

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

Les journées francophones d'Ingénierie des Connaissances –
PFIA 2023

Auteurs :

Karim EL HAFF (ICube, Archimède)

Agnès BRAUD (ICube)

Florence LE BER (ENGEES, ICube)

Véronique PITCHON (Archimède)



Université



de Strasbourg

Sommaire

1 | Motivations

2 | Description des données

3 | Questions de modélisation

4 | Conclusion

Sommaire

1 | Motivations

2 | Description des données

3 | Questions de modélisation

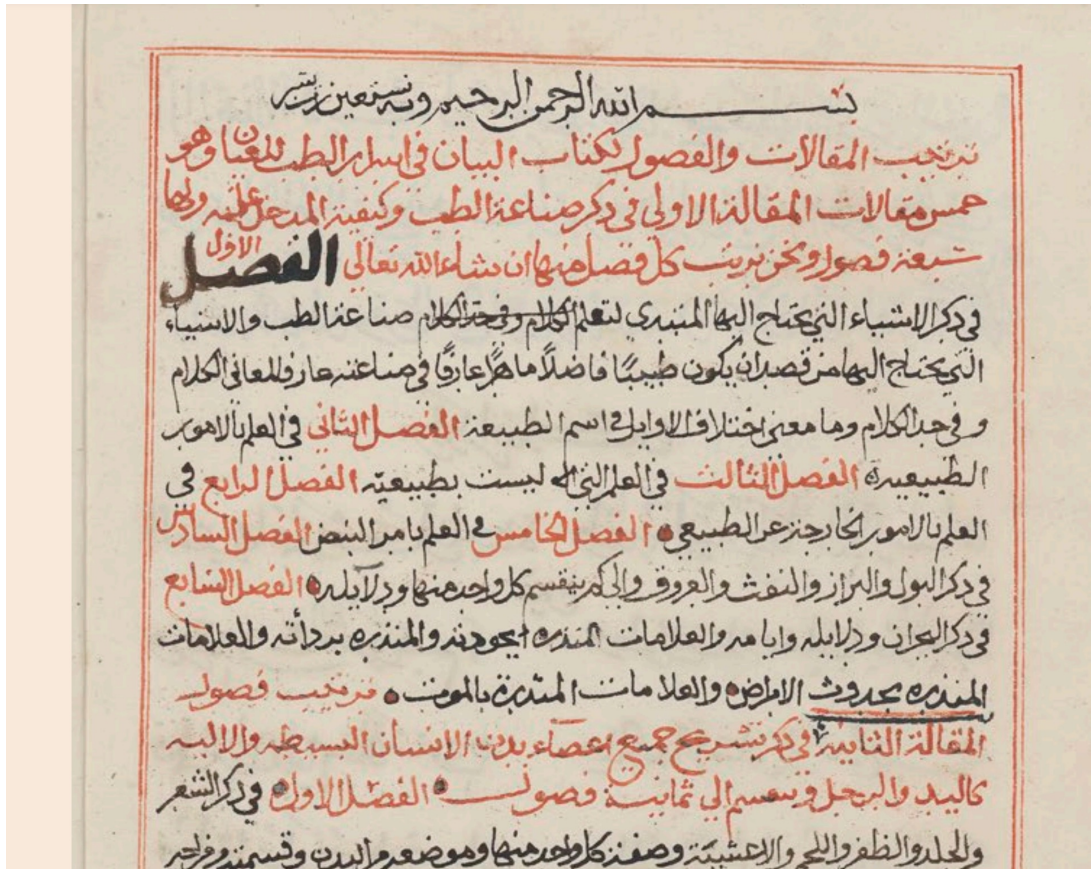
4 | Conclusion

Pourquoi modéliser les connaissances sur la médecine médiévale ?

- ◆ La science médiévale souvent négligée
- ◆ L'âge d'or de la médecine arabe = l'époque médiévale

→ Le but de conception de nouveaux médicaments nécessite l'extraction et la modélisation des données non-structurées

Pourquoi modéliser les connaissances sur la médecine médiévale ?

