

GESTION DES ANNOTATIONS SEMANTIQUES EN SANTE

Le projet **elisa** 

Journées MaDICS – Atelier QUALIMADOS

Cédric PRUSKI

Marseille, le 23 juin 2017

UNIVERSITÄT LEIPZIG



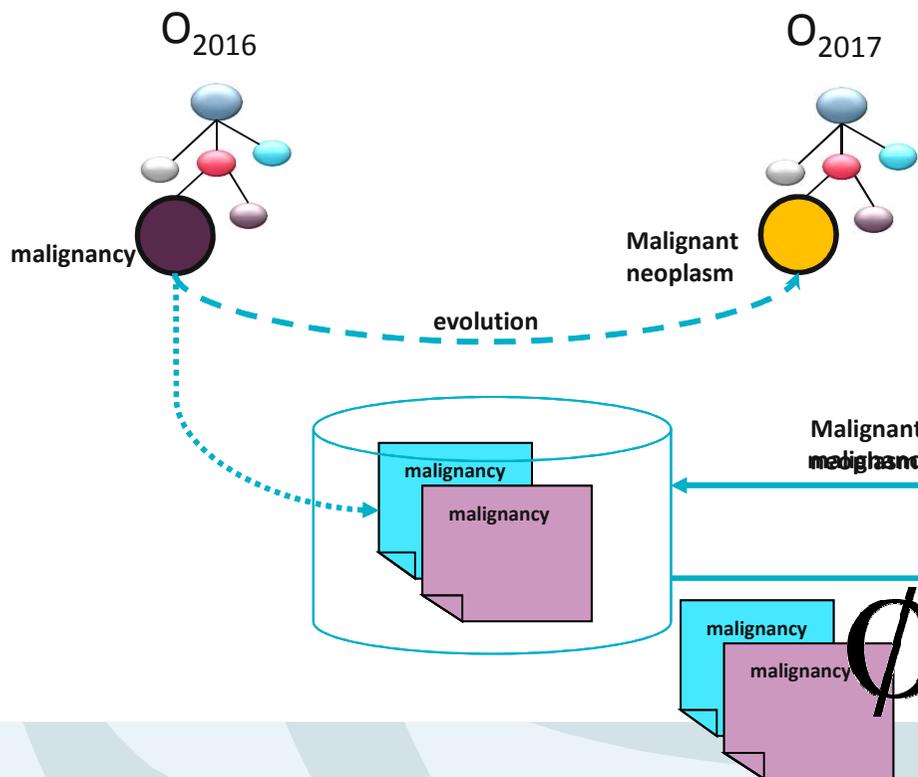
DFG



LUXEMBOURG
INSTITUTE OF SCIENCE
AND TECHNOLOGY | **LIST** 

The LIST logo is a colorful, abstract sphere composed of multiple overlapping lines in red, blue, and white.

ILLUSTRATION DU PROBLÈME



Importance de générer des annotations cohérentes et de bien les faire évoluer au cours du temps



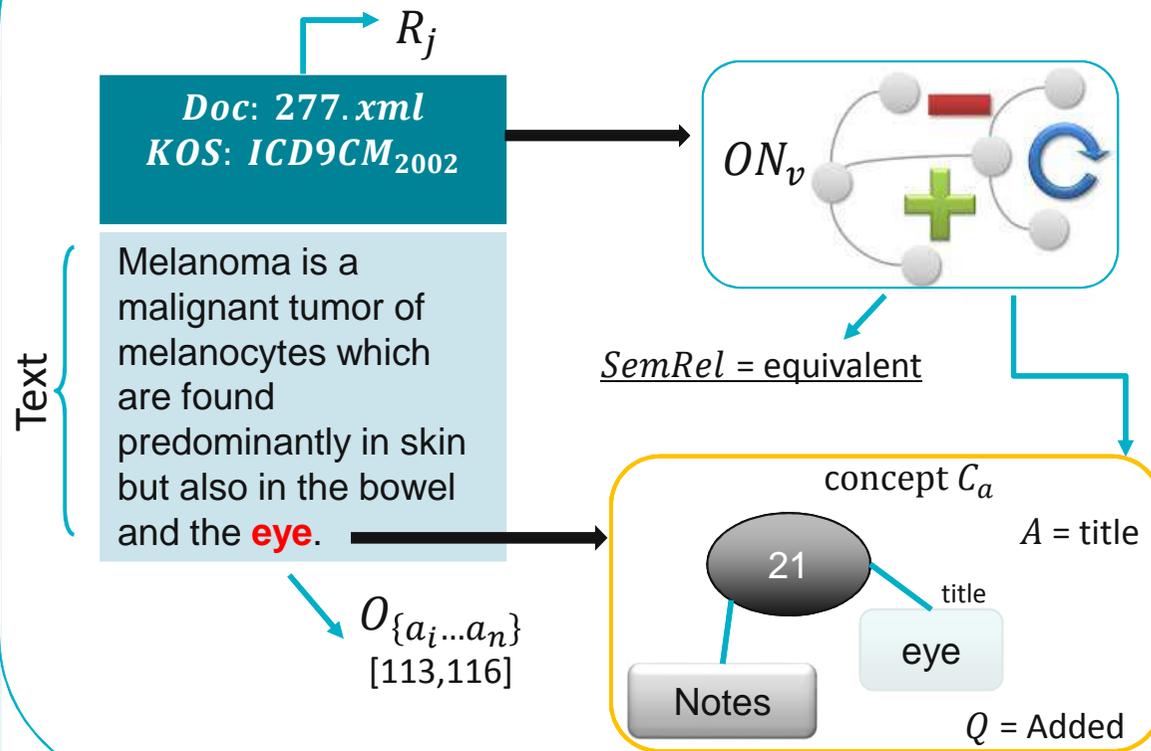
PLAN DE L'EXPOSÉ

1. Modèle d'annotation sémantique
2. Méthode pour la génération d'annotations sémantiques
3. Mesure de l'impact de l'évolution des RTOs sur les annotations sémantiques
 - a) Méthode
 - b) Résultats expérimentaux
4. Travaux existants
5. L'approche de maintenance des annotations sémantiques
6. Travaux en cours
7. Conclusion

