Modélisation des ingrédients de remèdes issus de pharmacopées arabes médiévales dans une base de données graphe

Karim El Haff[1,2], Agnès Braud[1], Florence Le Ber[1], Véronique Pitchon[2]

[1] Université de Strasbourg, CNRS, ENGEES, ICube UMR 7357, F67000 Strasbourg
[2] Université de Strasbourg, CNRS, Archimède UMR 7044, F67000 Strasbourg
{kelhaff, agnes.braud, florence.le-ber, pitchon}@unistra.fr

Résumé

Cet article présente le travail de modélisation engagé dans le cadre d'un projet interdisciplinaire, dont le but est d'étudier les remèdes décrits dans les pharmacopées arabes médiévales. Les informations extraites d'un texte ancien, traduit en anglais, ont été représentées dans une base de données graphe. Cette étude de cas a mis en évidence plusieurs problèmes de modélisation, notamment pour la représentation d'ingrédients qui sont des sous-parties de plantes et dont l'appellation comporte des ambiguïtés. L'article détaille ces problèmes et les solutions apportées.

Mots-clés

Base de données graphe, modélisation de connaissances, textes anciens.

Abstract

This paper presents the modelling work undertaken as part of an interdisciplinary project, whose aim is to study the remedies described in medieval Arabic pharmacopoeias. Information collected from an ancient medical text, translated in English, were represented in a graph database. This study highlighted several modelling issues, including the representation of ingredients that are plant subparts, and whose name are often ambiguous. This paper details these problems and the proposed solutions.

Keywords

Graph database, knowledge modelling, ancient texts.

1 Introduction

Les textes médicaux anciens contiennent une mine de connaissances sur les maladies, les traitements et les pratiques de guérison. Ces dernières années, ces textes ont suscité un intérêt croissant [7, 16] et certains ont fait l'objet d'approches de type fouille de données [4].

Le projet PARADISE (MITI80 CNRS, 2021) s'inscrit dans cette lignée. Il a pour objectif de développer des approches informatiques pour exploiter les informations contenues dans les textes des pharmacopées arabes médiévales : il s'agit d'extraire des informations concernant principalement les remèdes associés aux maladies infectieuses, les organiser et les interroger afin de mettre en évidence puis de tester des ingrédients qui pourraient être utilisés dans la création de médicaments pouvant, en particulier, être une alternative aux antibiotiques. Le projet PARADISE regroupe pour cela des historiens, pharmaco-botanistes, biologistes et informaticiens.

Le projet se décline en plusieurs étapes. La première consiste à extraire des textes anciens (pour le moment des traductions anglaises de ces textes) les descriptions des remèdes d'intérêt pour les biologistes. Dans cette première étape, les termes utiles sont annotés et servent à l'apprentissage d'un système de reconnaissance d'entités nommées [5]. Dans la deuxième étape, les termes extraits et des informations annexes sont représentés dans une base de données, afin de permettre leur interrogation.

La majorité des remèdes extraits du corpus comporte des ingrédients à base de plantes, qui sont utilisées en tout ou partie, avec ou sans préparation ou transformation. Dans cet article, nous nous focalisons sur la modélisation de ces ingrédients, de sorte à faciliter leur interrogation. En particulier, il faut à la fois pouvoir retrouver des plantes utilisées dans différents remèdes, mais aussi les parties de plantes concernées, qui peuvent avoir un effet spécifique. Pour répondre à ces besoins, et en l'absence d'un modèle du domaine préexistant, nous avons fait le choix d'utiliser une base de données orientée graphe, pour la souplesse de représentation qu'elle offre. Néanmoins, différentes problématiques de modélisation sont apparues, que nous présentons et discutons.

L'article est organisé comme suit : après cette première partie introductive, la section 2 décrit brièvement les principes des bases de données orientées graphes, puis présente des travaux de modélisation voisins du nôtre. La section 3 présente les données et connaissances mobilisées, la section 4 décrit les différents problèmes de modélisation rencontrés, ainsi que les choix réalisés et leurs limites. La dernière section dresse quelques conclusions et perspectives.

