

# Conception et réalisation d'un système de recherche sémantique de maladies

Rapport de projet de session  
GLO-7001 — Conception de systèmes intelligents

Rémi CÉRÈS      Mattéo DELABRE

19 décembre 2019

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Étude du contexte du projet de système de recherche</b>	<b>2</b>
2.1	Intelligence artificielle en médecine et projet Wikimédica . . . . .	2
2.2	Tâche de recherche sémantique de maladies . . . . .	3
2.3	Médecin et base de connaissances, agents du système . . . . .	4
<b>3</b>	<b>Conception du système de recherche basé sur les données de Wikimédica et Wikidata</b>	<b>4</b>
3.1	Modèle et sources de connaissances médicales . . . . .	4
3.2	Interactions entre le médecin et le système de recherche . . . . .	6
<b>4</b>	<b>Réalisation d'un prototype du système</b>	<b>8</b>
4.1	Stratégie de récupération automatique des connaissances . . . . .	8
4.2	Alignement des connaissances récupérées . . . . .	9
4.3	Stockage durable des connaissances alignées . . . . .	10
4.4	Algorithmes de recherche approximative . . . . .	10
4.5	Interface d'accès frontal aux connaissances . . . . .	11
<b>5</b>	<b>Conclusion</b>	<b>11</b>

## 1 Introduction

Afin de mettre en pratique les connaissances acquises en suivant les cours de conception de systèmes intelligents, nous avons eu l'opportunité de réaliser un projet en rapport avec Wikimédica<sup>1</sup>. Cette plate-forme, lancée en 2017 et gérée par une OBNL, est une ressource basée sur le principe du *wiki*, placée sous licence libre [2], qui rend disponible à toutes et à tous des informations médicales théoriques, cliniques

---

<sup>1</sup><https://wikimedi.ca/>

et pédagogiques. Destinée en priorité aux professionnels de la santé et aux étudiants en médecine, elle leur permet de partager leurs connaissances et d'en améliorer collectivement les données [8]. Sa particularité est d'être un wiki de type sémantique, dont les différentes pages sont connectées par des métadonnées exploitables par un programme informatique. Notre projet consiste à mettre au point un outil permettant d'assister à la réalisation de diagnostics différentiels en naviguant parmi les maladies renseignées dans le wiki sur la base des propriétés sémantiques qui les lient à leurs signes et symptômes.

Ce rapport se compose de trois parties, commençant par une étude du contexte du projet, de ses différents acteurs et actions souhaitées. La suite du rapport s'intéresse à la conceptualisation des connaissances nécessaires à la réalisation du projet et aux moyens de communication entre ses agents. Enfin, la dernière partie présente les détails de conception et d'implémentation d'un premier prototype du système.

## 2 Étude du contexte du projet de système de recherche

### 2.1 Intelligence artificielle en médecine et projet Wikimédica

Dans le cadre de leur pratique, les médecins sont régulièrement amenés à prendre des décisions informées relatives à leurs patients, basées sur la considération méticuleuse d'un ensemble d'indices et en s'appuyant sur les consensus scientifiques existants ainsi que sur leur expérience. La quantité d'information à caractère non-structuré à la disposition des professionnels de médecine fait qu'il leur est difficile de s'assurer de prendre en compte en tout temps tous les cas possibles et les dernières bonnes pratiques. Une façon d'exploiter plus efficacement ces données est d'utiliser des machines pour reproduire le raisonnement des médecins sur les problématiques qui leur sont présentées; autrement dit, il s'agit de construire des systèmes d'intelligence artificielle [10]. En médecine, comme dans les autres champs d'application critiques de l'intelligence artificielle, de tels systèmes font face à la contrainte supplémentaire de fiabilité, de transparence et de reddition de comptes. Ces contraintes n'empêchent pas le domaine d'être l'objet de plus en plus de recherches, de projets et financements [3], dans la mesure où ces systèmes ont un large potentiel d'application, avec à l'horizon la réduction du nombre d'erreurs médicales, le recentrage de l'activité des praticiens sur l'écoute et l'accompagnement des patients, ou encore la réduction des consultations inutiles permettant le désengorgement des services médicaux.

L'un des objectifs du projet Wikimédica est de recenser et fournir des données structurées sous la forme d'une ontologie exploitable par des machines. De par cet objectif, ce wiki a pour vocation de devenir une source de données utile à la réalisation d'applications en intelligence artificielle médicale. Une application possible est la réalisation de diagnostics différentiels, une activité du médecin consistant à établir une liste de maladies qui pourraient plausiblement être la cause de signes et de symptômes exprimés par un patient. Cette liste ne doit exclure aucune possibilité, en particulier les plus graves et les plus probables, sous peine de mettre potentiellement en danger le patient et d'engager pénalement la responsabilité du médecin [1].

