

BioPortal

Ecosistema

- ▶ datahub.io
- ▶ Bio2RDF, LOD
- ▶ NLM (<https://www.nlm.nih.gov/>)
 - ▶ Medline, Mesh, PubMed, PubMed Central, Entrez
 - ▶ UMLS
- ▶ OBO Foundry
- ▶ WHO Classifications
- ▶ NCBO y BioPortal

datahub.io

- ▶ *Free Data management platform* de Open Knowledge Foundation
- ▶ Basada en CKAN: herramienta para el manejo y publicación de colecciones de datos.
- ▶ Organizaciones e individuos registran los Datasets que quieren dar a conocer. (ej: Bio2RDF, BioPortal, Canada, ...)
- ▶ +10.000 datasets registrados.
- ▶ Facilita búsqueda y obtención de metadata, links de descarga, SPARQL endpoints, acceso web, ejemplos de datos, OWL Schemas, etc.
- ▶ acceso desde portal web y API web services.

Bio2RDF

- ▶ Proyecto *open-source* de Universidad Laval y otras (Canada) iniciado en 2008
- ▶ Objetivo: RDFizar bases de datos biológicas de diferentes orígenes y formatos, y publicarlas como Open Linked Data (LOD)
- ▶ Define una “RDFization Guide” y provee herramientas y scripts para transformaciones.
- ▶ 378 datasets en datahub.io
- ▶ SPARQL endpoints para algunos de ellos.
- ▶ API de consulta, descargas de n-quads, metadata, esquemas OWL, estadísticas, mappings, faceted browser, etc.

NLM – NCBI

National Library of Medicine – Center for Biotechnology Information

- ▶ Medline
 - ▶ Base de datos bibliográfica de NLM
 - ▶ Referencias a publicaciones desde 1950
- ▶ MeSH (Medical Subject Headings)
 - ▶ Tesauro controlado de la NLM
 - ▶ Indexa artículos médicos, ClinalTrials.gov, y otras publicaciones, audio y video.
- ▶ PubMed
 - ▶ Portal online de acceso a la base de datos de Medline
 - ▶ Links a abstracts y artículos completos (PubMed Central casi 4M)
- ▶ Entrez
 - ▶ Buscador cross-database (40): literatura, moleculares, genéticas, etc.

UMLS (Unified Medical Language System)

- ▶ Iniciado en 1986 por NLM
- ▶ Objetivo: Asistir a profesionales de la salud e investigadores en la recuperación en integración de información biomédica.
- ▶ Obstáculo: diferentes vocabularios, unificación impracticable.
- ▶ Propuesta: crear vínculos entre los diferentes vocabularios (ontology alignment)
 - ▶ CUI: Concept Unique Identifier. Vincula *strings* con el mismo significado.
- ▶ Gratuito (requiere licencia UMLS)
- ▶ Acceso desde Browser, Web Service API, o instalación local
- ▶ 3 herramientas principales
 - ▶ Metathesaurus
 - ▶ Semantic Network
 - ▶ SPECIALIST Lexicon and Lexical Tools

OBO Foundry

- ▶ The Open Biological and Biomedical Ontologies Foundry
- ▶ Comunidad de desarrolladores de ontologías.
- ▶ Revisan y mantienen ontologías y vocabularios de acuerdo a un conjunto de principios (apertura, formatos, versionado, documentación, propiedades, etc)
- ▶ Con el objetivo de desarrollar una familia de ontologías interoperables que sean **lógicamente** bien formadas y **científicamente** precisas.

WHO Classifications

- ▶ Familia de clasificaciones internacionales (WHO-FIC)
 - ▶ ICD: **Diseases**. Version actual ICD-10
 - ▶ ICF: **Functioning, Disability and Health**
 - ▶ ICHI: **Health Interventions**

NCBO (National Center for Biomedical Ontology)

- ▶ Financiada por NIH
- ▶ Visión: promover el uso de ontologías para conseguir interoperabilidad semántica de todo el conocimiento biomédico
- ▶ Soporte a la comunidad de investigadores con herramientas, capacitación, recursos, foros, publicaciones, etc.
- ▶ En el desarrollo de las herramientas están involucradas las universidades de Stanford, de Victoria y de Buffalo, y la clínica Mayo.