2 Travaux connexes

2.1 Bases de données orientées graphes

Dans le domaine des bases de données, depuis la fin des années 1960, la structure de table interconnectée de la base de données relationnelle a été le modèle dominant de stockage et d'interrogation de données. Avec la croissance des données produites par les réseaux sociaux et la nécessité de trai-

ter efficacement de telles données, de nouveaux systèmes de gestion de données ont été développés. Un système de gestion de bases de données orienté graphe [1] permet de gérer des données en s'appuyant sur une représentation sous forme de graphe.

Différents travaux ont été menés autour des bases de données orientées graphes, par exemple pour adapter les algorithmes de parcours de graphes aux différentes requêtes possibles [14]. D'un point de vue applicatif, ces bases sont largement utilisées pour modéliser les réseaux sociaux. Des applications plus spécifiques ont été développées pour par exemple analyser les termes de requêtes et de leurs reformulations [12] ou pour représenter des référentiels de modèles en génie logiciel [11]. Ce type de base de données est également largement utilisé dans le domaine biomédical [17].

2.2 Modélisation de connaissances

Les bases de connaissances, ou ontologies [6], ont été développées depuis de nombreuses années dans différents domaines. Elles permettent de formaliser les connaissances d'un domaine, de les partager et de les utiliser dans des raisonnements formels. L'avantage des ontologies est aussi la possibilité de les réutiliser et de les étendre à des sous-domaines. En particulier, les ontologies développées dans les domaines végétal et médical pourraient être des supports à nos travaux de modélisation. Toutefois chaque ontologie étant développée avec un objectif spécifique, sa réutilisation nécessitera toujours un travail de modélisation important et l'appel à différentes sources de données et de connaissances, comme explicité par [3] qui présente une ontologie pour la médecine d'urgence.

Plant Ontology (PO) [18] est une ressource communautaire qui comprend des termes normalisés, des définitions et des relations décrivant les structures et les stades de développement des plantes. Cette ontologie est complétée par une base de données d'annotations provenant d'études génomiques et phénomiques. Une ontologie spécifique [9] a été construite pour la plante *Arabidopsis Thaliana* qui est utilisée comme plante modèle pour la botanique et d'autres sciences végétales. Cette ontologie vise à décrire les caractéristiques physiques, biochimiques et génétiques d'*Arabidopsis Thaliana* et les relations entre ces caractéristiques. La base de connaissances Knomana (*KNOwledge MANAgement on pesticide plants in Africa*) recense les connaissances actuelles sur les plantes utilisées comme pesticides en Afrique [15].

Ces différentes ontologies du domaine végétal décrivent les caractéristiques globales des plantes, et dans une visée (à part la base Knomana) plutôt génétique ou productive, alors que dans les remèdes issus des pharmacopées anciennes apparaissent des parties de plantes (par exemple les coques de glands ou des pépins de melon) qui peuvent avoir été transformées (grillées, séchées, etc.). C'est la modélisation de ces ingrédients que nous présentons dans la suite de l'article.

3 Présentation des données

Dans cette section nous présentons le type d'informations que nous manipulons et la base constituée.

3.1 La collecte des données initiales

Le corpus exploré est un manuscrit médical qui décrit 292 remèdes ou préparations. C'est une partie de la traduction anglaise par Oliver Kahl de l'ouvrage « Dispensatory in the Recension of the ʻAḍudī Hospital » écrit par Sābūr ibn Sahl au IXe siècle [8]. Le corpus est constitué de 36 961 *tokens* qui ont été annotés dans le but de servir à l'entraînement d'un modèle de reconnaissance d'entités nommées (*Named Entity Recognition*, NER) [5].

Les remèdes concernent principalement des maladies infectieuses, touchant différents organes du corps (poumon, peau, etc.) et générant des symptômes (toux, saignement, douleurs etc.). La description d'un remède est composée d'une liste d'ingrédients avec des quantités et une description de la préparation. Toutes les informations contenues dans les remèdes ne sont pas exploitées dans ce travail. Pour effectuer l'annotation, seules quatre étiquettes ont été utilisées : Type (pour la forme du remède), Sym (pour les symptômes), Ing (pour les ingrédients) et Org (pour les organes). Les données sont annotées dans un fichier CSV dans le format IOB2 [1] L'annotation et le nettoyage du corpus ont été effectués par le premier auteur pendant un mois et revus en profondeur par la dernière autrice, historienne experte en médecine arabe médiévale.