*Comment les techniques d'intelligence artificielle peuvent-elles être combinées aux données de Wikimédecin pour aider à poser un diagnostic différentiel ?*

L'objectif de notre projet est d'apporter des éléments de réponse à cette question en concevant un système capable d'assister un médecin dans l'établissement de diagnostics différentiels pour ses patients qui lui permette de naviguer parmi les différentes possibilités correspondant aux données dont il dispose sur leur situation.

## 2.2 Tâche de recherche sémantique de maladies

Le médecin utilise le système avec pour but d'identifier une maladie, à savoir une « altération de la santé caractérisée par des causes, des symptômes, une évolution et des possibilités thérapeutiques propres » [5], correspondant à la situation de son patient. Remarquons que les requêtes formulées par le médecin dans le système sont toujours des conjonctions de termes et jamais des disjonctions. La difficulté de cette tâche réside dans le fait que le médecin cherche à obtenir un diagnostic à partir des signes (constatés objectivement sur le patient par un observateur), symptômes (perçus et décrits par le malade), facteurs de risque (variables associées à un risque accru de contracter une maladie) et antécédents de son patient, alors que les connaissances du wiki décrivent plutôt ces informations à partir de chaque maladie. En inversant cette relation, le résultat comportera en effet souvent plusieurs maladies possibles dans la mesure où beaucoup d'entre elles ont des présentations qui se recoupent. Par conséquent, le système doit être interactif : après un retour d'une liste initiale de candidats, le médecin peut choisir de tester des signes supplémentaires ou d'interroger le patient sur la présence éventuelle d'autres symptômes pour éliminer des maladies incompatibles. Après récupération de ces nouvelles données, l'utilisateur soumet une nouvelle requête modifiée au système et la procédure reprend du début. Cette interaction est représentée conceptuellement dans la figure 1.

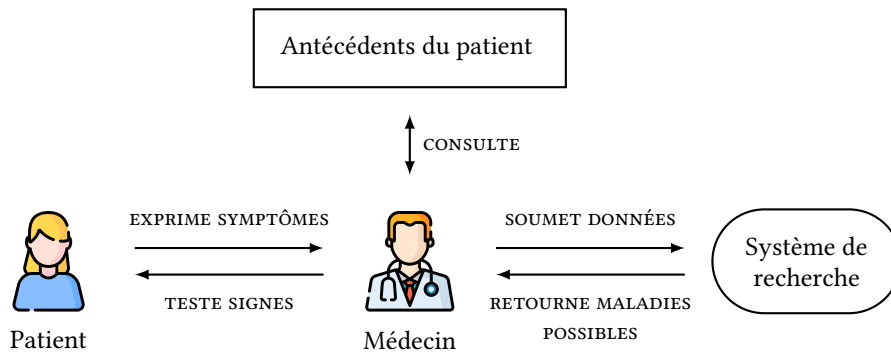


FIG. 1 : Vue conceptuelle de la tâche de recherche sémantique de maladies impliquant le médecin, le patient et le système de recherche.

Wikimédecin offre déjà un formulaire permettant de rechercher parmi les maladies qui sont y référencées<sup>2</sup>. Nous avons utilisé cette page existante comme base pour

<sup>2</sup>[https://wikimedi.ca/wiki/Spécial:RunQuery/Recherche\\_de\\_maladies](https://wikimedi.ca/wiki/Spécial:RunQuery/Recherche_de_maladies)

concevoir notre système.

### 2.3 Médecin et base de connaissances, agents du système

Les bases de connaissances ne constituent pas par elles-mêmes des systèmes de recherche adaptés pour les utilisateurs humains car elles foisonnent de données et ne leur permettent pas de se focaliser précisément sur leur activité de recherche. Le système de recherche de maladies par signes et symptômes se place donc comme un intermédiaire entre le médecin et la base de connaissances offerte par Wikimédica renfermant de la connaissance académique sur différents aspects de la médecine.

Les médecins utilisateurs du système sont des professionnels de santé possédant une formation et des connaissances de base comprenant notamment le vocabulaire médical nécessaire et la capacité de réaliser un diagnostic différentiel. Ils communiquent d'une part avec leurs patients physiquement au travers de consultations, et d'autre part avec le système de recherche en utilisant un ordinateur connecté à Internet.

La base de connaissance est constituée des pages de Wikimédica et de leurs relations telles que maintenues par les contributeurs du projet, qui sont des professionnels de santé reconnus comme tels par le Collège des médecins du Québec.

## 3 Conception du système de recherche basé sur les données de Wikimédica et Wikidata

### 3.1 Modèle et sources de connaissances médicales

Le système doit connaître d'une part des informations sur les maladies que le médecin est susceptible de rechercher, et d'autre part les liens qui existent entre ces maladies et leurs signes, symptômes, facteurs de risque et antécédents éventuels. En outre, puisque le nombre de maladies possibles pour un ensemble de symptômes peut être grand, le système doit orienter le diagnostic sur les éléments les plus importants en priorité. Pour ce faire, l'application doit pouvoir évaluer la gravité et la probabilité d'occurrence chez un patient donné de chaque maladie et trier les résultats par gravité et probabilité décroissantes.

