

Procesamiento de lenguaje natural (PLN) supervisado con codificación automática para una mejor atención clínica

Xavier Pastor, Artur Conesa, Raimundo Lozano y Elisa Asensio.
Hospital Clínic – Universitat de Barcelona

Infoday sobre tecnologías del lenguaje en sanidad y biomedicina.
2 de Diciembre, 2019

Plan TL

Plan de Impulso de las
Tecnologías del Lenguaje



UNIVERSITAT DE
BARCELONA

CLÍNIC
BARCELONA

Hospital Universitari

Hospital Clínic: modelo "dual", un verdadero desafío

■ Hospital universitario de alta tecnología (datos de 2018)

- Camas: 713; quirófanos: 31; salas de partos: 7; consultorios ambulatorios: 240
- Hospital de cuidados agudos (44.700 ingresos / año)
- Clínica ambulatoria de especialidades (529.134 visitas / año)
- Cirugía (23.127 / año) (Trasplantes de órgano: 458 / año)

Para mantener el umbral alcanzado como un hospital de alta tecnología

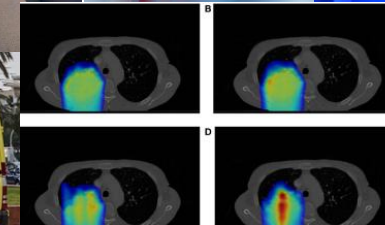
Y...
Para enfrentar y liderar el desafío de la "Atención continuada del paciente"

- Médicos: 1.130
- Enfermeras / enfermeras auxiliares: 2.485
- Personal administrativo: 634
- Servicios generales: 263
- Consejo de administración: 9
- Presupuesto de investigación : 75.799.253 euros

- Liderando el proceso de reingeniería entre Atención Primaria y Especializada para brindar una verdadera "Atención Integrada" que incluye atención domiciliaria y programas para pacientes crónicos (EPOC, ICC, SIDA, Cáncer), Atención a Largo Plazo y conexión de Salud Mental en una población de aproximadamente 520.000 habitantes.

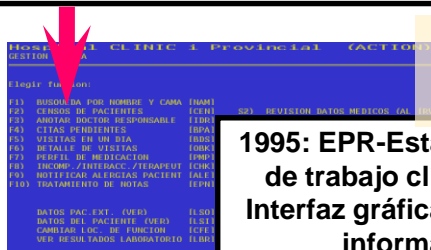


INTELLIGENT CONNECTIVITY



Implementación de la HCE en el Hospital Clínic

1984: HIS - interfaz textual sin información estructurada – S. Financiero



Datos administrativos

1995: EPR-Estación de trabajo clínico. Interfaz gráfica con información estructurada.



1997: ERP.

Documentación clínica



2003: ERP + EPR: Solución clínica. Interfaz gráfica con información estructurada y de flujo de trabajo.

Proceso de integración

2007: Historia clínica completa

2010: eMedication & data-based "CDSS"

2011-2017: Mejor personalización orientada las necesidades profesionales y a la seguridad del paciente.

2018-.....: Mejora funcional y "CDSS" basado en el conocimiento

Hospital Clínic

2011 → HIMSS EMRAM SCORE 6 / 7

Stage	Cumulative Capabilities
Stage 7	Comprehensive EMR/CDSS integration to share data. Data supporting leading outcomes reports, quality assurance, and business intelligence. Data continuity with IT, laboratory, etc.
Stage 6	Physician documentation interface with full CDSS (prioritized warnings, alerts, clinical decision support, evidence & compliance alerts) and clinical and medication administration.
Stage 5	Full implementation of CDSS (clinical alert, non-alert images).
Stage 4	CDSS in at least one clinical service area under for medication (i.e. a Pharmacy) may have Clinical Decision Support based on clinical guidelines.
Stage 3	Non-prioritized documentation. Blue alerts. May have Clinical Decision Support for one checking drug-drug and/or drug-PCSS alerts (single ratings).
Stage 2	Clinical Data Repository (CDR). (Electronic Patient Record may have Controlled Medical Vocabulary). Clinical Decision Support (CDSS) for preliminary certified findings. Document imaging and health information exchange (HIE) capability.
Stage 1	Advanced I-CD. Settings. Pharmacy. All installed CDSS processing (i.e. HIE). HIE data input (data from external service providers).
Stage 0	All Three Accidents (i.e. HIE, PHO) not installed. CDSS not processing I-CD. Settings. Pharmacy data input (data from external service providers).

