

R.I. et ontologies en médecine

Retour sur 5 expériences

2^e journée RI & IA
MSH Paris Nord – 12 décembre 2017

Jean Charlet^{1,2}

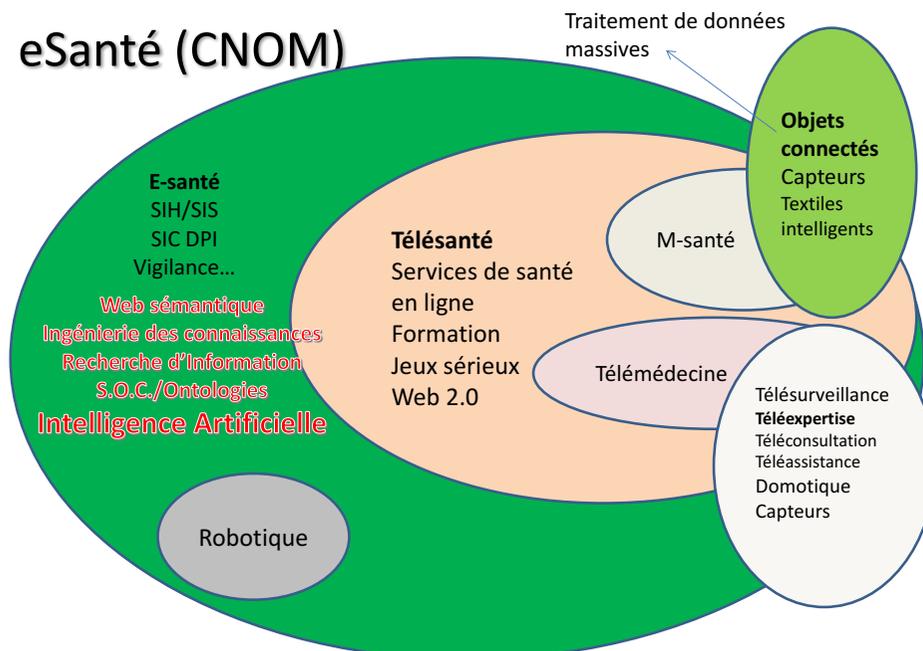
¹AP-HP/DRCI, Paris, France
²INSERM U1142/LIMICS, Paris, France

Plan

- eSanté
 - Ontologies
 - Des problèmes de partage de données et de connaissances
- Et des focus sur :
 - Menelas
 - PERTOMed
 - LERUDI
 - COVALMO
 - PARON
 - Watson
- Discussion



eSanté (CNOM)



Nota

- Il y a eu d'autres projets...
- ... d'ingénierie documentaire, de dossier patient textuel (XML/CDA), d'intégration de données, d'aide à la décision, de gestion des connaissances, de pharmaco-vigilance, de développement de plateforme (DAFOE)...



Menelas



Nom : Menelas
 Années : 1992-1995
 Type : Européen
 Personnes du labo : PZ, JB, BB, JC
 Objectif : reconnaissance du contenu du CRH
 Domaine : chirurgie cardiaque
 Quantité de raisonnement : ++++
 Résultat : ontologie + preuve de concept

Classes Menelas

artery > coronary_artery.
 coronary_artery > bisecting_artery.
 coronary_artery > circumflex_coronary_artery.
 coronary_artery > diagonalis_artery.
 coronary_artery > interventricular_artery.
 interventricular_artery > interventricular_anterior_artery.
 interventricular_artery > interventricular_posterior_artery.

Relations Menelas

```
rel >> DEFxPhyFctObjxPhyFctObj.  
  DEFxPhyFctObjxPhyFctObj >> sub_functional_object.  
    sub_functional_object >> induced_morphe.  
    sub_functional_object >> loc.  
      loc >> loc_above.  
      loc >> loc_after.  
      loc >> loc_at.  
      loc >> loc_before.  
rel >> linguistic_rel.  
  linguistic_rel >> ainsi_que_f.  
  linguistic_rel >> apres_f.  
  linguistic_rel >> assoc_f.
```

Domaine et portée dans Menelas

```
Positive_catalog_R_AB agt is  
  [intentional_change]--(agt)->[human_being].  
Positive_catalog_R_AB begin_at is  
  [temporal_interval]--(begin_at)->[date_time_stamp].
```