**Sources de connaissances médicales** À terme, la source de données principale sera Wikimédica, cependant son ontologie est en encore cours de construction et ne contient présentement que 157 maladies, signes et symptômes, ce qui n'est pas suffisant pour notre application. Nous combinons donc les informations structurées de Wikimédica avec celles de Wikidata. En outre, bien que les deux wikis précédents disposent de données épidémiologiques pour chaque maladie (par exemple, sur Wikidata, les propriétés P1193, P1603 et P2844 décrivent respectivement la prévalence, l'incidence et le nombre de cas d'une maladie) qui pourraient être utilisées pour évaluer sa probabilité d'occurrence, ces données sont pour le moment trop éparses pour pouvoir être utilisables. Nous avons choisi transitoirement de baser le critère d'importance sur des données de nature différente, à savoir le flux de visites sur la page

Wikipédia associée<sup>3</sup>. Ces données apportent une information nettement moins objective et beaucoup plus approximative sur l'importance relative des différentes maladies, mais nous considérons cette première approximation comme raisonnable car elle sert uniquement à ordonner les résultats. En outre, le flux de visite d'une page est variable en fonction de l'année, ce qui permet de refléter dans une certaine limite l'évolution de l'importance de maladies aux symptômes similaires au cours de l'année (voir la figure 2).

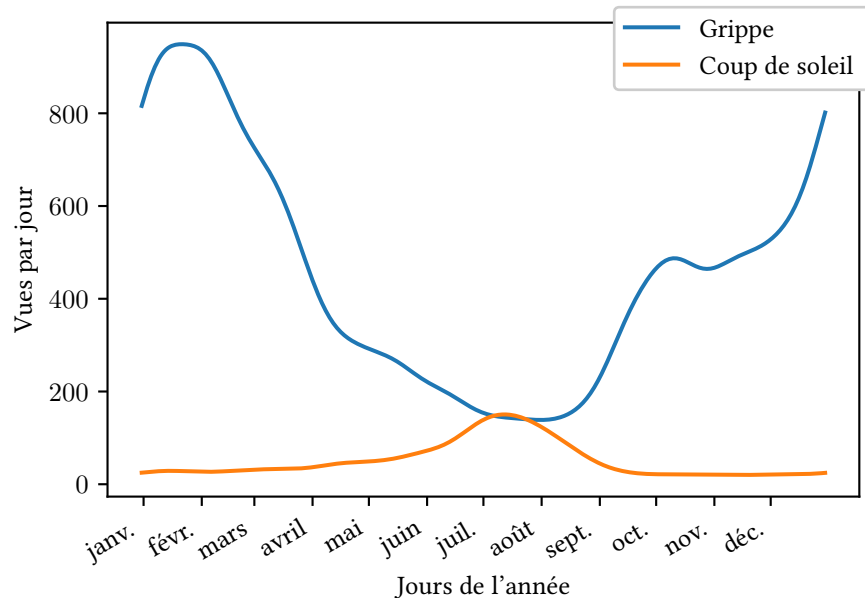


FIG. 2 : Nombre de visites au cours de l'année sur les articles du Wikipédia francophone « Coup de soleil » et « Grippe », deux conditions pouvant présenter des symptômes similaires. Ce nombre de visite est la moyenne des visites sur les années 2015–2019 lissée par un noyau gaussien d'écart-type dix jours.

**Modèle unifié de connaissances du domaine** Chaque maladie et chaque signe, symptôme ou facteur de risque est caractérisé par un nom, une courte description, une image d'illustration, un lien vers un l'article de wiki qui le décrit et un facteur d'importance calculé comme décrit précédemment. Les maladies sont reliées à leurs signes, symptômes ou facteurs de risque (qui peuvent être d'autres maladies) par une propriété qui regroupe sans distinction tous ces types de relation. La correspondance entre ces structures de connaissances et les deux sources utilisées est donnée par la table 1.

<sup>3</sup>Accessibles et librement réutilisables depuis l'adresse <https://dumps.wikimedia.org/other/pageviews/readme.html>.

Classes d'éléments	Wikimedia		Wikidata	
	Classe	Instances	Classe	Instances
Maladie	Maladie	66	Q12136	23 785
Signe	Signe	69	Q1441305	576
Symptôme	Symptôme	22	Q169872	1 076
Propriétés	Propriété	Utilisations	Propriété	Utilisations
Signe	Signe	41	P780 et P1542	4 125
Symptôme	Symptôme	30		
Facteur de risque	Facteur de risque	42	P5642	407

TAB. 1 : Correspondance entre le modèle de connaissance du système et ceux des sources de données Wikimedia et Wikidata.