2018 → fully paperless

eCPOE

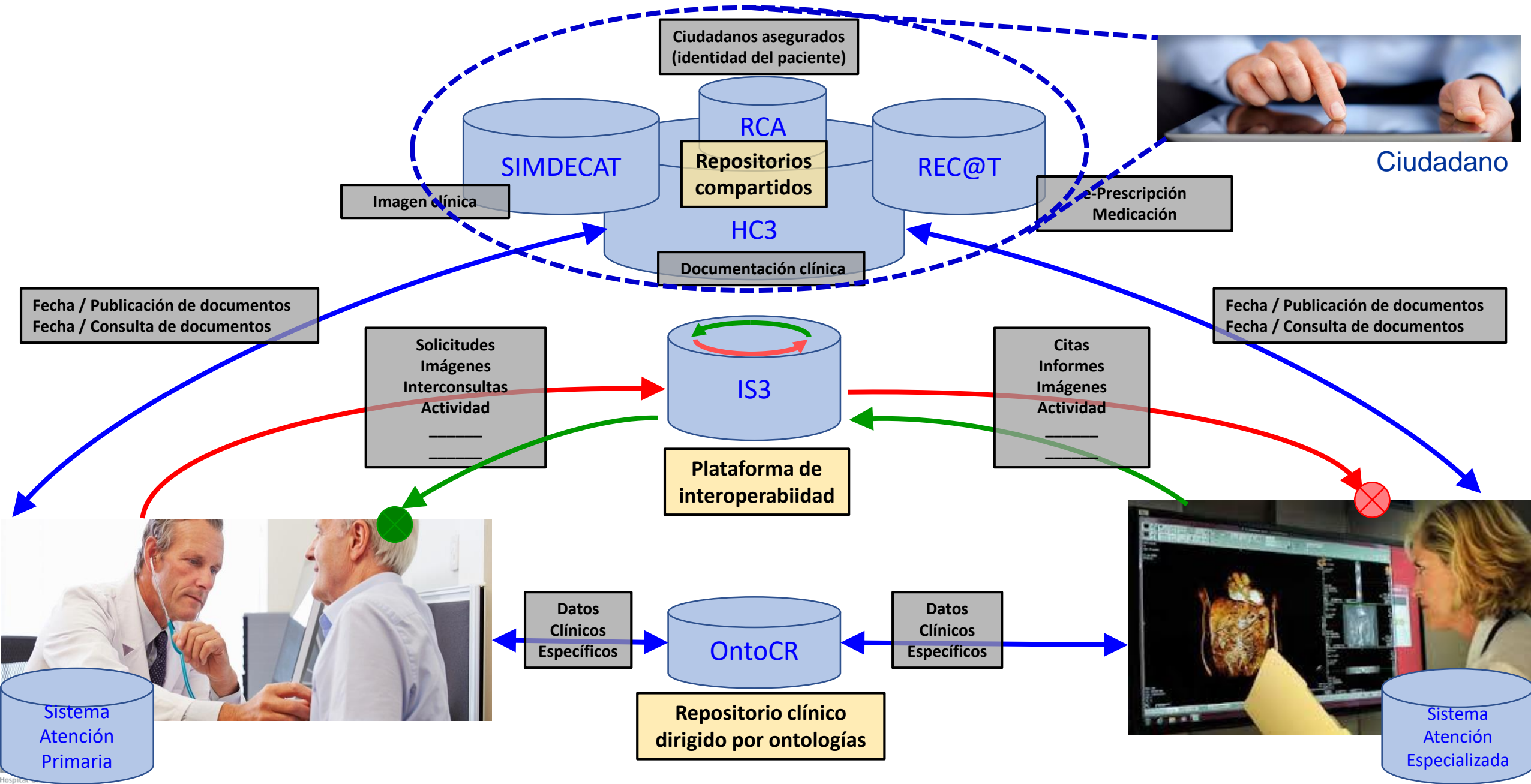
CDSS



CPOE: Computerized Physician Order Entry

CDSS: Clinical Decision Support System

Modelo de trabajo colaborativo en el SISCAT



Siguiendo los "lemas" de la WHO-FIC...

Una forma de llevar la codificación al punto de atención.