Connaissances encyclopédiques Menelas

```
Model admission(_x) is [admission: _x]-  
(pat)-->[human_being:_pat]  
(agt)-->[human_being:_doc]-  
  (state_of)--[knowledge_state:_ks]  
  (state_of)--[want_state:_ws] %  
(motive)-->[knowledge_state:_ks]  
  -->(content)-->[out_hospital]--(state_of)-->[human_being:_pat]  
(motivated_by)-->[state_of_physical_object:_path]--(state_of)-->[human_being:_pat]  
(result)-->[in_hospital:_ih]  
(performative_goal)-->[intentional_change]--(purported_obj)-->[human_being:_pat]  
(descriptive_goal)-->[state_of_physical_object:_ps]-  
  (intentional)  
  (state_of)-->[human_being:_pat]--(defines_area)-->[spatial_object:_sp3]  
  --(loc)-->[spatial_object:_sp3]--(defines_area)--[ward:_w]  
  (consists_loc)-->[spatial_object:_sp3]%  
(reason)-->[want_state:_ws]  
  -->(content)-->[in_hospital:_ih]--(state_of)-->[human_being:_p]  
(descriptive_result)-->[state_of_physical_object:_t]--(real)  
(attr)-->[admission_performance_attr]--(val_equal)-->[admission_performance_val] %
```

Menelas

- 1^{re} étapes de TAL avec IBM
- Top ontology + ontologie de domaine (développées par un informaticien philosophe et un médecin à partir, principalement, de UMLS)
- Très nombreuses relations (très structurées – 1832 classes, 436 relations)
- RdC : graphes conceptuels (Sowa)
- Modèles pour le raisonnement (~Scripts de Schank)
- R, P, F-M calculées concept par concept

PMID: 8130568, 7889741, 9357694



PERTOMed

Production et Evaluation de Ressources Terminologiques et Ontologiques dans le domaine de la Médecine

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017



Nom : PERTOMed
Années : 2003-2006
Type : TCAN/CNRS
Personnes du labo : AB, JC, MCJ
Objectif : reconnaissance du contenu du CRH (entre autres)
Domaine : pneumologie (entre autres)
Quantité de raisonnement : ++
Résultat : ontologie + preuve de concept (+ autres ressources)

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

PERTOMed

- Ontologie médicale construite à partir de corpus (CRHs AP-HP) et avec les cliniciens
- Anonymisation par principe
- 1150 classes, 22 relations
- RdC : dans les langages du WS (OWL/RDF)
- Prise en compte de la partie pneumo de la CIM-10
- TAL par Automate à États finis (FSA – Unitex) : preuve de concept
- Création de TopMenelas dans Bioportal

PMID: 16797227, 17238295, 20863975

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

15

code	label	valider?
J46	asthme aigu grave	<input checked="" type="checkbox"/>
J45.9	asthme	<input checked="" type="checkbox"/>
Z91.0	allergie	<input checked="" type="checkbox"/>
F17.2	tabagisme	<input checked="" type="checkbox"/>
R05	toux	<input checked="" type="checkbox"/>

Et il faut construire ces ontologies...

- *Top-down* et *bottom-up* stratégies
 - Avec une top-ontologie
 - Avec des méthodes de fouille de texte pour extraire puis valider des candidats termes¹

¹ Bachimont 2000

- De façon modulaire
- Avec des h*m non négligeables (>12)