**Inférences réalisées pour répondre aux requêtes** À partir d'un ensemble de signes et symptômes indiqué par le médecin, le système sélectionne toutes les maladies ayant une relation avec chacun des termes saisis par l'utilisateur. Cette relation est considérée de façon transitive : si un symptôme d'une maladie  $M_1$ , qui est elle-même le symptôme d'une autre maladie  $M_2$ , est saisi, la maladie  $M_2$  est également sélectionnée. Enfin, tous les autres symptômes des maladies ainsi sélectionnées sont également renvoyés, puisqu'ils constituent des éléments pouvant être testés par le médecin pour raffiner son diagnostic.

### 3.2 Interactions entre le médecin et le système de recherche

Notre projet se place dans le cadre des systèmes de recherche sémantique, qui utilisent les connaissances du domaine cible afin de comprendre et de répondre aux requêtes de l'utilisateur. Ils s'opposent aux systèmes de recherche lexicale qui tentent de mettre en correspondance chaque requête avec des documents par des techniques statistiques [4]. Le modèle de communication du système s'intéresse par conséquent aux interactions entre l'utilisateur (le médecin) et la base de connaissances (issue de Wikimedia et Wikidata) telles que spécifiées en figure 3. Il s'inscrit dans le domaine des interfaces personnes-machines et plus spécifiquement dans le contexte des interfaces de recherche d'information.

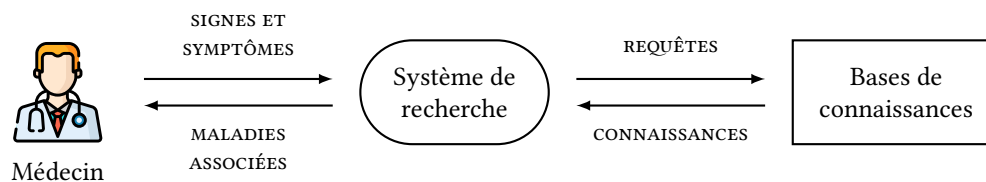


FIG. 3 : Synthèse des interactions entre les différents agents du système lors de la réalisation d'une tâche de recherche de maladie.

**Contraintes sur l'interface utilisateur** Dans la conception de l'interface, nous devons nous efforcer de respecter les principes fondamentaux des interfaces personne-machine pour les systèmes de recherche [7, p. 258] :

1. *Donner un retour d'information à l'utilisateur.* Au cours de l'utilisation du système, l'interface doit s'efforcer d'afficher prioritairement les informations pertinentes pour la poursuite de la recherche. Les relations entre les différentes maladies retournées et les signes et symptômes recherchés doivent être clairement explicités.
2. *Réduire la charge sur la mémoire de travail de l'utilisateur.* Le système doit rappeler à l'utilisateur les choix qu'il a fait au cours du processus de recherche et lui permettre de revenir en arrière sur certains choix et d'essayer de nouvelles stratégies. Il peut aussi suggérer à l'utilisateur des façons de poursuivre la recherche actuelle. Pour ce faire, lorsque l'utilisateur saisit un terme de recherche dans le système, celui-ci doit suggérer les signes, symptômes et maladies les plus proches dans l'ontologie, en s'efforçant d'ôter toute ambiguïté entre plusieurs sens d'un même terme. Par exemple, à la saisie de « hypot », le système doit suggérer, dans l'ordre, Hypotension, Hypotonie et Hyponatrémie, car ce dernier terme est plus distant de la saisie que les premiers.
3. *Être adapté aussi bien aux novices qu'aux experts.* Dans le contexte de notre application, nous considérons en priorité l'utilisation par un professionnel de la santé expert du domaine, mais une utilisation par un patient non-initié qui recherche une maladie spécifique est envisageable à terme.

Enfin, de façon générale, « lorsqu'un système interactif est bien conçu, l'interface disparaît quasiment et permet aux utilisateurs de se concentrer sur leur travail, leur exploration ou leur loisir »<sup>4</sup>.

**Solution proposée** Pour tenter de répondre à ces critères, nous envisageons une interface dotée d'un champ de recherche et montrant en tout temps une visualisation d'un sous-graphe de l'ontologie, mis à jour dynamiquement au fur et à mesure que l'utilisateur ajoute des termes pour préciser sa requête (voir figure 4). Initialement, seuls les nœuds les plus importants (en terme de nombre de liens ou de pertinence générale) de l'ontologie et leurs relations sont affichées. Pour formuler sa requête, l'utilisateur peut soit cliquer sur des éléments du graphe affiché, soit saisir des termes dans le champ de saisie (ces termes sont complétés en recherche approximative).

Le champ de recherche permet à l'utilisateur de garder en vue l'évolution de son exploration du graphe et d'éditer sa requête pour éventuellement revenir sur ses pas. Le graphe intègre simultanément les informations de résultats de la recherche, les liens entre les résultats et les termes de la recherche et des suggestions de poursuite de la recherche.