BioPortal

- ▶ “El repositorio mas exhaustivo de ontologías biomédicas”
- ▶ Portal Web y API REST
- ▶ Gratuito. Requiere registro y api key.
- ▶ NCBO Virtual Appliance (incluye source) previa solicitud.
- ▶ Releases mensuales (2016) con bugfixes y mejoras.
- ▶ BioPortal permite que la bio-comunidad pueda publicar, explorar, descubrir, evaluar, colaborar, criticar, etc., las bio-ontologías recopiladas.

Operaciones básicas

- ▶ Browse
 - ▶ buscar ontologías por nombre o descripción, y facetado
- ▶ Submit
 - ▶ Registrar una nueva ontología.
 - ▶ Pública o Privada (con licencia o autorizaciones), o “Solo resumen”.
- ▶ Search
 - ▶ Buscar clases o términos en las ontologías del bioportal.

Ontology page

- ▶ Resumen
- ▶ Clases
 - ▶ Detalles
 - ▶ Visualización
 - ▶ Notas
 - ▶ Mappings (de clase)
- ▶ Propiedades
- ▶ Notas
- ▶ Mappings (de ontología)
- ▶ Widgets

Mappings (Ontology Alignment)

▶ UMLS

- ▶ Cada nuevo término es revisado **primero por algoritmos** de *lexical matching* y luego **manualmente** por editores de NLM para determinar si es sinónimo de algún término existente.

▶ BioPortal

- ▶ Los términos se analizan **únicamente en forma automática**.
 - ▶ Algoritmos LOOM (Lexical OWL Ontology Matcher)
 - ▶ UMLS CUIs y OBO xrefs
- ▶ La **comunidad** de BioPortal pueden agregar mappings, a través de la web de BioPortal, o por la API Rest.
- ▶ Esto puede generar incoherencias.

Annotator

Anota texto libre
con términos de las
ontologías
seleccionadas

Annotator
Get annotations for biomedical text with concepts from the ontologies [?](#)

insert sample text

Melanoma is a malignant tumor of melanocytes which are found predominantly in skin but also in the bowel and the eye.

Select Ontologies

SNOMEDCT

[clear selection](#) [select from list](#)

Annotations total results 10 (direct 10 / ancestor 0 / mapping 0)

CLASS filter	ONTOLOGY filter	TYPE filter	CONTEXT	MATCHED CLASS filter	MATCHED ONTOLOGY filter
Malignant melanoma	Systematized Nomenclature of Medicine - Clinical Terms	direct	Melanoma is a malignant ...	Malignant melanoma	Systematized Nomenclature of Medicine - Clinical Terms
Is a	Systematized Nomenclature of Medicine - Clinical Terms	direct	Melanoma is a malignant tumor of ...	Is a	Systematized Nomenclature of Medicine - Clinical Terms
Malignant neoplastic disease	Systematized Nomenclature of Medicine - Clinical Terms	direct	... is a malignant tumor of melanocytes which ...	Malignant neoplastic disease	Systematized Nomenclature of Medicine - Clinical Terms
Malignant	Systematized Nomenclature of Medicine - Clinical Terms	direct	... is a malignant tumor of melanocytes ...	Malignant	Systematized Nomenclature of Medicine - Clinical Terms
Neoplasm	Systematized Nomenclature of Medicine - Clinical Terms	direct	... a malignant tumor of melanocytes which ...	Neoplasm	Systematized Nomenclature of Medicine - Clinical Terms
Skin structure	Systematized Nomenclature of Medicine - Clinical Terms	direct	... predominantly in skin but also in ...	Skin structure	Systematized Nomenclature of Medicine - Clinical Terms

Recommender

Utiliza servicio de Annotator y luego recomienda que ontología/s se adaptan mejor al texto de acuerdo a 4 criterios:

- ▶ Coverage
- ▶ Acceptance
- ▶ Knowledge Detail
- ▶ Specialization

Input

Text Keywords (separated by commas)

