

## La semàntica a salut

Per Albert Graupera i Ariadna Rius (OFSTI- CCI)

Un dels objectius que es persegueixen a diferents àmbits, entre ells l'estatal, és la compartició d'informació mèdica. Per assolir aquesta fita cal un treball intensiu d'estandardització en diferents fronts, amb la finalitat d'aconseguir l'anomenada interoperabilitat. Es treballa per tenir accés a la informació d'un pacient des de qualsevol servei de salut, de manera que les decisions clíniques puguin tenir en compte totes les dades disponibles d'aquella persona. Garantir la qualitat assistencial passa per assolir la independència de la informació a nivell departamental, territorial, lingüístic, dels proveïdors de salut i dels processos de treball.

En aquest escenari els estàndards són l'eina per assolir la interoperabilitat. La capacitat de compartició es desglossa en diferents nivells, existint molts tipus d'interoperabilitat (legal, tècnica, funcional, de procés, semàntica, etc.). En aquesta entrada es tindrà en compte el nivell **operatiu** (interoperabilitat tècnica i funcional) i el **semàntic**.

En un idioma es pot diferenciar la sintaxi de la semàntica. Per tal que la comunicació sigui efectiva, els dos elements s'han de transmetre amb èxit. No n'hi ha prou amb llegir un text, cal entendre què explica. Amb la interoperabilitat passa el mateix, el sistema d'informació necessita els dos nivells de comprensió. La interoperabilitat operativa garanteix l'intercanvi de missatges i s'assoleix amb estàndards que defineixen com han de ser, per exemple, els fitxers de transmissió de la informació: quina estructura, contingut o format han de tenir. La interoperabilitat semàntica fa referència a les dades clíniques dels arxius. Els estàndards d'aquest àmbit defineixen la codificació a utilitzar pel domini concret de l'intercanvi. Reprenent el cas de l'idioma, dues persones s'han d'expressar en una llengua que coneguin per entendre's i que la comunicació sigui efectiva.

### **Tipus de vocabulari**

Es diferencien dos tipus principals de vocabulari: el natural i el controlat. El primer llenguatge és el que utilitzem les persones per comunicar-nos entre nosaltres: per un sistema d'informació clínica, seria el que usen els metges quan parlen amb els pacients i amb d'altres professionals. El llenguatge natural sorgeix espontàniament i depèn del context en el qual s'utilitza. És molt ric però esdevé ambigu, sense definicions rigoroses, i sovint és especialitzat, presenta acrònims i paraules d'argot. Degut a aquestes característiques, no es pot utilitzar d'una manera previsible a un sistema, és un vocabulari per interlocutors humans. El llenguatge natural no és interoperable, ja que no es pot compartir d'una forma coherent, ni comparable, perquè no és possible establir equivalències entre diferents fonts d'informació. També cal destacar les mancances que presenta en quant a qualitat de les dades, en termes d'integritat, exactitud i precisió.

Per resoldre els problemes citats és necessari transformar el llenguatge natural a vocabulari controlat, codificant-lo. Amb aquest pas s'especifica i identifica el significat de la informació i s'estableixen regles semàntiques per interpretar-lo. A més de solucionar els inconvenients citats, el resultat de la conversió es pot usar com a base de sistemes experts o d'ajuda a la presa de decisions. Tenir les dades formalitzades també facilita la traducció a diferents idiomes, l'explotació per part de diferents departaments (p.e clínics i administratius) i permet considerar informació a nivell global (de més d'un centre assistencial, per exemple). Els estàndards semàntics mencionats anteriorment són vocabularis controlats i normalment estan acotats a un domini, pel qual s'han dissenyat específicament. N'hi ha de tres tipus, en funció del seu ús al sistema d'informació: d'interfase, de referència i de sortida.

