

Textes cliniques : produire, traiter et exploiter les données

Natalia Grabar

CNRS, Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France
natalia.grabar@univ-lille.fr

13/06/2022, Atelier Langage et Santé - JEP 2022, l'île de Noirmoutier

Plan

- Introduction
- Approches et méthodes
- Thématiques
- Conclusion

Contexte

- **Domaine médical :**
 - un grand producteur de données
- **Types de données :**
 - structurées : formulaires, ressources terminologiques, tableaux...
 - non structurées : texte
 - résumés de sortie, comptes-rendus de consultation, d'imagerie ou d'hospitalisation, comptes-rendus d'imagerie, etc.
 - orales (dictée vocale, appels SAMU/hôpitaux, enregistrements de patients...)
 - imagerie
 - produits chimiques, pharmacie
 - séquences de gènes
 - ...
- **Dimensions de données :**
 - médicales : processus de soins, contexte clinique
 - de santé : pour le patient

Informatisation des hôpitaux

- Systématisation lors de la création et collecte
- Stockage
 - entrepôts de données : I2B2, Agfa Healthcare, Dedalus France...
- Unité : patient
- Accumulation des données
- Enrichissement
 - sources de données
 - ressources terminologiques (normalisation, interopérabilité...)
 - données liées
 - ...
- Possibilité de leur exploitation pour la recherche
 - aspects légaux
 - aspects éthiques
 - ouverture vers le contexte interdisciplinaire

Centralité du patient

- HPST : hôpital, patient, santé, territoire
 - loi de 2009
- Prise de décisions
- Patient : propriétaire des données
- Consultation, compréhension

Besoins

- Professionnels de santé :
 - accéder rapidement à une information donnée
 - différents niveaux (document, paragraphe, phrase...)
 - besoins principaux dans le traitement de l'information :
 - trouver
 - extraire
 - catégoriser
 - structurer
 - lier...
- Patients :
 - trouver l'information correcte
 - comprendre l'information
 - [Williams *et al.*, 1995, Patel *et al.*, 2002, Jucks & Bromme, 2007]
 - besoins principaux dans le traitement de l'information :
 - rechercher/trouver
 - comprendre

Objectif

- Proposer une vue de travaux de recherche autour des données médicales :
 - non structurées
 - textes écrits
 - à destination des professionnels de santé
- Approches et méthodes
- Thématiques et applications

Approches et méthodes

- Recherche d'information
- Extraction d'information
- Reconnaissance vocale
- Similarité sémantique
- Résumé automatique
- Questions-Réponses
- Simplification
- ...

Systèmes basés sur des règles et par apprentissage
BERTisation

Approches et méthodes

BERTisation :

- fine-tuning de modèles existants

[Flamholz *et al.*, 2022, Noh & Kavuluru, 2021, Bear Don't Walk Iv *et al.*, 2021, Michalopoulos *et al.*, 2021, Liu *et al.*, 2021a, Amir *et al.*, 2021]

- adaptation au domaine [Wang *et al.*, 2021, Liao *et al.*, 2021]

- transfer learning

[Legrand *et al.*, 2021, Hussain *et al.*, 2021, Bear Don't Walk Iv *et al.*, 2021]

- self-training [Liao *et al.*, 2021]

- enrichissement avec des informations :

- orthographe et lexique [Ding *et al.*, 2021]
- classes syntaxico-sémantiques [Majewska *et al.*, 2021]
- connaissances médicales [Roy & Pan, 2021]

- subword embeddings [Lauriola *et al.*, 2021, Kim *et al.*, 2021b]

- vector retrofitting [Ding *et al.*, 2021, Majewska *et al.*, 2021]