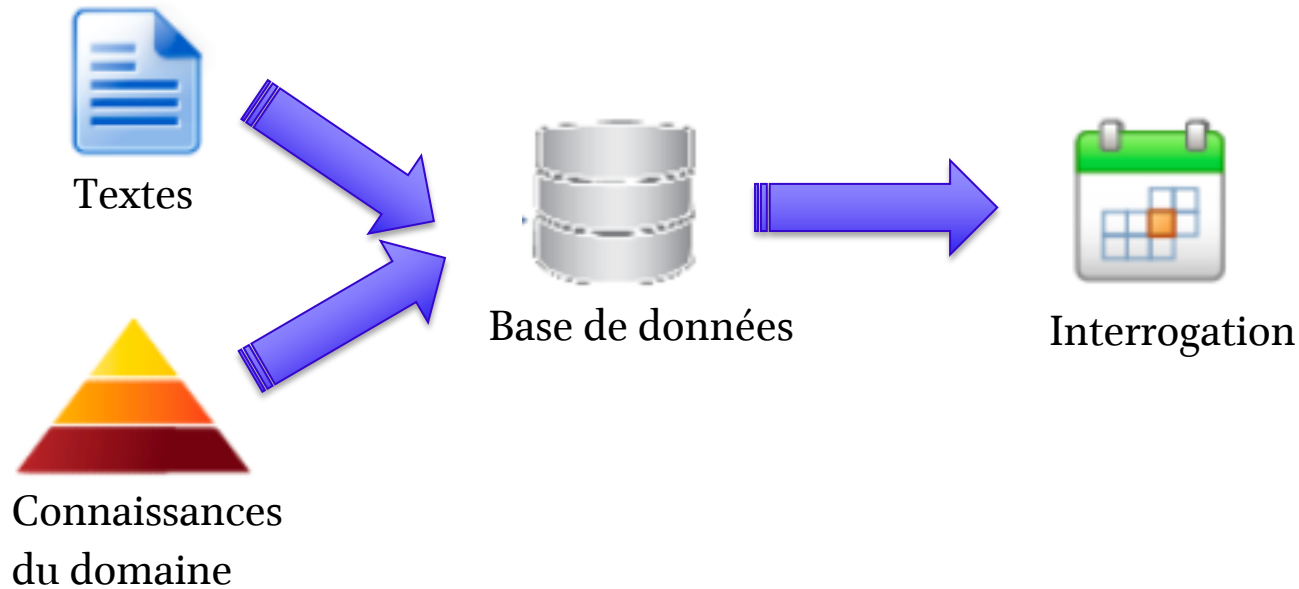


Pharmacopée, *nom*

- Recueil officiel des pharmaciens contenant la nomenclature des médicaments et leur description (composition, effets, etc.) (Larousse)

Exemple du manuscrit d'Al Hamaoui à la BNU, Strasbourg

Pourquoi modéliser les connaissances sur la médecine médiévale ?



Notre article se focalise sur l'étape de la modélisation des connaissances de cette chaîne de traitements.

1 | Motivations

2 | **Description des données**

3 | Questions de modélisation

4 | Conclusion

Texte choisi pour tester l'approche

- ◆ « Sābūr ibn Sahl's Dispensatory in the Recension of the 'Aḍudī Hospital »
- ◆ Traduit vers l'anglais par Oliver Kahl
- ◆ Manuellement annoté (environ 35,000 tokens, 292 remèdes) pour identifier les entités utiles

Ce que nous devons savoir : les entités

- ◆ Type : la forme du remède (pastille, pilule, etc.)
- ◆ Sym : un symptôme de maladie
- ◆ Ing : un ingrédient utilisé
- ◆ Org : un organe mentionné

Notre premier corpus

Chapter One on the Preparation of Pastilles

[1]

The prescription of the *tabasheer* pastille with *alhagi* which is useful against thirst, acute fevers, and cramp. *Tabasheer* four *dirham*; starch one *dirham*; *alhagi* five *dirham*; serpent melon seeds and gourd (seeds) three *dirham* of each, light-coloured poppy one *dirham*; tragacanth and gum-arabic one *dirham* of each. Every single (ingredient) is pounded and strained separately. (This) is kneaded with the maceration of fleawort seeds, formed into pastilles of one *dirham* each, dried in the shade on the back of a sieve, (and) a potion (may be made by using) one pastille with barley-water.

