

**SERVIDOR DE TERMINOLOGÍAS ORBITS
PARA LA CODIFICACIÓN DE FENOTIPOS
CLÍNICOS EN CÓDIGOS HPO.**

RECUPERACION DE CASOS MEDIANTE PLN
EN ESCENARIO DE ENFERMEDADES RARAS.



OBJETIVOS DEL DEMOSTRADOR

Extraer y codificar a la terminología HPO de manera automática la información fenotípica contenida en los reportes clínicos no estructurados en el escenario de las Enfermedades raras utilizando Procesamiento de Lenguaje Natural.

En este contexto, las ontologías fenotípicas como HPO u ORDO tienen como objetivo estandarizar la recopilación de datos fenotípicos haciéndolos accesibles para el análisis informático. Específicamente, HPO se puede combinar con diferentes herramientas computacionales de priorización de genes basadas en fenotipos, como Phen-Gen, etc.

Desafíos del proyecto

- Todas estas herramientas requieren la entrada manual de los fenotipos clínicos del paciente.
- A menudo, solo se proporciona información limitada sobre el fenotipo de un individuo en un formulario de solicitud de prueba.
- Actualmente, el uso de HPO no está completamente integrado en el entorno clínico, por lo que la identificación de variantes causales depende del analista, lo que se traduce en términos muy costosos en términos de tiempo y posibles errores humanos.



**Collaboration project with
Hospital Clinic of Barcelona**

Gloria González Gacio, PhD.
Mireia Rodríguez Naque
Toni Mas Mota
Victoria Bérez Baldrich

**bitac. CLÍNIC
BARCELONA**
Hospital Universitari

Elisa Asensio
Santiago Frid
Artur Conesa, PhD.
Xavier Pastor, PhD.
Clinical Documentation Department
Medical Informatics Department

IMPACT  

BARCELONA HEALTH HUB
RECOGNIZED AS ONE OF THE TOP DIGITAL HEALTH ORGANIZATIONS

SNOMED International

3

El estudio piloto se ha realizado en el Hospital Clínico de Barcelona.

DESCRIPCIÓN DE LA SOLUCIÓN

Implementación de APIs de tipo RESTful del servidor de terminologías ORBITS de BITAC para la codificación de fenotipos clínicos con códigos HPO en Enfermedades Raras a partir de informes clínicos no estructurados, utilizando procesos de PLN basados en SNOMED CT. En este trabajo se diseñó un pipeline completo para extraer conceptos médicos de informes clínicos con su posterior asociación a los términos HPO correspondientes, mediante mapeo manual y su comparación con el mapeo automático utilizando PLN.

Case study Workflow

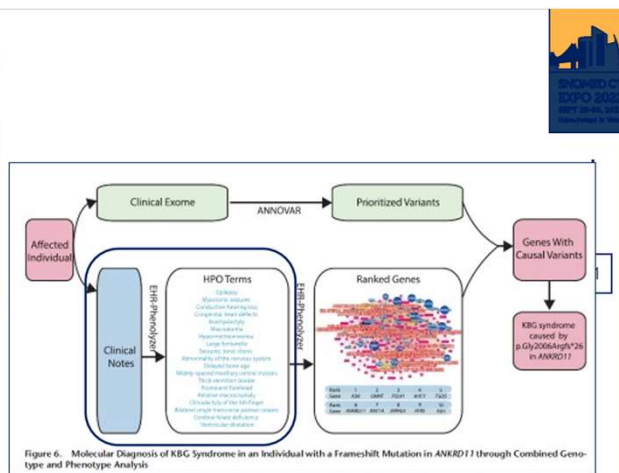
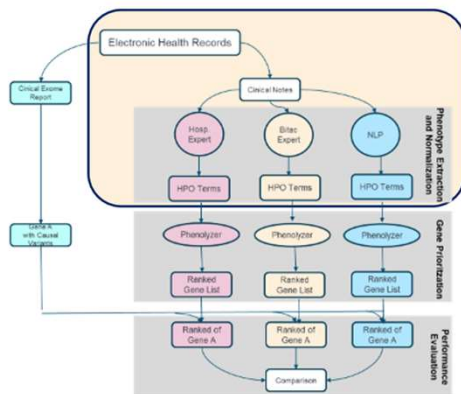


Figure 6. Molecular Diagnosis of KBG Syndrome in an Individual with a Frameshift Mutation in ANKRD11 through Combined Genotype and Phenotype Analysis

NLP orbiTS

APIS Access

```

{
  "entidades": [
    {
      "categoria": "Hallazgo",
      "conceptid": "248200007",
      "descripcion": "dismorfia facial",
      "fsn": "faces dismórfica",
      "Score": 100
    },
    {
      "categoria": "Trastorno",
      "conceptid": "22066008",
      "descripcion": "hipertelorismo",
      "fsn": "hipertelorismo ocular",
      "Score": 100
    },
    {
      "categoria": "Trastorno",
      "conceptid": "391987006",
      "descripcion": "pectus excavatum",
      "fsn": "tórax en embudo",
      "Score": 100
    },
    ...
  ],
  "entidades negadas": [
    {
      "label": "NEG_HALLAZGO",
      "pattern": "386661006"
    },
    {
      "label": "NEG_HALLAZGO",
      "pattern": "111306001"
    }
  ]
}

