos anonimizados y otra para la extracción automática de información clínica de textos, contando con colaboradores, expertos, hospitales y representantes de empresas relacionadas con este sector. Este evento también tiene como objetivo contribuir a la formación y difusión de tecnologías del lenguaje aplicadas al sector de salud, tanto para el entorno de la industria como el sector académico y expertos sanitarios e investigadores.

- **CUÁNDO:** 13 de enero de 2023.
- **DÓNDE:** Aula Magna de la Facultad de Medicina de la Universidad de Barcelona. Carrer de Casanova, 143, 08036 Barcelona
- **AUDIENCIA:** Este evento va dirigido a personal sanitario, profesionales clínicos, investigadores en el ámbito de la salud y biomedicina, sociedades médicas, empresas del sector farmacéutico, empresas sector IA y tecnologías del lenguaje, informática hospitalaria, especialistas y académicos de PLN, asociaciones de pacientes, entre otros.

Agenda

10:00-11:30

Acreditaciones y registro de participantes

11:30-11:50

Bienvenida

- Alfonso Valencia, BSC

- Dr. Antoni Castells, Director Médico del Hospital Clínic
- Jesús Pinilla, Secretaria de Estado de Digitalización e Inteligencia Artificial (SEDIA)

11:50-12:00

Plan Nacional de TL - Nueva economía del Lenguaje Avance y actividades SEDIA?

Ponentes: Jesús Pinilla, Secretaria de Estado de Digitalización e Inteligencia Artificial (SEDIA)

Título tentativo: Presentación del Plan Nacional de TL y Nueva economía del lenguaje: aplicaciones en salud, biomedicina. El valor de los datos para el desarrollo tecnológico.

12:00-12:45

Sesión 1: Anonimización de datos en Salud: protocolos, recursos y corpus con datos anonimizados

Moderador: - Maria Antonia Marti - UB (5 min)

- Impacto, relevancia y estrategias de anonimización de textos clínicos. Ponente: Martin Krallinger - BSC (5 min)
- Creación del corpus CARMEN-I: estructura de información, protocolo y guías de procesos de anotación y anonimización, esquemas de anotación y control de calidad. Ponente: Salvador Lima - BSC (10 min)
- Aspectos clínicos de relevancia para la creación de datos anonimizados en el contexto del corpus CARMEN-I. Ponente: Xavier Pastor - Hospital Clínic de Barcelona (10 min)

Preguntas de asistentes y conclusiones (15 min)

12:45-13:30

Mesa Redonda 1: Relevancia, necesidades, estrategias e impacto de datos anonimizados en Salud.

Moderador: Xavier Pastor, Hospital Clínic de Barcelona

Temas a tratar incluyen:

- Importancia de sistemas basados en IA en el ámbito clínico para generar datos accesibles
- Relevancia de datos no estructurados y anonimización mediante técnicas de PLN
- Usos y aplicaciones de los datos clínicos anonimizados en el

entorno clínico

- Dificultades, retos y aspectos éticos y legales para la generación de datos anonimizados en el entorno de salud

Participantes:

- Carlos Parra, Hospital Virgen del Rocío
- Miguel Ángel Maciá Martínez, Coordinador del Programa BIFAP Agencia Española de Medicamentos y Productos Sanitarios (AEMPS)
- Adolfo Muñoz Carrero, Instituto de Salud Carlos III
- Artur Conesa, Jefe de Documentación Clínica del Hospital Clínic de Barcelona

13:30-14:30

Comida

Networking y mesas de difusión/interacción con empresas y grupos de investigación demostraciones de aplicaciones de tecnologías del lenguaje

14:30 - 15:15

Sesión 2: Desarrollo y sistemas de PLN clínico: recursos y herramientas de IA para el aprovechamiento de historia clínica

Moderador: Martin Krallinger, BSC

- Generación de datos anotados para el desarrollo de sistemas avanzados de PLN clínico de calidad. Ponente: Salvador Lima - BSC (10 min)
- Integrando el conocimiento clínico en sistemas de IA: aprendiendo del experto. Ponente: Santiago Frid - Hospital Clínic de Barcelona (10min)
- Armonización y normalización de datos clínicos extraídos mediante técnicas de PLN: generación automática de datos estructurados. Ponente: Luis Gasco - BSC (10 min)

Preguntas de asistentes y conclusiones (15 min)

15:15-15:30

Pausa-Café

15:30 - 16:30

Mesa Redonda 2: Aplicaciones y desarrollo de IA de calidad: retos, necesidades y oportunidades de tecnologías del lenguaje aplicados al dominio de salud y biomedicina

Moderador: Isaac Cano, IDIBAPS-Hospital Clínic

Temas a tratar incluyen:

- La historia clínica electrónica como fuente de datos para uso secundario, investigación y modelos predictivos en el entorno clínico.
- Oportunidades, aplicaciones y transferencia de recursos de IA/PLN: academia, empresas y sector de salud.
- Importancia de escenarios de evaluación de calidad, reproducibilidad y adaptación de herramientas de PLN.
- Mecanismos de formación y divulgación de los avances tecnológicos y PLN en el ámbito de la salud.
- Importancia y necesidades de expertos en salud: PLN como herramienta y recurso de acceso, gestión y análisis de información.
- Necesidades de la academia e industria de tecnologías del lenguaje e IA para el desarrollo de sistemas de calidad, competitivos y útiles.

Participantes:

- Miguel Pedrera, Hospital Universitario 12 de Octubre
- Joan Calvet, Institut d'Investigació i Innovació Parc Taulí, Hospital de Sabadell
- Joaquín Grande Baos, Servicio Canario de la Salud
- Andrea Guala, Vall d'Hebron Instituto de Investigación
- Antonio José Moreno Rojas, Coordinador de Interoperabilidad y Estrategia de Salud Digital, Servei Salut Illes Balears

Preguntas de asistentes y conclusiones (10 min)

16:30 - 17:30

Mesa Redonda 3: Oportunidades, retos y tendencias de desarrollos innovadores basados en IA y Procesamiento del Lenguaje Natural para el sector de la salud.

Moderador: Jesús Pinilla, Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA)

Temas a tratar incluyen:

- Tendencias del sector tecnológico de procesamiento de datos de salud: nuevos desarrollos de productos basados en IA y PLN
- Retos y barreras para la industria de tecnologías del lenguaje y software inteligente para mejorar el aprovechamiento de la historia clínica electrónica
- Principales necesidades de empresas que quieren aprovechar recursos basados en los recientes avances de PLN
- Oportunidades, casos de éxito y escenarios de uso de IA en salud.

Participantes:

- Ricardo Farreras, W4KIT
- Juli Climent, ASHO
- Gabriel de Maeztu, IOMED
- Ander Martínez, Fujitsu Research of Europe
- Pilar López-Úbeda, HT medica

Preguntas de asistentes y conclusiones (10 min)

17:30-18:00

Conclusiones, clausura y reflexión de la jornada

- Alfonso Valencia, BSC
- Xavier Pastor, Hospital Clínic de Barcelona
- Jesús Pinilla, SEDIA