



Infoday sobre Tecnologías del Lenguaje en Sanidad

9 de Septiembre 2018
Hospital Universitario Virgen del Rocío (Sevilla)

bitac. Clinical terminology
for e-health



INNOVATIVE SME

Valid until Dec 31st 2018



BITAC es una empresa global con sede en Barcelona dedicada a la prestación de servicios terminológicos especializados.

Nuestra razón de ser pasa por **ayudar a los profesionales de la salud** a compartir su información en entornos digitales.

Tenemos una amplia **experiencia** en el desarrollo de soluciones relacionadas con los principales estándares de interoperabilidad semántica, como CIE-g/10, LOINC o SNOMED-CT.



Diagnostics



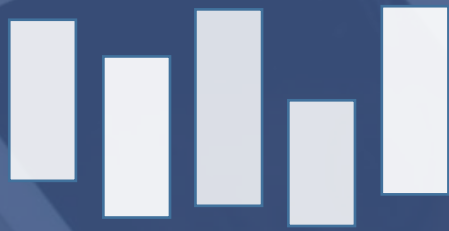
Desde BITAC hemos desarrollado **CTMAP**, una herramienta **de soporte a la codificación de textos biomédicos**

CTMAP facilita la codificación automática de los términos locales a términos normalizados a través de estándares clínicos

Características principales:

- Entorno de **programación como servicios** (SAAS, Software As A Service).
- Integrable con cualquier Sistema de información vía **Servicios Web**
- **Multiterminología** (CIE-10-MC/PCS, CIE-9, LOINC, SNOMED CT, OMIM)
- **Multiidioma**
- **Autoaprendizaje**

Elementos clave de CTMAP



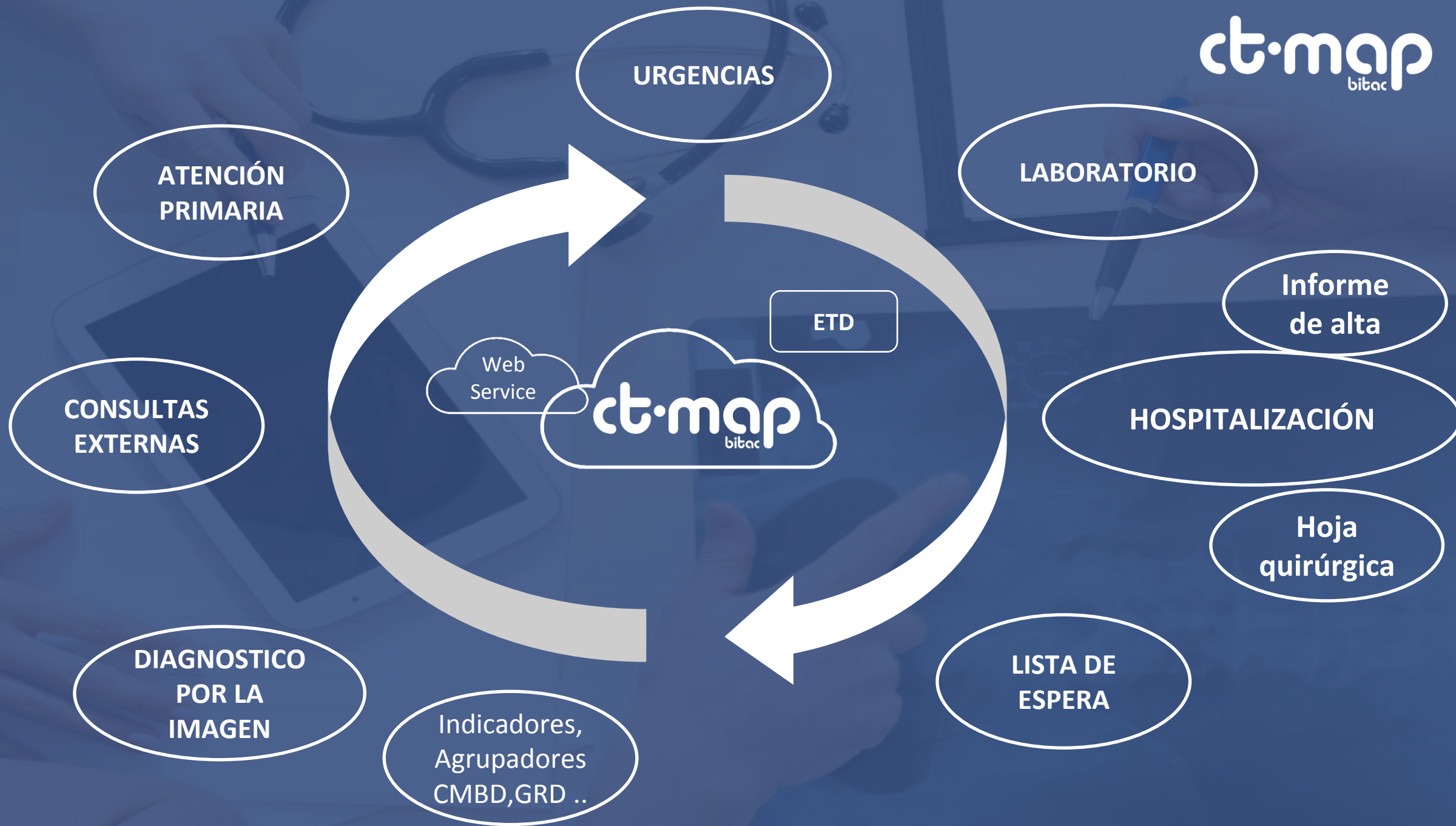
Una base terminológica formada por más de 2,000.000 de sinónimos relacionados con estándares



