

BiTeM / SIB Text Mining group Geneva, Switzerland

IHE World Summit
Amsterdam, June 7

The group

- **Patrick Ruch, Fund raising** 😊
- Arnaud Gaudinat, Web IT
- Julien Gobeill, Computer sciences
- Igor Mihlit, Information sciences
- Anais Mottaz, MD
- Luc Mottin, Bioinformatics
- Emilie Pasche, Bioinformatics
- Dina Vishnyakova, Computer sciences
- Douglas Teodoro, Computer sciences
- Romain Tanzer, Statistics

EU MD-Paedigree Project

NATURAL LANGUAGE PROCESSING IN A CLINICAL DIGITAL REPOSITORY

MD-PAEDIGREE

MD-PAEDIGREE

ICT-2011.5.2

VIRTUAL PHYSIOLOGICAL HUMAN PROJECT
Number 600932

THE CLINICALLY-LED VPH PROJECT

/ developing a model-driven data- and workflow-based digital repository

/ leveraging an information processing and knowledge discovery evolutionary framework

for personalised and predictive treatment in paediatrics.

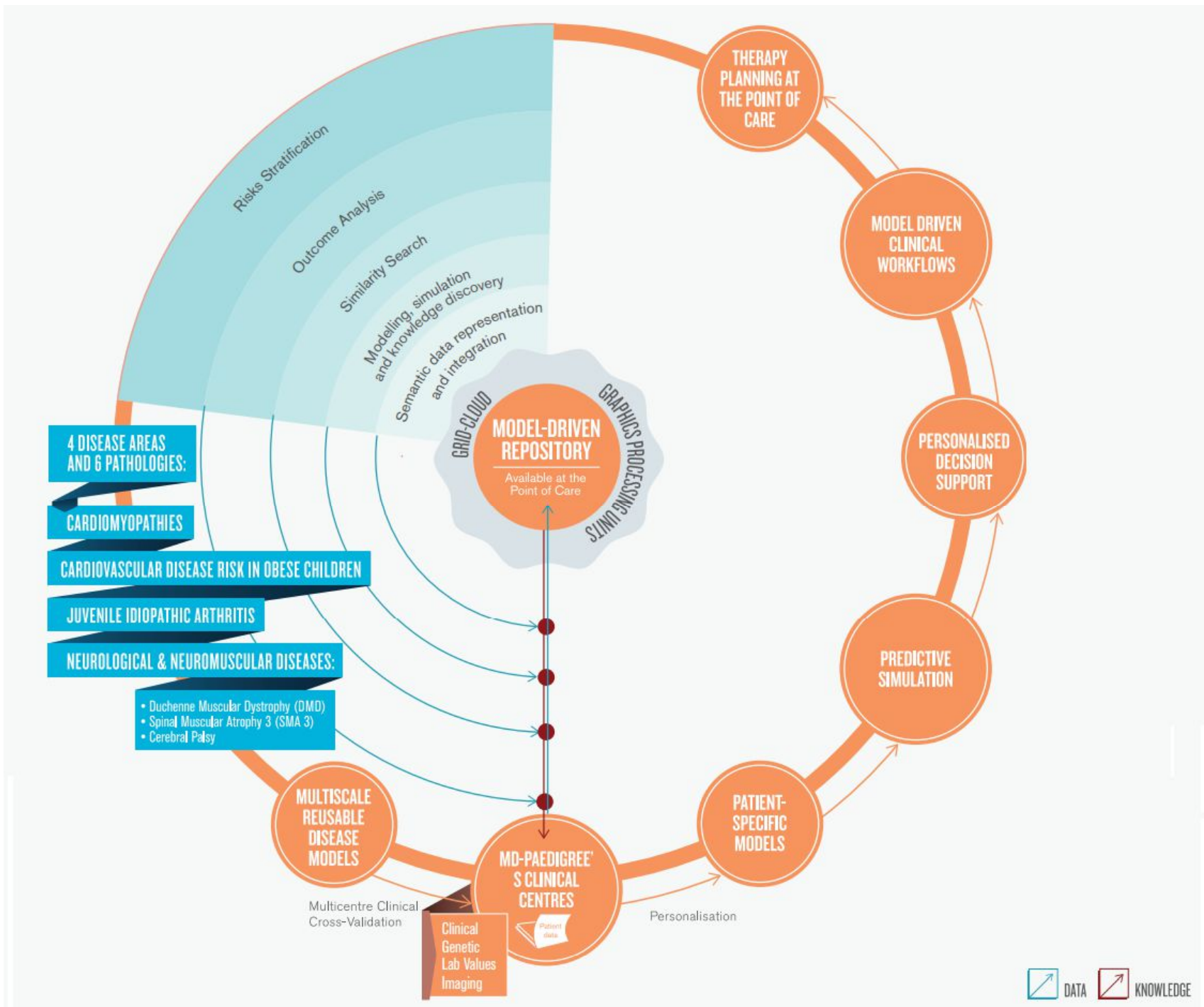
MD-Paedigree: a 48-months Integrated Project co-funded by the European Commission under the 7th Framework Programme.