Ce corpus annoté a donc deux utilités : il constitue une ensemble de données d'entraînement pour la NER, dont le résultat pourra être utilisé pour l'analyse d'autres manuscrits ; il constitue également un ensemble permettant la construction d'un premier modèle de données.

3.2 L'enrichissement des données

Après l'annotation du texte à l'aide de IOB2, nous avons identifié les ingrédients uniques présents dans le texte, soit 957 ingrédients. Les termes désignant les ingrédients végétaux ont ensuite été vérifiés avec l'aide d'un botaniste pour identifier la plante concernée et y rattacher ses propriétés. Un tableur (fichier de type CSV) a été créé et complété avec un ensemble d'informations obtenu à partir des ressources The World Flora Online [2] pour la botanique, CHEMnetBASE [3] et Reaxys [4] pour les molécules naturelles.

Le tableau 1 en présente un extrait : il contient le nom de l'ingrédient tel qu'il est écrit dans le livre (exemple : absinthe leaves), le nom vernaculaire de la plante ainsi que des synonymes courants (absinthe, wormwood [...]), la partie de la plante utilisée (leaf), le nom scientifique de la plante (*Artemisia absinthium*), les synonymes scientifiques de son nom (*Absinthium bipedale* [...]), la famille de la plante (*Asteraceae*), son origine géographique (Afghanistan, Albania,

1. IOB2, abréviation de « inside, outside, beginning » est un format de marquage commun en traitement automatique des langues [13].
2. `http://www.worldfloraonline.org/`
3. `https://dnp.chemnetbase.com/`
4. `https://www.reaxys.com/`

Ingrédient	Nom vernaculaire	Partie de la plante	Nom Scientifique	Synonymes	Famille	Origine géographique	Molécules actives	Toxicité
absinthe leaves	absinthe, wormwood [...]	leaf-_absinthe	*Artemisia absinthium*	*Absinthium bipedale* [...]	*Asteraceae*	Afghanistan, Albania, Algeria [...]	polyphenol, monoterpene	whole plant
absinthe sap	absinthe, wormwood [...]	sap_absinthe	*Artemisia absinthium*	*Absinthium bipedale* [...]	*Asteraceae*	Afghanistan, Albania, Algeria [...]	polyphenol, monoterpene	whole plant
citrons from Susa	citron tree from susa	fruit_citron-tree	*Citrus Medica*	*Citrus acida* [...]	*Rutaceae*	Assam, Bangladesh, East Himalaya [...]	polyphenol, coumarin, flavonoid, terpene	none
peels of celery roots	celery	peel_root-_celery	*Apium Graveolens*	*Apium australe var. latisectum* [...]	*Apiaceae*	Afghanistan, Albania, Algeria, [...]	glycosid, polyphenol, furocoumarin	whole plant low

TABLE 1 – Extrait du tableau de données : les termes issus du texte sont complétés par des informations sur les plantes correspondantes

Algeria [...]), les molécules actives trouvées dans la plante (polyphenol, monoterpene), et sa toxicité (whole plant).

3.3 Le modèle des données

Nous avons développé un modèle pour représenter dans une base de données orientée graphe les relations complexes entre les remèdes, les ingrédients, les plantes et leurs parties, ainsi que les transformations d'ingrédients. Une représentation de ce modèle est présenté en figure 1. Nous utilisons l'outil Neo4j [5], qui offre une interface d'interrogation et de visualisation des données.