Una serie de algoritmos de *string matching*



Un conjunto de reglas que permiten refinar la búsqueda y ajustar la coincidencia



Automatic Mapping Tool of Local Laboratory Terminologies to LOINC

Estibaliz Parceró, Jose Alberto Maldonado,
Luis Marco, Montserrat Robles
IBIME. Universitat Politècnica de València.
Valencia, Spain
{esparig, luimarr1, jamaldo, mrobles}@upv.es

Victoria Bérez, Toni Más, Mireia Rodríguez
Bitac S.L.
Barcelona, Spain.
{vberez, tonimm, mireiam}@bitac.com

Abstract

Clinical concepts and terminologies usually differ from one software system to another, making the exchange of data between them a very arduous and difficult task. It mainly involves finding the correspondences between the local identifiers of the concepts to be exchanged. In this paper, we describe a software tool called SAM (Automatic Mapping System), a tool for terminological normalization of laboratory tests to LOINC. SAM helps technicians in this task through simplification of the terminological mapping process. It uses approximate string matching techniques along with a thorough mining of previous mappings and LOINC parts.

1. Introduction

Standardized vocabularies are vital for ensuring the good use, communication and integration of clinical information

Spanish, and Catalan), typographical errors, abbreviations, synonyms, etc. forming an enriched corpus of alternatives names. Therefore, finding a map means to establish a binding between local and standardized terms, what bring us to a Record Linkage problem [5].

The candidate terms resulting from the search are ranked using a score function based on approximate string matching techniques and on mining previous mappings and LOINC parts. The technician has to choose one from this set of candidates or use the automatic mapping mechanism that SAM provides.

2. Materials and methods

Tecnología utilizada para la especificación declarativa de predicados de selección aproximada

Implementación y evaluación de la técnicas probabilísticas y comparación con técnicas fuzzy.	T3.1. Implementación de técnicas probabilísticas de búsqueda aproximada T3.2. Incorporación de reglas probabilísticas de búsqueda aproximada T3.3. Evaluación resultados técnicas probabilísticas T3.4. Comparación resultados técnicas fuzzy-probabilísticas	R3.1. Motor de búsqueda con pool de técnicas probabilísticas R3.2. Incorporación de evaluación de la técnicas probabilísticas según sistema terminológico R3.3. Informe de comparación resultados técnicas fuzzy con probabilísticas según sistema terminológico
--	--	--

Potencia del motor de búsqueda LOINC :

- Algoritmos de String matching
- Mapeos previos validados (BC) – Machine learning
- Recursos lexicos, diccionarios de sinónimos

Revisión Lotes

Lote	Fecha Creación	Terminología
Lote prueba_3	2017-03-22 10:00:22	LOINC

Finalizar Revisión

Filtrar por descripción:

Similitud mínima	Diferencia autoeval.	Autovalidar	Deshechar Autovalidar
<input type="text" value="1"/>	<input type="text" value="0.2"/>	<input type="button"/>	<input type="button"/>