```

NLP API JSON result example

```

{
  "mapeo": {
    "interminologigen": "SNOMED CT",
    "interminologidestino": "HPO",
    "terminologidestino": "HPO - Human Phenotype Ontology",
    "version": "1.1",
    "impp": "2022",
    "cat": "EQ",
    "tipo": "OTO",
    "accion": "Ene",
    "descripcionterminologigen": [
      {
        "codigolocal": "391987006",
        "descripcionpre": "tórax en embudo (trastorno)"
      }
    ],
    "descripcionterminologidestino": [
      {
        "codigolocal": "HPO:0000767",
        "descripcionpre": "pectus excavatum"
      }
    ]
  }
}

```

Mapping HPO→SNOMED CT API JSON example

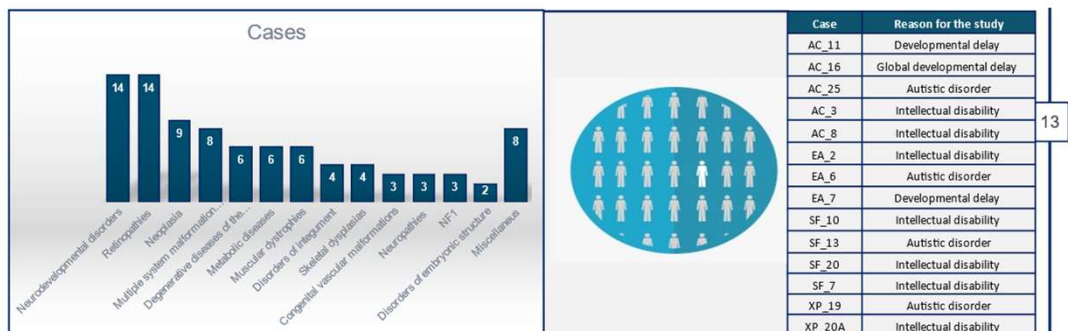
DESARROLLO DE LA SOLUCION UNA DE LAS INSTITUCIONES QUE PERTENECEN A IMPaCT-Data

Para realizar las pruebas correspondientes, se seleccionaron 90 casos de estudio por parte de los departamentos de informática médica y documentación del Hospital Clínic de Barcelona. El criterio era tener una enfermedad rara y haber realizado un test genético, preferiblemente un test masivo.

Los casos fueron anonimizados previamente por el servicio de informática médica y proporcionados en formato de texto libre.

Estos casos incluyen trastornos del neurodesarrollo, retinopatías, neoplasias, neuropatías, distrofias musculares, entre otros. Para este estudio piloto seleccionamos 14 casos relacionados con trastornos del neurodesarrollo. La razón principal para la inclusión fueron los diagnósticos principales: retraso en el desarrollo, discapacidad intelectual o autismo.

Case study Workflow. Cohort



CONCLUSIONES SOBRE EL RESULTADO

- ✓ Las anotaciones fenotípicas son relevantes para interpretar los resultados de la secuenciación del genoma y el exoma completo en pacientes con enfermedades raras.
- ✓ La mayoría de las veces, las notas clínicas se escriben como texto libre en los sistemas electrónicos que las computadoras no pueden usar para el análisis.

- ✓ Las aplicaciones ORBITS NLP proporcionan anotaciones lingüísticas y semánticas a partir de texto libre no estructurado utilizando diccionarios SNOMED CT e incluyen API Restful para obtener códigos relacionados con HPO.
- ✓ Estos conceptos y relaciones se pueden usar para codificar informes clínicos y, por lo tanto, pueden reducir sustancialmente los esfuerzos humanos que consumen mucho tiempo, ahorrando tiempo a los clínicos, facultativos del laboratorio e investigadores.
- ✓ Las ontologías fenotípicas permiten la interoperabilidad entre registros y otros recursos, como biobancos o bases de datos ómicas

CONCLUSIONES HOSPITAL CLÍNIC

Desde el punto de vista del Hospital Clínic de Barcelona, este proyecto ha sido muy beneficioso a modo de demostrador. Las notas clínicas de nuestros sanitarios contienen datos muy ricos a nivel fenotípico, lo cual puede ser particularmente importante en casos de uso como los de enfermedades raras, pero al estar en un formato no estructurado su reutilización se ve comprometida. De hecho, este trabajo prácticamente no se realizaba localmente, desaprovechando el conocimiento generado y el trabajo llevado a cabo. El reconocimiento automático de entidades y su codificación con terminologías como HPO permite evitar trabajos manuales intensivos por parte de los profesionales, y pone a su disposición información semánticamente interoperable para dar soporte a la toma de decisiones clínicas, así como para su uso secundario.

Una de las dificultades que nos hemos encontrado es que, para agilizar el proyecto tras la aprobación del comité, el proceso para la identificación de documentos clínicos pertinentes, así como para su correcta anonimización, fue finalmente manual, requiriendo un gran esfuerzo por parte del grupo de investigación y limitando la cantidad de informes disponibles para utilizar en el demostrador.

En caso de que hubiéramos integrado la solución en el HIS como se había planteado inicialmente, nos habríamos ahorrado mucho tiempo y recursos humanos especializados, supliendo los trabajos manuales intensivos y a su vez permitiendo recuperar mayor número de casos.
