

Closing the gap between SNOMED CT and unstructured medical texts

**Reflections on practical implementations in
French-speaking hospitals in Belgium**

Céline Andris



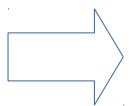
EarlyTracks
From Text To Knowledge



Introduction

- 50-80% unstructured data in electronic medical records
- Natural, flexible way of expression for practitioners

	Unstructured	Structured
Computer processable data	-	+
Intuitive data entry	+	-
Flexibility	+	-
Ergonomy	+	-
Time efficiency	+	+/-



Low end-user adoption of structured data entry systems, unstructured data is likely to remain



Introduction

1. Identifying medical information in free texts

- Methodology
- Challenges and solutions
- Results

2. Matching the medical information to concepts

- Methodology
- Challenges and solutions
- Results



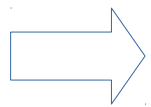
1. Identifying medical information in free texts

- Methodology
 - Information extraction : identify meaningful pieces of information in texts
 - Machine-learning model trained on manually annotated data
 - Selected categories : procedures, findings, disorders
 - Context :
 - Negation (“no meningitis”)
 - History (“history of melanoma”)
 - Uncertainty (“possible tuberculosis”)



1. Identifying medical information in free texts

- Challenges
 - Lack of standard structure in medical documents
 - Sections, headings, ...
 - Lack of standard grammar
 - Texts vs lists
 - Ungrammatical vs grammatical
 - ...



Solution : a flexible model capable to deal with different document structures and different types of textual data



1. Identifying medical information in free texts

Cher Confrère,

Votre patient, Monsieur X, âgé de xx ans, a été hospitalisé dans le service de Néphrologie du XX/XX/XXXX au XX/XX/XXXX.

Motif d'admission :

Diarrhées

Anamnèse :

Pas de douleur abdominale

Pas de fièvre.

Notion de dyspnée à l'effort sans notion de douleurs thoraciques

Antécédents médicaux :

- Insuffisance rénale chronique
- Hypertension artérielle
- Œsophagite de reflux
- Goutte
- Pneumopathies à répétitions

Antécédents chirurgicaux :

- Transplantation rénale en XXXX
- Néphrectomie gauche
- Parathyroïdectomie

Histoire clinique:

Patiente âgée de xx ans, hospitalisée pour une chute survenue le XX.XX.XX dans les escaliers avec réception sur les deux genoux. Cette chute lui a engendré une impotence fonctionnelle des deux membres ainsi qu'une douleur importante au niveau des deux genoux après le traumatisme. La patiente explique que ses deux rotules n'étaient plus en place mais qu'elles se seraient remises en place de manière spontanée lors de l'extension des deux genoux pour mettre ses jambes droites. La patiente a été vue le XX.XX.XX aux Urgences à X et l'examen clinique a démontré à gauche la présence d'un tiroir antérieur et d'une extension active impossible et à droite une instabilité du genou avec une laxité en valgus importante, un tiroir antérieur, un genu recurvatum et une extension active impossible. Des radiographies réalisées à xxxxxxxx démontraient l'absence de lésion osseuse post-traumatique mais on remarquait que sur les radios des genoux, aussi bien à gauche et à droite, il y a la présence d'une patella alta. Une échographie a été également réalisée à X démontrant au niveau du genou gauche, la présence d'une rupture du ligament rotulien à son insertion proximale, une rupture du ligament latéral interne et d'un arrachement de la capsule interne et au niveau du genou droit, la présence d'une rupture du ligament rotulien à son insertion tibiale, une rupture du ligament latéral interne et latéral externe. Un traitement a été instauré aux Urgences à X par la mise en place d'une attelle postérieure plâtrée cruro-pedieuse bilatérale.

1. Identifying medical information in free texts

- Results

	Precision	Recall
Procedures	80.0%	85.8%
Disorders	85.8%	89.3%
Findings	88%	71.5%

Precision = number of correct results / number of all returned results

Recall = number of correct results / number of results that should have been returned



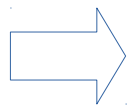
2. Matching the medical information to concepts

- Methodology
 - Link extracted medical information to concepts in a controlled vocabulary or standard terminology
 - SNOMED CT as a standard terminology to ensure consistency and interoperability
 - Development of resources and algorithms to support the matching process



2. Matching the medical information to concepts

- Challenges
 - No French edition of SNOMED CT



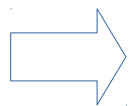
Development of our own resources in French



2. Matching the medical information to concepts

- Gap between terms extracted and standard descriptions
 - Misspellings, anglicisms, wrong use of accents or punctuation, use of symbols, shorthand notations, abbreviations, acronyms, abuse of terms,...

Syndrome de Kalman	Syndrome de Kallman
Dystrophie neuro-axonale	Dystrophie neuroaxonale
d+ abdo	Douleur abdominale
# humérus	Fracture de l'humérus
Allergie pdci	Allergie au produit de contraste iodé



Development of a customized dictionary



2. Matching the medical information to concepts

- Synonymy : different terms can refer to the same medical concept
 - SNOMED CT includes synonyms, but the variation in EPR is wider

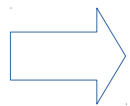
SNOMED CT concept	French description	Reality in EPR
65147003 Twin pregnancy	Grossesse gémellaire	Grossesse gémellaire Enceinte de jumeaux Enceinte de jumelles Grossesse double X attend des jumeaux X attend des jumelles

 Synonymy enrichment



2. Matching the medical information to concepts

- Polysemy : one term can refer to different medical concepts
 - Ambiguity of acronyms and abbreviations
 - Example : “IR” 723188008|Renal insufficiency
409623005|Respiratory insufficiency
 - Ambiguity of full terms
 - Example : “anorexie” 249468005|Anorexia symptom (finding)
56882008|Anorexia nervosa (disorder)

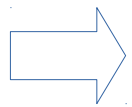


Proposed solutions: real time disambiguation



2. Matching the medical information to concepts

- No corresponding concept in SNOMED CT
Example : “BPCO Gold I”
- Too many possible corresponding concepts in SNOMED CT
Example : “tabagisme”



Proposed solution: concept-related templates



2. Matching the medical information to concepts

- Results

	Precision Perfect match	Precision Partial match	Recall
Procedures	81.9%	15.9%	90.4%
Disorders	86.9%	12.1%	91.6%
Findings	78.1%	19.7%	73.2%

Precision = number of correct results / number of all returned results

Recall = number of correct results / number of results that should have been returned



Conclusion

- It is possible to extract and structure medical information from free texts
 - Our results are quickly improving
 - The more input, the more accurate our customized resources
- Improvement of clinical documentation
 - Ambiguity is resolved in real time
- Coexistence between structured and unstructured data entry in EPR
 - Concept-related templates
 - Narratives remain



Céline Andris



celine@earlytracks.com

