

Du langage naturel au langage structuré : application au dossier médical

Place de la SNOMED

Antoine Buemi



Table ronde

L'aide à la décision dans le SI

HIT 2010 - Paris – 21 mai 2010

Écueils à franchir

■ Textes en langage naturel

- Nombreuses ambiguïtés
 - polysémie, homographie
 - fréquent recours aux sigles, aux éponymes
- Situations contextuelles
 - pré-supposés, non-dits
- Variabilité interpersonnelle

■ Terminologies de référence

- Choix des termes préférentiels et des synonymes
- Validation des traductions
- Exhaustivité du domaine couvert

■ Outils d'indexation automatique

- Analyse syntaxique
- Auto-apprentissage



**La liberté du poète
est le désespoir
de la science**

Jean-Pierre Tétreault
Actualités, mai 1975

En pratique

- Le patient présente des cervicalgies dans un contexte de gastro-entérite avec des douleurs abdominales, avec une perte de connaissance très brève, sans mouvements anormaux, sans douleur thoracique et palpitation.

G-C006 dans
D5-41706 gastro-entérite
G-C008 avec
G-C008 avec
F-96700 connaissance
L-A1100 brève
G-C009 sans
G-C009 sans
F-37000 douleur thoracique

- Cervicalgie chez un patient affecté par une gastro-entérite avec douleur abdominale avec brève perte de connaissance sans mouvement anormal, sans douleur thoracique, sans palpitations.

F-16250 cervicalgie
G-0126 patient affecté
D5-41706 gastro-entérite
G-C008 avec
F-50860 douleur abdominale
G-C008 avec
F-A5574 brève perte de connaissance
G-C009 sans
A-80450 mouvement
G-A210 anormal
G-C009 sans
F-37000 douleur thoracique
G-C009 sans
F-37150 palpitations

Corps et queue du pancréas – exérèse : glucagonome malin bien différencié
Ganglion lymphatique – exérèse : métastase d'un glucagonome malin
Rate – exérèse : aspect normal

CT	3.5			
40133006	T-65200	corps du pancréas	pankreas gövdesi	胰体
73239005	T-65300	queue du pancréas	pankreas kuyruğu	胰尾
65801008	P1-03000	excision	eksizyon	切除术
66515009	M-81523	glucagonome malin	glukagonoma, malign	胰高血糖素瘤
41746004	G-F501	bien différencié	iyi diferansiye	高分化
59441001	T-C4000	ganglion lymphatique	lenf düğümü	淋巴结
65801008	P1-03000	excision	eksizyon	切除术
14799000	M-80006	tumeur métastatique	tümör, metastatik	转移瘤
66515009	M-81523	glucagonome malin	glukagonoma, malign	胰高血糖素瘤
78961009	T-C3000	rate	dalak	脾
65801008	P1-03000	excision	eksizyon	切除术
17621005	G-A460	normal	normal	正常

Structuration *a priori* : les formulaires

LUMIERE BRONCHIQUE

ASPECT

<input type="checkbox"/> ·Normal	T-26060		
<input type="checkbox"/> ·Remaniement positionnel	M-00000		
<input type="checkbox"/> ·Attraction des orifices bronchiques	T-26060	G-A460	
<input type="checkbox"/> ·Distorsion des orifices bronchiques	T-26060	M-01000	
<input type="checkbox"/> ·Coudure	T-26060	M-02595	
<input type="checkbox"/> ·Irrégulière	T-26060	M-02010	
<input type="checkbox"/> ·Excroissance cartilagineuse	T-26060	M-34140	
<input type="checkbox"/> ·Fistule broncho-médiastinale	T-26000	G-A402	
<input type="checkbox"/> ·Fistule ganglio-bronchique	T-26000	M-71420	
		T-D3300	M-39306
		T-C4000	M-39306

Couverture terminologique

■ Langage naturel

- 20 à 60 %

■ Langage naturel reformulé

- SNOMED 3.5 VF : 80 à 85 %
- SNOMED 3.5 VF + SNOMED CT : 90 à 95 %
- 5 à 10 % d'informations manquantes

Structure de maintenance
- nationale
- coordination internationale

**Convergence 3.5 – CT
+ traduction**

Avancées attendues

- **Traitement automatique du langage**
- **Analyse syntaxique**
- **Analyse contextuelle**
- **Algorithmes de structuration du langage**

- **Et surtout**

**amélioration de la qualité descriptive
du langage des médecins**

... le rêve !