Exemple d'un remède du livre « Sābūr ibn Sahl's Dispensatory in the Recension of the 'Aḍudī Hospital » traduit par Oliver Kahl

Données annotées

Un exemple d'annotation →

	A	B	C
1	Remedy	Word	Tag
2	Remedy: 1	The	O
3		prescription	O
4		of	O
5		the	O
6		tabasheer	B-ing
7		pastille	B-type
8		with	O
9		alhagi	B-ing
10		which	O
11		is	O
12		useful	O
13		against	O
14		thirst	B-sym
15		,	O
16		acute	B-sym
17		fevers	I-sym
18		,	O
19		and	O
20		cramp	B-sym
21		Tabasheer	B-ing
22		four	O
23		dirham	O
24		;	O
25		starch	B-ing

Le Format IOB2

maux B-sym

de I-sym

ventre I-sym

mettre O

une O

pomme B-ing

dans O

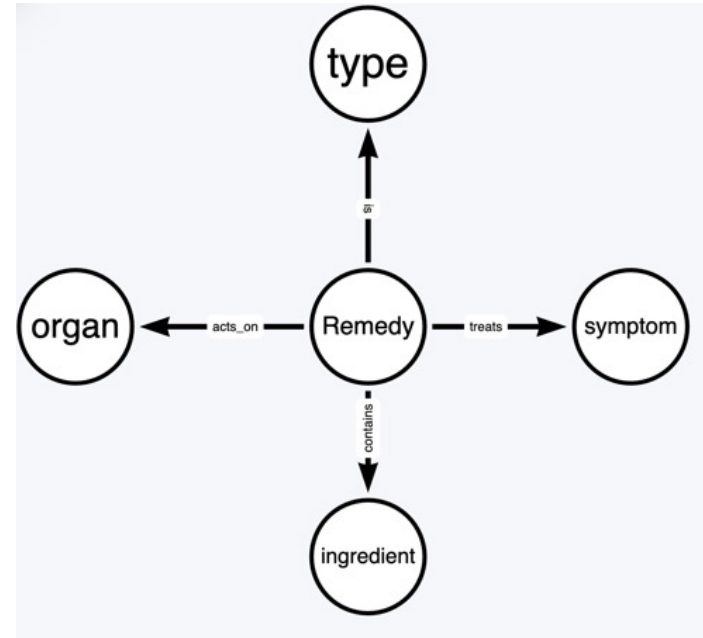
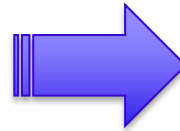
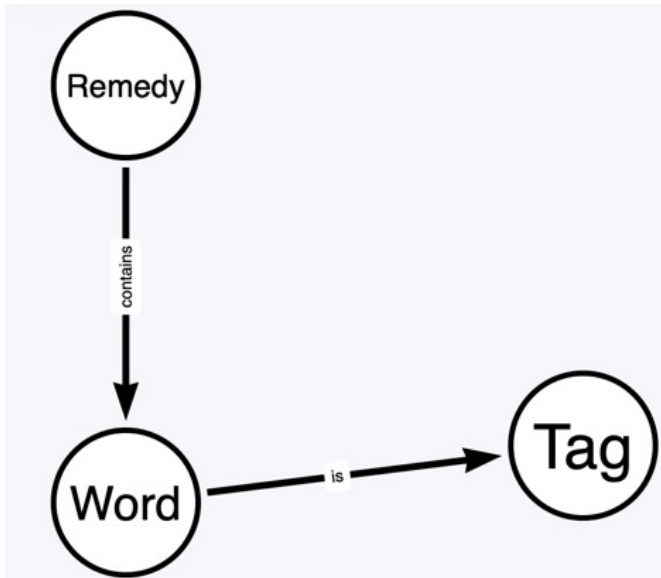
	A	B	C
1	Remedy	Word	Tag
2	Remedy: 1	The	O
3		prescription	O
4		of	O
5		the	O
6		tabasheer	B-ing
7		pastille	B-type
8		with	O
9		alhagi	B-ing
10		which	O
11		is	O
12		useful	O
13		against	O
14		thirst	B-sym
15		,	O
16		acute	B-sym
17		fevers	I-sym
18		,	O
19		and	O
20		cramp	B-sym
21		Tabasheer	B-ing
22		four	O
23		dirham	O
24		;	O
25		starch	B-ing