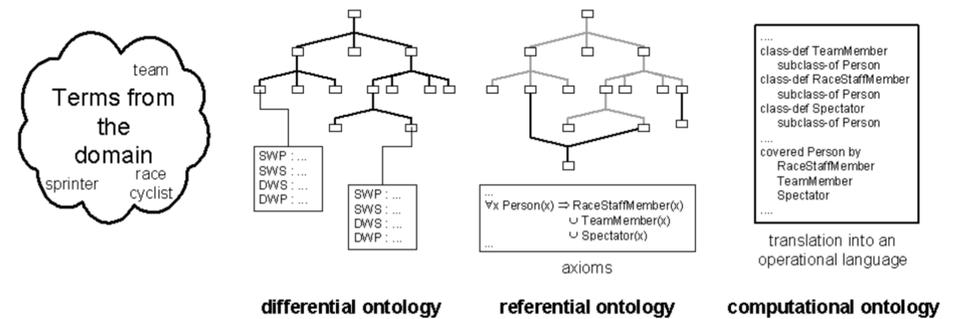
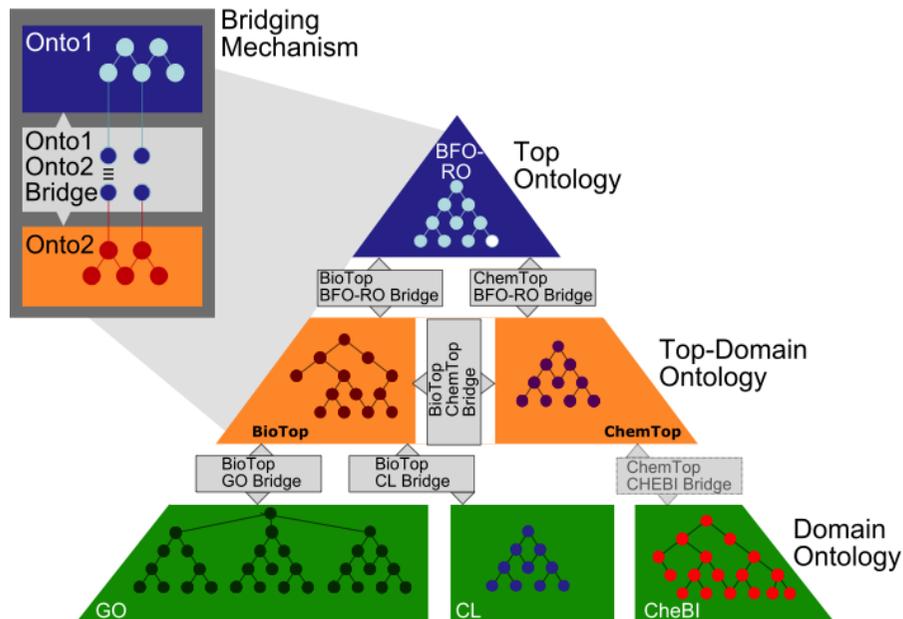


Fig. 1. The 3 steps of the differential methodology for building ontologies



Mise à disposition des ontologies

<http://pertomed.limics.upmc.fr/~onto/ontologies/>
<http://pertomed.limics.upmc.fr/~onto/ontologies/public/OntoMenelas/>

<http://bioportal.bioontology.org/ontologies/TOP-MENELAS>

<http://bioportal.bioontology.org/ontologies/ONTOPNEUMO>



LERUDI

LEcture Rapide en Urgence du Dossier Informatisé

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017



Nom : LERUDI
 Années : 2008-2011
 Type : ASIP-Santé
 Personnes du labo : JC, SJD, AB, FD, PYV
 Objectif : reconnaissance du contenu des textes du DPI
 Domaine : urgences
 Quantité de raisonnement : +++
 Résultat : ontologie + preuve de concept et architecture sémantique

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

LERUDI Ecran de synthèse

Documents récents

Document	Date	Statut
Cro bloc d'urgence viscéral et digestion	24/08/2009	CONSULT
Cr réanimation	17/11/2009	EXAMS
Crh chirurgie viscérale	26/10/2009	BIO
Cr echocardi	21/10/2009	EXAMS
Cr bio sortie	03/09/2009	BIO

Pathologies

Pathologie	Nombre de docs
Abcès du foie	13 docs
Cholécystite	11 docs
Péritonite	9 docs
Diagnostic de processus tumoral	8 docs
Adénocarcinome biliaire	7 docs
Hypertension artérielle	7 docs
Abcès sous-hépatique	6 docs
Malaria immunitaire	6 docs

Actes médicaux

Acte	Nombre de docs
Cholécystectomie	13 docs
Cholangiographie	11 docs
Antibiothérapie	9 docs
Scanner abdominal	6 docs
Créatininémie	5 docs
Échographie cardiaque	5 docs
Lipasémie	4 docs
TP	4 docs

Traitements

Traitement
Cérépime
Énoxaparine
Hydroxyzine
Nicotine
Pantoprazole
Zopiclone

LERUDI

- Ontologie médicale construite à partir de corpus (articles congrès), à partir de SNOMED, CCAM, CIM-10... et avec les cliniciens
- Anonymisation par respect réglementation
- 10031 classes, 60 relations
- Dans les langages du WS (OWL/RDF/SKOS)
- Utilisation de top-Menelas

<http://bioportal.bioontology.org/ontologies/ONTOLURGENCES>

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

24

LERUDI

- Évaluation sophistiquée avec plusieurs experts, R=0.98, P=0.8, Taux de contresens=0.02
- Applications développée par SWORD
- => ConSoRe

