

Grouping the Pharmacovigilance Terms with a Hybrid Approach

Marie Dupuch¹, Laëtitia Dupuch², Amandine Périnet³,
Thierry Hamon³, Natalia Grabar¹

¹CNRS UMR 8163 STL, Université Lille 3, France

²Université Toulouse III Paul Sabatier, France

³ LIM&BIO (EA3969) UFR SMBH, Université Paris 13, France

Plan

- Context
- Material and Methods
- Results and Discussion
- Limitations and Perspectives

Context

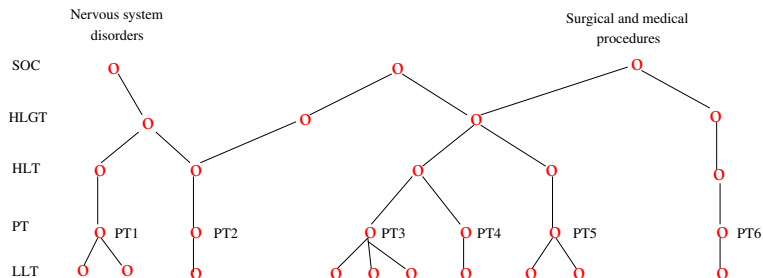
Pharmacovigilance

- Pharmacovigilance:
 - Collection, *analysis* and prevention of Adverse Drug Reactions (ADRs)
- Grouping ADRs reports:
 - Based on the MedDRA term coding
 - Medical Dictionary for Drug Regulatory Activities (Brown et al, 1999)
 - Detection of the safety alerts (signal detection)
 - Statistical methods (ROR, IC, CHI2, PRR, YuleQ, EBGm...)
 - ⇒ Groups of terms
 - Fine-grained terminology
 - *hepatitis infectious, hepatitis infectious mononucleosis, hepatitis viral...*

Context

MedDRA Terminology

- Coding Adverse Drug Reactions
- Signs and symptoms, diagnosis, medical and surgical procedures, etc.
- Hierarchically structured



Context

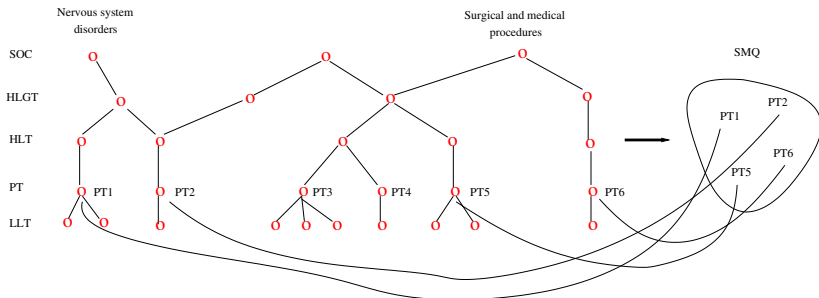
Objective: Group semantically similar terms

- MedDRA
 - HLT terms
 - SMQs (Standardized MedDRA Queries)
 - Manually built by experts (CIOMS, 2004)
 - MedDRA structure
 - Scientific literature
 - ⇒ Long and meticulous work
- Other resources (ontoEIM):
 - Hierarchical subsumption (Alecú et al,2008; Jaulent et al,2009)
 - Semantic distance (Bousquet et al, 2005; lavindrasana, 2006)
 - Subsets of MedDRA terms, no evaluation
- *Semantic methods to systematically group MedDRA terms*
 - Semantic similarity methods
 - Terminology structuring methods: synonymy, hyperonymy
 - Evaluation:
 - comparison with the SMQs
 - expert evaluation

Context

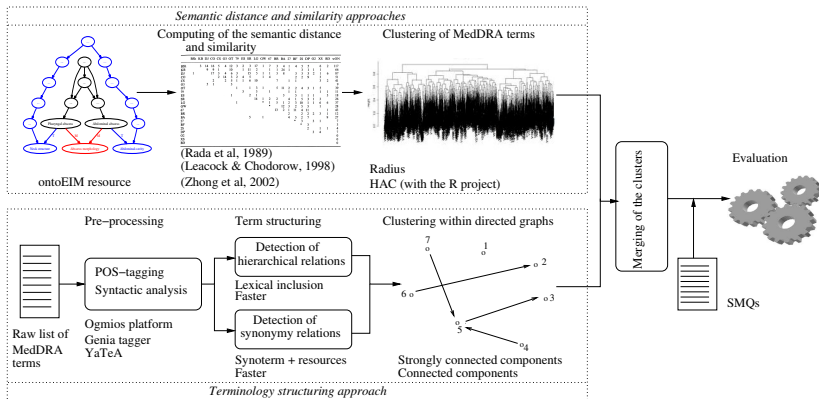
SMQs or Standardised MedDRA Queries

- Groups of MedDRA terms related to a diagnosis (84 SMQs)
 - *Acute renal failure, Hepatic disorders, Thrombocytopenia*