 Progreso validación: /

#	Id técnica	R#	Componente	Unidad	Sistema	Tempo	Escala	Método	Área	#	Mejor Candidato	Revisado	Adendum
	BALFO1		NEISSERIA GONORRHOICAE AFIN en UI		Urina		cuantitativa	PLM (monoclon)	INMUNOLOGIA	1	21416-3 - Neisseria gonorrhoeae DNA:PT:In:Urine:PT:Urd:Probe amp tar	<input type="checkbox"/>	
	C3C		COMPLEMENTO C3 en Suero	mg/dL	Suero		cuantitativa	Turbidimetría	INMUNOLOGIA	1	4485-9 - Complement C3:MCnc:Ser:Plas:Pt:Qn:	<input type="checkbox"/>	
	C4C		COMPLEMENTO C4 en Suero	mg/dL	Suero		cuantitativa	Turbidimetría	INMUNOLOGIA	1	4498-2 - Complement C4:MCnc:Ser:Plas:Pt:Qn:	<input type="checkbox"/>	
	CARDG		CARDIOLIPINA [ACA] Ac IgG en Suero	GPL U/mL	Suero		cuantitativa	ELISA	INMUNOLOGIA	1	3181-5 - Cardiolipin Ab IgG:ACnc:Ser:Pt:Qn:IA	<input type="checkbox"/>	
	CARDM		CARDIOLIPINA [ACA] Ac IgM en Suero	MPL U/mL	Suero		cuantitativa	ELISA	INMUNOLOGIA	1	3182-3 - Cardiolipin Ab IgM:ACnc:Ser:Pt:Qn:IA	<input type="checkbox"/>	
	CMVG		CITOMEGALOVIRUS IgG AC	aUI/ml	Suero		cuantitativa	ELFA	INMUNOLOGIA	5	5124-3 - Cytomegalovirus Ab IgG:ACnc:Ser:Plas:Pt:Qn:IA	<input checked="" type="checkbox"/>	
	CMVM		CITOMEGALOVIRUS IgM AC				cuantitativa	ELFA	INMUNOLOGIA	5	24119-0 - Cytomegalovirus Ab IgM:Pt:Thr:Ser:Plas:Pt:Ord:IA	<input type="checkbox"/>	
	AUST		HEPATITIS B- Ag superficie		Suero		Cualitativa	ECLIA	INMUNOLOGIA	5	10900-9 - Hepatitis B virus surface Ab:Pt:Thr:Ser:Pt:Ord:IA	<input type="checkbox"/>	
	BALA13		HEPATITIS C ARN VIRAL en Plasma		Plasma EDTA congel		cuantitativa	PCR (hibridación)	INMUNOLOGIA	5	11259-9 - Hepatitis C virus RNA:Pt:Thr:Ser:Plas:Pt:Ord:Probe amp tar	<input type="checkbox"/>	
	CUANTI		HEPATITIS B Antic HBs - CUANTIFICAD	U/mL	Suero		cuantitativa	ECLIA	INMUNOLOGIA	5	63557-3 - Hepatitis B virus surface Ag:ACnc:Ser:Plas:Pt:Qn:IA	<input type="checkbox"/>	

Código manual	Incidencias	Observaciones
<input type="text"/>	<input type="text"/>	<input type="text"/>

 Buscar por:

Código	Componente	Unidad	Propiedad	Muestra	Tempo	Tipo	Método	Área	Bitac	Status	Rank	Order/obs	SM	Incidencia	Opciones
5124-3	Cytomegalovirus Ab IgG		ACnc	Ser/Plas	Pt	Qn	IA	MICRO	1005124	ACTIVE	673	Both	0,84		
7852-7	Cytomegalovirus Ab IgG		ACnc	Ser/Plas	Pt	Qn		MICRO	1007852	ACTIVE	0	Both	0,84		
5125-8	Cytomegalovirus Ab IgM		ACnc	Ser/Plas	Pt	Qn	IA	MICRO	1005125	ACTIVE	968	Both	0,76		
7853-5	Cytomegalovirus Ab IgM		ACnc	Ser/Plas	Pt	Qn		MICRO	1007853	ACTIVE	0	Both	0,76		
5124-3	Cytomegalovirus Ab IgG:ACnc:Ser:Plas:Pt:Qn:IA		ACnc	Ser/Plas	Pt	Qn	IA	MICRO	1005124	ACTIVE	673	Both	0,84		

Revisión Lotes

Lote	Fecha Creación	Terminología
Lote prueba_3	2017-03-22 10:00:22	LOINC

Finalizar Revisión

Filtrar por descripción:

Similaridad mínima Diferencia autoeval.

