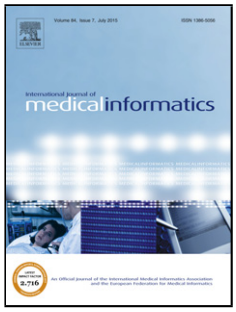
Accepted Manuscript



Title: Accuracy of Using Natural Language Processing Methods for Identifying Healthcare-associated Infections

Authors: Nastassia Tvardik, Ivan Kergourlay, André Bittar, Frédérique Segond, Stefan Darmoni, Marie-Hélène Metzger

PII: S1386-5056(18)30436-2
DOI: https://doi.org/10.1016/j.ijmedinf.2018.06.002
Reference: IJB 3710

To appear in: *International Journal of Medical Informatics*

Received date: 18-12-2017
Revised date: 27-4-2018
Accepted date: 3-6-2018

Please cite this article as: Tvardik N, Kergourlay I, Bittar A, Segond F, Darmoni S, Metzger M-Hélène, Accuracy of Using Natural Language Processing Methods for Identifying Healthcare-associated Infections, *International Journal of Medical Informatics* (2018), https://doi.org/10.1016/j.ijmedinf.2018.06.002

Accuracy of Using Natural Language Processing Methods for Identifying Healthcare-associated Infections

Nastassia Tvardik[a], Ivan Kergourlay[b], André Bittar[c], Frédérique Segond[d], Stefan Darmoni[e,f], Marie-Hélène Metzger[a,g*]

[a] Université de Lyon, F-69000, Lyon ; Université Lyon 1 ; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France (metzger.marie-helene@orange.fr)

[b] University Hospital of Rouen, Department of Biomedical Informatics, CISMeF, Rouen, France. (ivan.kergourlay@chu-rouen.fr)

[c] Holmes Semantic Solutions, Grenoble, France (andre.bittar@gmail.com)

[d] Viseo Technologies, Grenoble, France (frederique.segond@viseo.com)

[e] TIBS, LITIS EA 4108, Normandy University, France (stefan.darmoni@chu-rouen.fr)

[f] INSERM, U1142, LIMICS, Paris, France

[g] Hospices Civils de Lyon, Hôpital de la Croix-Rousse, Unité d'hygiène et d'épidémiologie, F-69317, Lyon, France

Corresponding author
* Marie-Hélène METZGER,
UMR UCBL - CNRS 5558
HCL, Unité d'hygiène et d'épidémiologie
Hôpital de la Croix-Rousse
103 Grande-Rue de la Croix-Rousse
F-69317 Lyon cedex 04
France
Tel: 00 33 472 07 19 82 - Fax: 00 33 472 07 19 85
E-mail : metzger.marie-helene@orange.fr

Highlights

- This study confirmed the feasibility of using natural language processing for detection of healthcare-associated infections (HAI)
- The performances of HAI detection were satisfactory: overall sensitivity of 83.9% and specificity of 84.2%
- The most frequent cause of medical record misclassification was an inaccurate temporal labeling of medical events
- Improvements in semantic analysis algorithms would increase the detection performance
- NLP could offer a new standardized case-finding process for the HAI monitoring in hospitals