MODÈLE POUR DÉCRIRE LES ANNOTATIONS SÉMANTIQUES

MODÈLE D'ANNOTATION SÉMANTIQUE

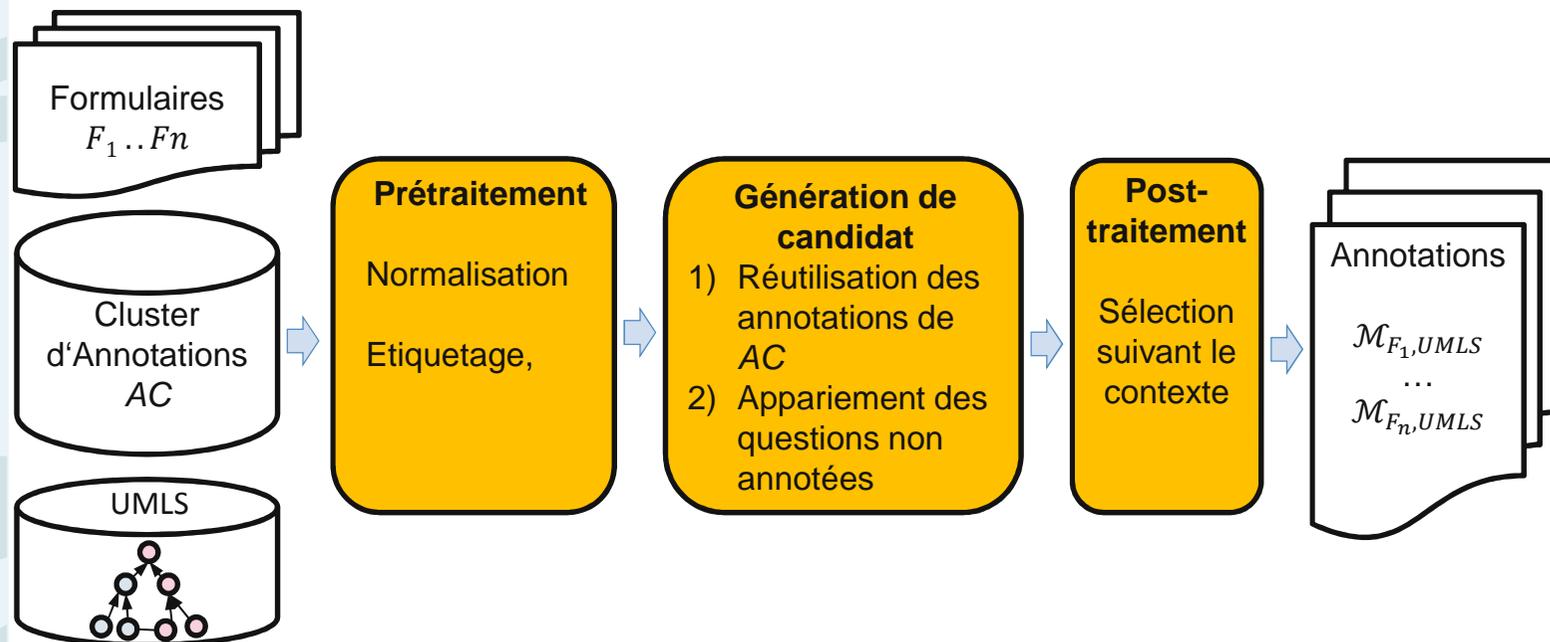
$$SAM = (I_u, ON_v, R_a, Offset, Q, A, SemRel, U_f)$$



$I_u = 231$

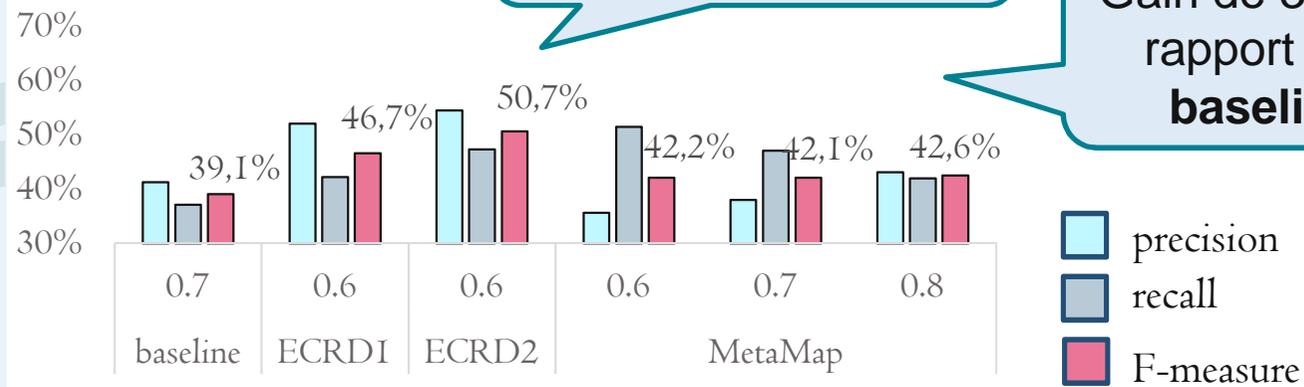
GÉNÉRATION DES ANNOTATIONS SÉMANTIQUES

MÉTHODE POUR LA GÉNÉRATION D'ANNOTATIONS SÉMANTIQUES

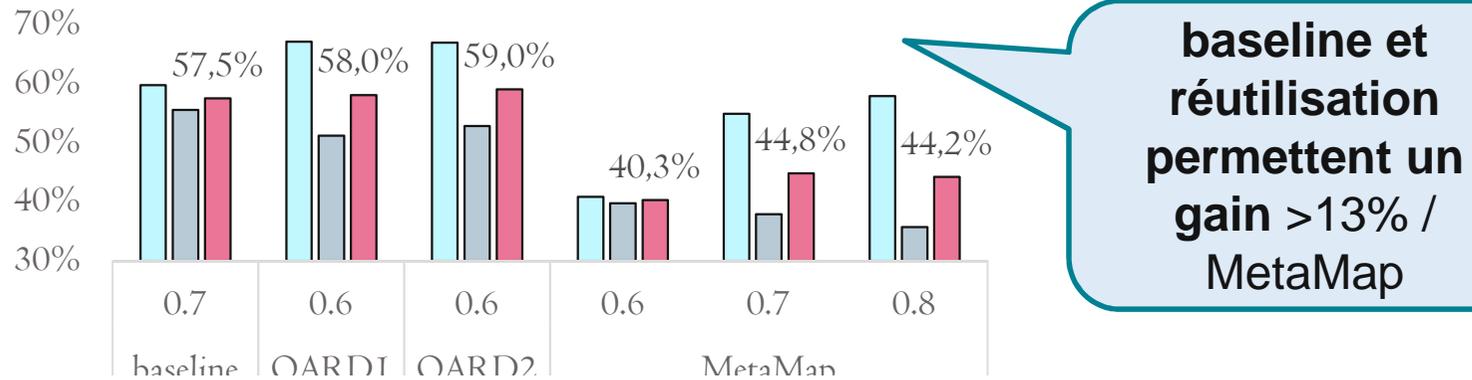


Christen, V., Groß, A., Rahm, E. A Reuse-based Annotation Approach for Medical Documents. In International Semantic Web Conference, ISWC 2016, Kobe, Japan.