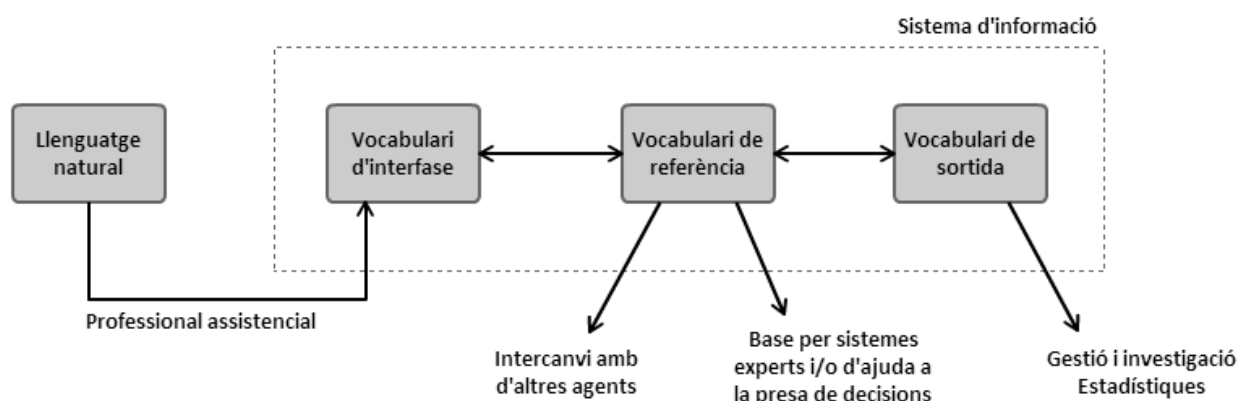
El vocabulari d'interfase és el llenguatge que utilitza el professional assistencial per entrar les dades al sistema d'informació. Aquest primer nivell de registre permet concretar i formalitzar les dades clíniques que rep el sistema. Sovint succeeix que les codificacions que utilitzen els centres no són estàndards, de manera que la informació registrada no es pot compartir ni comparar amb d'altres agents. És com si ens inventéssim un idioma propi, ningú ens entendria i no ens podríem comunicar, en aquella llengua, amb d'altres persones.

De la necessitat, relativament recent, de compartir informació mèdica, va sorgir la de comptar amb un vocabulari que permetés entendre les codificacions pròpies: el de referència. S'utilitza com a base per comparar dades i representa els termes registrats en llenguatge d'interfase amb el màxim nivell de detall. El vocabulari de referència té relacions entre els conceptes que el componen i amb els termes del

d'interfase. La interoperabilitat semàntica s'assoleix amb aquesta mena de llenguatge. Si ens volem comunicar amb una persona que parla xinès, caldrà que nosaltres en sapiguem o que tots dos utilitzem un tercer idioma. Imaginem que no sabem xinès però sí anglès, i que el nostre interlocutor també en sap: ens podrem comunicar. En aquest exemple, estaríem utilitzant l'anglès com a llenguatge de referència. En la informació mèdica pot ser que dos sistemes utilitzin codificacions diferents pels mateixos dominis de manera que, a priori, no es podrien comunicar. Si relacionem cadascun d'aquests vocabularis a un tercer (el de referència), no només entendrem el que es transmet sinó que, a més a més, ho podrem comparar, identificar i utilitzar. El vocabulari de referència també millora el registre i l'explotació d'informació, ja que aporta dades addicionals i un major grau de detall.

El darrer tipus de vocabulari és el de sortida, el qual permet analitzar la informació des de diferents perspectives (p.e clínica i financera). El llenguatge de sortida és el que s'utilitza en la gestió i en la investigació, sobretot amb finalitats estadístiques.

La imatge següent mostra els tipus de vocabulari explicats:



Aquests tipus de vocabulari no són excloents, sovint el d'interfase i el de sortida coincideixen i es pot donar el cas que el de referència i el d'interfase siguin el mateix, tot i que no és freqüent. El vocabulari d'interfase i el de sortida, solen tenir finalitats agrupadores que facilitin l'explotació econòmica i estadística de les dades. El llenguatge de referència, en canvi, està dissenyat, específicament, per representar informació mèdica, motiu pel qual és granular, fins arribar a un gran nivell de detall.

### **Classificació del vocabulari en funció del seu objectiu i estructura**

En el vocabulari controlat cada concepte té assignat un identificador únic (ID), que no es repeteix en cap altre cas, com un DNI per les persones. S'entén com a concepte cadascuna de les unitats de significat que té la codificació. L'assignació d'aquests IDs es realitza de manera diferent depenent del tipus de codificació que s'utilitza, pot ser seqüencial, mitjançant algorismes, per grups i subgrups, per zones geogràfiques, etc. o per una combinació d'aquestes, per exemple:

01 007 00241

- 2 dígits pel grup (01)
- 3 dígits pel subgrup (007)
- 5 dígits seqüencials (00241)

A l'identificador es pot assignar un dígit de control, el qual es genera a partir d'un algorisme fixat, i servirà per verificar que la cadena de dígits introduïda és correcta. Per exemple afegir a l'ID el primer decimal resultant de l'operació de sumar tots els nombres de l'identificador i dividir-lo entre 3:

01 007 00241 = 15      15/3=5,0

ID final: 01 007 00241 0

Els conceptes poden estar relacionats entre ells per tal de formar nous significats, el grau de complexitat d'aquestes relacions vindrà donat pel tipus de codificació que s'usi. El més senzill seran llistats sense relacions, on els termes poden estar agrupats a través dels identificadors. En els complexes, cada concepte estarà relacionat amb un o varis antecessors i/o predecessors (el que es coneix com estructura d'arbre), i permetran a l'usuari la unió de diferents termes per especificar o per la creació de nous significats.

