

# Anotación de casos clínicos con Snomed CT

Mesa Redonda: Acceso, distribución y relevancia de datos anotados y corpus para tecnologías del lenguaje



VeraTech está reconocida como SPIN-OFF de la Universitat Politècnica de València



**VeraTech**  
FOR HEALTH



**PYME INNOVADORA**  
15 de mayo al 31 de octubre de 2021



VeraTech está reconocida como PYME INNOVADORA por el Ministerio de Economía y Competitividad

# Experiencia de partida

- Anotación de 1000 casos clínicos narrativos en castellano en el dominio oncológico.
  - Casos sin ningún tipo de estructura.
  - Variabilidad entre los casos: distintos orígenes, vocabulario, longitud.
  - Tipos de entidades: síntomas, enfermedades, entidades observables, fármacos y procedimientos.
  - Cada mención (evidencia textual) se anota con uno o varios tipos de entidades y uno o varios concepto Snomed CT.

# Proceso de anotación

- Dos anotadores independientes.
- Fases principales:
  1. Desarrollo de una guía de anotación. Tarea compleja y continúa.
  2. Descubrimiento de menciones (evidencia textual) y asociación del tipo de entidad.
  3. Mapeo de las menciones con Snomed CT.
- Se optó por un proceso en dos fases para:
  - Reducir el trabajo de mapeo con Snomed CT (una misma evidencia textual puede aparecer múltiples veces).
  - Asegurar el tener un único mapeo a Snomed CT para cada evidencia textual.

# Algunas lecciones aprendidas

- La guía de anotación es básica.
- Los anotadores debe conocer con detalle la guía y deben participar en su mejorar durante el proceso de anotación.
- La separación entre tipos de categorías debe estar clara y si no es posible establecer prioridades. Por ejemplo, etiquetar el fármaco y no el procedimiento de administración del fármaco.
- La terminología utilizada tiene consecuencias directas sobre el proceso de anotación y la guía de anotación. Por ejemplo, lo calificadores que soporta (lateralidad, abordaje, dispositivos, severidad, etc.).
- Debe incluir ejemplos.
- Las reglas de selección del concepto de la terminología deben ser precisas.

# Subconjunto de conceptos Snomed CT por tipo de entidad

TIPO DE ENTIDAD	DEFINICIÓN DEL SUBCONJUNTO	NUMERO DE CONCEPTOS
SINTOMA	<404684003   hallazgo clínico (hallazgo)   <u>minus</u> <64572001   enfermedad (trastorno)	35719
ENFERMEDAD	<64572001   enfermedad (trastorno)	75361
PROCEDIMIENTO	<71388002   procedimiento (procedimiento)	57805
FARMACO	<373873005   producto biológico/farmacéutico (producto)	26651
ENTIDAD OBSERVABLE	< 363787002   entidad observable (entidad observable)	8877

Tabla 1. Subconjunto de SNOMED CT para cada grupo semántica

# Explotación e interoperabilidad del corpus anotado

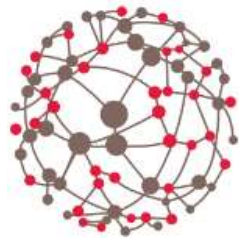
- El uso de Snomed CT en particular como terminología de anotación puede suponer diversos beneficios y retos:
  - Facilita la compartición del corpus anotado: terminología multi-idioma, incluye gran cantidad de sinónimos para los conceptos.
  - Descubrir enlaces entre las anotaciones utilizando los enlaces definatorios entre los conceptos Snomed CT.
  - Explotar el lenguaje de consulta basado en el significado: Identificar casos clínicos, validar las anotaciones.
    - “muestra los casos clínicos sobre neoplasias benignas del abdomen”

<<20376005 | enfermedad neoplásica benigna (trastorno) | :  
363698007 | sitio del hallazgo (atributo) | =<<113345001 |  
estructura abdominopélvica (estructura corporal) |



CONCEPTO	IDENTIFICADOR DEL CONCEPTO
seudofibroma de ligamento ancho (trastorno)	254879003
papiloma benigno de vejiga (trastorno)	281700003
neoplasia benigna del divertículo de Meckel (trastorno)	92212001
adenoma de duodeno (trastorno)	447731000
neoplasia benigna de la médula suprarrenal (trastorno)	91968002
neoplasia benigna del intestino delgado (trastorno)	92385005
neoplasia benigna del trigono vesical (trastorno)	92452001

# Gracias por su atención



**VeraTech**  
FOR HEALTH



**VeraTech**  
FOR HEALTH



[www.veratech.es](http://www.veratech.es)



[veratech@veratech.es](mailto:veratech@veratech.es)



+34 961 071 863



@veratech