Jeu de données 1:

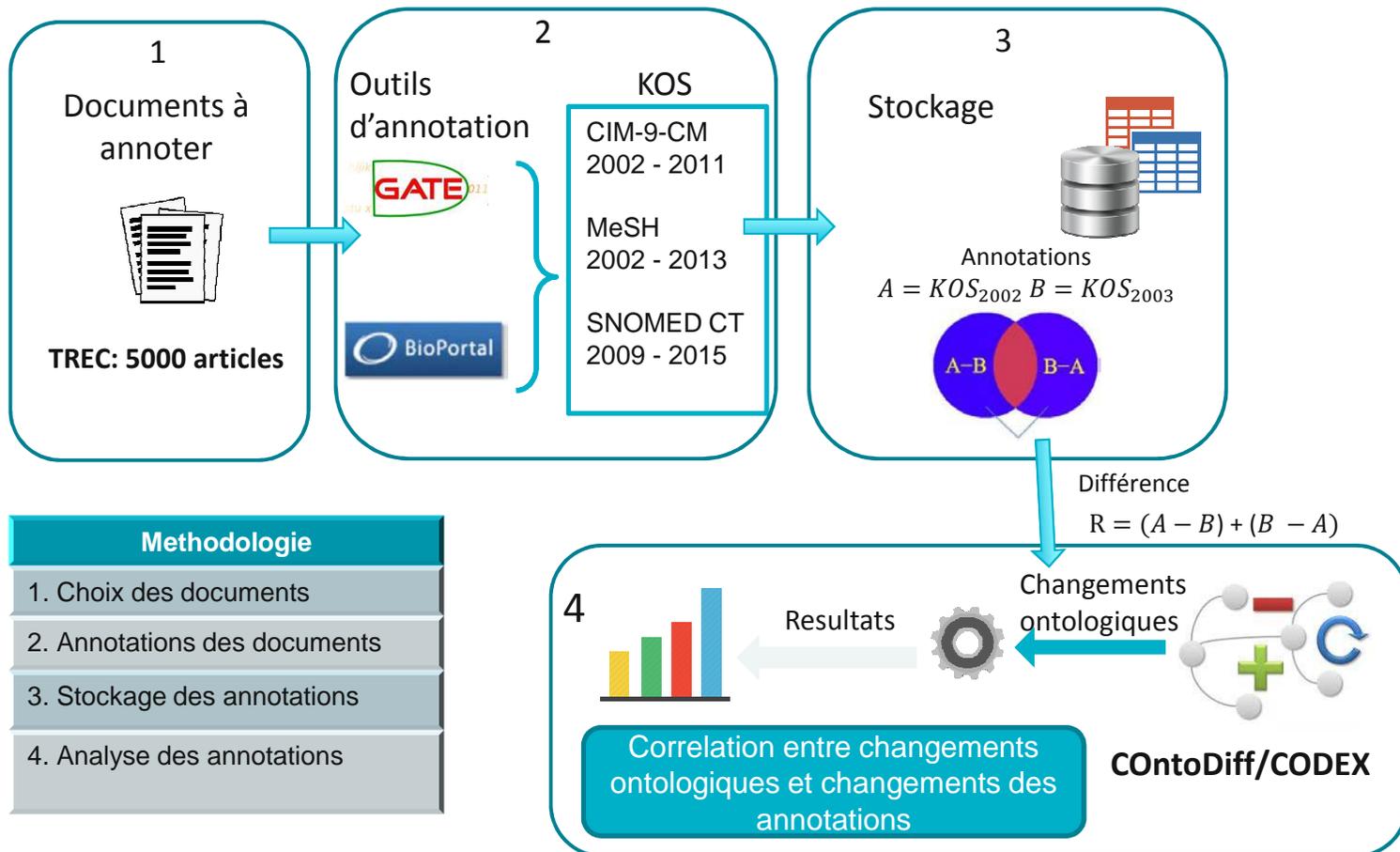


Jeu de données 2:



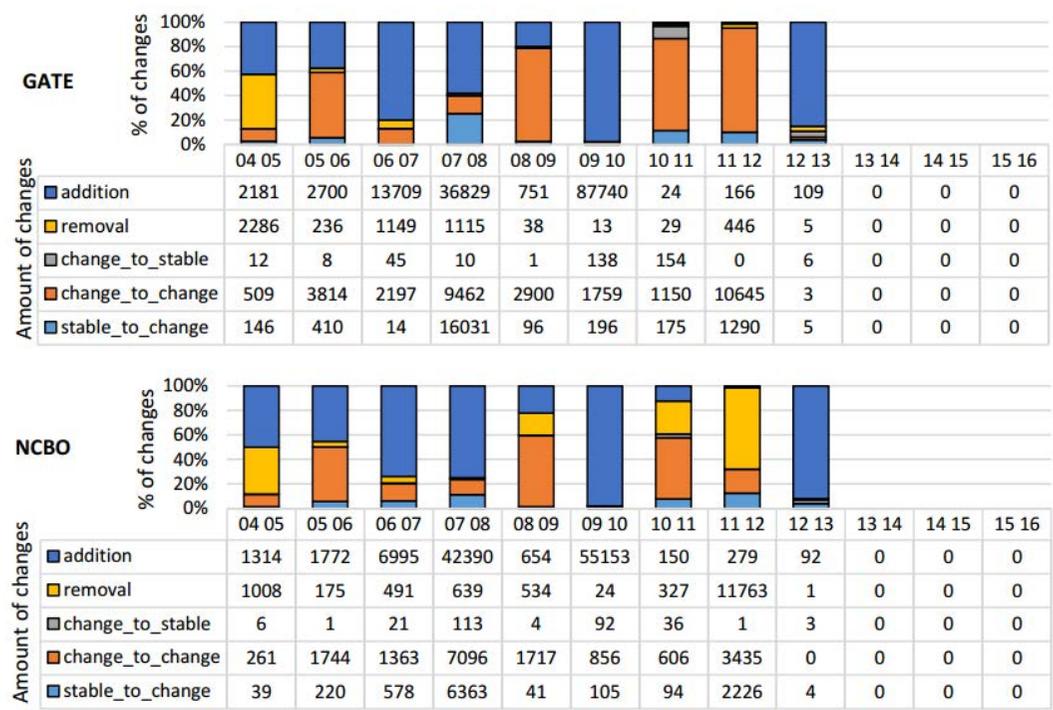
MESURE DE L'IMPACT DE L'ÉVOLUTION DES RTO SUR LES ANNOTATIONS SÉMANTIQUES

IMPACT DE L'ÉVOLUTION DES RTO SUR LES ANNOTATIONS SÉMANTIQUES

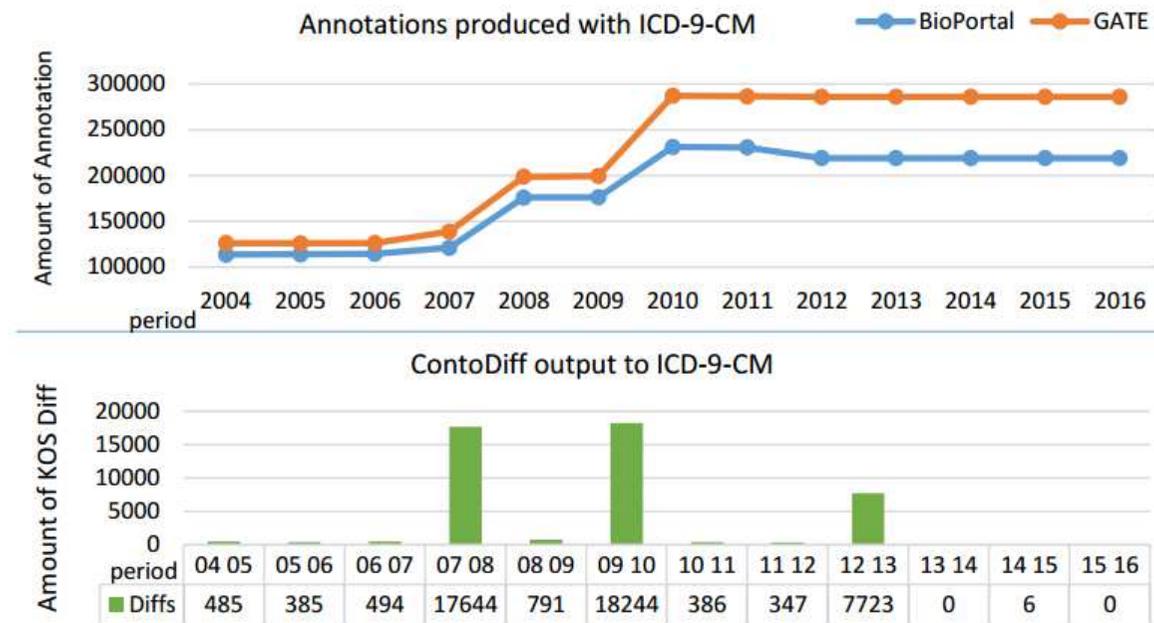


Methodologie
1. Choix des documents
2. Annotations des documents
3. Stockage des annotations
4. Analyse des annotations

IMPACT DE L'ÉVOLUTION DES RTO SUR LES ANNOTATIONS SÉMANTIQUES



IMPACT DE L'ÉVOLUTION DES RTO SUR LES ANNOTATIONS SÉMANTIQUES



Cardoso, S. D., et al. (2016). Leveraging the impact of ontology evolution on semantic annotations. In Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, pages 68–82.

PROBLÈMATIQUE

- **Principales contraintes à prendre en compte:**
 - Une approche manuelle pour la maintenance est irréaliste vu le nombre d'annotations à adapter
 - Les documents annotés ne sont pas tous accessibles (ex: dossier patient)
 - Dans certains cas les annotations ne sont pas modifiables
 - Certaines RTOs ne sont pas exprimées dans des langages formels pour utiliser des techniques de raisonnement (ex: CIM 10, MeSH ...)

Question de recherche

Comment adapter automatiquement les annotations sémantiques impactées par l'évolution des RTOs sous-jacentes sans ré-annoter la totalité des documents?

ÉTAT DE L'ART

TRAVAUX EXISTANTS

Détection automatique des annotations incohérentes

- Identification des annotations invalides
- Pas de corrections de ces annotations

(Eilbeck et al., 2009)
(Qin and Atluri, 2009)
(Kopke and Eder, 2011)
(Zavalina et al., 2015)

Détection automatique et correction manuelle des annotations

- Seuls les changements ontologiques simples sont pris en compte

(Maynard et al., 2007)
(Auer and Herre, 2007)
(Burger et al., 2010)
(Abgaz, 2013)

Détection et correction automatique des annotations

- Spécifiques au domaine
- Utilisation des techniques de raisonnement

(Luong and Dieng-Kuntz, 2006)
(Tissaoui et al., 2011)
(Park et al., 2011)
(Frost and Moore, 2014)