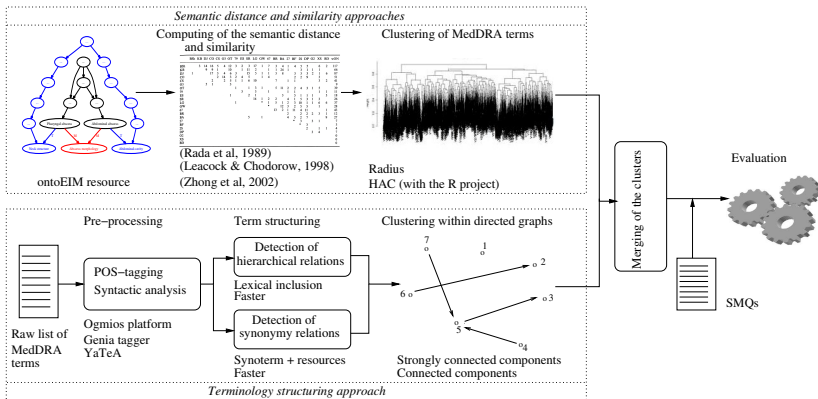


- PTs within SMQs \in different SOCs
- Number of SOCs per SMQ:
 - Varies between 4 to 25
 - The average is: 8.26 SOCs per SMQ

General schema of the method

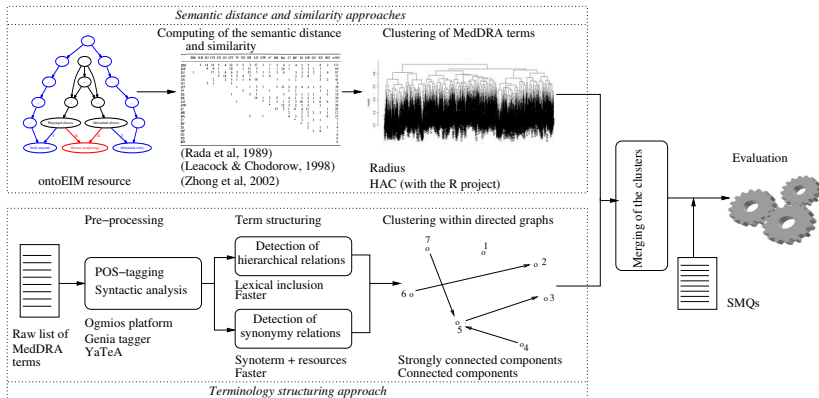


General schema of the method



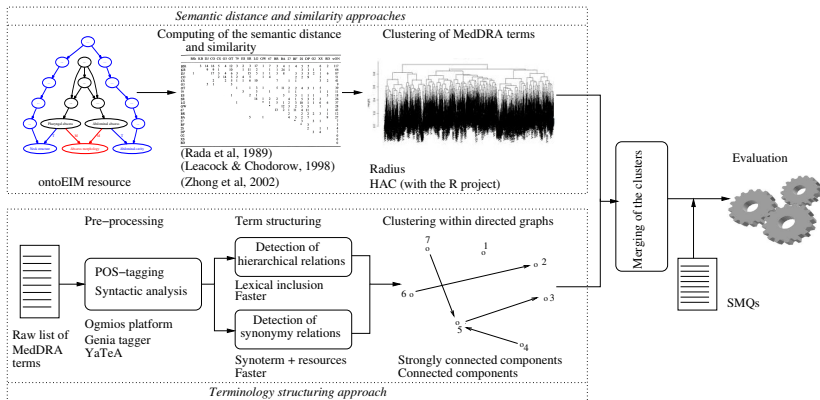
Semantic distance between two terms (4)
and between their formal definitions (10 and 0)

General schema of the method



$$dist_{ontoEIM}(A, B) = \frac{\sum_{i \in \{ADR, M, T\}} W_i * sp(A_i, B_i)}{\sum_{j \in \{ADR, M, T\}} W_j} = 3.5$$