TOTAL COSTS: € 16,432,513.00

EU CONTRIBUTION: € 11,869,000.00

Starting date: February 1st, 2013.

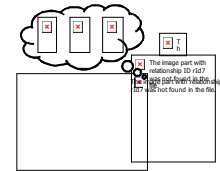




Why ?

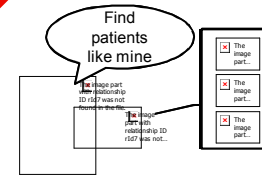
Physicians, who are facing complex diseases (e.g. pediatric cardiac diseases), show a great interest in finding cohorts of patients similar to their patients. Thus, they can observe the response of a particular treatment and learn about the outcomes at different points in time in a given clinical pathway. In this context, a case-based retrieval service has been developed.

1



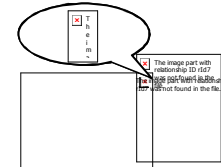
The physician has to choose a treatment for his patient.

2



The system retrieves similar patients and the physician can browse his patient file to see treatments and outcomes.

3



The physician selects a treatment with the help of the collected information.

Evaluation

n

38 episodes of care of existing patients have been randomly selected. For each case, two medical experts evaluated the top-10 answers. For 8 queries, no similar case was found among the top-10. Such queries might artificially decrease the precision of the search engine because some situation might simply be unique.

All queries (38)

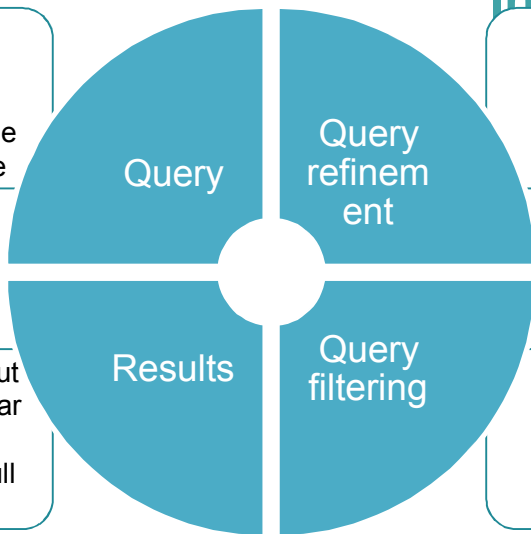
P0	0.5
P5	0.44
P10	0.42

Queries with at least one relevant case (30)

P0	0.63
P5	0.55
P10	0.54

User Interface

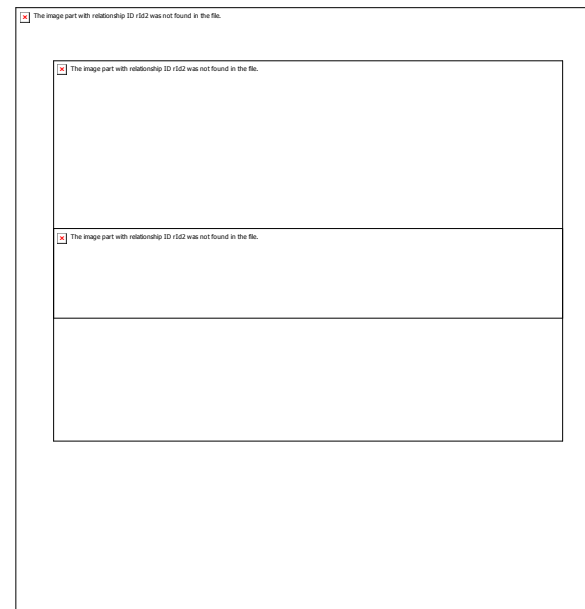
By patient identifier
By free-text description of the episode of care



Information about the ranked similar cases
Access to the full patient history

MeSH
Rocchio
...