Le modèle est structuré autour de la relation `Contains` (contient, 1700 instances) entre les nœuds `Remedy` (remède, 287 inst.) et `Ingredient` (ingrédient, 715 inst.). Cette relation est dotée de deux attributs : le nom original de l'ingrédient dans le manuscrit, ainsi qu'une origine géographique spécifiée dans le nom le cas échéant, comme par exemple *Antioch* pour l'ingrédient *Antioch scammony* (scammonée d'Antioche). Les ingrédients peuvent être une plante entière ou une partie de plante : le nœud `Ingredient` est alors relié par la relation `PartOf` (partie de, 259 inst.) à un autre nœud de type `Ingredient` (plante ou partie de plante). Un nœud `Ingredient` est également relié à un nœud `Taxon` (311 inst.), grâce à la relation `HasTaxon` (a comme taxon, 503 inst.). Ce nœud possède 5 attributs : le nom scientifique de la plante (espèce ou genre), la liste des synonymes scientifiques, la liste des origines géographiques possibles, la liste des principes actifs et la toxicité. De plus, chaque nœud `Taxon` est relié par `FromFamily` (de la famille, 312 inst.) à un nœud `Family` (famille, 114 inst.), qui représente une famille de plantes. Enfin, notre modèle prend en compte les transformations d'ingrédients dans les remèdes. La relation `ContainsTransformed` (contient transformé, 560 inst.) relie les remèdes aux ingrédients transformés (nœud `TransformedIngredient`, 259 inst.) et possède les mêmes attributs que la relation `Contains`. La relation `IsFrom` (provient de, 258 inst.) permet de relier les ingrédients transformés à l'ingrédient d'origine, avec l'attribut `transformation_type` qui permet de catégoriser la transformation.

5. `https://neo4j.com/fr/`

4 Questions de modélisation

Définir le modèle de données pour enregistrer les remèdes issus du corpus étudié dans la BD graphe a soulevé plusieurs questions que nous développons ci-dessous.

4.1 Les ingrédients sous-parties de plantes

La première question portait sur la façon de modéliser les parties de plantes utilisées comme ingrédients dans les remèdes. En effet, nombre d'ingrédients utilisés dans les remèdes ne sont des parties de plantes, telles que des graines, des fruits, des racines, des feuilles, etc.

Dans le tableau 1, la colonne "Partie de la plante" précise la partie correspondant à chaque ingrédient mentionné dans un remède. Par exemple, si l'ingrédient original est *apple seeds* (pépins de pomme), la colonne "Partie de la plante" contiendra "seed_fruit_apple tree". Pour formaliser ces informations, il est intéressant de décomposer la chaîne menant d'une sous-partie à la partie entière de la plante. Par exemple, pour l'ingrédient "apple seeds", la chaîne va de "Seed : seed_fruit_apple tree" à la plante "Plant : apple tree", en passant par "Fruit : fruit_apple tree". On voit ici trois catégories de parties de plantes, la graine, le fruit et l'arbre (plante entière).

Chaque partie d'une plante a été modélisée comme un nœud dans un arbre hiérarchique. La racine de l'arbre est le nœud "plante entière", qui est lié au nom scientifique de la plante dans notre modèle.

Les sous-parties de plantes ont été regroupées en 21 types, sous-types du nœud `Ingredient` : `Fruit` (fruit), `Pulp of fruit` (pulpe de fruit), `Inner skin` (peau intérieure : couche interne d'un fruit, souvent mince et fibreuse), `Peel` (pelure : couche externe du fruit), `Shell` (coquille : couche dure et extérieure de certains fruits), `Seed` (graine), `Pulp of seed` (pulpe de graine), `Seed core` (cœur de graine), `Seed vessel` (enveloppe de la graine), `Stem` (tige), `Leaf` (feuille), `Twig` (rameau), `Stalk` (pédoncule : tige qui porte la fleur puis le fruit), `Flower` (fleur), `Flower buds` (bourgeons floraux), `Root` (racine), `Root peel` (pelure de racine : couche externe de la racine), `Bark` (écorce), `Mucilage` (mucilage : substance visqueuse et épaisse produite par certaines plantes), `Sap` (sève), `Gall` (galle : excroissance anormale de la plante causée par une infection bactérienne