Ressources terminologiques

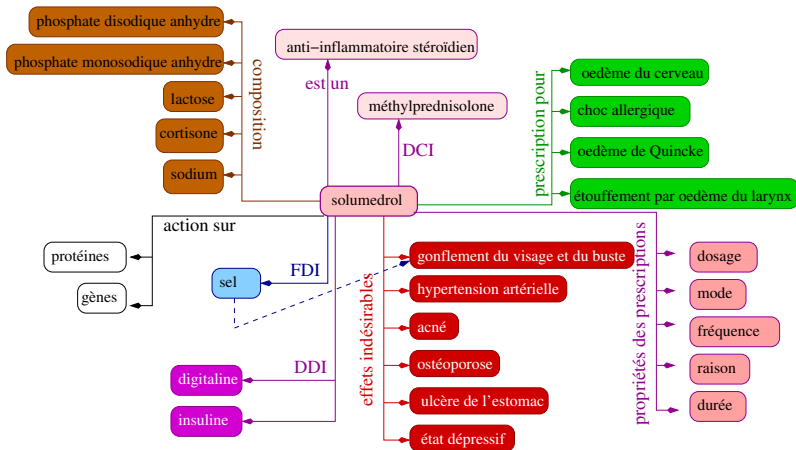
Base pour plusieurs thématiques et applications

- extraction de termes
- structuration de termes
- fusion de ressources terminologiques [Lindberg *et al.*, 1993]
- mise à jour, adaptation, enrichissement
- données liées : apprentissage supervisé distant
[Riedel *et al.*, 2010, Surdeanu *et al.*, 2012, Mintz *et al.*, 2009]
- transfert
 - domaines [Blitzer *et al.*, 2006, Jiang & Zhai, 2007, Plank & Moschitti, 2013, Legrand *et al.*, 2021]
 - langues [Hamon & Grabar, 2016]

Applications :

- codage PMSI de l'activité des hôpitaux
[Ruch *et al.*, 2008, Zhou *et al.*, 2021, Liu *et al.*, 2021c, Dong *et al.*, 2021]
- interopérabilité sémantique [Degoulet *et al.*, 1997, Walker *et al.*, 2005, DGME, 2009, Stroetmann *et al.*, 2009, Wajsbürt *et al.*, 2021]

Extraction d'information



● recrutement de patients pour les essais cliniques

[Oleynik *et al.*, 2019, Percha *et al.*, 2021, Du *et al.*, 2021, Liu *et al.*, 2021b]

Maladies neurodégénératives

Détection de patients atteints (AZH) :

- données orales (entretiens, interviews)
- transcription
- analyse du discours :
 - interaction verbale [Boyé *et al.*, 2014]
 - répétitions, disfluences [Duong *et al.*, 2003, Berrewaerts *et al.*, 2003, Ska & Duong, 2005, Lee, 2011, Boyé *et al.*, 2014]
 - syntaxe [Kynette & Kemper, 1986, Kemper *et al.*, 1990, Nef & Hupet, 1992]
 - lexique [Kynette & Kemper, 1986, Boyé *et al.*, 2014]
 - sémantique [Croisile *et al.*, 1996, Rousseau *et al.*, 2009]
 - pragmatique [Ska & Duong, 2005, Lee, 2011, Gaspers *et al.*, 2012]

⇒ détection à des stades précoces de la maladie

Santé mentale

- Maladies neurologiques et psychiatriques
 - COVID-19 et phases de confinement
- Thématiques récentes :
 - amélioration du diagnostic [Shiner *et al.*, 2021]
 - évolution temporelle des phases de la maladies [Viani *et al.*, 2021]
 - idées suicidaires, auto-mutilation [Rozova *et al.*, 2022, Cliffe *et al.*, 2021]
 - mésusage de substances [Cameron *et al.*, 2013, Kalyanam *et al.*, 2017, Bigeard *et al.*, 2019, Cox *et al.*, 2021]
 - assistance en ligne [Leung *et al.*, 2021, Hassan *et al.*, 2021]
- Provenance des données :
 - réseaux sociaux
 - dossiers cliniques