<http://esante.gouv.fr/actus/services/le-projet-lerudi-fiche-siglaetique>
<http://bioportal.bioontology.org/ontologies/ONTOLURGENCES>
<https://link.springer.com/article/10.1007/s13341-017-0755-z>

ConSoRe



Big data: Unicancer déploie un logiciel d'analyse sémantique de ses bases de données

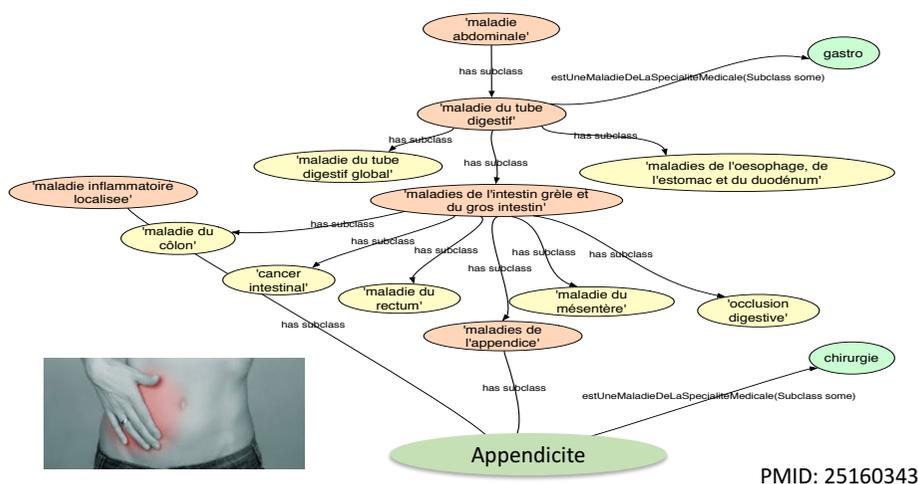
PARIS, 18 juin 2015 (TICSanté) - La fédération Unicancer vient de lancer le déploiement d'un outil d'analyse sémantique des bases de données de ses centres de lutte contre le cancer (CLCC) afin notamment de simplifier la mise en place de programmes de recherche, a expliqué Emmanuel Reyrat, directeur des systèmes d'information (DSI) d'Unicancer, le 21 mai lors des salons santé autonomie, à Paris.

Cet outil, appelé ConSoRe (Continuum soins-recherche) a été présenté par Emmanuel Reyrat à l'occasion d'une conférence de la Fédération hospitalière de France (FHF) sur le thème "Données de santé: quelles ressources exploiter, pour quoi faire?".

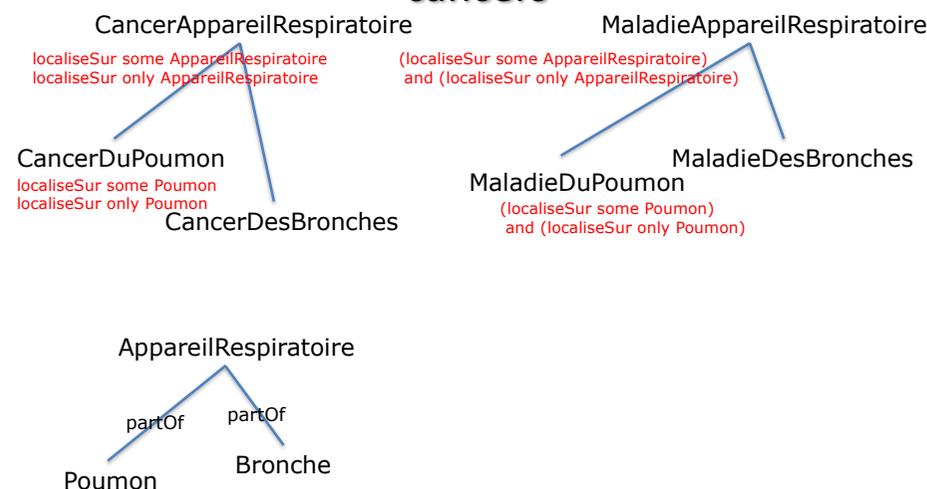
Les 18 CLCC possèdent des bases de données riches "de plusieurs dizaines de millions de documents", a souligné le DSI. Ces données structurées (PMSI, pharmacie, administration, etc.) et à 85% textuelles (comptes rendus opératoires, de consultation, etc.) représentent un volume de "big data"

http://www.ticsante.com/Big-data-Unicancer-deploie-un-logiciel-d-analyse-semantique-de-ses-bases-de-donnees-NS_2408.html