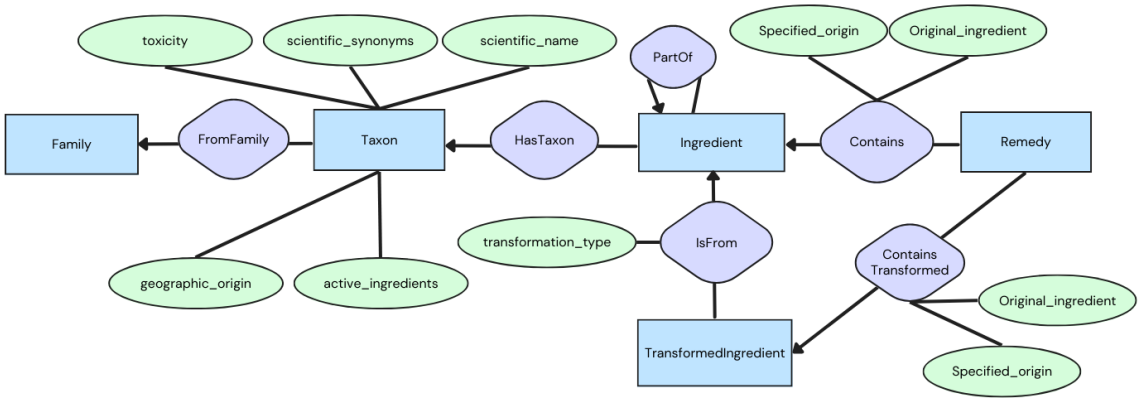

FIGURE 1 – Le modèle de la base de données, les nœuds sont symbolisés par des rectangles, les relations par des losanges et les attributs par des ellipses

ou fongique).

Finalement chaque ingrédient est décomposé selon la chaîne des sous-parties qui le constituent. Cette décomposition a été faite manuellement et un script python permet ensuite de remplir la base de données selon le modèle de la figure 1 en s'appuyant sur la hiérarchie des types de sous-parties de plantes. Cela a permis de modéliser les parties de plantes de manière plus précise et de relier les remèdes aux plantes dont leurs ingrédients sont issus pour faciliter la recherche d'informations. Par exemple, la figure 2 montre le graphe reliant l'ingrédient *citron* aux remèdes 41 et 252, et l'ingrédient *citron peels* au remède 252.

4.2 Les ingrédients transformés

Une deuxième question concerne la représentation des ingrédients transformés, dont la la composition chimique et les propriétés médicinales ont ainsi été modifiés par rapport à l'ingrédient original. Pour représenter cette information dans la base de données graphe, on utilise un type de nœud appelé `TransformedIngredient`, lié à l'ingrédient non transformé par une relation `IsFrom`, qui a comme attribut le type de transformation (séchage, broyage, etc.). Le nœud `TransformedIngredient` est ensuite lié au remède avec la relation `ContainsTransformed`. Cela permet de représenter l'ingrédient végétal original et l'ingrédient transformé, ainsi que les détails du processus de transformation.

Cette approche présente plusieurs avantages. Premièrement, en reliant l'ingrédient végétal original et l'ingrédient transformé qui en résulte, elle permet de retrouver par une seule requête (exploitant la structure de graphe) les remèdes contenant ce végétal, transformé ou non. Deuxièmement, en représentant explicitement l'ingrédient transformé, elle permet aux biologistes d'intégrer les effets spécifiques du processus de transformation sur la composition chimique de l'ingrédient et son effet potentiel. Enfin, d'un point de vue formel, cette représentation permet d'inclure des ingrédients dont la description comporte une négation ou une soustraction, ce qui n'est pas représentable directement par des graphes. Par exemple, l'ingrédient *seedless barberries* (baies d'épine-vinette sans pépins) est difficile à représenter sauf si nous considérons *l'épépinage* comme une transformation (voir figure 3).

4.3 Questions ouvertes

4.3.1 Problèmes sémantiques

Les données ont été extraites d'un texte en anglais, traduit d'un texte arabe ancien, et à ce titre, elles sont entachées d'ambiguïtés sémantiques, que nous ne pouvons lever directement et qui induisent des imprécisions dans la représentation. Nous en donnons deux exemples ci-dessous. Dans les deux cas, pour lever les ambiguïtés, il sera nécessaire de revenir au texte originel, en langue arabe, avec l'aide de spécialistes du lexique de la médecine arabe médiévale et avec l'appui de botanistes.