---

<sup>4</sup> « When an interactive system is well-designed, the interface almost disappears, enabling users to concentrate on their work, exploration, or pleasure. », traduction personnelle [11].

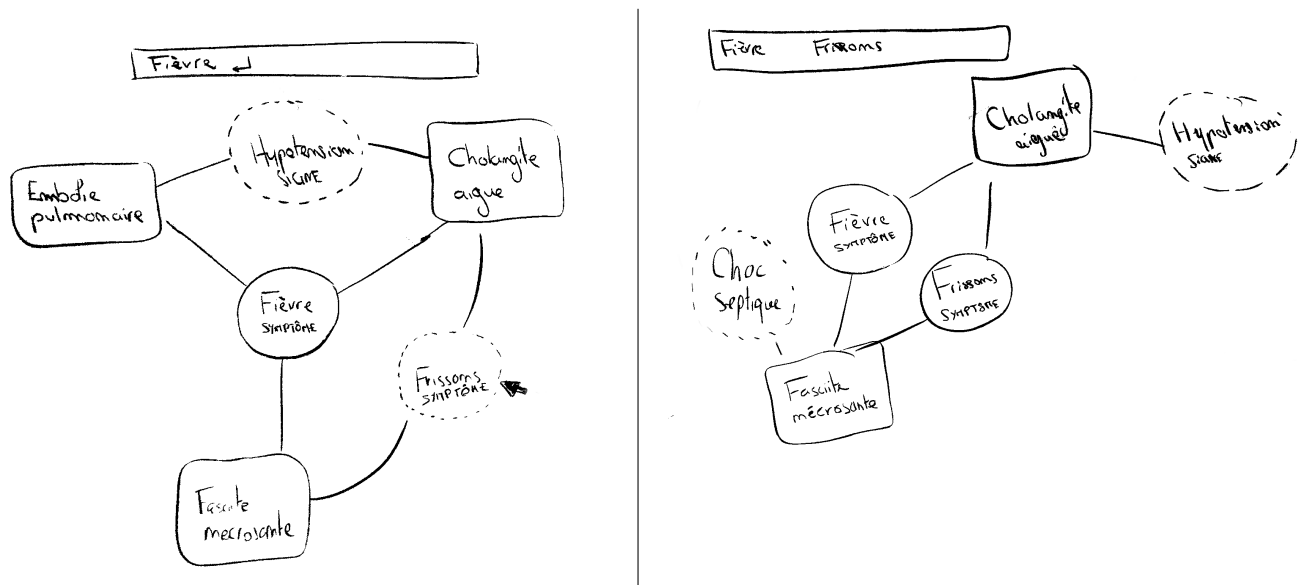


FIG. 4 : Maquette d'interface pour le système de recherche. À gauche, suite à la saisie du symptôme « Fièvre » par le médecin dans la zone de recherche, le système affiche un graphe des trois maladies (nœuds rectangulaires) associées à ce symptôme. Les autres signes et symptômes (nœuds ovales) associés à ces maladies sont également affichés comme des liens entre les maladies. En cliquant sur le symptôme « Frissons », ce nouveau symptôme est ajouté à la requête et le graphe mis à jour (à droite) pour filtrer les résultats en conséquence.

## 4 Réalisation d'un prototype du système

Cette partie explique les détails d'implémentation, les procédures techniques et les différents choix de technologies réalisés pour construire un premier prototype du système.

### 4.1 Stratégie de récupération automatique des connaissances

La première étape d'implémentation consiste à mettre en place un système permettant de récupérer automatiquement les connaissances décrites précédemment. Ces connaissances sont issues de diverses sources. Les signes, les symptômes, les maladies et leurs relations sont obtenues depuis Wikimedia et Wikidata. Le facteur d'importance est dérivé des statistiques de visite des pages Wikipédia. Chacune de ces sources mettent à disposition leurs données de façon qui leur est spécifique.

La récupération des données est programmée avec le langage de scripts Python. Ce langage jouit d'une communauté importante, de nombreuses bibliothèques ainsi que d'une documentation très complète.



**Wikimedia** Le site web Wikimedia dispose d'une fonctionnalité d'export des données structurées de son wiki décrites sous le formalisme *Resource Description Framework* (RDF) et la syntaxe RDF/XML. Le RDF est un modèle de graphe capable de décrire de façon formelle les ressources du web sémantique et leurs métadonnées de manière exploitable par des ordinateurs. Ce formalisme a été développé par le *World Wide Web Consortium* (W3C) en 1999 et est aujourd'hui à la base du Web sémantique. Le RDF/XML est une syntaxe, également définie par le W3C, permettant de sérialiser un graphe RDF comme un document XML. Pour récupérer les données des maladies, les signes et les symptômes de Wikimedia, le prototype télécharge directement les fichiers RDF/XML nécessaires depuis le site.