Demographics:
age
gender
...



Interface of case based retrieval

New query

1 Query 2 Refinement 3 Filters 4 Results

MY PATIENT

Clinical synthesis: Insufficienza mitralica di grado moderato con apici dei lembi valvolari ispessiti e piccolo prolasso del lembo posteriore in regione. A2-P2. Tracce di rigurgito aortico, di nessuna rilevanza. Per il resto esame nei limiti: Normali i ritorni venosi sistemici e polmonari. Normale anatomia intracardiaca. Normale flussimetria delle restanti valvole. Normali dimensioni endocavitarie. Normale la cinesi biventricolare. Non dotto arterioso pervio. Arco aortico normale.

Age: 10 years
Gender: NA

2 QUERY REFINEMENT

MeSH terms

<input checked="" type="checkbox"/>	D001022	Aortic Valve Insufficiency	<input type="checkbox"/>
<input checked="" type="checkbox"/>	D004374	Ductus Arteriosus, Patent	<input type="checkbox"/>
<input checked="" type="checkbox"/>	D012212	Rheology	<input type="checkbox"/>
<input type="checkbox"/>	D045268	Flowmeters	<input type="checkbox"/>
<input type="checkbox"/>	D008944	Mitral Valve Insufficiency	<input type="checkbox"/>
<input type="checkbox"/>	D004373	Ductus Arteriosus	<input type="checkbox"/>
<input type="checkbox"/>	D000715	Anatomy	<input type="checkbox"/>
<input type="checkbox"/>	D054814	Anatomists	<input type="checkbox"/>
<input type="checkbox"/>	D011391	Prolapse	<input type="checkbox"/>
<input type="checkbox"/>	D007698	Kinesis	<input type="checkbox"/>
<input type="checkbox"/>	D001167	Arteritis	<input type="checkbox"/>
<input type="checkbox"/>	D016011	Normal Distribution	<input type="checkbox"/>
<input type="checkbox"/>	D010920	Placenta	<input type="checkbox"/>
<input type="checkbox"/>	D001021	Aortic Valve	<input type="checkbox"/>
<input type="checkbox"/>	D013524	Surgical Flaps	<input type="checkbox"/>
<input type="checkbox"/>	D001023	Aortic Valve Prolapse	<input type="checkbox"/>
<input type="checkbox"/>	D008943	Mitral Valve	<input type="checkbox"/>
<input type="checkbox"/>	D016292	Conscious Sedation	<input type="checkbox"/>
<input type="checkbox"/>	D055422	Venous Valves	<input type="checkbox"/>