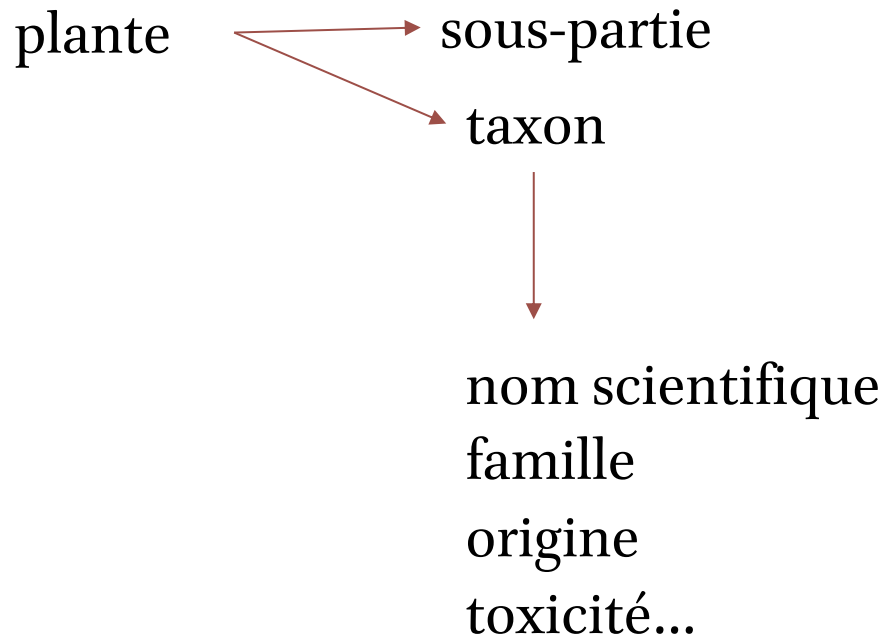
1 | Motivations

2 | Description des données

3 | **Questions de modélisation**

4 | Conclusion





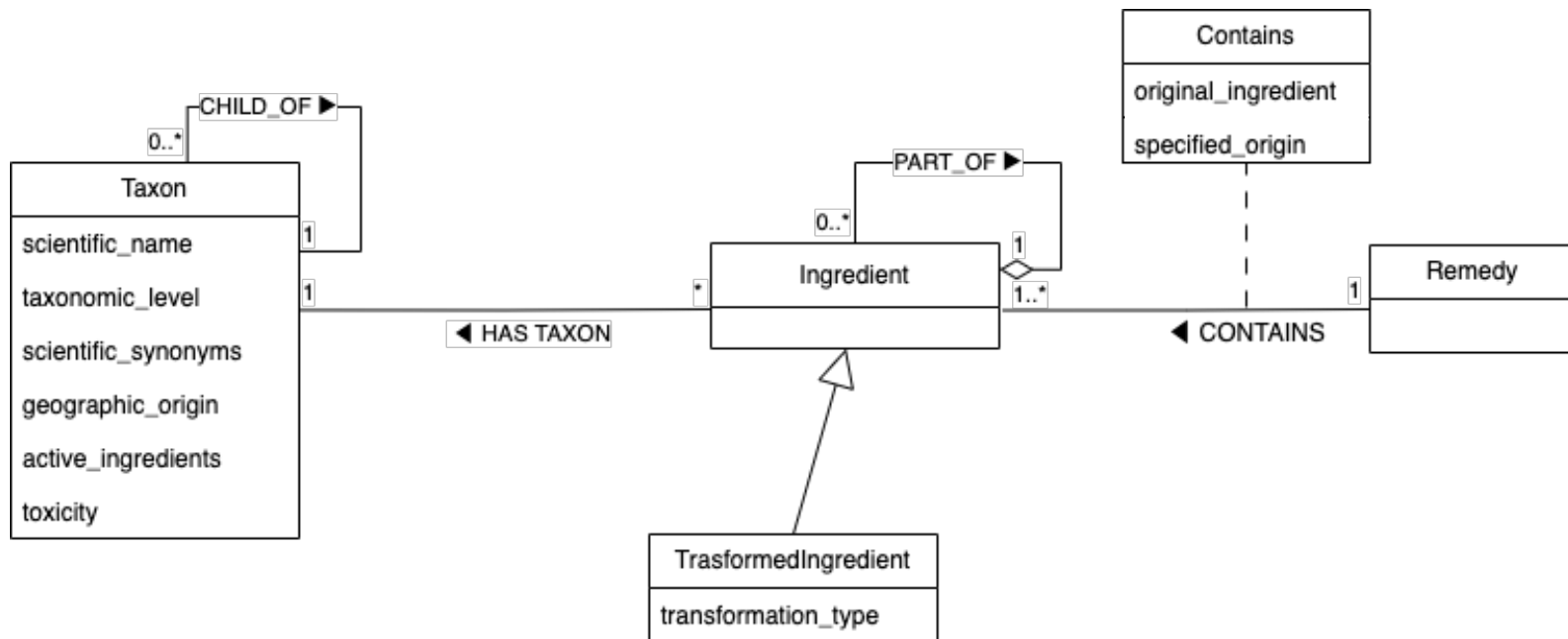
Enrichissement des données