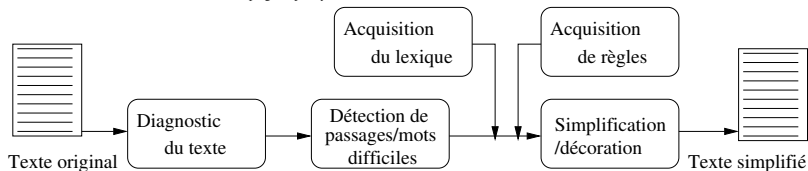
Caractériser l'urgence

- Appels du SAMU :
 - des milliers d'heures d'enregistrements
 - signal audio
- Aider à réguler les appels :
 - gravité et urgence [Kim *et al.*, 2021a]
- Indicateurs :
 - lexique
 - intonation, prosodie
 - qualité de la voix
 - ...
- Difficultés :
 - différencier les rôles
 - transcription
 - bruitage, accents...
 - annotation

Simplification

hypotension – diminution de la tension
myocarde – muscle du coeur
blépharospasme – mouvement involontaire de la pupille
acinésie – incapacité de faire certains mouvements
monoplégie – paralysie d'un seul membre

Alignement de phrases
Similarité sémantique



Modèles

lexique, morphologie,
syntaxe, structure
contexte
oculométrie

- une **hypotension artérielle** peut être observée en cas d'administration intraveineuse trop rapide, inférieure à 60 minutes (voir rubrique 4.2)
- une **diminution de la tension artérielle** peut être observée en cas d'administration intraveineuse trop rapide, inférieure à 60 minutes

[Koptient & Grabar, 2020, Cardon & Grabar, 2020a, Cardon & Grabar, 2020b]

Difficultés

- Disponibilité et accessibilité des données :
 - accès réglementé et difficile
 - corpus de cas cliniques [Grabar *et al.*, 2020]
- Corpus et données annotés :
 - documents cliniques
 - fiabilité des annotations
 - expertise des annotateurs
 - accord, consensus

Difficultés

- Données rares :
 - maladies rares (1/2 000, 1/10 000, 1/100 000...)
- Données incomplètes :
 - histoire de la maladie, analyses de laboratoire, prescriptions...
- Données incertaines [Simianu *et al.*, 2016, Reiner, 2018] :
 - incertitude scientifique
 - incertitude diagnostique
- Données distribuées :
 - entre différentes sources ou différents documents
- Fiabilité :
 - sources
 - modèles, résultats
- Explicabilité :
 - pour la prise de décisions

Conclusion

- Données de différents types
- Besoin de collaborations pour les traiter et valoriser
- Besoin de gros volumes de données
 - approches modernes



AMIR, S., VAN DE MEENT, J.-W. & WALLACE, B. C. (2021).

On the impact of random seeds on the fairness of clinical classifiers.

In *Proc of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, pp. 3808–3823 : Association for Computational Linguistics.



BEAR DON'T WALK IV, O. J., SUN, T., PEROTTE, A. & ELHADAD, N. (2021).

Clinically relevant pretraining is all you need.

J Am Med Inform Assoc, 28(9), 1970–1976.



BERREWAERTS, J., HUPET, M. & FEYEREISEN, P. (2003).

Langage et démente : examen des capacités pragmatiques dans la maladie d'Alzheimer.

Revue de Neuropsychologie, 13(2), 165–207.



BIGEARD, E., THIESSARD, F. & GRABAR, N. (2019).

Detecting drug non-compliance in internet fora using information retrieval and machine learning approaches.

In *MEDINFO 2019*, pp. 1–6.



BLITZER, J., McDONALD, R. & PEREIRA, F. (2006).

Domain adaptation with structural correspondence learning.

In *Proc of the 2006 Conf on Empirical Methods in Natural Language Processing*, pp. 120–128 : Association for Computational Linguistics.



BOYÉ, M., TRAN, T. & GRABAR, N. (2014).

Nlp-oriented contrastive study of linguistic productions of alzheimer and control people.

In *POLTAL*, pp. 412–424 : Springer, Advances in Natural Language Processing, LNCS 8686.