Rocchio refinement

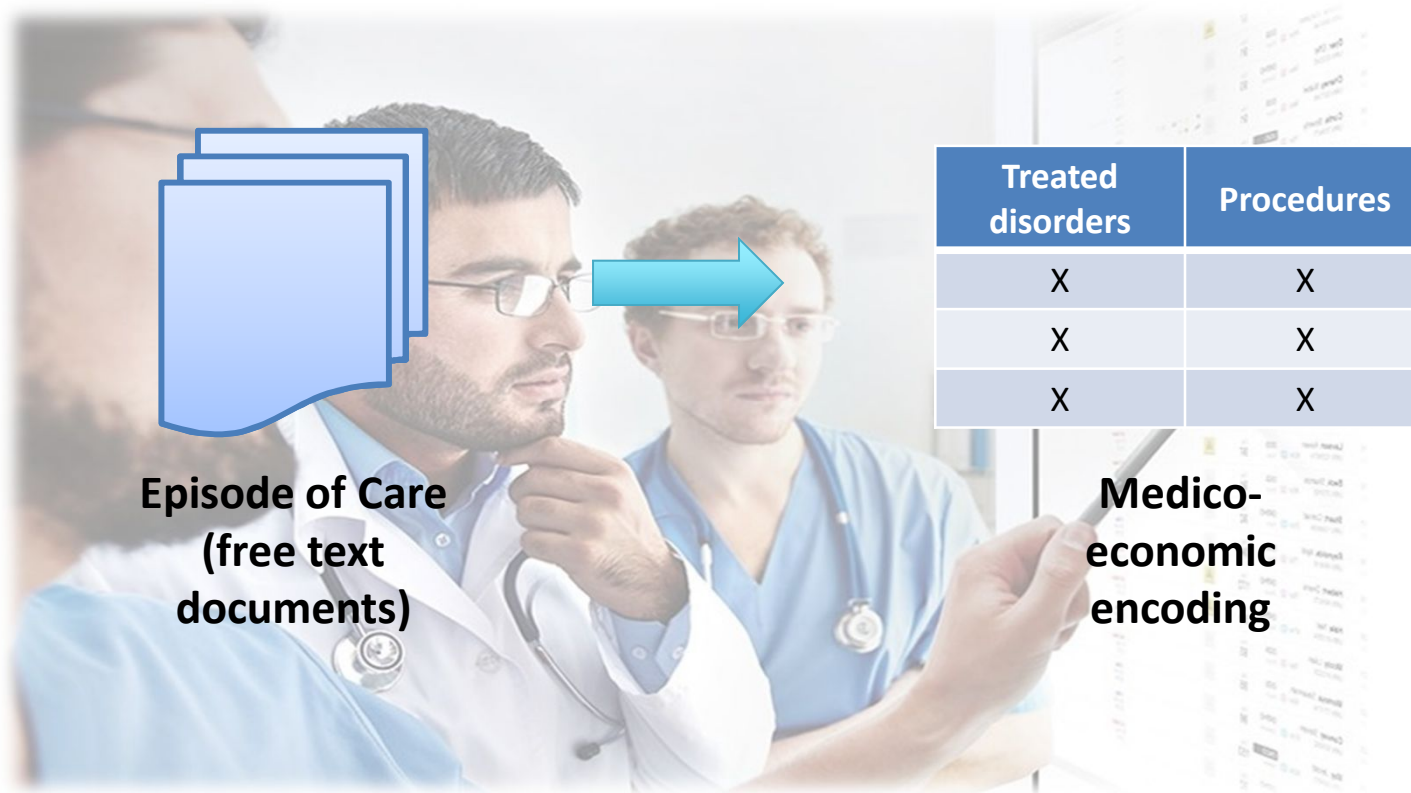
<input checked="" type="checkbox"/>	lieve
<input checked="" type="checkbox"/>	atrioventricolari
<input checked="" type="checkbox"/>	principale
<input checked="" type="checkbox"/>	nessun
<input checked="" type="checkbox"/>	semilunari
<input type="checkbox"/>	emodinamico
<input type="checkbox"/>	stimata
<input type="checkbox"/>	assente
<input type="checkbox"/>	escluso
<input type="checkbox"/>	adeguatamente
<input type="checkbox"/>	versamento
<input type="checkbox"/>	rilievo
<input type="checkbox"/>	anteriore
<input type="checkbox"/>	lieve-moderato/moderato
<input type="checkbox"/>	complessivamente
<input type="checkbox"/>	circolo
<input type="checkbox"/>	jets
<input type="checkbox"/>	significato
<input type="checkbox"/>	it

Previous Next

ODIN Project

NATURAL LANGUAGE PROCESSING IN EPISODES OF CARE

Task



ICD codes (treated disorders)

	A	B
1	code	label
2	I10.90	Hypertension essentielle, non précisée: Sans mention de crise hypertensive
3	Z37.0	Naissance unique, enfant vivant
4	O09.6	Durée de la grossesse: 37e semaine à 41 semaines révolues
5	X59.9	Autre accident et accident non précisé
6	Z38.0	Enfant unique, né à l'hôpital
7	E11.90	Diabète sucré, type 2: Sans complication: Non désigné comme décompensé
8	E78.0	Hypercholestérolémie essentielle
9	N19	Insuffisance rénale, sans précision
10	E55.9	Carence en vitamine D, sans précision
11	Y84.9	Incidents dus à des mesures médicales, sans précision
12	I67.10	Anévrisme cérébral (acquis)
13	A41.9	Septicémie, sans précision
14	R41.8	Symptômes et signes relatifs aux fonctions cognitives et à la conscience, autres et non précisés
15	J96.00	Insuffisance respiratoire aiguë: type I [hypoxique]
16	Y57.9	Complications dues à des médicaments ou à des drogues
17	R65.1	Syndrome de réponse inflammatoire systémique [SIRS] d'origine infectieuse avec complications org
18	E86	Hypovolémie
19	J18.1	Pneumonie lobaire, sans précision

CHOP codes (procedures)