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

Le rajout d'informations sur les ingrédients

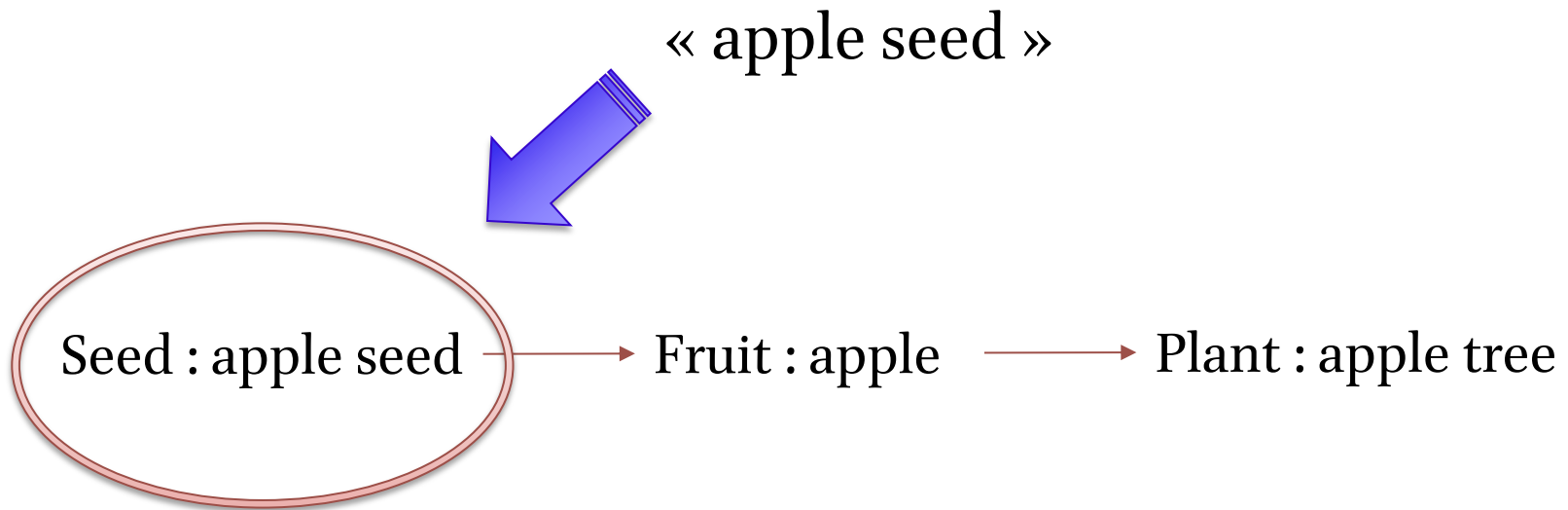
Ressources : *The World Flora Online* pour la botanique, *CHEMnet-BASE* et *Reaxys* pour les molécules naturelles.