Els principals tipus de codificació són:

### **Catàleg**

Llistat agrupat i enumerat de manera lògica dels conceptes que componen el catàleg. Cada dígit indica a quin grup pertany aquell terme. Descriurà vacunes, documents, objectes...

Per exemple, CatSalut utilitza un catàleg per les vacunes (immunitzacions), amb la següent estructura:

D;Diftèria  
T;Tètanus  
P;Tos ferina  
Polio;Poliomielitis  
FG;Febre groga  
(...)

### **Classificació**

Sistema de categories a les que s'assigna conceptes d'acord amb els criteris establerts. Cada concepte té un codi alfanumèric que facilita la seva explotació. En el cas de la CIM, el primer dígit indica a quin grup pertany cada concepte, i la resta l'identifiquen:

A33 Malalties bacterianes, Tètanus  
A36 Malalties bacterianes, Diftèria  
A37 Malalties bacterianes, Tos ferina  
(...)  
Z24.0 Necessitat immunitzacions bacteriana, Poliomièlitis  
Z24.3 Necessitat immunitzacions bacteriana, Febre groga  
(...)

### **Tesaurus**

Llista amb els termes utilitzats per representar conceptes, temes o continguts dels documents, intentant utilitzar una normalització terminològica per facilitar la comunicació entre usuaris i unitats d'informació. El contingut d'un tesaurus està relacionat mitjançant jerarquies, equivalències i/o relacions associatives (estructura d'arbre com és el cas de la codificació COSTART):

Malalties bacterianes  
    Tètanus  
    Diftèria  
    Tos ferina  
Necessitat immunitzacions bacteriana  
    Poliomièlitis  
    Febre groga

### Ontologia

Llistat de conceptes classificats i lligats entre si, per ser tractats i manipulats informàticament. Dit d'altre manera, permetre el canvi de dades entre programes i simplificar la unificació de diferents representacions. La ontologia es diferencia del tesaurus per la normalització dels conceptes, reutilització d'aquests, i aplicable fàcilment a una BBDD.

Un exemple d'ontologia és el MeSH:

Malalties bacterianes [C01.252]
Tètanus [C01.252.410]
Diftèria [C01.252.440]
Tos ferina [C01.252.400]
Necessitat immunitzacions bacteriana [N06.850.780.680]
Poliomielitis [N06.850.780.680.160]
Febre groga [N06.850.780.680.380]

### Terminologia clínica

Conjunt específic de coneixement presentats de manera sistemàtica i d'un mateix àmbit. Ve a ser com un diccionari especialitzat o un glossari, on cada concepte estarà relacionat amb altres conceptes i amb diferents descripcions. Un exemple n'és SNOMED CT:

138875005   SNOMED CT
87628006   enfermetat infecciosa bacteriana
397428000   Diftèria
76902006   tètanus
59994004   tos ferina
40733004   enfermetat infecciosa
398102009   poliomièlitis
16541001   febre groga

### Sistema d'addició de terminologia

Sistema amb diverses terminologies relacionades, amb la possibilitat d'afegir-ne de noves relacionant-les amb les existents en forma de mapeig. Com a exemple es pot trobar el UMLS (Unified Medical Language System):

C0000005 | ENG | X | L0000005 | X | S0007492 | X | A7755565 | | M0019694 | D012711 | MSH  
| PEN | D012711 | (131)I-Macroaggregated Albumin | 0 | N | |

En comparació amb la primera classificació del llenguatge, es solen utilitzar:

- Com a vocabulari **d'interface** i de sortida: classificacions, catàlegs com CIM-9-MC, CIM-10, ATC, NANDA, NIC, NOC, etc.
- Com a llenguatge de **referència**: terminologies clíniques com LOINC o SNOMED CT.

Els tesaurus, les ontologies i els sistemes d'addició de terminologies, es solen utilitzar com a base o suport als serveis terminològics, així com per completar la representació de dominis.