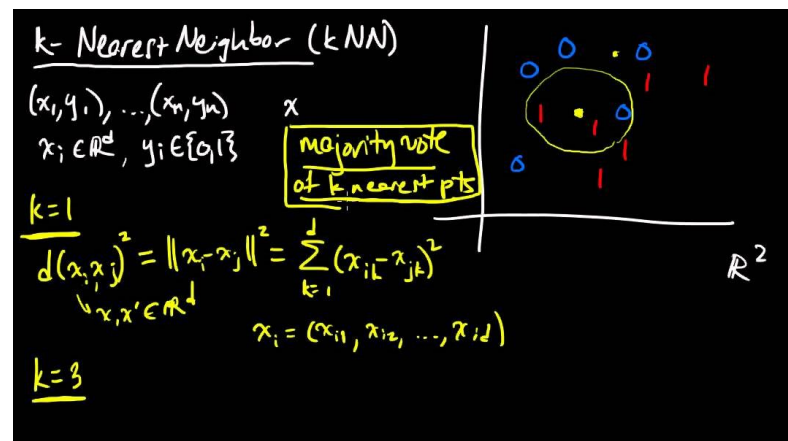
	A	B
1	code	label
2	03.91	Injection d'anesthésique dans le canal rachidien pour analgésie
3	73.59	Autre assistance manuelle de l'accouchement
4	54.91	Ponction abdominale percutanée (drainage)
5	99.04.10	Transfusion de globules rouges concentrés, 1 à 5 UT
6	99.BA.11	Traitement neurologique complexe de l'AVC aigu en stroke center (SC), de 24 heures au
7	73.4	Induction médicamenteuse du travail
8	75.69	Réparation d'autre déchirure obstétricale récente
9	96.6	Perfusion entérale de substances nutritives concentrées
10	89.14.00	Électroencéphalographie, SAP
11	93.22.99	Rééducation à la marche, autre
12	00.99.10	Réopération
13	93.83	Ergothérapie
14	99.15	Perfusion parentérale de substances nutritives concentrées
15	99.B7.11	Traitement complexe de médecine intensive, 60 à 119 points de dépense
16	00.9A.14	Agrafage linéaire par laparoscopie ou thoracoscopie
17	00.35	Chirurgie assistée par ordinateur avec sources de données multiples
18	88.91.10	Imagerie par résonance magnétique (IRM), cerveau et tronc cérébral
19	39.95.B2	Hémodiafiltration intermittente, avec anticoagulation par d'autres substances
20	93.96.99	Autre oxygénation, autre

Building cohorts

- Not treated disorders are not encoded
 - «Renal failures due to hypertension» ?
 - «Renal failures without hypertension» ?
- Some information only in free text
- Let data talk. Let data work.

Machine Learning

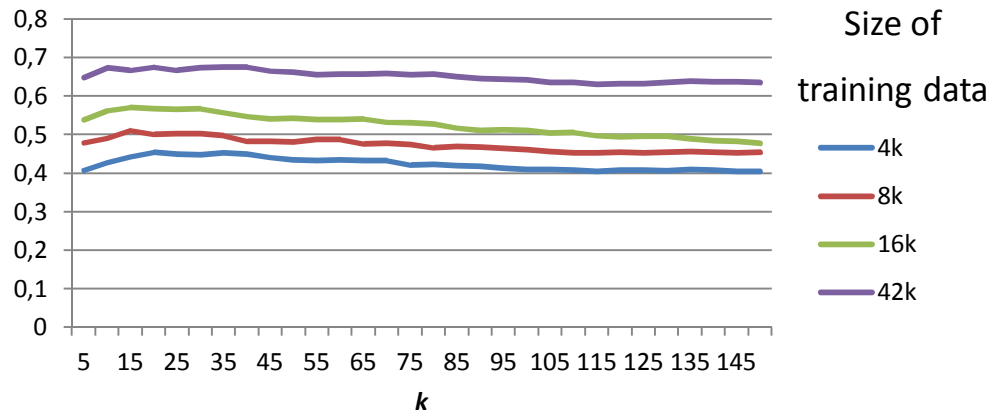
- 40,000 episodes of care encoded each year
- K-Nearest Neighbors
 - The most lexically similar documents are more likely to share the same disorders