Código	Componente	Unidad	Propiedad	Muestra	Tiempo	Tipo	Método	Área	Bitac	Status	Rank	Order/obs	SIM	Incidencia	Opciones
1401	4011	Citomegalovirus, IgG	U/ml	Suero	Puntual	Qn	QLIA	MICROBIO							
320	B0955	Citomegalovirus IgG An.	AU/mL	SVSUER-Suero		Cuantitativo	435-Quimioluminiscencia	Bioquímica							
700	04784596190_SA	CMV IgG	U/mL	Suero	Pt	Qn	ECLIA	INMUNO							
700	04784596190_PA	CMV IgG	U/mL	Plasma	Pt	Qn	ECLIA	INMUNO							
901	5003570	Cytomegalovirus Ab.IgG	U/mL	ACnc	Ser/Plas	Pt	EIA	INMUNO							
1501	8259	Citomegalovirus anticuepos IgG ratio	ratio	suolo	Pt	Qn	Quimioluminiscencia	SEROLOGI							
1502	4600	CMV anticuerpos IgG	UA/mL	SUERO	Pt	Ord	Immunoensayo - Cuantitativo	SEROLOGI							
1602	9094	Citomegalovirus IgG		Suero				Servicio de							
1603	529	CMV IgG		Suero		N	ELFA (VIDAS)	Microbiolog							
1505	3020	Citomegalovirus IgG (CMEIA)		Suero	Pt	Qn	Quimioluminiscencia	SEROLOGI							
1601	7752	CMV IgG		SUERO, PLASMA		Cuantitativo	CLIA	Microbiolog							
1605	529	CMV IgG		Suero			ELFA (VIDAS)	Microbiolog							
1506	4305	Citomegalovirus IgG (CLIA)	AU/mL	Suero		P/D/N	CLIA	Microbiolog							
1604	1415	Ac CITOMEGALOVIRUS IgG		SUERO				MICROBIO							

Código manual Incidencias Observaciones

 Buscar por: Buscar en LOINC

Código	Componente	Unidad	Propiedad	Muestra	Tiempo	Tipo	Método	Área	Bitac	Status	Rank	Order/obs	SIM	Incidencia	Opciones
5124-3	Cytomegalovirus Ab.IgG		ACnc	Ser/Plas	Pt	Qn	IA	MICRO	1005124	ACTIVE	673	Both	0,84		
7852-7	Cytomegalovirus Ab.IgG		ACnc	Ser/Plas	Pt	Qn		MICRO	1007852	ACTIVE	0	Both	0,84		
5126-8	Cytomegalovirus Ab.IgM		ACnc	Ser/Plas	Pt	Qn	IA	MICRO	1005126	ACTIVE	968	Both	0,76		
7853-5	Cytomegalovirus Ab.IgM		ACnc	Ser/Plas	Pt	Qn		MICRO	1007853	ACTIVE	0	Both	0,76		
16716-3	Cytomegalovirus Ab.IgG?2nd specimen		ACnc	Ser/Plas	Pt	Qn	IA	MICRO	1016716	ACTIVE	0	Both	0,645		

Gestión de Sinónimos

Selecciona Terminología

Muestra Método Propiedad Unidad Tiempo Asociado Ver asignaciones usuarios

Seleccione Muestra

Muestra Disponible

'plasma citrato congelado
(calcul)
*
*^patient
-

Sinónimos

espermatozoides
fluido seminal
I seminal
I. espermatico
I.espermatico
lavado seminal

Filtrado:

Muestra Método Propiedad Unidad Tiempo Asociado Ver asignaciones usuarios

Seleccione Metodo

Metodo Disponible

""spun westergren""
#n/a
(destilacion/titulacion) afnor t90-015
(espectrometria de masas con fuente de plasma acoplado)
(mdrd) fomula basada en creatinina
(ms)-mlpa

Sinónimos

10-aglutinacion directa
aggl
agglutinazione
aglut
aglutinacio
aglutinacion

Filtrado:



The screenshot shows a software interface for medical diagnosis. On the left, under 'Diagnostico Principal', the text 'colecistitis aguda litiásica' is entered. Below it, 'Otros diagnósticos' lists 'D 1 Enfermedad reanl crónica estadio 3' and 'D 2 infección por SAMR'. A search dropdown menu is open, showing results like 'K80.00 - Cálculo de vesícula biliar con co...', 'K81.0 - Colecistitis aguda (0.9)', etc. A settings window titled 'Mejoras para CIE10 Diagnosticos' is overlaid on the right, containing various checkboxes for search and display options.