CAMERON, D., SMITH, G. A., DANIULAITYTE, R., SHETH, A. P., DAVE, D., CHEN, L., ANAND, G.,

CARLSON, R., WATKINS, K. Z. & FALCK, R. (2013).

PREDOSE : A semantic web platform for drug abuse epidemiology using social media.

Journal of Biomedical Informatics, 46, 985–997.



CARDON, R. & GRABAR, N. (2020a).

Construction d'un corpus parallèle à partir de corpus comparables pour la simplification de textes médicaux en français.

volume 61, pp. 15–39.



CARDON, R. & GRABAR, N. (2020b).

French biomedical text simplification : When small and precise helps.
In *COLING 2020*, pp. 1–8.



CLIFFE, C., SEYEDSALEHI, A., VARDAVOULIA, K., BITTAR, A., VELUPILLAI, S., SHETTY, H., SCHMIDT, U. & DUTTA, R. (2021).

Using natural language processing to extract self-harm and suicidality data from a clinical sample of patients with eating disorders : a retrospective cohort study.
BMJ Open, 11(12), e053808.



COX, D. J., GARCIA-ROMEU, A. & JOHNSON, M. W. (2021).

Predicting changes in substance use following psychedelic experiences : natural language processing of psychedelic session narratives.
Am J Drug Alcohol Abuse, 47(4), 444–454.



CROISILE, B., SKA, B., BRABANT, M., DUCHENE, A., LEPAGE, Y., AIMARD, G. & TRILLET, M. (1996).

Comparative study of oral and written picture description in patients with Alzheimer's disease.
Brain and language, 53, 1–19.



DEGOULET, P., FIESCHI, M. & ATTALI, C. (1997).

Les enjeux de l'interopérabilité sémantique dans les systèmes d'information de santé.
Informatique et Santé, 9, 203–12.



DGME (2009).

Référentiel général de l'interopérabilité. RGI.
Technical report, Direction générale de la modernisation de l'état.



DING, X., MOWER, J., SUBRAMANIAN, D. & COHEN, T. (2021).

Augmenting aer2vec : Enriching distributed representations of adverse event report data with orthographic and lexical information.
Journal of Biomedical Informatics, 119, 103833.



DONG, H., SUÁREZ-PANIAGUA, V., WHITLEY, W. & WU, H. (2021).

Textes cliniques : produire, traiter et exploiter les données

Explainable automated coding of clinical notes using hierarchical label-wise attention networks and label embedding initialisation.

J Biomed Inform, **116**, 103728.



DU, J., WANG, Q., WANG, J., RAMESH, P., XIANG, Y., JIANG, X. & TAO, C. (2021).

COVID-19 trial graph : a linked graph for covid-19 clinical trials.

J Am Med Inform Assoc, **28**(9), 1964–1969.



DUONG, A., TARDIF, A. & SKA, B. (2003).

Discourse about discourse : What is it and how does it progress in Alzheimer's disease ?

Brain and cognition, **53**, 177–180.



FLAMHOLZ, Z. N., CRANE-DROESCH, A., UNGAR, L. H. & WEISSMAN, G. E. (2022).

Word embeddings trained on published case reports are lightweight, effective for clinical tasks, and free of protected health information.

J Biomed Inform, **125**, 103971.



GASPERS, J., THIELE, K., CIMIANO, P., FOLTZ, A., STENNEKEN, P. & TSCHEREPANOW, M. (2012).

An evaluation of measures to dissociate language and communication disorders from healthy controls using machine learning techniques.

In *IHI 2012*, pp. 209–218.



GRABAR, N., DALLOUX, C. & CLAVEAU, V. (2020).

CAS : corpus of clinical cases in French.

Journal of BioMedical Semantics, **11**(1), 1–7.



HAMON, T. & GRABAR, N. (2016).

Adaptation of cross-lingual transfer methods for the building of medical terminology in ukrainian.

In *Computational Linguistics and Intelligent Text Processing*, pp. 1–12.



HASSAN, A., ALI, M. D. I., AHAMMED, R., BOUROUIS, S. & KHAN, M. M. (2021).