Appellation courante et nom scientifique. L'un des problèmes que nous avons rencontrés lors de la modélisation est celui des noms de plantes, ou de parties de plantes, ambigus. Dans de nombreux cas, les noms des plantes apparaissant dans le manuscrit traduit ne peuvent être mis en correspondance univoque avec une dénomination scientifique, car ce sont des appellations courantes, qui peuvent correspondre à différentes espèces.

Par exemple, le terme *acorn* (gland), peut désigner en langage courant soit le fruit d'une espèce *Quercus sp* (*oak*), soit celui d'une espèce *Lithocarpus sp* (*stone oak*). Sans une référence plus spécifique à l'espèce en question, il est impossible de catégoriser les données avec précision. Pour ce cas, nous avons donc choisi d'assigner le double nom scientifique *Quercus sp ;Lithocarpus sp*.

Indication d'origine dans une appellation courante. Les désignations des ingrédients peuvent porter une information géographique, comme par exemple *Syrian apples* ou *Chinese rhubarb*. C'est le cas pour 50 ingrédients dans la base de données actuelle. Nous avons choisi de représenter cette information dans un attribut de la relation `Contains` reliant le remède et l'ingrédient. De cette manière, nous cherchons à généraliser les nœuds qui désignent les ingrédients ayant un même nom scientifique pour faciliter les requêtes dans un premier temps, tout en conservant l'information sur l'origine géographique. Toutefois, la sémantique

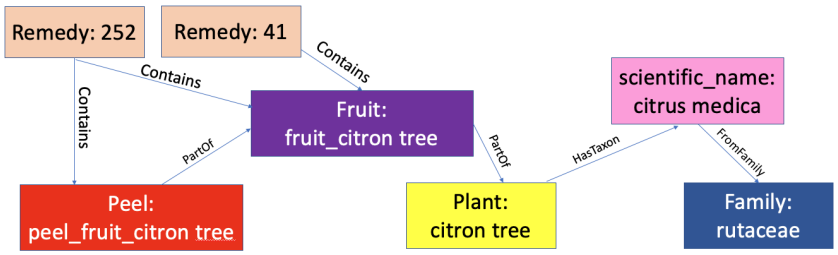

FIGURE 2 – Graphe reliant des remèdes contenant des ingrédients correspondant à des sous-parties du citronnier : le remède 252 a pour ingrédients à la fois le fruit et des pelures (zestes) du fruit.

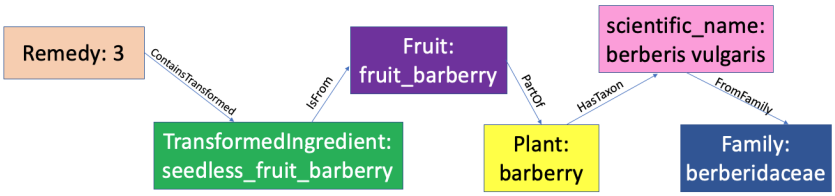

FIGURE 3 – Graphe d'un remède contenant des baies d'épine-vinette épépinées (*seedless barberries*)

portée par cette information peut être variable : elle peut effectivement désigner une origine géographique spécifique, choisie de préférence (le vin de Falerne a certaines propriétés, les dattes de Hairūn sont réputées) ou bien relever d'une appellation courante. Par exemple, le terme *Greek absinthe* pourrait suggérer qu'il s'agit d'une absinthe originaire de Grèce, mais nous ne pouvons pas en être certains sans recherches supplémentaires. De fait, distinguer une origine spécifique d'un terme utilisé couramment est important car une même plante peut avoir des propriétés différentes selon la région où elle est cultivée [10].

4.3.2 Les doubles ingrédients

Certains termes étiquetés comme ingrédients, extraits du corpus, recouvrent en fait plusieurs ingrédients. Par exemple, l'ingrédient *seed vessels and flowers of the pomegranate* (enveloppes de graines et fleurs de grenade) est une combinaison de deux parties différentes du grenadier. L'ingrédient *celery- and fennel-water* (l'eau de céleri et de fenouil) est un exemple plus complexe, où intervient une transformation de deux plantes distinctes.