Mejoras para CIE10 Diagnosticos	
<input checked="" type="checkbox"/> Cadenas Excluyentes	Excluir Sinónimos
<input type="checkbox"/> Filtrado del corpus	<input type="checkbox"/> CIE10
<input checked="" type="checkbox"/> Sustitución Abreviaturas	<input type="checkbox"/> Corpus BITAC
<input type="checkbox"/> Normalización anatómicas	<input type="checkbox"/> CIE9
	<input type="checkbox"/> SNOMED Español termir
	<input type="checkbox"/> SNOMED Inglés termina
	<input type="checkbox"/> SNOMED Extensión Min
	<input type="checkbox"/> Corpus Bitac 2
	<input type="checkbox"/> Corpus Bitac 3
	<input checked="" type="checkbox"/> Snomed Español no terr

Corpus terminológico >600.000 sinónimos
Mejoras: Abreviaturas y epónimos médicos, filtrado de corpus por palabra clave, reglas de exclusión y negación, normalizaciones anatómicas.

ESTRUCTURA:

Alfanumérico

Todos constan siempre de 7 caracteres

Cada carácter del código ocupa una “posición”

Cada posición tiene un significado según la sección

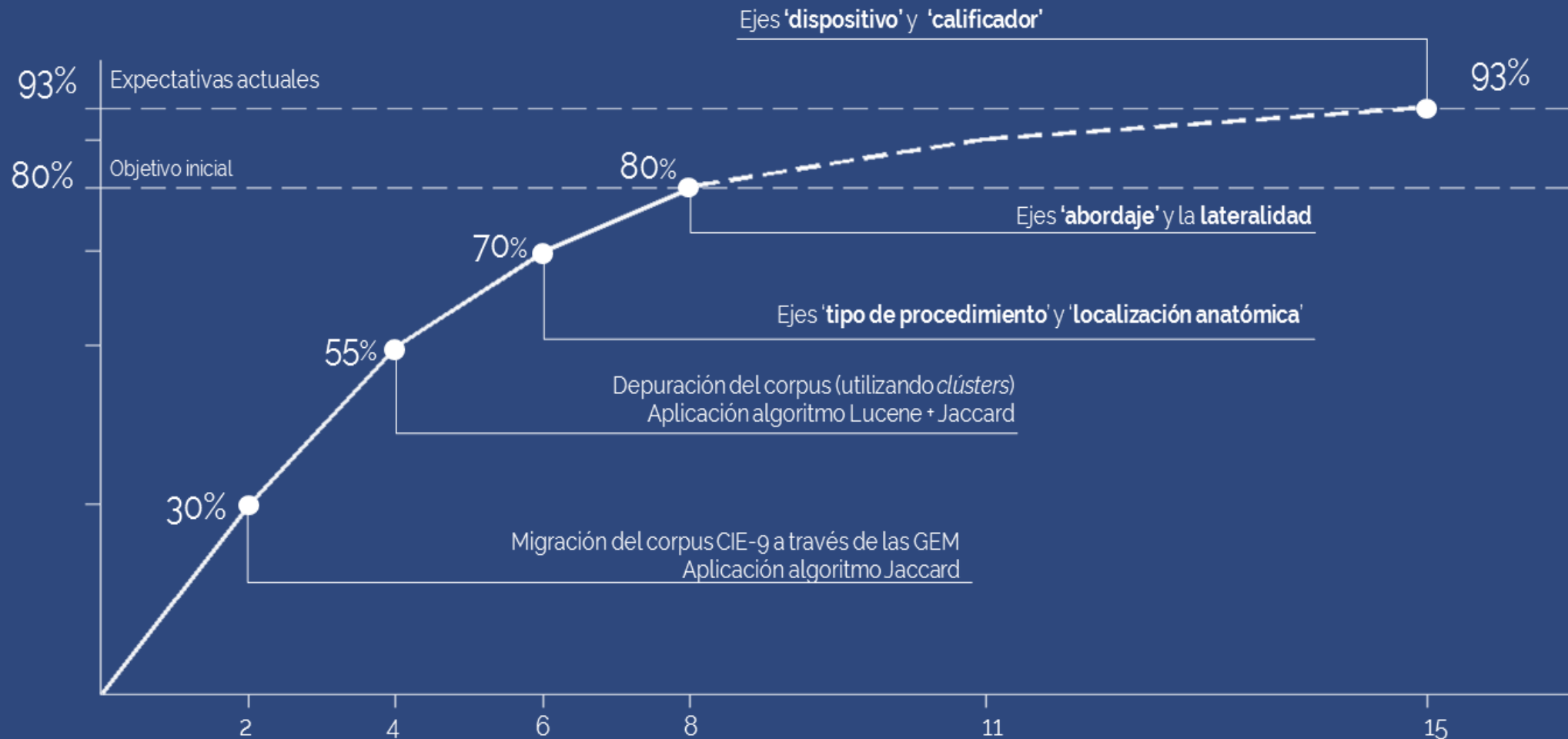
Existen 34 posibles valores: 0-9 y todas las letras excepto I y O