J
u
s
t
i
f
i
c
a
c
i
ó
n

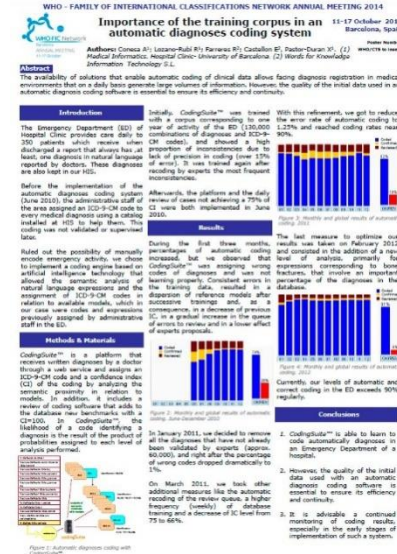
- La **narrativa textual** es la forma de comunicarse entre los profesionales de la salud.
- **Los problemas de salud** son impulsores esenciales para tomar decisiones clínicas.
- **Una lista de problemas de salud** sólida es crucial en un escenario de salud integrado para que los médicos compartan y administren adecuadamente.
- **Las terminologías** son fundamentales para aplicar **las tecnologías de la información y la comunicación** en tiempo real en el punto de atención.
- **CIE y SNOMED-CT** son dos terminologías comúnmente utilizadas en la asistencia sanitaria.
- **El procesamiento del lenguaje natural** es una nueva tecnología basada en la computación de datos pesados con algoritmos de **inteligencia artificial**, que podría aplicarse con éxito en tiempo real al lenguaje clínico para **obtener el código adecuado** con supervisión humana experta.

2015 Manchester: "Interoperability: Meaningful Exchange of Health Information"


2016 Tokyo: "Health Information in the New Era"

2017 Mexico City: "Sharing Data for Health"

2019 Banff: "Health information meets Health Informatics"



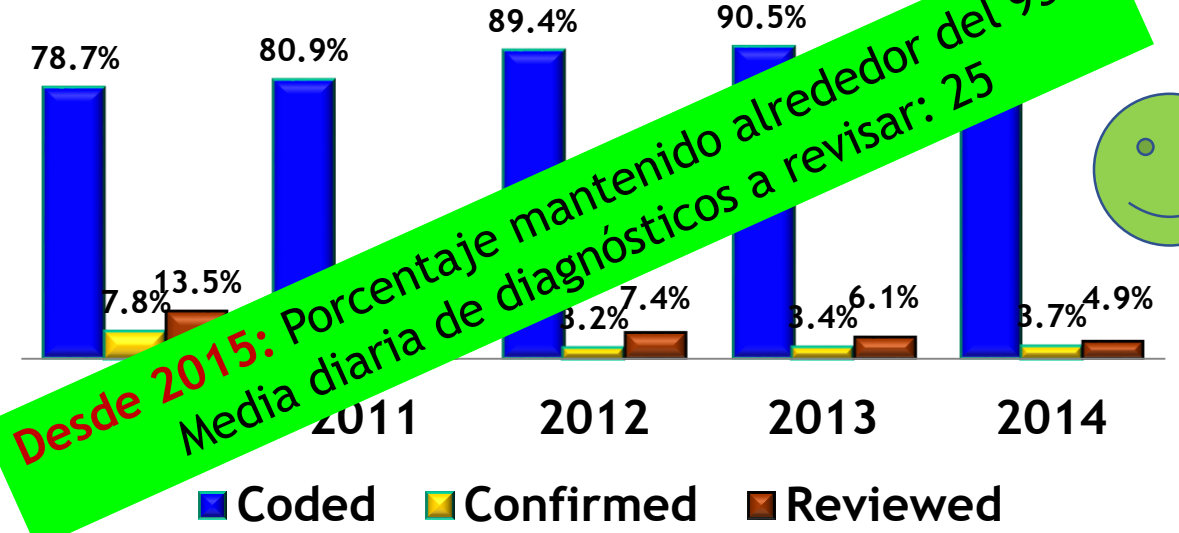
2014: resultados iniciales. Algunos éxitos y muchas lecciones aprendidas

- 2010: Entrenamiento inicial con códigos no supervisados.
- Aprendizaje positivo pero con menor velocidad de la esperada. 
- Detección de asignación de códigos erróneos (15% de inconsistencias)

Acciones correctoras

- Eliminación de diagnósticos no validados.
- Disminución del índice de confianza (IC) del 75% al 66% para forzar la revisión.
- Recodificación de la cola de revisión acumulada.
- Entrenamiento semanal de la base de datos.