**Wikidata** La base de connaissances Wikidata offre la possibilité d'accéder à ses données au travers d'un *endpoint SPARQL Protocol and RDF Query Language* (ou simplement SPARQL). Le SPARQL est un langage de requêtes permettant de décrire, de récupérer et de manipuler des ressources stockées au format RDF. Un *endpoint* SPARQL est un point d'accès disponible sur le Web permettant de répondre aux requêtes SPARQL vis-à-vis d'un ensemble de ressources RDF d'un site, ici Wikidata. Pour obtenir les données nécessaires depuis Wikidata, notre projet récupère les données en réalisant des requêtes HTTP vers cet *endpoint*.

**Statistiques de visites Wikipédia** Wikimedia met à disposition ses statistiques des nombres de vues de pages Wikipédia au travers d'une API au style « Representational state transfer » (REST) nommée *pageview statistics*. REST est un style d'architecture d'interface de programmation suivant un ensemble de contraintes pour créer des services web. Pour récupérer les informations sur le nombre de visites d'une page Wikipédia, le prototype s'adresse directement à cette API.

## 4.2 Alignement des connaissances récupérées

Une fois les données récupérées, la deuxième étape consiste à aligner les termes de Wikimedia sur ceux de Wikidata. Puisque Wikimedia est pour le moment une plus petite ontologie médicale que la partie de Wikidata considérée, nous avons décidé d'aligner les termes de Wikimedia sur ceux de Wikidata et non l'inverse. Pour déterminer l'*item* de Wikidata correspondant à un *item* de Wikimedia, nous utilisons dans l'ordre l'algorithme suivant :

1. Si le terme de Wikimedia possède la propriété `property:Wikidata_id`, l'alignement est déjà renseigné manuellement. L'application se base donc exclusivement sur cette information.
2. Si la propriété `property:Wikidata_id` n'est pas présente, le script d'alignement réalise alors une recherche par nom dans la base de connaissances de Wikidata à l'aide d'une requête SPARQL. Si un seul résultat correspond, alors l'alignement est effectué avec cet item.
3. Si aucun ou plusieurs résultats sont trouvés, l'alignement ne peut pas se faire automatiquement et l'intervention d'un opérateur est demandée.

En ce qui concerne les relations entre les éléments de l'ontologie, nous avons choisi de les accumuler sans distinction. Le script d'alignement a également été développé en Python. Il parvient à aligner automatiquement près de la moitié des termes de Wikimé-dica. Sur 100 termes, 32 sont alignés à l'aide de la propriété `property:Wikidata_id`, 18 par leur nom et les 43 restants à la main.

### 4.3 Stockage durable des connaissances alignées

Une fois les connaissances récupérées et alignées entre nos différentes sources, la troisième étape a pour objectif de les sauvegarder de manière pérenne et accessible dans une base de données. Les format RDF de données manipulés ici se présente sous forme de graphe. Il semble donc naturel de sauvegarder les informations dans une base de type graphe.

Nous avons fait le choix d'utiliser la technologie de base de graphe Neo4j. Neo4j est un projet *open-source* créé en 2007 et implémenté en Java. C'est aujourd'hui la base de données de type graphe la plus populaire [6].

### 4.4 Algorithmes de recherche approximative

Pour opérationnaliser cette définition, nous devons choisir une mesure de distance permettant de trier les différents termes candidats. Dans notre cas, la saisie de l'utilisateur sera rarement un mot complet et plus souvent un préfixe voire un sous-mot du terme recherché. Occasionnellement, cette saisie peut aussi contenir des fautes de frappe ou d'orthographe. Plusieurs heuristiques peuvent être envisagées pour prendre en compte ces considérations, en ordre croissant de complexité :

- conserver les résultats commençant par le préfixe de recherche saisi, en normalisant la casse et les accents. Cette heuristique n'inclut pas les résultats dont le mot saisi est un sous-mot ou qui diffèrent de la saisie par quelques insertions ou délétions. Pour implémenter efficacement cette heuristique, un arbre des préfixes peut être utilisé.
- conserver les meilleurs résultats en triant les termes par ordre décroissant du nombre de lettres en commun avec la recherche. Implémentée ainsi, cette solution ne prend pas en compte l'ordre des lettres dans les termes ce qui est pourtant un élément important. Pour pallier cela, nous pouvons trier les résultats selon un score qui pénalise les lettres mal placées.
- filtrer les meilleurs résultats en triant les termes par distance de Levenshtein [9] croissante avec la recherche. Cette distance permet de mesurer l'éloignement entre deux mots complets en termes du nombre de caractères devant être insérés, supprimés ou substitués pour passer de l'un à l'autre. Des structures de données efficaces existent pour implémenter une recherche par distance de Levenshtein, notamment les BK-trees et les automates de Levenshtein. Pour adapter la distance aux spécificités du problème de recherche approximative, nous pouvons imaginer donner un poids de 0 pour l'ajout de lettres en début ou en fin de recherche et de 1 pour les autres opérations.