CÓDIGO

POSICIÓN 1	POSICIÓN 2	POSICIÓN 3	POSICIÓN 4	POSICIÓN 5	POSICIÓN 6	POSICIÓN 7
SECCIÓN	SISTEMA ORGÁNICO	TIPO PROCEDIM.	LOCALIZ. ANATÓMICA	ABORDAJE	DISPOSITIVO	CALIFICADOR
				CONTRASTE	SUSTANCIA	

Evolución de la eficacia del CT-MAP (CIE-10 Procedimientos)



Configuración: gestión de sinónimos y reglas

Gestión terminología | Revisión Lotes | Cerrar Sesión | Cambiar idioma

- Terminología
- Campos Terminológicos
- Sinónimos
- Técnicas CIE 10 Diagnosticos**
- Técnicas CIE-10 Proc

Alta | Editar | Borrar

ES

Cadena	Tipo	Filtros	Habilitado
fracaso consolidacion	Palabra completa	M, N, F, K, G, P, R, D, E, H, J, Q	true
fractura patologica	Palabra completa	Capítulo 13, Z87.311-Z87.311	true
gestacion	Palabra completa	Capítulo 16, Capítulo 15, Capítulo 21	true
glaucoma	Palabra completa	Capítulo 7, B73.02-B73.02, Q15.0-Q15.0, Z98.83-Z98.83, Z83.511-Z83.511	true
glomerul	Prefijo	Capítulo 14, M31.14-M32.14	true

Gestión terminología | Revisión Lotes | Cerrar Sesión | Cambiar idioma

- Terminología
- Campos Terminológicos
- Sinónimos
- Técnicas CIE 10 Diagnosticos
- Técnicas CIE-10 Proc**

Alta | Editar | Borrar

ES

Código	Traducción	Habilitado
endurant	dispositivo intraluminal	true
eo	dispositivo intraluminal	true
espiral oclusion terapeutica	dispositivo intraluminal	true
esteem	audifono	true
express	dispositivo intraluminal	true

Observables
atributos

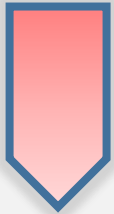
LOINC
From Regenstrief

85362-2 Mycobacterium tuberculosis complex DNA [Presence] in Sputum or Bronchial by NAA with probe detection

539-7 Mycobacterium sp identified in Sputum by Organism specific culture

COMPONENT: Mycobacterium
SYSTEM: Sputum, BAL...

agente causal → complejo Mycobacterium
sitio del hallazgo → estructura del pulmón
morfología asociada → inflamación granulomatosa
proceso patológico → proceso infeccioso



LIS

113861009 Mycobacterium tuberculosis

113858008 Mycobacterium tuberculosis complex

SNOMED
International
Leading healthcare
terminology worldwide



SCTID 154283005
tuberculosis pulmonar (trastorno)