Resultados

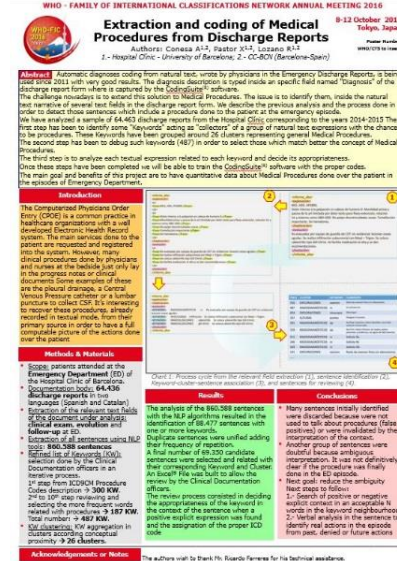


Enseñanzas: Se requieren datos de calidad y una supervisión continua por parte de codificadores expertos

2016: un reto. Identificación de concepto mediante PLN en el texto libre antes de codificar

El hándicap

- Los procedimientos médicos en Urgencias están integrados dentro de la narrativa del texto natural en el formulario de informe de alta.
- El problema era identificarlos adecuadamente antes de aplicar la herramienta de codificación.



La idea

- Corpus: 64.463 informes de urgencias
- Identificación de "Palabras clave" que pueden ser procedimientos.
- Agrupación de "Palabras clave" en 26 grupos que representan procedimientos médicos generales.
- Depuración de las palabras clave (487) para seleccionar la mejor coincidencia.
- Análisis de pertinencia de los pares "expresión textual" + "Palabra clave".

XML Snippets:

```
<informe_alta>
  exploración
  <tag>BEG, HDE, AFEBRIL</tag>
  </tag>
  <tag>Dolor intenso a la palpación en cabeza de humero D.</tag>
  <tag>Movilidad activa y pasiva de la art limitada por dolor tanto para flexo-extensión, rotación int y externa como ABD-ADD.</tag>
  <tag>No palpo discontinuidades óseas.</tag>
  <tag>Medicación importante:</tag>
  <tag>Sin hematoma.</tag>
  </tag>
  <tag>Re evaluada por equipo de guardia de COT sin evidenciar lesiones óseas agudas.</tag>
  <tag>Se realiza infiltración subacromial con Mepi + Trigen.</tag>
  <tag>Se coloca cabestrillo tipo Gill-Christ.</tag>
  <tag>Se facilita medicación al alta y se dan recomendaciones.</tag>
  </tag>
</informe_alta>
```

FREQ	CLUSTER	KEYWORD	CANDIDATE
550	EXPLORACIONES	examen	Resto de examen físico sin alteraciones.
487	RADIOLOGIAGNÓSTICOS	rx	Se solicita Rx.
454	EXPLORACIONES	otoscopia	Otoscopia.
337	SUTURAS	puntos	Girange 15 puntos.
331	RADIOLOGIAGNÓSTICOS	ap	Ap Resp expirado, vértice desdoblado, mamario visible controlado.
329	RADIOLOGIAGNÓSTICOS	ap	Ap C1c: tonos rítmicos, sin soplos, pulso presente y simétrico; no signo de feto derecho.
311	RADIOLOGIAGNÓSTICOS	rx	Solicito Rx.
246	RADIOLOGIAGNÓSTICOS	rx	Solicito Rx.
183	RADIOLOGIAGNÓSTICOS	rx	Solicito RX.
141	EXPLORACIONES	examen	Resto de examen físico sin alteraciones.

KEYWORDS: RADIOLOGIAGNÓSTICOS, INYecciones, INNOVACIONES, gill-christ

Enseñanzas: La identificación de conceptos en texto libre es posible con el refinamiento de la herramienta de PLN.

Resultados

- Asignación correcta en la mayoría de los casos.
- Algunos otros casos rechazados por ambigüedad o negatividad.
- Necesidad de buscar el contexto explícito y hacer un análisis verbal en el entorno de las palabras clave.