Dans la base de données actuelle, il y a 28 cas de tels ingrédients combinés. Pour faciliter l'interrogation de la base – par exemple retrouver tous les remèdes contenant des fleurs de grenadier – il est intéressant de décomposer ces ingrédients combinés en les ingrédients qui les composent.

Cependant, les désignations d'ingrédients combinés peuvent s'interpréter de deux façons : soit la combinaison est liée à la syntaxe du langage naturel qui a tendance à éviter la répétition de l'objet – graines et fleurs de grenade – facile à décomposer par un traitement syntaxique à base de règles (*A et B de C* devient *A de C et B de C* : graines de grenade et fleurs de grenade) ; soit elle est liée à la fabrication, par transformation, de l'ingrédient – l'eau de céleri et de fenouil : dans ce cas, savoir si les deux ingrédients sont transformés (par infusion ou décoction) ensemble ou séparément nécessite une expertise du domaine ainsi que le retour au texte dans sa langue originale.

Pour garder la double information, de la combinaison des ingrédients et de leur individualité, le modèle de la base pourrait être modifié en introduisant un type de nœud `CombinedIngredient`, lié aux nœuds `Ingredient` qui le composent. Cela permettrait de saisir des informations plus détaillées tout en maintenant le lien entre la source originale et les données modélisées.

5 Conclusion et perspectives

Le travail présenté ici s'inscrit dans un projet pluridisciplinaire visant à exploiter les pharmacopées anciennes dans le but de créer de nouveaux médicaments, en particulier en alternative aux antibiotiques. Nous avons dans un premier temps travaillé sur un texte arabe médiéval, traduit en anglais, dont nous avons extrait les descriptions des remèdes, comprenant les symptômes et les organes traités ainsi que les ingrédients qui les composent.

Nous avons choisi de stocker ces éléments dans une base de données orientée graphe. Nous rendons compte dans cet article des différents problèmes de modélisation que nous avons rencontrés et traités comme la prise en compte des sous-parties des plantes et des ingrédients transformées, ou qui nécessitent un approfondissement tel que les ambiguïtés sémantiques et syntaxiques.

Le cadre offert par le modèle de graphe s'est révélé pertinent à la fois pour faciliter la représentation des informations complexes que nous manipulons et l'interrogation des données, mais aussi pour visualiser les relations entre les différents éléments des remèdes extraits du manuscrit. Cette première base de données sur les remèdes anciens suscite beaucoup d'intérêt de la part des biologistes, mais nécessite d'être complétée pour qu'ils puissent l'utiliser dans leur recherche d'ingrédients utiles à la création de médicaments.

Le travail réalisé ouvre de nouvelles perspectives pour approfondir l'exploration des connaissances contenues dans

les manuscrits anciens de médecine arabe. Dans un premier temps, il s'agit de développer des modes d'interrogation de la base de données permettant de mettre en évidence des connaissances pertinentes en utilisant l'analyse de concepts formels, comme suggéré par Braud et *al.* [2]. Parallèlement, la base de données sera complétée à partir d'autres textes, dont l'annotation est en cours. Enfin, l'analyse des remèdes pourra être enrichie par une représentation explicite des différentes taxonomies utilisées, telles que les plantes, les parties de plantes, ainsi que les transformations des ingrédients. Ces taxonomies devront être reliées à des ressources externes afin de rendre ces données réutilisables et interopérables. Ainsi complétée, la base de données pourra alors être publiée.

Remerciements

Cette étude a été financée par le CNRS dans le cadre de l'appel à projets MITI80 (2021). Nous remercions tous ceux qui ont contribué à la réalisation de ce travail : nos collègues biologistes, pharmacologues, ainsi que Anthony Masiala (Master 1 en Botanique de l'Université de Strasbourg) qui a collecté les connaissances sur les plantes.

Références

[1] Amitabha Bhattacharyya and Durgapada Chakravarty. Graph database : A survey. In *2020 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, pages 1–8, 2020.

[2] Agnès Braud, Xavier Dolques, Pierre Fechter, Nicolas Lachiche, Florence Le Ber, and Veronique Pitchon. Analyzing the composition of remedies in ancient pharmacopeias with FCA. In *RealDataFCA'2021, ICFCA Workshop, Strasbourg, France*, CEUR Workshop Proc. 3151, pages 28–35, 2021.