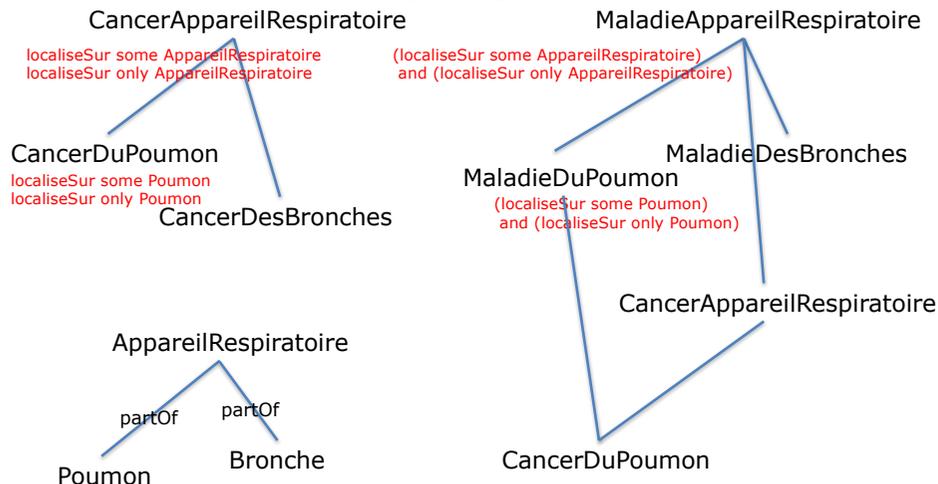
Et reprise de l'anatomie de la FMA



Raisonnement sur la localisation des cancers



Raisonnement sur la localisation des cancers



Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017 2/9



COVALMO

COConstruction et VALidation d'Ontologies Médicales

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017



Nom : COVALMO
 Années : 2012-2017
 Type : CHSA + UPMC
 Personnes du labo : MR, XA, JC
 Objectif : reconnaissance du contenu du CRH
 Domaine : psychiatrie
 Quantité de raisonnement : ++
 Résultat : ontologie = preuve de concept pour les psychiatres

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

COVALMO

- Ontologie construite à partir de corpus (CRHs CHSA) et avec les cliniciens
- Anonymisation (réglementation CNIL)
- Ontologie sociale finalisée (1476 classes, 200 relations)
- Prise en compte de la partie psychiatrie de la CIM-10, du DSM-IV, du DSM-5
- Test usage par prototype
- Mais surtout point d'étape avec les cliniciens

<http://bioportal.bioontology.org/ontologies/ONTOPSYCHIA>

PMID: 16797227, 17238295, 20863975

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017 32



PARON

Repérage et analyse des ruptures de PARcours de soin grâce à des ONTologies

OntoPsychia_VieSociale (http://www.limics.fr/OntoPsychia_VieSociale) : [Volumes/MyRaid/Users/jc/Documents/Djinfichiertravail/ontologies/OntoPsychia/OntoPsychia_Sociale_Novembre2016.owl]

Active Ontology x Entites x Classes x Object Properties x Individuals by class x DL Query x

Class hierarchy: Individu_Male

Annotations: Individu_Male

Annotations

- skos:prefLabel [language: fr]
- individu mâle
- skos:prefLabel [language: en]
- male individual
- skos:definition [language: fr]
- Personne, individu de sexe masculin
- skos:altLabel [language: fr]
- mâle

Description: Individu_Male

Equivalent To

SubClasses Of

- a_pour_sexe_de_naissance some Male
- Individu_Selon_Sexe

General class axioms

SubClass Of (Anonymous Ancestor)

- a_parent_biologique exactly 2 Parent_Biologique

Instances

Target for Key

Disjoint With

Disjoint Union

To use the reasoner click Reasoner > Start reasoner Show Inferences



Réseau SLA : Caractériser les ruptures de parcours de soin

- Existence de réseaux de coordination de soin SLA, (1) pour les patients, l'entourage et les professionnels de proximité intervenant dans la prise en charge des patients, et (2) faisant le lien entre la ville et l'hôpital et de façon spécifique avec le centre expert SLA
- Comprendre et décrire les parcours de santé des patients atteints de pathologies neurodégénératives afin de comprendre, identifier les points de fragilité et les leviers pour améliorer la continuité des parcours et prévenir les ruptures
- Via les données du réseau (description événements), les données PMSI, ...



