



École d'été interdisciplinaire en numérique de la santé du 10 au 14 juillet 2023

Entrepôt de données hospitalier pour un système de santé apprenant

Plan d'activité pédagogique

Type de l'activité : ☐ Atelier (2 h 20 min) ☒ Présentation (45 min + 5 min de questions)

Objectif : Le présent document décrit le contenu scientifique et l'organisation de l'activité pédagogique « Entrepôt de données hospitalier pour un système de santé apprenant » présentée lors de l'école d'été.

1 Introduction

L'activité pédagogique « Entrepôt de données hospitalier pour un système de santé apprenant » se déroulera dans le cadre de l'école d'été interdisciplinaire numérique de la santé (EINS) et s'inscrit dans le thème « Analyse des données en santé ».

Cet évènement scientifique s'adresse aux personnes effectuant des études universitaires, aux personnes professionnelles ainsi qu'aux patientes partenaires et patients partenaires qui désirent s'initier aux défis de mise en place de projets en numérique de la santé.

1.1 Renseignements sur le présentateur

Prénom, nom : Nicolas Garcelon

Affiliation principale :

Responsable de la plateforme Data science de l'Institut des maladies génétiques Imagine

Courriel : nicolas.garcelon@institutimagine.org

Site Web : <https://www.institutimagine.org/fr/nicolas-garcelon-757>

Biographie

Nicolas Garcelon a un diplôme d'ingénieur agronome ainsi qu'un doctorat de santé publique en informatique biomédicale. Il a travaillé au Département d'information médicale du Centre hospitalier de Rennes et dirige maintenant la plateforme *Data science* de l'Institut Imagine. Il est aussi chercheur dans l'équipe INSERM (*Information Sciences to support Personalized Medicine*) au Centre de recherche des Cordeliers à Paris. Il s'intéresse particulièrement à l'intrication entre la recherche et les soins. Ses travaux portent sur la conception et le développement de logiciels hospitaliers pour les cliniciens et les chercheurs afin de faciliter la réutilisation des données

médicales. Il a notamment créé Dr Warehouse, déposé sous licence à code source libre en septembre 2017.

2 Description

Cette section présente le contenu de l'activité et les principaux objectifs.

2.1 Contenu

L'objectif de cette présentation est de décrire les travaux menés à l'hôpital Necker-Enfants malades à Paris autour des maladies rares et la création d'un entrepôt de données hospitalier (Dr Warehouse) pour à la fois la recherche et le soin. Nous montrerons dans un premier temps la complexité des typologies de données hospitalières et la nécessité de ne pas se limiter aux données structurées, de considérer également les données non structurées (comptes rendus hospitaliers) plus complètes et plus riches. L'intégration de ces données, en raison de leur nature même, nécessite des développements particuliers au niveau du traitement des données en appliquant des méthodes de traitement du langage naturel, des outils de visualisation et de validation. Puis, nous présenterons les quatre principales fonctionnalités permettant la réutilisation de ces données pour faciliter la prise en charge des patients et accélérer la recherche : un moteur de recherche multimodal (pour les tests de faisabilité et le recrutement des patients); un outil de fouille de données (génération d'hypothèse et exploration des données); un outil d'aide au diagnostic par similarité entre patients; et, un outil d'alimentation de registre. L'objectif est aussi de montrer l'importance de conserver la contextualisation des données restructurées. Cette présentation s'applique à toutes les personnes qui travailleront sur les données hospitalières ou sur les registres médicaux. Il s'agit ici de montrer l'importance de conserver les données brutes tout au long du processus de collecte pour tirer le maximum d'information de ces données.

La présentation illustrera les cas d'utilisation dans le contexte des maladies rares ; la possibilité de retrouver des patients en errance diagnostique grâce à l'analyse automatique des comptes rendus sera présentée.