HIS

ICD-10-CM A15.0
Tuberculosis del pulmón



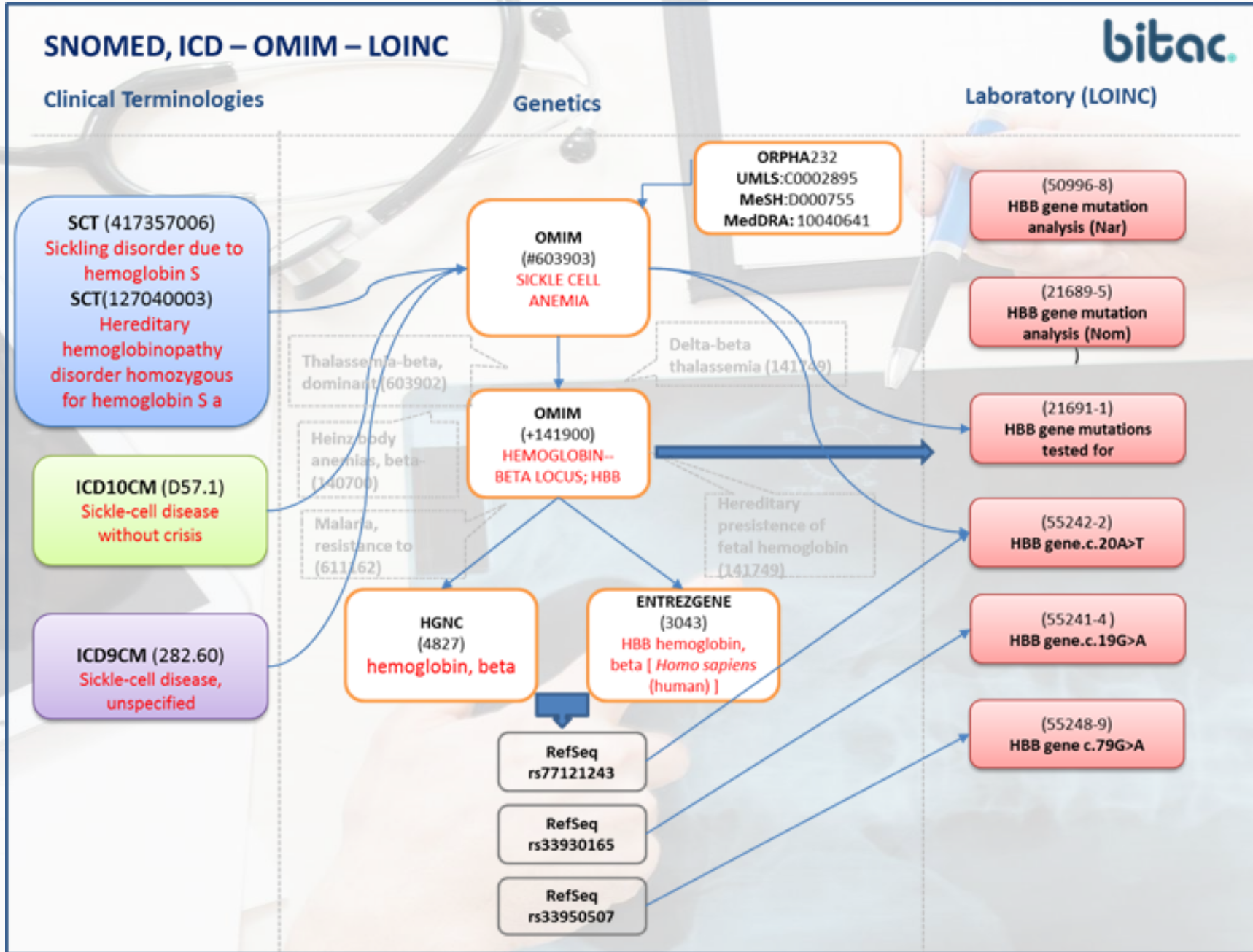
EDO

DATOS DE LABORATORIO
Fecha de diagnóstico de laboratorio: ___/___/___

Agente causal* (marcar una de las sigue:
 Mycobacterium africanum
 Mycobacterium bovis
 Mycobacterium caprae
 Mycobacterium tuberculosis
 Mycobacterium tuberculosis, complex
 Mycobacterium, otras especies del con
Muestra (marcar la muestra principal con
 Aspirado respiratorio: broncoaspirado
 Biopsia pulmonar: bronquial, transbro
 Biopsia sin especificar
 Biopsia linfática
 Espato
 Exudado nasofaríngeo

Resultados de Aislamiento: Positivo Negativo No realizado
Resultados de Visualización: Positivo Negativo No realizado
Otras pruebas con resultado positivo (marcar todas las pruebas con resultado positivo):
 Ácido Nucleico, detección
 Granulomas, presencia histológica
Resultados de VIH: Positivo Negativo No realizado
Envío de muestra al Laboratorio Nacional de Referencia (LNR): SI No
 Identificador de muestra del declarante al LNR: _____
 Identificador de muestra en el LNR: _____
Resultados de pruebas de susceptibilidad antimicrobiana: SI No

	Sensible	Intermedio	Resistente
Amikacina	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



SELECTING TERMINOLOGIES FOR OMICS HEALTH RECORDS

We selected the key vocabularies, ontologies, classifications we would like to study in a first line. The clinical terminologies with broader environment than genomics were out of the scope of this project.