Nom : PARON
Années : 2016-2019
Type : ICM + UPMC
Personnes du labo : JC, SC
Objectif : reconnaissance du contenu des textes des événements du réseau de soin
Domaine : SLA
Quantité de raisonnement : ++/+++
Résultat : ontologie + étude rétrospective

PARON

- Ontologie construite à partir de corpus (événements réseau SLA) par le doctorant (professionnel) et avec les cliniciens
- Anonymisation (réglementation CNIL), support du début de l'annotation (remplacement des entités nommées par leur classe – Medina/LIMSI)
- Ontologie sociale + de coordination finalisées (3000 classes, x relations)
- Développement d'un outil de TAL à l'état de l'art (Gate)
- Résultats en cours très encourageants dans le repérage des critères de rupture...
- Ontologie médicale à développer (approche modulaire, *Parkinson ready*)

Réseau SLA Activité et modélisation ontologique

The screenshot displays a software interface for ontology management. On the left, a 'Class hierarchy' tree shows a path from 'RechercheDeKiné' to 'RechercheDeProfessionnel' and 'RechercheDeStructure'. On the right, the 'Annotations' panel shows 'altLabel [language: fr] recherche de kiné' and 'prefLabel [language: fr] recherche de kinésithérapeute'. Below, the 'Description' panel for 'RechercheDeKiné' lists various relationships like 'Equivalent To', 'SubClass Of', and 'General class axioms'.



WATSON/IBM

... et le cycle de la *Hype*

- https://fr.wikipedia.org/wiki/Cycle_du_hype
- Le Big Data ?
- Deep Learning = IA ?
- Et IBM/Watson ...



```
Nom : WATSON/IBM
Années : 2015...
Type : IBM
Personnes du labo : JC, SC
Objectif : reconnaissance du contenu des
textes de l'hôpital +++
Domaine : oncologie
Quantité de raisonnement : +++
Résultat : ???
```

Google ibm watson

Tous Actualités Vidéos Images Shopping Plus Paramètres Outils

Environ 18 100 000 résultats (0,66 secondes)

IBM Watson - Vous À La Puissance IBM - ibm.com
www.ibm.com/
 Osez L'Innovation Avec L'Intelligence Artificielle Pour Les Entreprises
 IBM Cloud · IBM Systems · Big Data & Analytics · Les Enjeux de la Fraude

IBM Watson
<https://www.ibm.com/watson/> · Traduire cette page
 Watson is the AI platform for business.
 Watson Conversation · Watson Discovery · Watson blog · Watson Media

IBM Watson Talent - France
<https://www.ibm.com/watson/fr-fr/talent/>
 Watson Talent étend l'expertise humaine, ce qui permet aux RH d'améliorer et d'accroître l'impact des personnes sur l'entreprise.

Watson (intelligence artificielle) — Wikipédia
[https://fr.wikipedia.org/wiki/Watson_\(intelligence_artificielle\)](https://fr.wikipedia.org/wiki/Watson_(intelligence_artificielle))
 Watson est un programme informatique d'intelligence artificielle conçu par IBM dans le but de répondre à des questions formulées en langage naturel.
 Participation à Jeopardy! · Applications actuelles et ... · Notes et références

Watson : comment marche l'IA d'IBM dans la santé, la banque... - JDN
www.journaldu.net.com · Web & Tech · Entreprises · 26 oct. 2017 - Watson, l'intelligence artificielle d'IBM, est capable de répondre aux questions en langage naturel que lui posent ses utilisateurs dans de ...

A la rencontre de Watson, l'intelligence artificielle star d'IBM - Le Monde
www.lemonde.fr.../a-la-rencontre-de-watson-l-intelligence-artificielle-star-d-ibm_50933...
 12 mars 2017 - Watson peut composer de la musique, conseiller des médecins ou même tenter d'évaluer l'humeur de son interlocuteur. Zoom sur cette ...

Comment, avec son programme Watson, IBM veut changer le monde
www.lemonde.fr/economie/article/.../watson-le-savant-calcul-d-ibm_5194573_3234.htm...
 1 oct. 2017 - Transport, santé, relation client... L'intelligence artificielle promet de changer le monde. Watson, le programme phare de Big Blue, peine à ...

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close



Last February, MD Anderson Cancer Center, part of the University of Texas, cancelled its partnership with Watson.