2.2 Objectifs de formation

Cette activité permettra à une personne participante de :

- O1. Comprendre la notion d'entrepôt de données hospitalier à partir d'un cas concret.
- O2. Voir un exemple d'utilisation de méthodes d'apprentissage profond pour le traitement du langage naturel dans un hôpital et des cas d'utilisation y afférents.
- O3. D'appréhender l'importance des données textuelles dans le contexte des maladies rares.

3 Références

Cette section présente les principales références documentaires utilisées pour construire l'activité et les références pour approfondir des concepts présentés.

3.1 Références essentielles

- N. Garcelon, A. Neuraz, R. Salomon, H. Faour, V. Benoit, A. Delapalme, A. Munnich, A. Burgun, B. Rance, A clinician friendly data warehouse oriented toward narrative reports: Dr Warehouse, J Biomed Inform. (2018). <https://doi.org/10.1016/j.jbi.2018.02.019>
- N. Garcelon, A. Neuraz, R. Salomon, N. Bahi-Buisson, J. Amiel, C. Picard, N. Mahlaoui, V. Benoit, A. Burgun, B. Rance, Next generation phenotyping using narrative reports in a rare disease clinical data warehouse, Orphanet J Rare Dis. 13 (2018). <https://doi.org/10.1186/s13023-018-0830-6>

- N. Garcelon, A. Neuraz, V. Benoit, R. Salomon, S. Kracker, F. Suarez, N. Bahi-Buisson, S. Hadj-Rabia, A. Fischer, A. Munnich, A. Burgun, Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse: Dr. Warehouse and the needle in the needle stack, *Journal of Biomedical Informatics*. 73 (2017) 51–61.
<https://doi.org/10.1016/j.jbi.2017.07.016>
- M. Vincent, M. Douillet, I. Lerner, A. Neuraz, A. Burgun, N. Garcelon, Using deep learning to improve phenotyping from clinical reports, *Studies in Health Technology and Informatics*. Accepted (2021).
- X. Chen, C. Faviez, M. Vincent, L. Briseño-Roa, H. Faour, J.-P. Annereau, S. Lyonnet, M. Zaidan, S. Saunier, N. Garcelon, A. Burgun, Patient-Patient Similarity-Based Screening of a Clinical Data Warehouse to Support Ciliopathy Diagnosis, *Frontiers in Pharmacology*. 13 (2022).
<https://www.frontiersin.org/article/10.3389/fphar.2022.786710> (accessed March 25, 2022).
- C. Faviez, M. Vincent, N. Garcelon, C. Michot, G. Baujat, V. Cormier-Daire, S. Saunier, X. Chen, A. Burgun, Enriching UMLS-Based Phenotyping of Rare Diseases Using Deep-Learning: Evaluation on Jeune Syndrome, Challenges of Trustable AI and Added-Value on Health. (2022) 844–848. <https://doi.org/10.3233/SHTI220604>

3.2 Références complémentaires

- Soldaini L, Goharian N. QuickUMLS: a fast, unsupervised approach for medical concept extraction. 2016;4.
- Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010 Oct;17(5):507–13.
- Luo L, Yan S, Lai P-T, Veltri D, Oler A, Xirasagar S, et al. PhenoTagger: A Hybrid Method for Phenotype Concept Recognition using Human Phenotype Ontology. *Bioinformatics*. 2021 Jan 20;btab019
- Habibi M, Weber L, Neves M, Wiegandt DL, Leser U. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics*. 2017 Jul 15;33(14):i37–48.
- Hully, M., Lo Barco, T., Kaminska, A., Barcia, G., Cances, C., Mignot, C., et al. (2021). Deep Phenotyping Unstructured Data Mining in an Extensive Pediatric Database to Unravel a Common KCNA2 Variant in Neurodevelopmental Syndromes. *Genet. Med*. 23, 968–971.
[doi:10.1038/s41436-020-01039-z](https://doi.org/10.1038/s41436-020-01039-z)