Output

Ontologies Ontology sets

insert sample input

Primary treatment of DCIS now includes 3 options: lumpectomy without lymph node surgery plus whole breast radiation (category 1); total mastectomy with or without sentinel node biopsy with or without reconstruction (category 2A); lumpectomy without lymph node surgery without radiation (category 2B). Workup for patients with clinical stage I, IIA, IIB, or T3,N1,M0 disease was reorganized to distinguish optional additional studies from those recommended for all of these patients. Recommendation for locoregional treatment for patients with clinical stage I, IIA, IIB, or T3,N1,M0 disease with 1-3 positive axillary nodes following total mastectomy was changed from "Consider" to "Strongly consider" postmastectomy radiation therapy.

advanced options

Edit Input

Recommended ontologies

POS.	ONTOLOGIES	FINAL SCORE	COVERAGE SCORE	ACCEPTANCE SCORE	DETAIL SCORE	SPECIALIZATION SCORE	ANNOTATIONS	HIGHLIGHT ANNOTATIONS
1	NCIT SNOMEDCT CADSRVS ONTOAD	71.6	95.0	77.7	64.9	64.2	42	<input checked="" type="checkbox"/>
2	NCIT SNOMEDCT CADSRVS MESH	70.2	92.6	82.3	64.4	64.4	39	<input type="checkbox"/>
3	NCIT SNOMEDCT CADSRVS HUPSON	69.6	91.1	81.6	63.1	66.7	40	<input type="checkbox"/>
4	NCIT CADSRVS RCD ONTOAD	69.5	88.7	64.1	65.6	72.5	40	<input type="checkbox"/>
5	NCIT SNOMEDCT CADSRVS LOINC	69.4	91.1	82.2	62.4	66.3	39	<input type="checkbox"/>
6	NCIT SNOMEDCT CADSRVS	69.2	90.3	82.1	63.1	67.1	39	<input type="checkbox"/>
7	NCIT SNOMEDCT LOINC	69.0	91.8	87.9	64.4	59.1	41	<input type="checkbox"/>

Criterios Recommender

- ▶ **Coverage:** el Recommender invoca al Annotator y usa el resultado para determinar la cobertura de cada ontología sobre el texto ingresado.
- ▶ **Acceptance:** la cantidad de visitas a la ontología en el Bioportal, y su **presencia o no en UMLS** se usan para computar que tan conocida y confiable es la ontología para la comunidad biomédica.
- ▶ **Detail of knowledge:** usa el número de definiciones, sinónimos y propiedades de las clases que “anotan” al texto para determinar el nivel de detalle que le provee la ontología.
- ▶ **Specialization:** determina que tan especializada es cada ontología para el dominio del texto ingresado. Se calcula usando el número y tipo de las anotaciones y la **posición de cada clase anotada en la jerarquía de la ontología**. El resultado se normaliza por el tamaño de la ontología para identificar las ontologías más chicas especializadas en el dominio.

Resource Index

- ▶ Utiliza el servicio de Annotator para anotar recursos disponibles públicamente.
- ▶ Se pueden proponer la indexación de recursos disponibles públicamente enviando un mail a support@bioontology.org
- ▶ “salida al exterior” del bioportal

API Rest

- ▶ <http://data.bioontology.org/documentation>
- ▶ Requiere API Key
- ▶ Search, Annotator, Recommender, Resource Index
- ▶ Batch endpoints
 - ▶ Recibe colección con varias consultas en una sola llamada.
 - ▶ Por ahora solo clases
- ▶ Ontology analytics
- ▶ Resource endpoint: Media Types, Hypermedia links

SPARQL Endpoint (beta)

- ▶ <http://sparql.bioontology.org/>
- ▶ Consultas sobre las 2 bases de datos: ontologías y mappings
- ▶ No funciona en Firefox
- ▶ La UI web no permite consultas federadas
- ▶ No soporta “property paths”



Gracias por su atención!

Preguntas?...

...pueden escribirme a diego.bortot@fing.edu.uy