## 4.5 Interface d'accès frontal aux connaissances

La dernière étape de réalisation du prototype avait pour objectif d'implémenter un frontal réalisant les concepts d'interface établis à la section précédente. Nous avons choisi d'implémenter ce frontal sur le web, pour le rendre accessible aisément à tous sans installation particulière et pour rester dans la philosophie des wiki qui sont entièrement basés sur le web. Par ailleurs, le JavaScript, langage au cœur des technologies du web, est un langage pour lequel nous disposons déjà de connaissances et d'expérience. Enfin, ce langage possède une large communauté très active, de nombreuses bibliothèques et une documentation très complète.

Pour créer la structure de l'application et définir les interactions, nous avons utilisé le cadre de travail React<sup>5</sup>. Cette bibliothèque *open source* est développée par Facebook depuis 2013 et permet de simplifier la création d'applications très demandeuses en interactions. Elle s'appuie sur le principe de la programmation réactive et permet ainsi un code déclaratif et plus efficace et flexible que la programmation web classique.

Enfin, le graphe est affiché sous forme de nœuds reliés par des arêtes déplaçables par l'utilisateur. La mise en forme du graphe est réalisée à l'aide d'un algorithme de simulation de force physique dans lequel les nœuds sont modélisés comme des masses ponctuelles et les liens comme des ressorts. Nous avons utilisé pour réaliser cette simulation physique simple la bibliothèque Springy.js.

## 5 Conclusion

L'objectif du projet était d'explorer les différentes façons dont les techniques d'intelligence artificielle peuvent aider les médecins dans la réalisation de diagnostics différentiels. Pour ce faire, nous avons conçu un système capable d'exploiter les données structurées de la plateforme Wikimedia afin de proposer une interface de recherche sémantique de maladies par signes, symptômes, antécédents et facteurs de risque. Cette interface se présente sous la forme d'un graphe interactif, montrant au médecin les relations existant entre les résultats obtenus et lui proposant par là même des suggestions de poursuite de sa recherche. Pour illustrer plus précisément notre concept, nous avons décidé de construire un prototype d'application qui réalise les fonctionnalités de base du système<sup>6</sup>. La figure 5 montre quelques impressions d'écran de ce prototype. Après présentation de ce prototype aux responsables de la plateforme Wikimedia, nous sommes convaincus qu'il a le potentiel de devenir un système d'aide poussé à la réalisation de diagnostics différentiels et également de servir d'illustration à la puissance des techniques du web sémantique appliquées à la médecine.

## Références

- [1] ASSOCIATION CANADIENNE DE PROTECTION MÉDICALE, éd. *Guide des bonnes pratiques. Diagnostic différentiel*. Avr. 2018. URL : <https://www.cmpa-acpm>.

<sup>5</sup><https://reactjs.org/>

<sup>6</sup>Voir <https://matteo.delab.re/projects/wikimedia-disease-search>.

ca/serve/docs/ela/goodpracticesguide/pages/manage\_risk/The\_diagnostic\_process/differential\_diagnosis-f.html (consulté le 12/11/2019).

- [2] Attribution-ShareAlike 4.0 International (CC-BY-SA 4.0). Creative Commons, déc. 2018. URL : <https://creativecommons.org/licenses/by-sa/4.0/legalcode> (consulté le 15/09/2019).
- [3] Varun H BUCH, Irfan AHMED et Mahiben MARUTHAPPU. « Artificial intelligence in medicine : current trends and future possibilities ». In : *British Journal of General Practice* 68.668 (2018), p. 143-144.
- [4] CONTRIBUTEURS DE WIKIPEDIA. *Semantic search*. 2019. URL : [https://en.wikipedia.org/w/index.php?title=Semantic\\_search&oldid=915042761](https://en.wikipedia.org/w/index.php?title=Semantic_search&oldid=915042761) (consulté le 18/11/2019).
- [5] CONTRIBUTEURS DE WIKIPÉDIA. *Maladie — Wikipédia, l'encyclopédie libre*. 2019. URL : <https://fr.wikipedia.org/w/index.php?title=Maladie&oldid=164883945> (consulté le 30/11/2019).
- [6] DB ENGINES. *DB-Engines Ranking of Graph DBMS*. 2019. URL : <https://db-engines.com/en/ranking/graph+dbms> (consulté le 18/12/2019).
- [7] Marti A. HEARST. « User interfaces and visualization ». In : Ricardo BAEZA-YATES, Berthier RIBEIRO-NETO et al. *Modern information retrieval*. ACM, 1999. Chap. 10.
- [8] Yvon LAROSE. « Un wiki pour les professionnels de la santé ». In : *ULaval Nouvelles* (oct. 2018). URL : <https://nouvelles.ulaval.ca/vie-universitaire/-6f975a18a06639cb8e48445d7f1f06c6> (consulté le 16/09/2019).
- [9] Vladimir Iosifovich LEVENSHTAIN. « Binary codes capable of correcting deletions, insertions, and reversals ». In : *Soviet Physics-Doklady*. T. 10. 8. 1966, p. 707-710.
- [10] Stuart J. RUSSELL et Peter NORVIG. *Artificial Intelligence. A Modern Approach*. 3<sup>e</sup> éd. Prentice Hall, 2010.
- [11] Ben SHNEIDERMAN et al. *Designing the User Interface : Strategies for Effective Human-Computer Interaction*. 6<sup>e</sup> éd. Pearson, 2016.