<https://www.statnews.com/2017/09/05/watson-ibm-cancer/>

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

Économie

ECONOMIE Les données du Monde · Economie mondiale · Economie française · Entreprises · Emploi

ARTICLE SÉLECTIONNÉ DANS LA MATINALE DU 01/10/2017

Comment, avec son programme Watson, IBM veut changer le monde

Transport, santé, relation client... L'intelligence artificielle promet de changer le monde. Watson, le programme phare de Big Blue, peine à relever le défi de l'industrialisation.

LE MONDE ECONOMIE | 02.10.2017 à 09h39 · Mis à jour le 02.10.2017 à 15h11 | Par Chloé Hecketsweiler et Sandrine Cassini

Abonnez vous à partir de 1 €

Kendall Square, c'est un peu la Silicon Valley de la Côte est américaine. Ce quartier de Cambridge — un coin de 2 kilomètres carrés — est le royaume du Massachusetts Institute of Technology (MIT), l'une des Meccas de la recherche aux États-Unis.

http://www.lemonde.fr/economie/article/2017/10/02/watson-le-savant-calcul-d-ibm_5194573_3234.html?xtmc=watson&xtrcr=13

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

L'“IA” Watson, le pour et le contre

Pour

- + Trouve les articles pertinents
- + Propose des protocoles de soins connus/prévus
- + Affiche toutes les options pour discuter

Contre

- Incapable de proposer un bon protocole en cas d'échec du premier
- En accord avec les professionnels dans 33% des cas (Danemark)
- Ne comprend pas bien les dossiers médicaux
- A du mal à contextualiser les propositions (par pays)
- Impossible, à ce jour, d'écrire un article médical démontrant l'avantage d'utiliser Watson

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

Watson et le cycle de la Hype

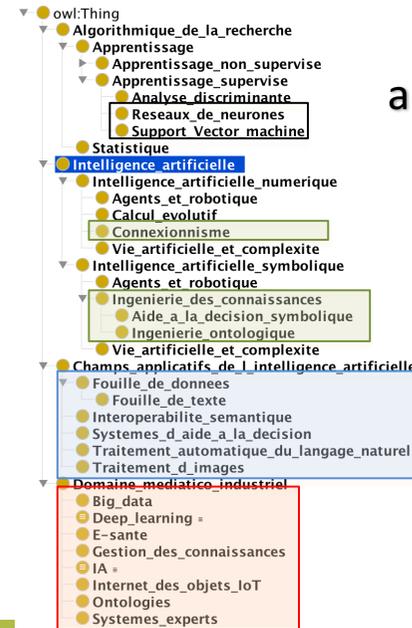
- https://fr.wikipedia.org/wiki/Cycle_du_hype
- Où est parti le Big Data ?
Dans le paquet des buzz inutiles ?
- Deep Learning = IA ?
- Et IBM/Watson fait peut-être des choses bien ?
- Il y a des modes médiatiques... et pendant ce temps on travaille

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

Un petit cours d'I.A. appliqué à la médecine (1)



- Dans les journaux, le Web
- Les domaines qu'aborde l'IA
- Les sous-disciplines de l'IA
- Les algorithmes

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

Un petit cours d'I.A. (2)

		ALL		
	Team	P	R	F
Official runs	SIBM-run1	.857	.689	.764
	LITL-run2	.666	.414	.510
	LIRMM-run1	.541	.480	.509
	LIRMM-run2	.540	.480	.508
	LITL-run1	.651	.404	.499
	TUC-MI-run2	.044	.026	.033
	TUC-MI-run1	.025	.015	.019
	average	.475	.358	.406
	median	.541	.414	.508
Non-official	LIMSI-run2	.872	.784	.825
	LIMSI-run1	.883	.760	.817
	TUC-MI-run1-corrected	.883	.539	.669
	TUC-MI-run2-corrected	.882	.536	.667
	UNIPD-run1	.629	.468	.537
	UNIPD-run2	.518	.384	.441
	Mondeca-run1	.375	.131	.194
	Frequency baseline	.339	.237	.279

- Retrouver des codes de maladies CIM-10 dans les textes des certificats de décès du CépiDc (et aussi sur un échantillon anglais)
- Un cas théoriquement très facile
- VP, FP, VN, FN, Précision (VP/(VP+FP)), rappel (VP/(VP+FN)), F-mesure (moyenne harmonique(Rappel, Précision))
- Sur cet exercice, les équipe françaises sont très bonnes dans les 2 langues (fr, en) mais spécialement bonnes en français (cette figure)
- Les meilleures équipes utilisent des techniques mixtes d'annotation sémantique et d'apprentissage (svm)