2017: un nuevo reto: Rastreo completo mediante PLN del informe de alta y codificación con CIE-10-MC

El hándicap

- Compromiso nacional para cambiar el estándar de codificación de CIE-9-MC a CIE-10-MC en Enero 2018.
- Cambios muy importantes: número de entradas, estructura de los códigos, más requisitos de información clínica y un lenguaje completamente nuevo para la codificación de procedimientos.

La solución

- Desarrollo de plantillas para los informes de alta.
- Escaneo de texto mediante PLN que identifica palabras clave correspondientes a diagnósticos y procedimientos.
- Propuesta de código para cada expresión seleccionada.
- Validación o revisión por codificadores expertos.

The screenshot displays a software interface for medical coding. It is divided into several panels:

- ANTECEDENTES:** Contains patient history text such as "mujer de 83 años sin alergias medicamentosas conocidas ni hábitos tóxicos" and "fractura peritrocantérea fémur izquierdo".
- SEGMENTOS:** Lists identified key terms like "fractura osteoporótica fémur izquierdo", "déficit de vitamina b12", "trastorno cognitivo leve", "hipertensión arterial", "cardiopatía isquémica", "stent", and "ap iam".
- DIAGNOSTICO:** Shows the mapping of these terms to ICD-10-MC codes. For example, "fractura osteoporótica f" is linked to M80.052A, "hiperparatiroidismo sec" to E21.1, and "osteosíntesis con clavo tñna femur izq" to Q0S706Z.
- PROCEDIMIENTO:** Lists surgical procedures like "osteosíntesis con clavo tñna femur izq" and "clavo tñna".

At the bottom, there are filters for "CLIQ" (1) and date ranges (07-04-2019 13:36).

Enseñanzas: La herramienta ofrece una propuesta completa de códigos y ahorra tiempo al codificador

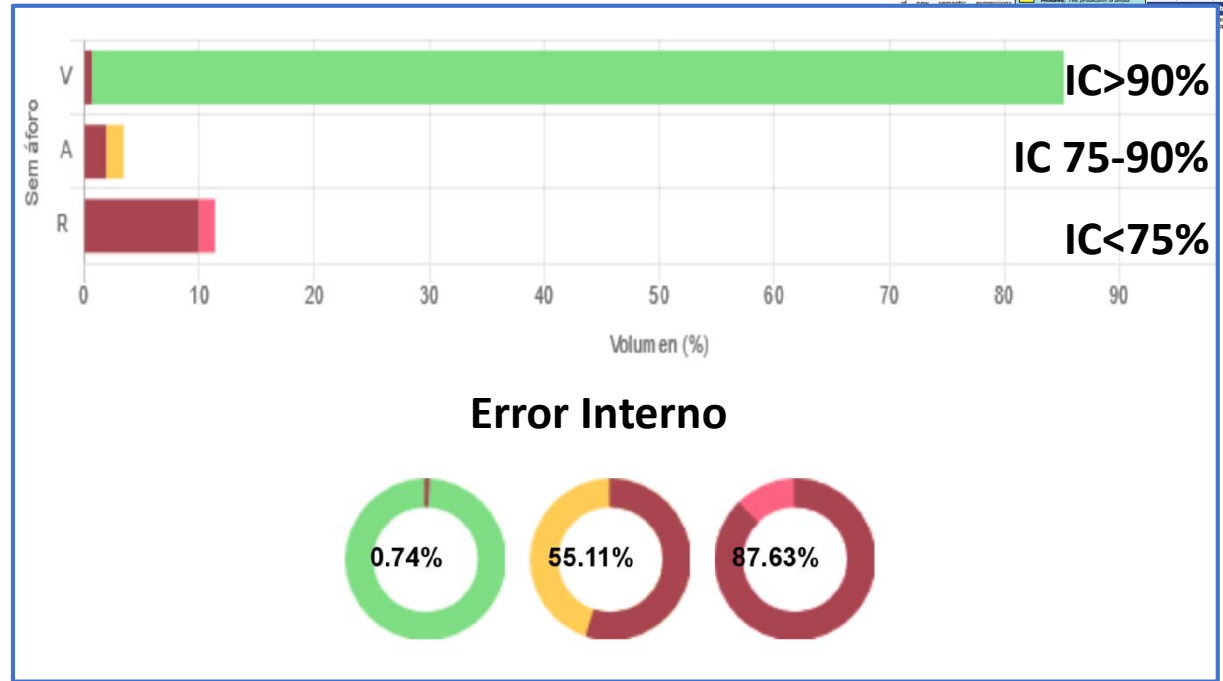
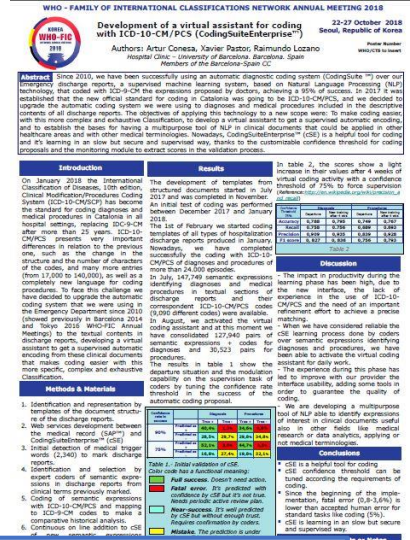
2018: Una preocupación: calidad y supervisión de las tecnologías de PLN