MAINTENANCE DES ANNOTATIONS SÉMANTIQUES

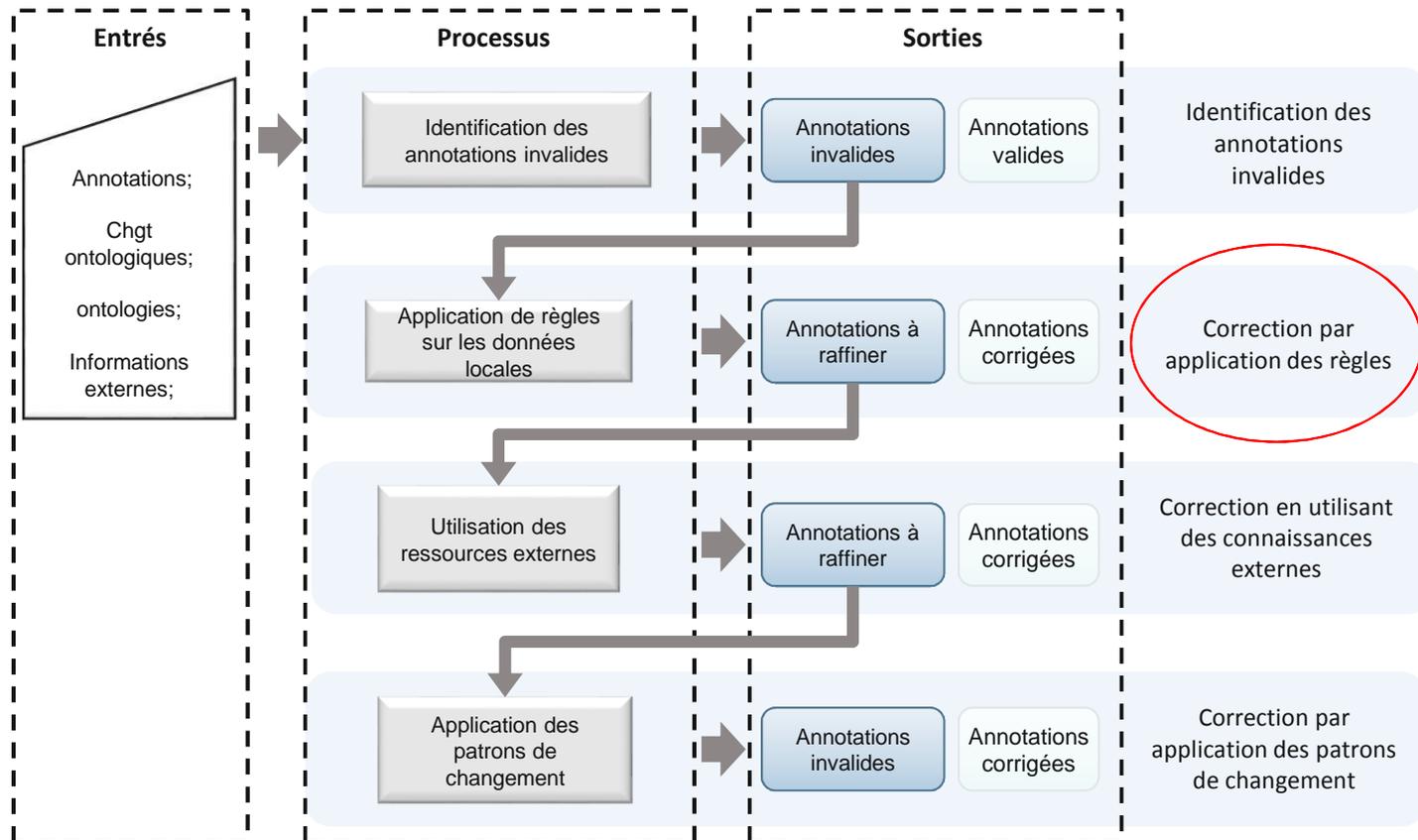
DEUX CAS DE FIGURE

- Maintenance directe des annotations:
 - Les annotations sont modifiables directement
 - Les documents annotés sont accessibles

 - Maintenance indirecte des annotations:
 - Les annotations sont accessibles mais ne peuvent pas être modifiées
 - Les documents annotés ne sont pas consultables
- **Préserver l'utilisabilité des données annotées**

Cardoso, S. D., et al. (2017). Towards a Multi-level Approach for the Maintenance of Semantic Annotation, In HEALTHINF 2017, Porto, Portugal, February 19-23, 2017

1^{ER} CAS: MAINTENANCE DIRECTE



RÈGLES

- *MergeAnnot*: Fusion de deux annotations dans la nouvelle version
- *IncreaseAnnot*: Augmentation de la quantité d'information à annoter
- *ResurrectAnnot*: Reconsidération d'une annotation supprimée dans la précédente version

- *PluralAnnot*: Considération des pluriels/singuliers

- *ChangeConceptAnnot*: Changement du concept pour annoter
- *SplitAnnot*: Eclatement de l'annotation

- *SuperClassAnnot*: Utilisation du super concept pour annoter

- L'application des règles se fait suivant le guide des bonnes pratiques pour l'annotation
Dogan et. al. "*NCBI disease corpus: a resource for disease name recognition and concept normalization.*"