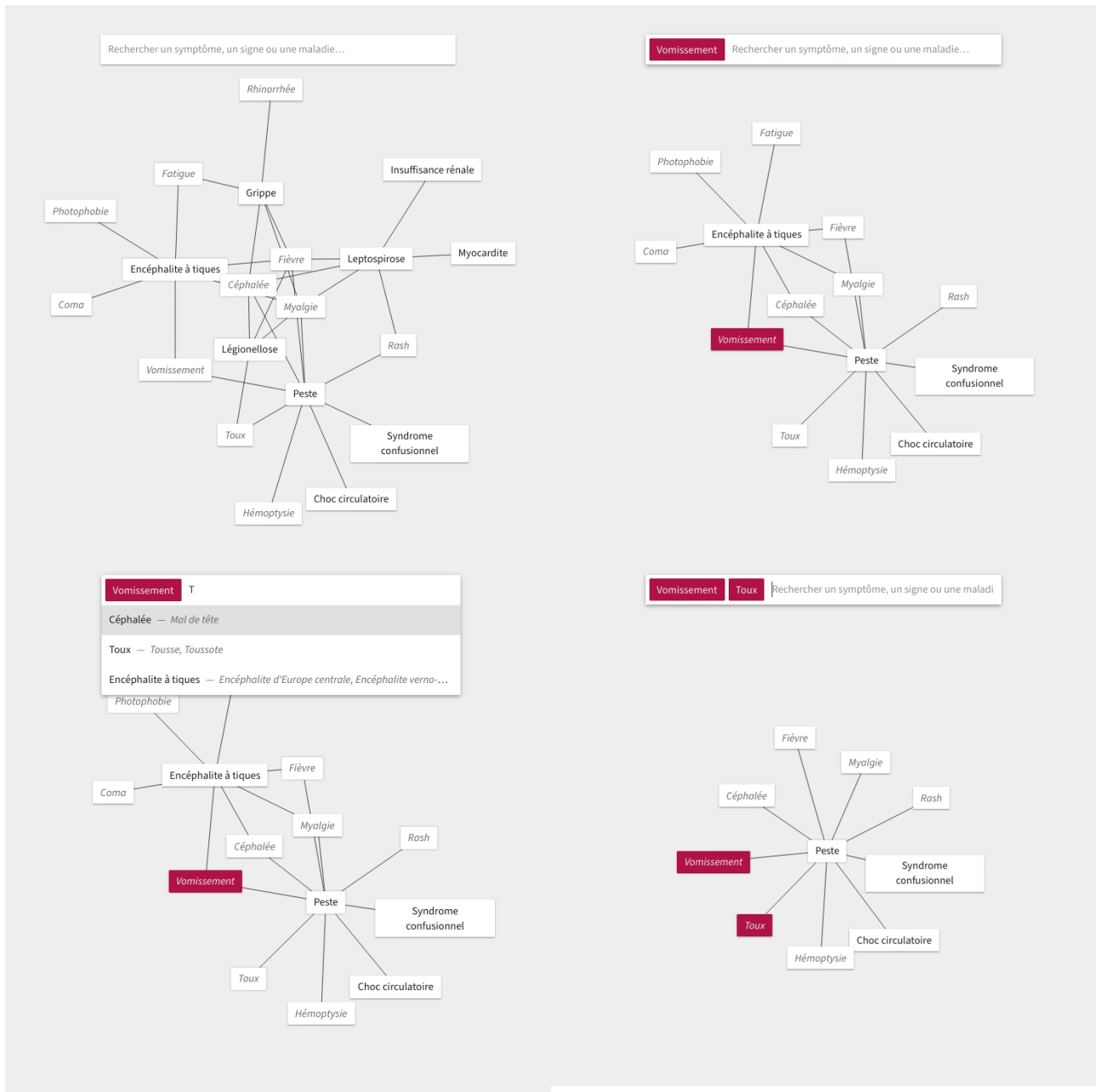


FIG. 5 : Capture d'écran illustrant le fonctionnement du projet. En haut à gauche, l'état initial de l'application où le champ de recherche est vide est le graphe affiché au complet. En haut à droite, le terme « Vomissement » est recherché, le graphe est mis à jour en conséquence. En bas à gauche, l'utilisateur est en train de saisir le préfixe « T » et le système lui propose plusieurs termes correspondants à sa saisie. Enfin, en bas à droite, l'utilisateur complète sa recherche avec le mot « Toux ».