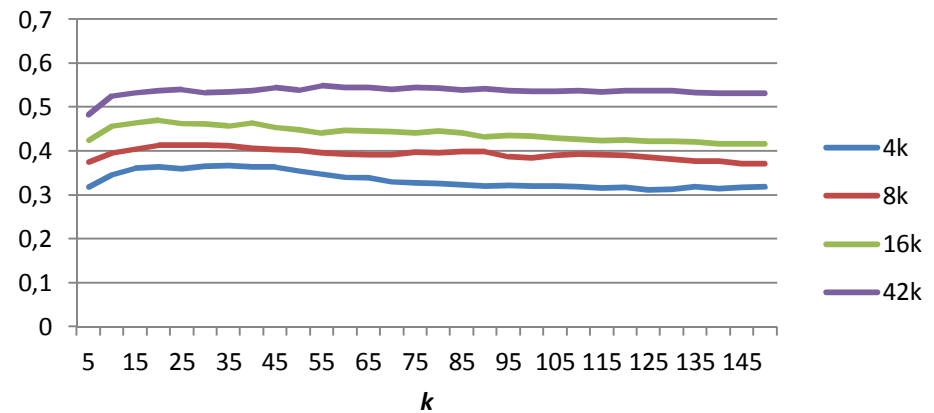
Some results

- The more data you have, the better it is

CIM P0



CIM R10



Semantic enrichment

- Mapping concepts thanks to multilingual terminologies (UMLS)

AVC hémisphérique
gauche dans le
contexte d'une
hypertension
artérielle et
fibrillation atriale
non anticoagulée.
....



Concepts MeSH

D020521 Accident vasculaire cérébral

D006973 Hypertension artérielle

D001281 Fibrillation auriculaire

One document

**Detected MeSH
concepts**

Detected disorders (on sample)

	A	B	C	D
1	Disorders			
2				
3	MeSH id	MeSH label	DF (sur 262)	Pourcentage
4	D004194	Maladie	172	65.6%
5	D004630	Urgences	143	54.6%
6	D010146	Douleur	136	51.9%
7	D013577	Syndrome	96	36.6%
8	D006973	Hypertension artérielle	88	33.6%
9	D006967	Hypersensibilité	85	32.4%
10	D006470	Hémorragie	80	30.5%
11	D003371	Toux	76	29.0%
12	D004417	Dyspnée	70	26.7%
13	D014947	Plaies et blessures	69	26.3%
14	D006547	Hernie	66	25.2%
15	D007239	Infection	65	24.8%
16	D009369	Tumeurs	62	23.7%
17	D003920	Diabète	55	21.0%
18	D005715	Jeu de hasard	51	19.5%
19	D007964	Hyperleucocytose	50	19.1%
20	D012008	Récidive	49	18.7%
21	D003076	Coït interrompu	49	18.7%
22	D000013	Malformations	45	17.2%
23	D010291	Parésie	45	17.2%

...

59	D011655	Embolie pulmonaire	23	8.8%
60	D013742	Tétanos	23	8.8%
61	D008019	Style de vie	23	8.8%
62	D006333	Défaillance cardiaque	23	8.8%
63	D006417	Hématurie	23	8.8%
64	D014549	Incontinence urinaire	23	8.8%
65	D007238	Infarctus	22	8.4%
66	D011014	Pneumopathie infectieuse	22	8.4%
67	D001281	Fibrillation auriculaire	22	8.4%
68	D000783	Anévrisme	22	8.4%
69	D001022	Insuffisance aortique	21	8.0%
70	D001284	Atrophie	21	8.0%
71	D001847	Maladies osseuses	21	8.0%
72	D017202	Ischémie myocardique	21	8.0%
73	D018805	Sepsie	21	8.0%
74	D001996	Bronchopneumonie	20	7.6%
75	D003556	Cystite	20	7.6%
76	D003967	Diarrhée	20	7.6%
77	D007746	Présentation foetale	20	7.6%
78	D013610	Tachycardie	20	7.6%

Semantic groups

Groupe sémantique	# moyen par doc	# types sémantiques	# concepts	# synonymes
Procedures	5.0	7	2262	9289
Disorders	4.4	12	4820	21258
Anatomy	4.1	11	1618	6127
Concepts & Ideas	2.5	12	1003	3934
Physiology	2.0	9	1065	3622
Chemicals & Drugs	1.7	26	9046	37984
Occupations	1.7	2	427	996
Living Beings	1.3	23	3914	13092
Activities & Behaviors	1.0	9	547	1926
Objects	0.6	5	456	1654
Phenomena	0.6	6	405	1392
Devices	0.3	3	256	1306
Geographic Areas	0.2	1	403	981
Organizations	0.0	4	190	1157
Genes & Molecular Sequences	0.0	5	169	1040

Evaluation

- Lack of benchmarks for evaluation...
- Difficult to provide a non-evaluated tool.