EXEMPLE DE RÈGLE

MergeAnnot

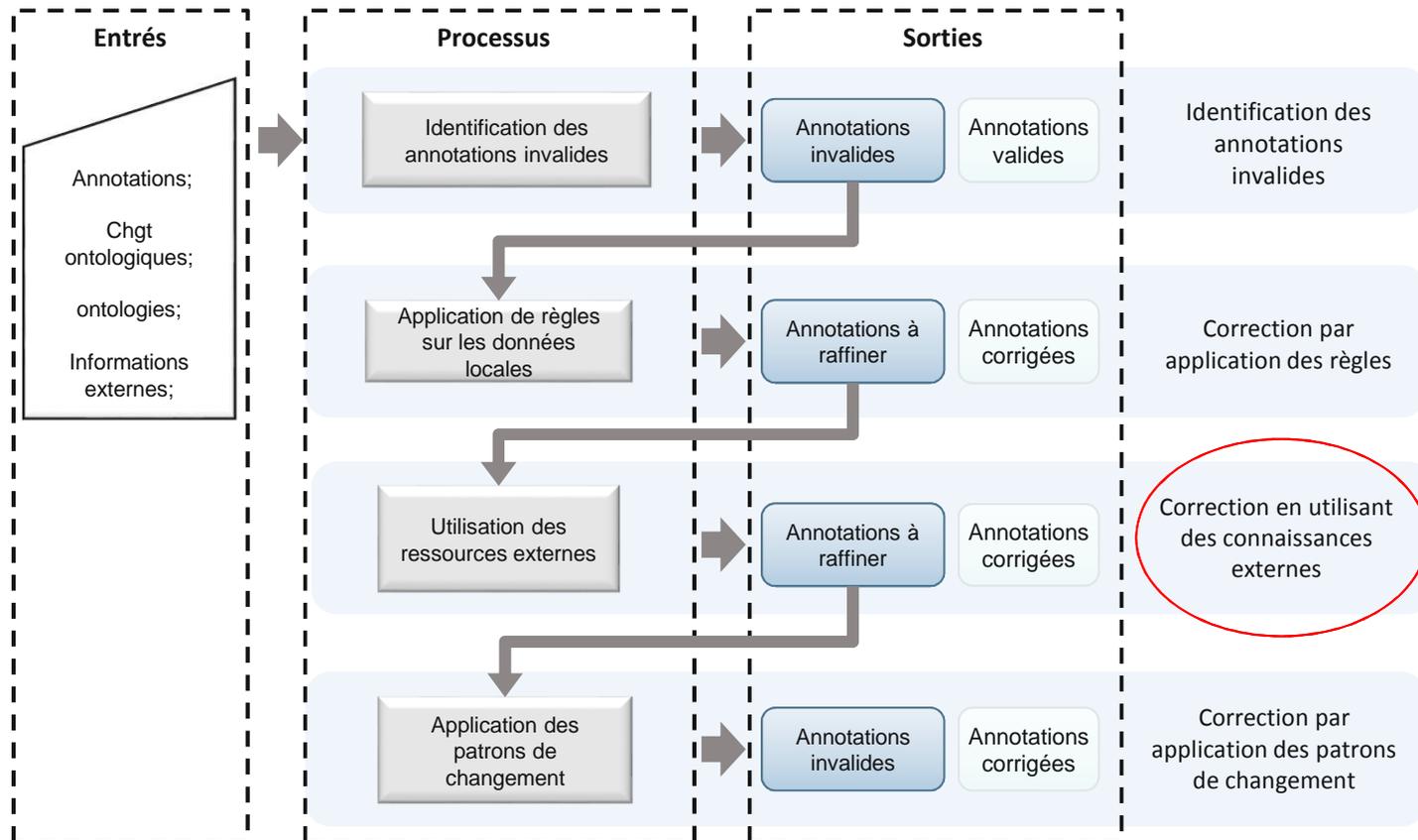
MeSH 2009

D011247 D006973
diabetes mellitus and pregnancy-induced hypertension . Apgars were