Development of NLP-integrated intelligent web system for e-mental health.

Comput Math Methods Med, **1**, 1546343.



HUSSAIN, M., SATTI, F. A., HUSSAIN, J., ALI, T., ALI, S. I., BILAL, H. S. M., PARK, G. H., LEE, S. & CHUNG, T. (2021).

A practical approach towards causality mining in clinical text using active transfer learning.
J Biomed Inform, 123, 103932.



JIANG, J. & ZHAI, C. (2007).

Instance weighting for domain adaptation in NLP.
In *Proc of the 45th Ann Meeting of the Assoc of Comp Linguistics*, pp. 264–271 : Association for Computational Linguistics.



JUCKS, R. & BROMME, R. (2007).

Choice of words in doctor-patient communication : an analysis of health-related internet sites.
Health Commun, 21(3), 267–77.



KALYANAM, J., KATSUKI, T., LANCKRIET, G. R. G. & MACKEY, T. K. (2017).

Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the twittersphere using unsupervised machine learning.
Addictive Behaviors, 65, 289–295.



KEMPER, S., RASH, S., KYNETTE, D. & NORMAN, S. (1990).

Telling stories : The structure of adults' narratives.
European Journal of Cognitive Psychology, 2, 205–228.



KIM, D., OH, J., IM, H., YOON, M., PARK, J. & LEE, J. (2021a).

Automatic classification of the korean triage acuity scale in simulated emergency rooms using speech recognition and natural language processing : a proof of concept study.
J Korean Med Sci, 36(27), e175.



KIM, T., HAN, S. W., KANG, M., LEE, S. H., KIM, J.-H., JOO, H. J. & SOHN, J. W. (2021b).

Similarity-based unsupervised spelling correction using biowordvec : Development and usability study of bacterial culture and antimicrobial susceptibility reports.
JMIR Med Inform, 9(2), e25530.



KOPTIENT, A. & GRABAR, N. (2020).

Rated lexicon for the simplification of medical texts.

In *Proc of HEALTHINFO 2020*, pp. 1–6.



KYNETTE, D. & KEMPER, S. (1986).

Aging and the loss of grammatical forms : A cross-sectional study of language performance.
Language and Communication, 6, 65–72.



LAURIOLA, I., AIOLLI, F., LAVELLI, A. & RINALDI, F. (2021).

Learning adaptive representations for entity recognition in the biomedical domain.
J Biomed Semant, 12(10), 1–16.



LEE, H. (2011).

Viellissement normal et maladie d'Alzheimer : analyse comparative de la narration semi-dirigée au niveau lexical.

In *Méthodes et analyses comparatives en sciences du langage*.



LEGRAND, J., TOUSSAINT, Y., RAÏSSI, C. & COULET, A. (2021).

Syntax-based transfer learning for the task of biomedical relation extraction.
J Biomed Semant, 12(16).



LEUNG, Y. W., WOUTERLOOT, E., ADIKARI, A., HIRST, G., DE SILVA, D., WONG, J., BENDER, J. L., GANCARZ, M., GRATZER, D., ALAHAKOON, D. & ESPLIN, M. J. (2021).

Natural language processing-based virtual cofacilitator for online cancer support groups : Protocol for an algorithm development and validation study.

JMIR Res Protoc, 10(1), e21453.



LIAO, S., KIROS, J., CHEN, J., ZHANG, Z. & CHEN, T. (2021).

Improving domain adaptation in de-identification of electronic health records through self-training.
J Am Med Inform Assoc, 28(10), 2093–2100.



LINDBERG, D., HUMPHREYS, B. & MCCRAY, A. (1993).

The Unified Medical Language System.

Methods Inf Med, 32(4), 281–291.



LIU, F., SHAREGHI, E., MENG, Z., BASALDELLA, M. & COLLIER, N. (2021a).

Self-alignment pretraining for biomedical entity representations.