CLEF eHealth 2017 Multilingual Information Extraction task overview: ICD10 coding of death certificates in English and French
<https://pdfs.semanticscholar.org/7726/a6eff024adee59c1bf21d88f5a0f75239f29.pdf>

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

Watson, qu'est-ce que c'est ? (1/2)

- Un programme utilisant des techniques de TALN et de fouille de données
- Un programme entraîné par des être humains donnant les « gold standards » des traitements (heureusement ?)
It's taken nearly six years of painstaking work by data engineers and doctors to train Watson in just seven types of cancer, and keep the system updated with the latest knowledge
- Fouillant les données de l'hôpital sur les patients

Jean Charlet

Enjeux de la eSanté...

12/ 12/ 2017

Watson, qu'est-ce que c'est ? (2/2)

- Un programme fouillant aussi (tous) les textes de l'hôpital pour *apprendre* des connaissances et pour *retrouver* d'autres données sur les patients
- Utilisant des technologies principalement numériques (et des technologies symboliques ?) pour le TALN (aucune publication sérieuse trouvée)
- Utilisant des technologies numériques pour la fouille de données (connexionnisme)
- Dans un contexte médiatique fort : *le Digital Canary* de l'IA en médecine

Watson, un échec ?

- Quelles que soient les méthodes, pour une annotation complexe, une F-mesure de 0,7 est un maximum
- Si dans les textes médicaux, il y a des erreurs ou des imprécisions ou des sources de confusion (EDC), il est difficile de les corriger
- **Aujourd'hui**, une analyse correcte (> 0.9) des éléments textuels d'un dossier standard est impossible sans intervention humaine
- Et IBM/Watson fait des choses **surement très bien... mais pas tout à fait ce qu'ils ont (sur)vendu**

Watson, un échec ?

- Quelles que soient les méthodes, pour une annotation complexe, une F-mesure de 0,7 est
- Si dans les textes médicaux, il y a des erreurs ou des imprécisions ou des sources de confusion (EDC), il est difficile de les corriger
- **Aujourd'hui**, une analyse correcte (> 0.9) des éléments textuels d'un dossier standard est impossible sans intervention humaine
- Et IBM/Watson fait des choses **surement très bien... mais pas tout à fait ce qu'ils ont (sur)vendu**

Episode Dépressif Caractérisé
Every Day Carry
Entrepôt de Données Cliniques
Electronic Data Capture

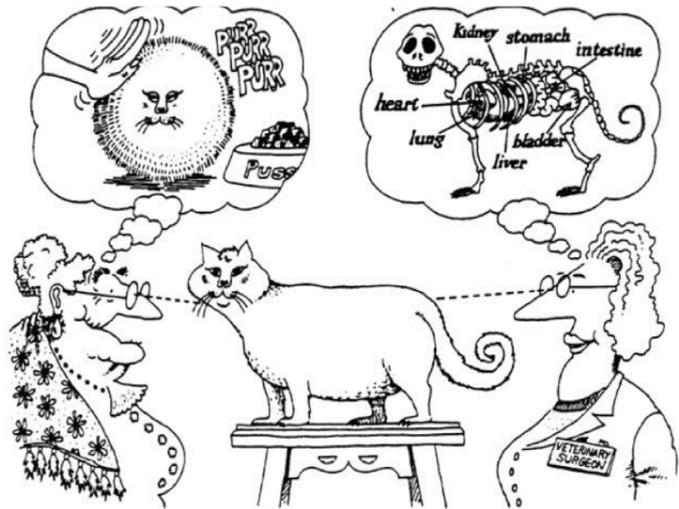
Discussion : des propositions

- **Continuer les recherches**
 - Approches mixtes, ontologies, apprentissage, etc.
 - **Ce que fait très probablement IBMWA**
- Faire intervenir les **professionnels de santé** dans l'annotation sémantique des textes (sur de "petits" échantillons – EDC)
- **Communiquer** sur la modélisation, les processus de TALN et d'apprentissage auprès de ceux qui écrivent les textes

Pour

- Une meilleure structuration des données en général
- Une amélioration du format et de l'expression et de la véracité des textes

⇒ Meilleure reconnaissance et résultats



Abstraction focuses upon the essential characteristics of some object, relative to the perspective of the viewer.

Je vous remercie de votre attention