General schema of the method

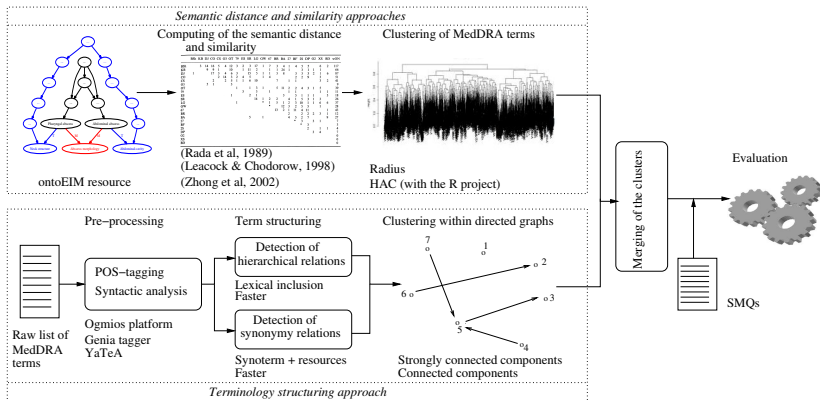


{*external ear lesion excision, lesion excision*}

{*malignant neoplasm of orbit, neoplasm of orbit*}

{*acute promyelocytic leukaemia, acute leukaemia*}

General schema of the method

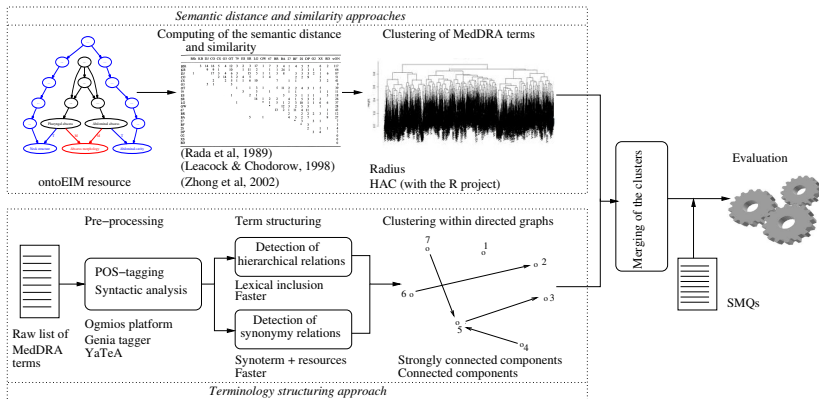


$\{\textit{sepsis, infection}\} \implies \{\textit{wound sepsis, wound infection}\}$

$\{\textit{arterial insufficiency, artery insufficiency}\}$

$\{\textit{eye penetration, penetrating eye}\}$

General schema of the method



non-disjoint clusters

Evaluation of the computed clusters

- Correctness of the acquired relations
- Quantitative evaluation (comparison with SMQs):
 - P (specificity)
 - number of relevant grouped terms as a percentage of the total number of the grouped terms
 - R (sensitivity)
 - number of relevant grouped terms as a percentage of the number of terms in the corresponding SMQ
 - F (f-measure)
 - the harmonic mean of P and R with $\beta = 1$
- Qualitative evaluation with an expert

Results

Terminology structuring methods

Methods and relationships	#relations
Hierarchical relations	
Lexical inclusions	3,366
Morpho-syntactic variants	743
Medical synonyms	
3 biomedical terminologies	1,879
UMLS/Filtered UMLS	190
Morpho-syntactic variants	100
Medical synonyms and WordNet	
3 biomedical terminologies	1,939
UMLS/Filtered UMLS	227

Results

Clustering of terms

Approach	SMQs		
	#clus	interval	mean
Semantic distance	2,931	[2; 546]	17
Structuring (hie)	748	[1; 117]	3.43
Structuring (hie+syn)	748	[1; 119]	3.82
Merging (hie)	2,998	[1; 563]	24.44
Merging (hie+syn)	2,998	[1; 594]	26.03

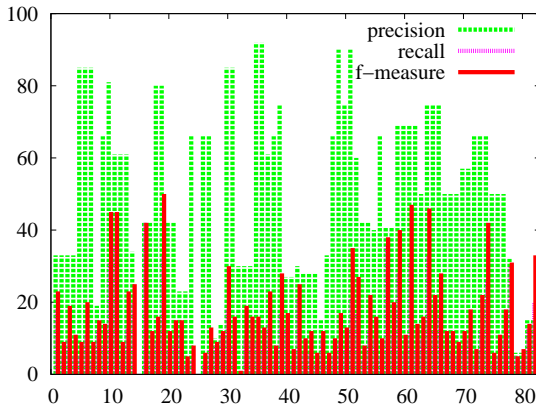
Results

Correctness of the acquired relations

- Manual analysis of the generated hierarchical relations
- 144 pairs (5%) with syntactic ambiguities:
 - *anticonvulsant drug level, anticonvulsant drug level*
 - *cranial nerve injury, cranial nerve injury*
 - *eye movement disorder, eye movement disorder*
- Semantic relations remain correct
 - the constraint involved through the syntactic analysis guarantees correct propositions