In *Proc of the 2021 Conf of the North American Chapter of the Ass for Comp Linguistics : Human Language Technologies*, pp. 4228–4238 : Association for Computational Linguistics.



LIU, H., CHI, Y., BUTLER, A., SUN, Y. & WENG, C. (2021b).

A knowledge base of clinical trial eligibility criteria.

J Biomed Inform, **117**, 103771.



LIU, Y., CHENG, H., KLOPFER, R., GORMLEY, M. R. & SCHAAF, T. (2021c).

Effective convolutional attention network for multi-label clinical document classification.

In *Proc of the 2021 Conf on Empirical Methods in Natural Language Processing*, pp. 5941–5953, Online and Punta Cana, Dominican Republic : Association for Computational Linguistics.



MAJEWSKA, O., COLLINS, C., BAKER, S., BJÖRNE, J., BROWN, S. W., KORHONEN, A. & PALMER, M. (2021).

BioVerbNet : a large semantic-syntactic classification of verbs in biomedicine.

J Biomed Semantics, **12**(1), 12.



MICHALOPOULOS, G., WANG, Y., KAKA, H., CHEN, H. & WONG, A. (2021).

UmlsBERT : Clinical domain knowledge augmentation of contextual embeddings using the Unified Medical Language System Metathesaurus.

In *Proc of the 2021 Conf of the North American Chapter of the Ass for Comp Linguistics : Human Language Technologies*, pp. 1744–1753 : Association for Computational Linguistics.



MINTZ, M., BILLS, S., SNOW, R. & JURAFSKY, D. (2009).

Distant supervision for relation extraction without labeled data.

In *Proc of the Joint Conf of the 47th Ann Meeting of the ACL and the 4th Int Joint Conf on Natural Language Processing of the AFNLP*, pp. 1003–1011, Suntec, Singapore : Association for Computational Linguistics.



NEF, F. & HUPET, M. (1992).

Les manifestations du vieillissement normal dans le langage spontané oral et écrit.

L'année psychologique, **9**(3), 393–419.



NOH, J. & KAVULURU, R. (2021).

Improved biomedical word embeddings in the transformer era.
J Biomed Inform, **120**, 103867.



OLEYNIK, M., KUGIC, A., KASÁČ, Z. & KREUZTHALER, M. (2019).

Evaluating shallow and deep learning strategies for the 2018 n2c2 shared task on clinical text classification.
J Am Med Inform Assoc, **26**(11), 1247–1254.



PATEL, V., BRANCH, T. & AROCHA, J. (2002).

Errors in interpreting quantities as procedures : The case of pharmaceutical labels.
Int Journ Med Inform, **65**(3), 193–211.



PERCHA, B., PISAPATI, K., GAO, C. & SCHMIDT, H. (2021).

Natural language inference for curation of structured clinical registries from unstructured text.
J Am Med Inform Assoc, **29**(1), 97–108.



PLANK, B. & MOSCHITTI, A. (2013).

Embedding semantic similarity in tree kernels for domain adaptation of relation extraction.
In Proc of the 51st Ann Meeting of the Assoc for Comp Linguistics, pp. 1498–1507.



REINER, B. (2018).

Quantitative analysis of uncertainty in medical reporting : Creating a standardized and objective methodology.
J Digit Imaging, **31**(2), 145–149.



RIEDEL, S., YAO, L. & MCCALLUM, A. (2010).

Modeling relations and their mentions without labeled text.
In Proc of ECML PKDD, pp. 148–163.



ROUSSEAU, T., DE SAINT-ANDRÉ, A. & GATIGNOL, P. (2009).

Évaluation pragmatique de la communication des personnes âgées saines.
Neurologie Psychiatrie Gériatrie, **9**, 271–280.



ROY, A. & PAN, S. (2021).

Textes cliniques : produire, traiter et exploiter les données

Incorporating medical knowledge in BERT for clinical relation extraction.

In *Proc of the 2021 Conf on Empirical Methods in Natural Language Processing*, pp. 5357–5366, Online and Punta Cana, Dominican Republic : Association for Computational Linguistics.