Le modèle de données

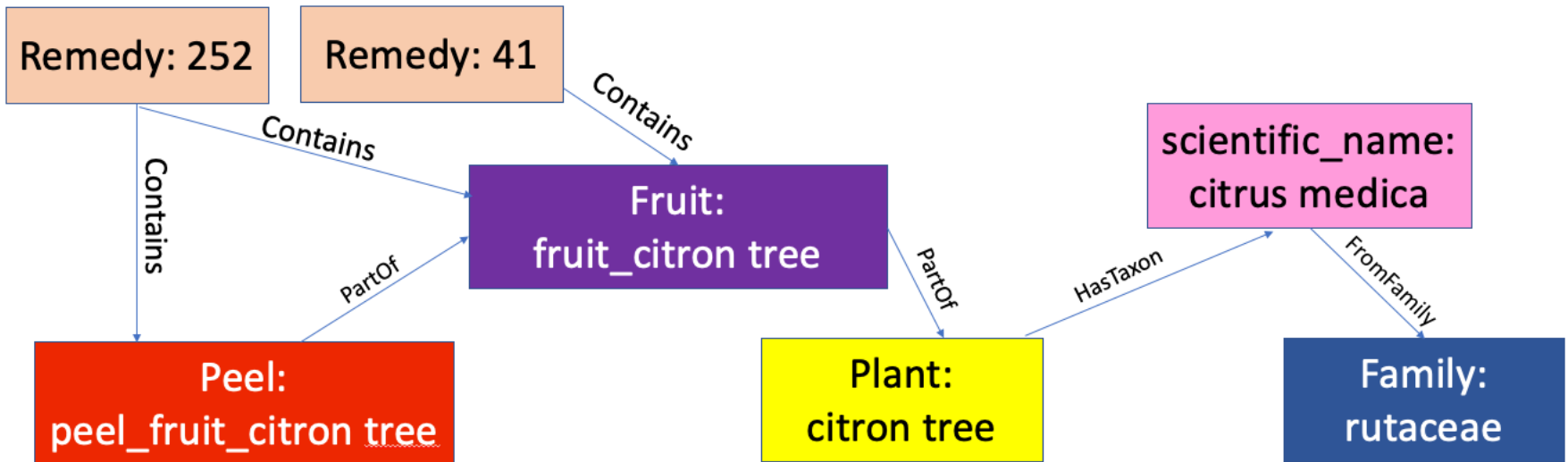


Les ingrédients sous-parties de plantes

Plusieurs ingrédients utilisés dans les remèdes sont des parties de plantes, telles que des graines, des fruits, des racines, des feuilles, etc.