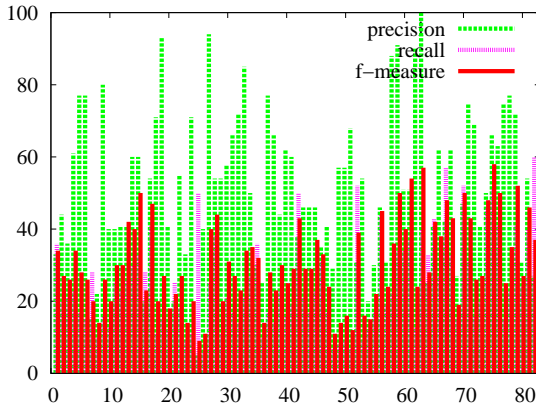
Results

Terminology structuring methods



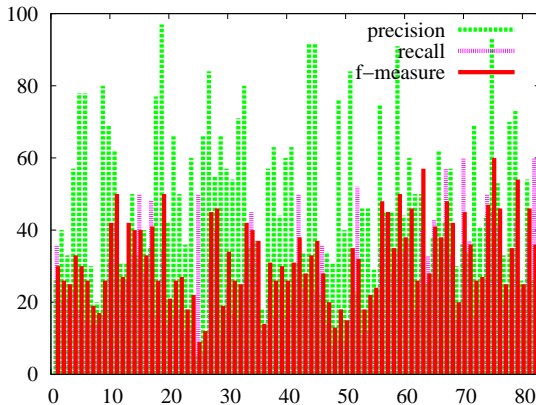
Results

Semantic similarity methods



Results

Merging of the two methods



Results

Analysis with an expert

SMQ *Embolic and thrombotic events, arterial*

- False positives:
 - *Cerebral arteriosclerosis, Cervical myelopathy, Intercostal neuralgia, Acute disseminated encephalomyelitis...*
- Too general terms:
 - *Ischaemia, Arterial disorder, Artery thrombosis, Shock, Vascular insufficiency...*
- Relevant terms missing in the SMQs:
 - *Varicose veins vulval, Cerebral ataxia, Hepatic artery stenosis, Renal artery stenosis...*

Results

Analysis with an expert

SMQs	Number of terms			Reference			After expertise		
	SMQ	clu	com	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
<i>Angioedema_{sd}</i>	52	32	13	40	25	30	43	26	33
<i>Angioedema_{st}</i>	52	31	19	61	36	45	61	36	45
<i>Angioedema_{mrg}</i>	52	33	21	63	42	50	71	48	57
<i>Embolitic and thrombotic..._{sd}</i>	132	159	48	30	36	32	32	39	35.2
<i>Embolitic and thrombotic..._{st}</i>	132	13	12	92	9	16	92	9	16
<i>Embolitic and thrombotic..._{mrg}</i>	132	130	49	38	37	37.5	47	46	46.5
<i>Haemodynamic oedema..._{sd}</i>	36	22	7	32	20	24	54	33	41
<i>Haemodynamic oedema..._{st}</i>	36	31	13	42	36	39	84	72	78
<i>Haemodynamic oedema..._{mrg}</i>	36	35	16	46	44	45	86	83	84.5

Discussion

- At least 2 hierarchical levels in the obtained groupings
 - Important contribution of lexical inclusions
- Hierarchical structure of MedDRA: very rough
 - ⇒ Intermediate hierarchical levels can be created
- Generation of non-disjoint clusters
- Quality of the gold standard
 - Important terms may be missing in the SMQs (Pearson 2009; Mozzicato 2007)
 - Evaluation by an expert is required
- Correctness vs. relevance of the semantic relations

Limitations and Perspectives

- Semantic similarity: 51% of the PTs only
 - Improvement of the alignment of the MedDRA terms (Nadkarni et al, 2011; Mougín et al, 2011)
- Two semantic methods exploited
 - Assumption: complementary contribution of different methods
 - Corpora for the detection of other semantic relations
 - (Resnik, 1999; Lin, 1998; Jian & Conrath, 1997)
- Only two types of semantic relations
 - Analysis of the types of semantic relations between the terms within the SMQs
- Exploitation of more sound clustering methods
 - Creation of hierarchies of clusters
- Variation of the results according to the SMQs
- Evaluation of this work for the creation of the SMQs
 - Systematic recruitment of the terms
 - Naming of the clusters
- Evaluation: SMQs vs. signal detection