- Enero 2018: Entrenamiento inicial con CIE-10-MC.
 - Agosto 2018: Activación del asistente virtual.
 - Septiembre 2019.
- Evaluación cualitativa del entrenamiento inicial.
Mejoría lenta.

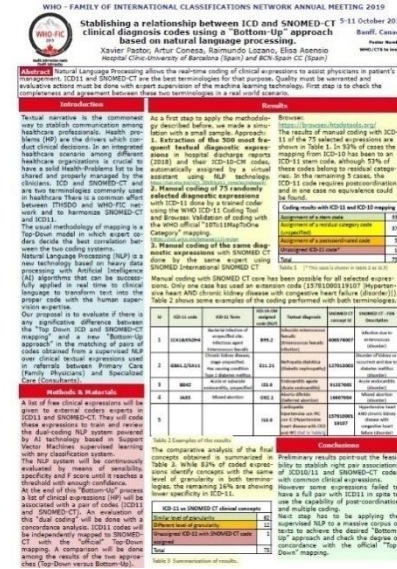
Evolution of scores (learning monitoring)

Confidence Threshold 75%	Diagnosis		Procedures	
	Departure	New training after 4 wks	Departure	New training after 4 wks
Accuracy	0,788	0,795	0,749	0,767
Recall	0.758	0.756	0,689	0.693
Precision	0,909	0,935	0,839	0,928
F1 score	0, 827	0, 836	0,756	0,793

http://en.wikipedia.org/wiki/precision_and_recall



Enseñanza: El conocimiento del error interno para cada IC permite una supervisión óptima del aprendizaje automático



2019: preparados para nuevos proyectos en el mundo clínico real paso a paso

- **Extender** la codificación automatizada a otros entornos: pacientes ambulatorios, hospital de día, hospitalización domiciliaria, ...
- Entrenar la herramienta de PLN con **otras palabras clave** para extraer datos para la atención médica y la investigación: problemas de salud, medicamentos, metabolismo, alergias, ...
- Actualizar la herramienta de PLN para codificar con **múltiples terminologías** : ICD10, ICD11, ICF, ICHI, SNOMED-CT...

Observaciones finales y propuestas

Observaciones

- El PLN es una tecnología útil para codificar.
- La codificación con PLN se puede aplicar de manera efectiva y eficiente en un entorno clínico actual.
- La codificación con PLN requiere una tecnología adecuada y la supervisión por codificadores expertos del proceso completo.
- La aplicación del PLN en cualquier otro entorno requiere evaluación y capacitación periódicas
- El PLN abre un camino para aplicaciones nuevas y sostenibles en el trabajo clínico y de investigación

Propuestas

- Continuar con la hoja de ruta:
 - Múltiples opciones de codificación: ICD10CM, ICD11, ICF, ICHI, SNOMED-CT, ...
 - Nuevos proyectos:
 - Codificación de problemas de salud entre atención primaria y especializada.
 - Concordancia de la codificación dual y evaluación de mapeo utilizando un enfoque ascendente (bottom-up)
 - Uso del PLN en investigación
- Desarrollar un modelo estándar para PLN aplicado a los textos clínicos y la codificación: objetivos, funciones, herramientas y usuarios



Muchas gracias por su atención !!!

xpastor@clinic.cat