Standard	Long Name	Owner/Organization	Type of Standard	Priority
HPO	Human Phenotype Ontology	Charité Berlin - Monarche Initiative	Phenotype-Genotype	1
OMIM	Online Mendelian Inheritance in Man	V. A. McKusick and colleagues Johns Hopkins Univ., NCBI	Phenotype-Genotype	1
DO	Disease Ontology	Northwestern University, Center for Genetic Medicine and the University of Maryland School of Medicine, Institute for Genome Sciences.	Phenotype-Genotype	1
GO	Gene Ontology	The Gene Ontology Consortium	Phenotype-Genotype	1
ClinVar	ClinVar	Intramural Research Program of the NIH, National Library of Medicine. NCBI	Clinical Association phenotypes_sequence variation	1
HGVS	Human Genome Variation Society	Human Genome Variation Society	Sequence/Genes	1
HGNC	Human Gene Nomenclature Committee	Human genome Organization (HUGO)	Sequence/Genes	1
dbSNP RefSNP	Database of single nucleotide polymorphisms	NCBI (National Center for Biotechnology Information)	Sequence/Genes	1
RefSeq	Database of Reference sequences	NCBI (National Center for Biotechnology Information)	Sequence/Genes	1
LRG	Locus Reference Genomic	European Bioinformatics Institute (EMBL-EBI).NCBI.	Sequence/Genes	1
ORPHANET	Portal for information on rare diseases and orphan drugs	INSERM ((French		2
MedGen	Medical Genetics	NCBI (National		2
NCI Thesaurus	The National Cancer Institute (NCI) Thesaurus	NCI Enterprise		2
CCDS	Consensus CDS protein set.	EBI, HGNC,		2
VEGA	The Vertebrate Genome Annotation	Wellcome T		2
COSMIC	Catalogue Of Somatic Mutations In Cancer	Wellcome T		2
Ensembl	genome browser for vertebrate genomes	European B		2
NCBI gene	NCBI genes identifiers	NCBI (National		2
RxNORM	Normalized names for clinical drugs	National Lib		3
LOINC	Logical Observation Identifiers Names and Codes	Regenstrief		3
SNOMED CT	Systematized Nomenclature of Medicine –Clinical Terms	The Internat (IHTSDO)		3
ICD10	International Classification of Diseases 10th Revision	World Heal		3
EFO	Experimental Factor Ontology	European B		3
ClinGen	Clinical Genome Resource	National Hu Program of		3
Uniprot	Universal Protein Resource	European B Bioinformat		3
UCSC	University of California Santa Cruz data base	University of Ca		3
SO	Sequence Ontology	Gene Ontology Consortium	Sequence/Genes	3

HIT Clinical Genomics Data Standards

Ensure transfer of data between systems ...

Health Level Seven –HL7

Ensure standard description of tests, results, and interpretations ...

LOINC, Logical Observation Identifiers Names and Codes
 HGVS Nomenclature, Human Genome Variation Society
 HGNC, Human Gene Nomenclature Committee
 RefSeq, Reference Sequences NCBI
 dbSNP, Single Nucleotide Polymorphism
 ISCN, International Society for Cytogenetics Nomenclature

Ensure standard context for interpretations (i.e. associations) ...

SNOMED & RxNORM

Other References

LRG, OMIM, COSMIC, PubMed...

Priority 1: first line study; Priority 2: second line study; Priority 3: out of the scope of this study.

Desde nuestra experiencia en sanidad, la **aplicación de tecnologías del lenguaje para la correcta codificación de textos biomédicos** debe ir acompañada de :

- Conocimiento previo de los estándares clínicos
- Base de conocimiento o base terminológica
- Bibliotecas de sinónimos, diccionarios y recursos léxicos de soporte al dominio clínico de aplicación.



bitac. clinical terminology
for e-health



INNOVATIVE SME

Valid until Dec 31st 2018