ROZOVA, V., WITT, K., ROBINSON, J., LI, Y. & VERSPOOR, K. (2022).

Detection of self-harm and suicidal ideation in emergency department triage notes.

J of the Am Med Inform Ass, **29**(3), 472–480.



RUCH, P., GOBEIL, J., TBAHRITI, I. & GEISSBÜHLER, A. (2008).

From episodes of care to diagnosis codes : automatic text categorization for medico-economic coding.

In *Ann Symp Am Med Inform Assoc (AMIA)*, pp. 636–40.



SHINER, B., LEVIS, M., DUFORT, V. M., PATTERSON, O. V., WATTS, B. V., DUVALL, S. L., RUSS, C. J. & MAGUEN, S. (2021).

Improvements to PTSD quality metrics with natural language processing.

J Eval Clin Pract.



SIMIANU, V. V., GROUNDS, M. A., JOSLYN, S. L., LECLERC, J. E., EHLERS, A. P., AGRAWAL, N., ALFONSO-CRISTANCHO, R., FLAXMAN, A. D., & FLUM, D. R. (2016).

Understanding clinical and non-clinical decisions under uncertainty : a scenario-based survey.

BMC Medical Informatics and Decision Making, **16**(153), 1–9.



SKA, B. & DUONG, A. (2005).

Communication, discours et démençe.

Psychologie et NeuroPsychiatrie du Vieillessement, **3**(2), 125–133.



STROETMANN, V. N., KALRA, D., LEWALLE, P., RECTOR, A., RODRIGUES, J.-M., STROETMANN, K. A., SURJAN, G., USTUN, B., VIRTANEN, M. & ZANSTRA, P. E. (2009).

Semantic interoperability for better health and safer healthcare.

Technical report, European commission, Information society and media.



SURDEANU, M., TIBSHIRANI, J., NALLAPATI, R. & MANNING, C. D. (2012).

Multi-instance multi-label learning for relation extraction.

In *Proc of the 2012 Joint Conf on Empirical Methods in Natural Language Processing and Comp Natural Language Learning*, pp. 455–465, Jeju Island, Korea : Association for Computational Linguistics.



VIANI, N., BOTELLE, R., KERWIN, J., YIN, L., PATEL, R., STEWART, R. & VELUPILLAI, S. (2021).
A natural language processing approach for identifying temporal disease onset information from mental healthcare text.
Sci Rep, **11**(1), 757.



WAJSBÜRT, P., SARFATI, A. & TANNIER, X. (2021).
Medical concept normalization in French using multilingual terminologies and contextual embeddings.
J Biomed Inform, **114**, 103684.



WALKER, J., PAN, E., JOHNSTON, D., ADLER-MILSTEIN, J., BATES, D. W. & MIDDLETON, B. (2005).
The Value of Healthcare Information Exchange and Interoperability.
Technical report, US Health Affairs.



WANG, J., ABU-EL-RUB, N., GRAY, J., PHAM, H. A., ZHOU, Y., MANION, F. J., LIU, M., SONG, X., XU, H., ROUHIZADEH, M. & ZHANG, Y. (2021).
COVID-19 SignSym : a fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to omop common data model.
J Am Med Inform Assoc, **28**(6), 1275–1283.



WILLIAMS, M., PARKER, R., BAKER, D., PARIKH, N., PITKIN, K., COATES, W. & NURSS, J. (1995).
Inadequate functional health literacy among patients at two public hospitals.
JAMA, **274**(21), 1677–1682.



ZHOU, T., CAO, P., CHEN, Y., LIU, K., ZHAO, J., NIU, K., CHONG, W. & LIU, S. (2021).
Automatic ICD coding via interactive shared representation networks with self-distillation mechanism.
In *ACL, Ed., Proc of the 59th Ann Meeting of the Assoc for Comp Linguistics and the 11th Inter Joint Conf on Natural Language Processing*, pp. 5948–5957.