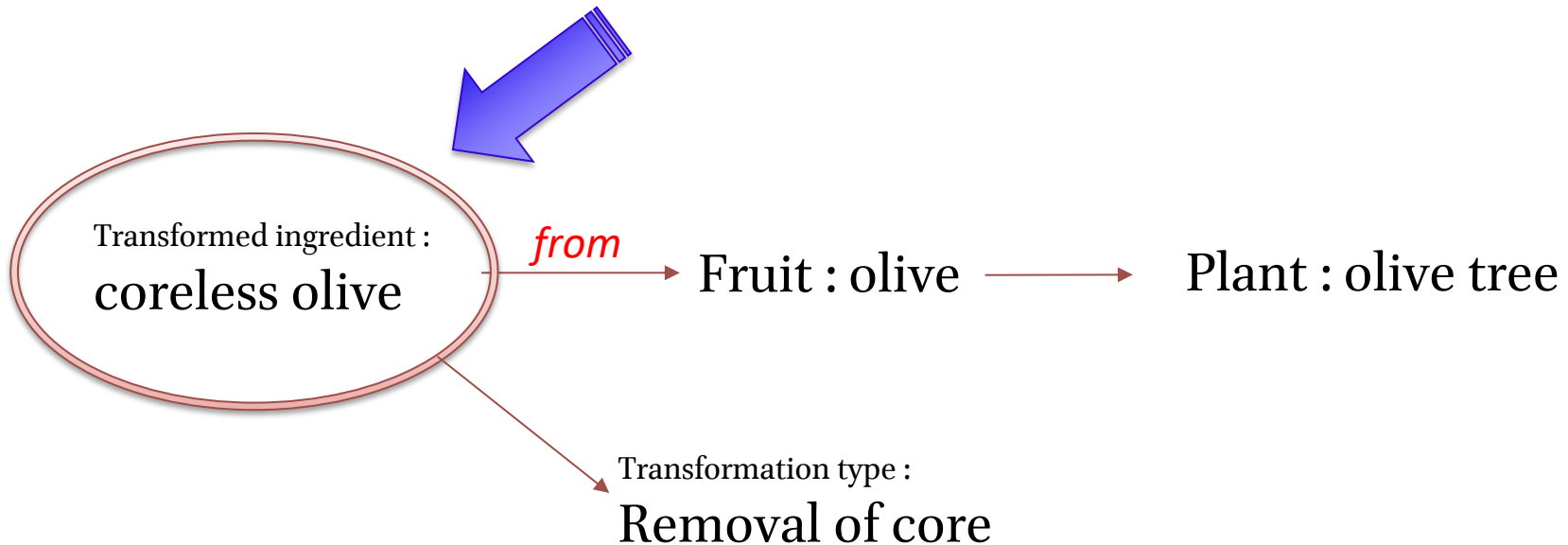
Les ingrédients sous-parties de plantes



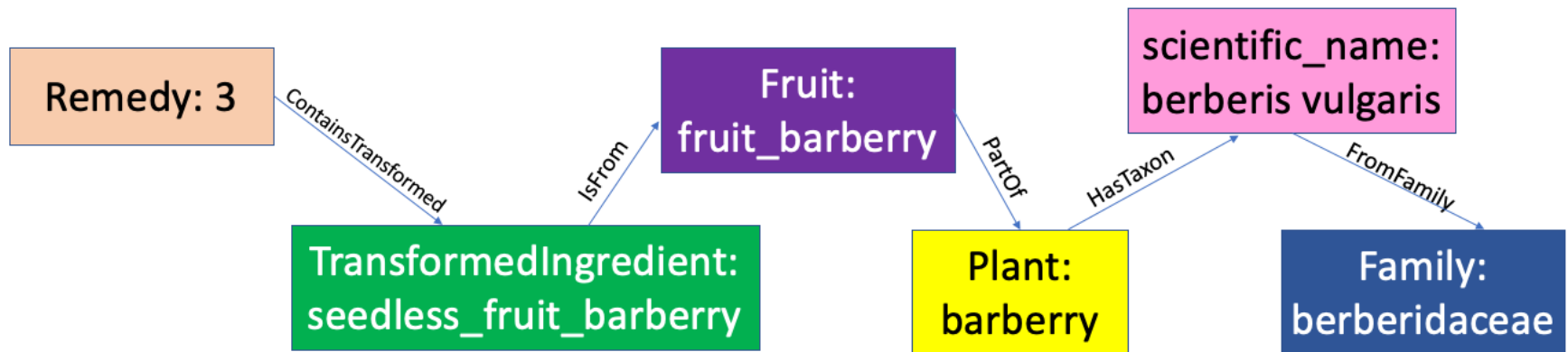
Les ingrédients transformés

Cette représentation permet d'inclure des ingrédients dont la description comporte une négation ou une soustraction.

« coreless olive »



Les ingrédients transformés



Questions ouvertes : ambiguïtés sémantiques

Appellation courante et nom scientifique :

Une appellation courante peut correspondre à différentes espèces.

Acorn (gland) : Quercus ? Lithocarpus ?

Questions ouvertes : ambiguïtés sémantiques

Indication d'origine dans une appellation courante :

Les désignations des ingrédients peuvent porter une potentielle information géographique

Syrian apples : d'origine syrienne ou nom courant?

Chinese rhubarb : d'origine chinoise ou nom courant ?

- 1 | Motivations
- 2 | Description des données
- 3 | Questions de modélisation
- 4 | Conclusion**

Une base de données exploitable



- ◆ Exploitable par des experts en biologie et en histoire
- ◆ Accessible (Neo4j)
- ◆ Flexible pour l'enrichissement

Un exemple de visualisation de base de données graphe avec Neo4j Browser

Pistes d'études

- ◆ La désambiguation des entités nommées
- ◆ Le liage des entités nommées à des ontologies
- ◆ La mise en place d'un système d'interrogation

وجميع الشجر والمخل والنخج واللبون ولذات القليل
 ووزن الثقل وهو من شجر الفحل والوقاح الربيع
 والنبات وهو الأتمل



ومن النابون من اسمه البادون وهو صلبان وذليل الأزمنة
 ماوزه شبهة بوزن السبرو وهو أكثر صوة ثابون حيرة
 من الأبل كبريه الزاجية وحيزه العتيبة صيرة و
 قرينه الغرضية كثر منبعا الطول ومن الشايب على



منه فاطمة طار ومنه منقذ صخر وجميع الشجر
 واللبون وهو ران البون واما حفاطه والذليل واللبان
 فترسنة بوجو عظمه مثل عظم الشجر ومنه ما يد حتر
 على عظم البان في حيزه انه كذا منقذ طابا الربيع حتر
 حيدر من بوزن ران وديال الحوسل وهو صلبان الحان
 حيزا طاب حيزه للمعارة واما حيزه فان حفاطه

Merci !

Contact

Karim EL HAFF

Doctorant en 2^{ème} année

Ouvert à la discussion !

ICUBE

ARCHIMEDE

Université de **Strasbourg**

karim.el-haff@etu.unistra.fr