diabetes mellitus and pregnancy-induced hypertension . Apgars were
D046110

MeSH 2010

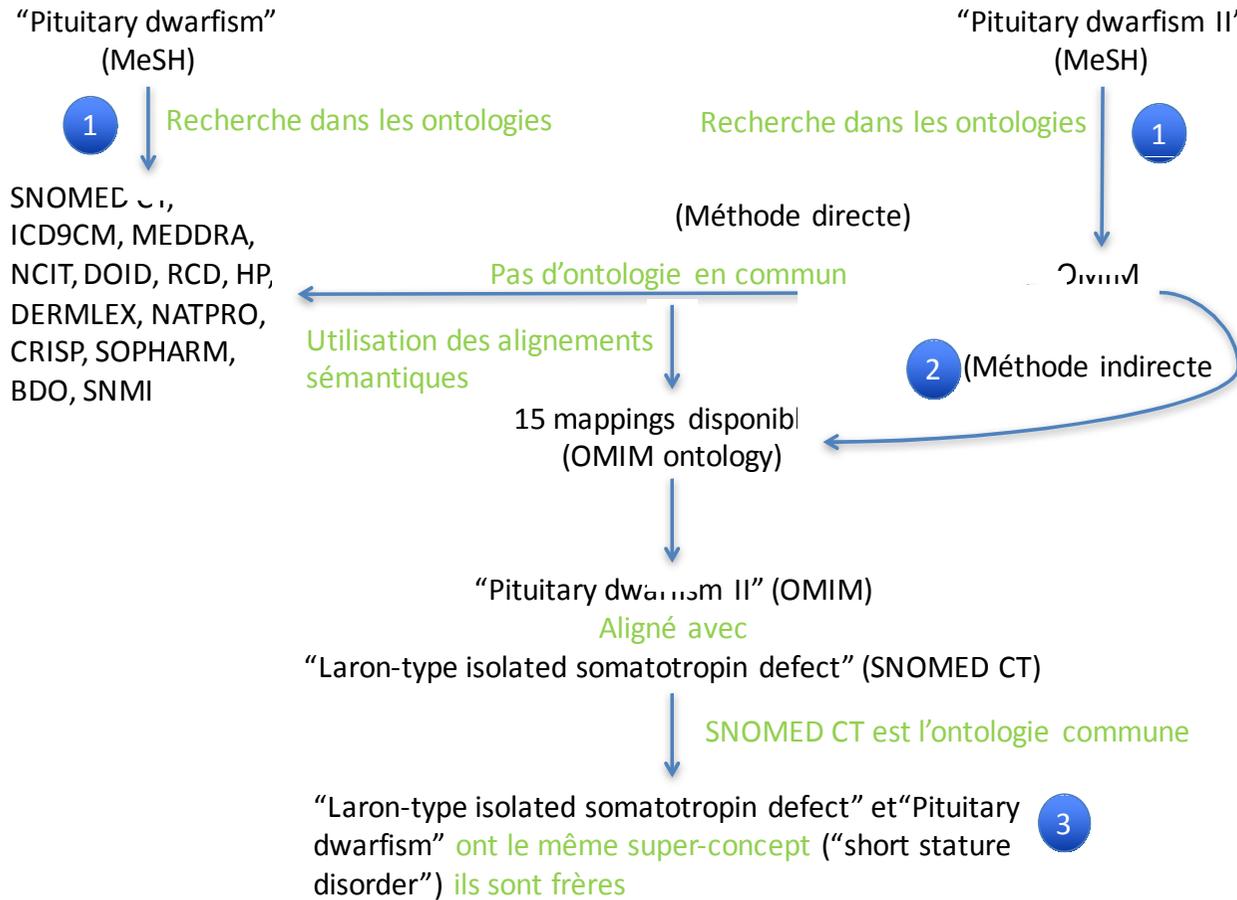
1^{ER} CAS: MAINTENANCE DIRECTE



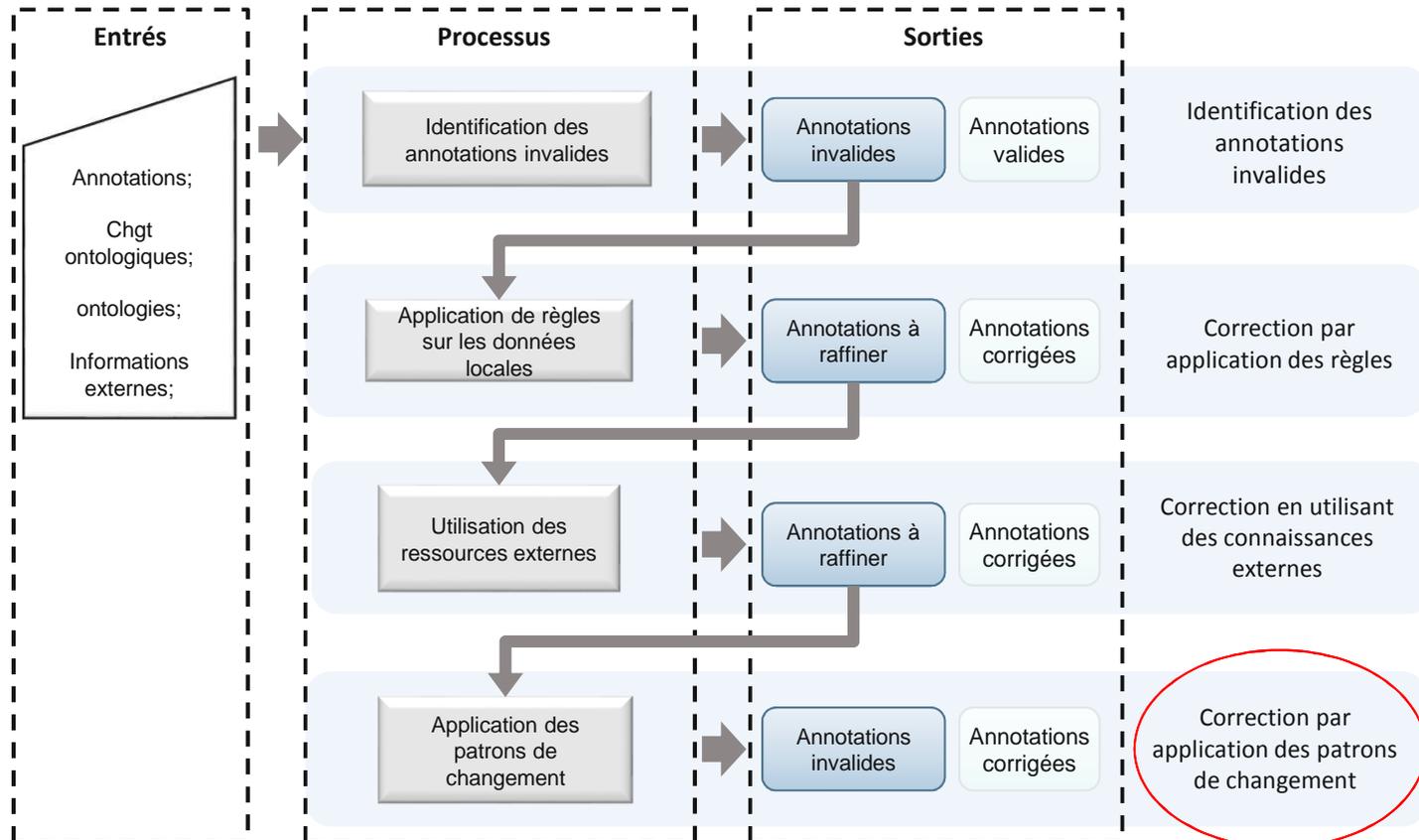
UTILISATION DE RESSOURCES EXTERNES

Données du Web Sémantique

- Les étiquettes des concepts peuvent ne pas avoir de lien syntaxique avant et après évolution
 - Ex: Cancer  Malignant neoplasm
- Besoin de considérer des aspects liés à la **sémantique** des concepts
 - Utilisation du Web Sémantique pour caractériser l'évolution des concepts
- Dans le domaine de la santé nous pouvons utiliser les ontologies et alignements contenus dans Bioportal
- Nous sommes capable d'identifier si un concept:
 - Reste équivalent, devient plus ou moins spécifique, sont reliés par une relation indéfinie



1^{ER} CAS: MAINTENANCE DIRECTE



1^{ER} CAS: MAINTENANCE DIRECTE

Patrons de changement

- Définition de patrons de changement:
 - Lexicaux
 - Sémantiques
- Le but est d'être capable d'identifier comment les concepts d'une ontologie ont évolué et quel est le résultat (en termes de sémantique) de cette évolution
- Analyse du contexte d'un concept avant et après évolution:
 - Super-concepts, sous-concepts et concepts frères
- Utilisation de cette information pour maintenir les annotations sémantiques

Dos Reis, J. C., et al. (2015). Recognizing lexical and semantic change patterns in evolving life science ontologies to inform mapping adaptation, In AIIM journal, 63(3), pp 153-170

1^{ER} CAS: MAINTENANCE DIRECTE

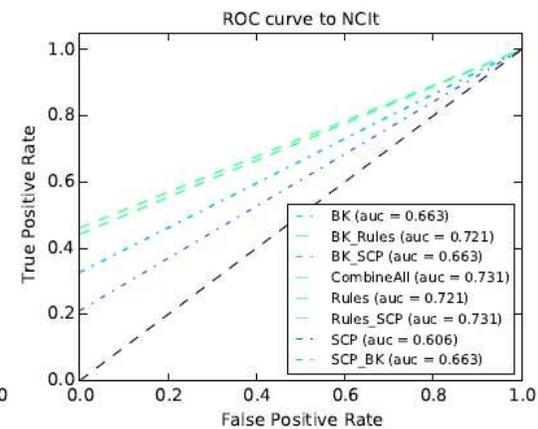
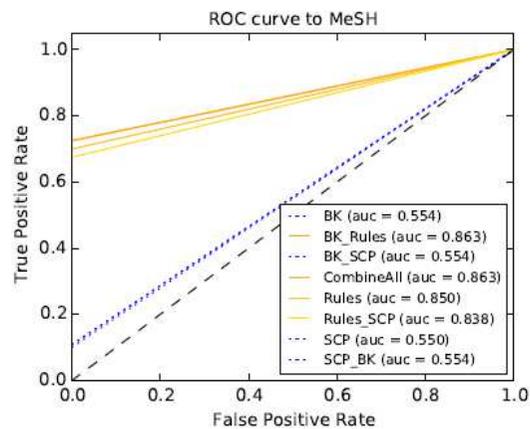
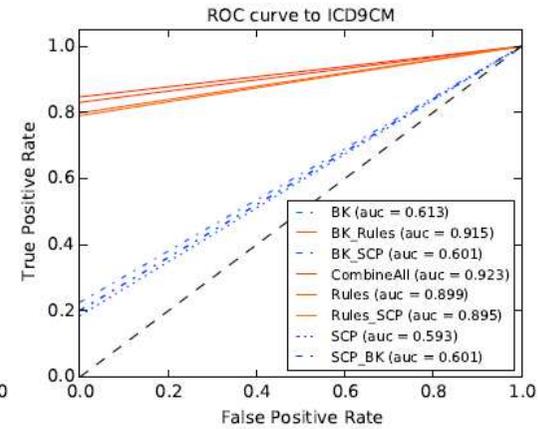
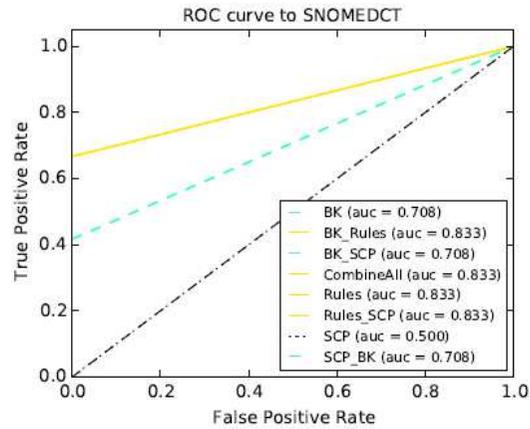
Evaluation