[3] Jean Charlet, Gunnar Declerck, Ferdinand Dhombres, Pierre Gayet, Patrick Miroux, and Pierre-Yves Vandenbussche. Construire une ontologie médicale pour la recherche d'information : problématiques terminologiques et de modélisation. In *23es journées francophones d'Ingénierie des connaissances, Paris, France*, pages 33–48, 2012.

[4] Erin Connelly, Charo I. del Genio, and Freya Harrison. Data mining a medieval medical text reveals patterns in ingredient choice that reflect biological activity against infectious agents. *mBio*, 11(1), 2020.

[5] Karim El Haff, Wissam Antoun, Florence Le Ber, and Véronique Pitchon. Reconnaissance des entités nommées pour l'analyse des pharmacopées médiévales. In *EGC 2023 - Extraction et Gestion des Connaissances, Lyon, France*, volume RNTI-E-39, pages 329–336, 2023.

[6] T. R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*. Kluwer Academic Publishers, 1993.

[7] Freya Harrison, Aled E. L. Roberts, Rebecca Gabrilska, Kendra P. Rumbaugh, Christina Lee, and Stephen P. Diggle. A 1,000-Year-Old Antimicrobial Remedy with Antistaphylococcal Activity. *mBio*, 6(4) :e01129–15, 2015.

[8] Oliver Kahl. *Sābūr Ibn Sahl's Dispensatory in the Recension of the 'Aḍudī Hospital*. BRILL, 2009. Arabic edition and English translation.

[9] Sara Hosseinzadeh Kassani and Peyman Hosseinzadeh Kassani. Building an Ontology for the Domain of Plant Science using protégé, 2018. arXiv :1810.04606.

[10] Wei Liu, Dongxue Yin, Na Li, Xiaogai Hou, Dongmei Wang, Dengwu Li, and Jianjun Liu. Influence of Environmental Factors on the Active Substance Production and Antioxidant Activity in Potentilla fruticosa L. and Its Quality Assessment. *Scientific Reports*, 6(1) :28591, 2016.

[11] Thierry Millan. Utilisation des bases de données orientées graphe comme référentiels de modèles. *Revue des Sciences et Technologies de l'Information - Série TSI*, 35(6) :695–719, 2017.

[12] Josiane Mothe and Sagun Pai. Mise en œuvre d'une base de données graphe pour l'analyse des logs de requêtes en recherche d'information. In *14eme Conference francophone en Recherche d'Information et Applications (CORIA 2017), Marseille, France*, pages pp. 43–58, 2017.

[13] Lance A. Ramshaw and Mitchell P. Marcus. Text chunking using transformation-based learning. *CoRR*, cmp-lg/9505040, 1995.

[14] Marko A. Rodriguez and Peter Neubauer. The Graph Traversal Pattern, 2010. arXiv :1004.1001 [cs].

[15] Pierre J. Silvie, Pierre Martin, Marianne Huchard, Priscilla Keip, Alain Gutierrez, and Samira Sarter. Prototyping a Knowledge-Based System to Identify Botanical Extracts for Plant Health in Sub-Saharan Africa. *Plants*, 10(5), 2021.

[16] Some, Borlli Michel Jonas, Georgeta Bordea, Frantz Thiessard, and Gayo Diallo. Enabling West African Herbal-Based Traditional Medicine Digitizing : The WATRIMed Knowledge Graph. In *MEDINFO 2019 : Health and Wellbeing e-Networks for All*, pages 1548–1549. IOS Press, 2019.

[17] Santiago Timón-Reina, Mariano Rincón, and Rafael Martínez-Tomás. An overview of graph databases and their applications in the biomedical domain. *Database*, 2021 :1–22, 2021.

[18] Ramona L. Walls, Laurel Cooper, Justin Elser, Maria Alejandra Gandolfo, Christopher J. Mungall, Barry Smith, Dennis W. Stevenson, and Pankaj Jaiswal. The plant ontology facilitates comparisons of plant development stages across species. *Frontiers in Plant Science*, 10, 2019.