Method	ICD9CM			MeSH			NCIT			SNOMEDCT		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BK	1	0,16	0,28	1	0,05	0,51	1	0,13	0,24	0,96	0,54	0,69
BK & Rules	1	0,98	0,99	0,98	0,96	0,95	0,98	0,88	0,93	1	0,79	0,88
BK & SCP	1	0,19	0,32	1	0,05	0,51	0,91	0,19	0,32	0,96	0,54	0,69
CombineAll	1	0,98	0,99	0,98	0,96	0,95	0,98	0,88	0,93	1	0,79	0,88
Rules	1	0,98	0,99	0,98	0,96	0,95	0,98	0,88	0,93	1	0,79	0,88
Rules & SCP	1	0,98	0,99	0,98	0,96	0,95	0,98	0,88	0,93	1	0,79	0,88
SCP	1	0,08	0,15	1	0,02	0,50	0,83	0,10	0,17	0	0	0
—	—	—	—	—	—	—	—	—	—	—	—	—

Precision (P), Recall (R) and F1-Score (F1) of impacted annotations computed by the developed heuristics. The red and orange color highlight low and medium recall, respectively.

1^{ER} CAS: MAINTENANCE DIRECTE

Evaluation



TRAVAUX EN COURS

2^{eme} cas: Maintenance indirecte

- Maintenance indirecte des annotations:
 - Les annotations sont accessibles mais ne peuvent pas être modifiées
 - Les documents annotés ne sont pas consultables

→ **Préserver l'utilisabilité des données annotées**
- Définir une approche permettant la recherche d'information annotée avec une ancienne version d'une ontologie
- Idée: Ajouter les informations provenant de l'évolution des ontologies au niveau des requêtes

CONCLUSION

- Importance de considérer des annotations capables de bien évoluer au cours du temps
- Une approche pour la maintenance des annotations sémantiques:
 - A base de règle,
 - Exploitant les données du Web Sémantique
 - Utilisant les patrons de changement
- Perspectives:
 - Conceptualisation de l'approche indirecte pour la maintenance des annotations sémantiques
 - Utilisation des techniques d'apprentissage pour optimiser les processus de génération et d'adaptation des annotations

RÉFÉRENCES

- Cardoso, S. D., Pruski, C., Silveira, M. D., Lin, Y., Groß, A., Rahm, E., and Reynaud-Delaître, C. (2016). Leveraging the impact of ontology evolution on semantic annotations. In Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, pages 68–82.
- Eilbeck, K., Moore, B., Holt, C., and Yandell, M. (2009). Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics*, 10(1):67.
- Qin, L. and Atluri, V. (2009). Evaluating the validity of data instances against ontology evolution over the semantic web. *Information and Software Technology*, 51(1):83 – 97.
- Köpke, J. and Eder, J. (2011). Semantic invalidation of annotations due to ontology evolution. In Meersman, R., Dillon, T., Herrero, P., Kumar, A., Reichert, M., Qing, L., Ooi, B.-C., Damiani, E., Schmidt, D., White, J., Hauswirth, M., Hitzler, P., and Mohania, M., editors, *On the Move to Meaningful Internet Systems: OTM 2011*, volume 7045 of Lecture Notes in Computer Science, pages 763–780. Springer Berlin Heidelberg.
- Zavalina, O. L., Kizhakkethil, P., Alemneh, D. G., Phillips, M. E., and Tarver, H. (2015). Building a framework of metadata change to support knowledge management. *Journal of Information & Knowledge Management*, 14(01):1550005.
- Maynard, D., Peters, W., and Sabou, M. (2007). Change management for metadata evolution.
- Auer, S. and Herre, H. (2007). A Versioning and Evolution Framework for RDF Knowledge Bases, pages 55–69. Springer Berlin Heidelberg, Berlin, Heidelberg.

RÉFÉRENCES

- Burger, T., Morozova, O., Zaihrayeu, I., Andrews, P., and Pane, J. (2010). Report on methods and algorithms for linking user-generated semantic annotations to semantic web and supporting their evolution in time.
- Abgaz, Y. M. (2013). Change impact analysis for evolving ontology-based content management. PhD thesis, Dublin City University.
- Luong, P.-H. and Dieng-Kuntz, R. (2006). A rule-based approach for semantic annotation evolution in the coswem system. In *Canadian Semantic Web*, volume 2 of *Semantic Web and Beyond*, pages 103–120. Springer US.
- Tissaoui, A., Aussenac-Gilles, N., Hernandez, N., and Laublet, P. (2011). Evonto - joint evolution of ontologies and semantic annotations. In Dietz, J., editor, *International Conference on Knowledge Engineering and Ontology Development (KEOD)*, Paris, 26/10/2011-29/10/2011, pages 226–231.
- Park, Y. R., Kim, J., Lee, H. W., Yoon, Y. J., and Kim, J. H. (2011). Gochase-ii: correcting semantic inconsistencies from gene ontology-based annotations for gene products. *BMC Bioinformatics*, 12(1):1–7.
- Frost, H. R. and Moore, J. H. (2014). Optimization of gene set annotations via entropy minimization over variable clusters (emvc). *Bioinformatics (Oxford, England)*, 30(12):1698–1706.

QUESTIONS



Cedric.pruski@list.lu

www.elisa-project.lu