Exploring Old Arabic Remedies with Formal and Relational Concept Analysis

Vanessa Fokou, Karim El Haff, Agnès Braud, Xavier Dolques, Florence Le Ber, Veronique Pitchon

Exploring Old Arabic Remedies with Formal and Relational Concept Analysis *

Vanessa Fokou[1][0009−0009−3778−2708], Karim El Haff[12][0009−0000−0519−6418], Agnès Braud[1][0000−0003−3614−9141], Xavier Dolques[1][0000−0002−5579−1714], Florence Le Ber[1][0000−0002−2415−7606], and Véronique Pitchon[2]

Université de Strasbourg, ENGEES, CNRS, ICube UMR 7357, F 67000 Strasbourg
{vfokou, agnes.braud, dolques, florence.le-ber}@unistra.fr
Université de Strasbourg, CNRS, Archimède UMR 7044, F 67000 Strasbourg
pitchon@unistra.fr, karim.el-haff@etu.unistra.fr

Abstract. Exploring old pharmacopoeias is a promising way to find active ingredients that can be useful to design new drugs. Nevertheless, studying these texts is a laborious task for biologists. Therefore, an interdisciplinary project was undertaken: texts have been annotated to extract relevant information and represent it within a graph database. Formal Concept Analysis (FCA) and Relational Concept Analysis (RCA) have then been used to explore this database, in order to answer questions regarding remedies and their ingredients. This paper presents the data and some results obtained with FCA and RCA. It highlights the suitability of these approaches to explore these data and answer the needs of biologists.

Keywords: Formal Concept Analysis · Relational Concept Analysis · Text Data · Graph Database · Old Pharmacopoeia.

1 Introduction

Exploring old pharmacopoeias is a promising way to find active ingredients that can be useful to design new drugs, e.g. to fight against antibioresistance. Indeed historical manuscripts, particularly those originating from the Abbasid era, offer a treasure of forgotten knowledge [10]. Exploring these manuscripts involves several steps and expertise. Historians select and check the manuscripts, biologists select and test the ingredients. In between, methods are needed to extract, represent and navigate through information from these manuscripts, in order to answer biologists' questions, e.g. which ingredients often appear together in remedies for similar symptoms?

Formal Concept Analysis (FCA) and derived methods such as Relational Concept Analysis (RCA) appear as particularly suitable for querying such qualitative/relational data, as discussed in [18]. A preliminary work on data from

* This research is supported by ANR 21-CE23-0023 SmartFCA and CNRS MITI'80 2021 PARADISE.

old pharmacopoeias using FCA is described in [2]. The aim was to find out frequent or co-occurring ingredients in remedies prescribed for urinary problems, extracted from 5 pharmacopoeias. A table of 35 remedies described by their ingredients has been considered. The analysis of both concepts and rules have shown promising results in answering questions from biologists.

The work presented here focuses on a single manuscript from the 9th century that was chosen since it gives a systematic representation of medical knowledge at this period. Information has been extracted from this book [7], then checked and completed before being represented in a graph database. FCA/RCA have been used on this dataset and various models have been experimented to answer various questions, following the ideas discussed in [18].

The paper is organized as follows. Section 2 briefly introduces FCA and RCA. Section 3 describes the original data and the structure of the database. In Sect. 4, we present modelling variants and results obtained with FCA and RCA. Related works are summarized in Sect. 5. Section 6 is a conclusion.

2 FCA and RCA Basics

Formal Concept Analysis [8] consists of extracting conceptual structures in binary tables describing objects by their attributes, called a formal context. A formal context is a triple $K = (O, A, I)$ where O and A are sets of objects and attributes respectively and I is a binary relation between O and A, i.e., $I \subseteq O \times A$. We define the functions $f : \mathcal{P}(O) \to \mathcal{P}(A)$ and $g : \mathcal{P}(A) \to \mathcal{P}(O)$ such that $f(X) = \{y \in A \,|\, X \times \{y\} \subseteq I\}$ and $g(Y) = \{x \in O \,|\, \{x\} \times Y \subseteq I\}$. The concept lattice L computed from K is the set of concepts $\{(X, Y) \,|\, X \subseteq O, Y \subseteq A, f(X) = Y \text{ and } g(Y) = X\}$, provided with a partial order relation based on inclusion. X is the concept extent, Y is the concept intent. Table 1 (left-hand-side) presents two formal contexts, `remedies` where object remedies are described by their form, and `ingredients` where objects ingredients (plants) are described by the part of plant used. Figure 1 (left-hand-side) shows the concept lattices built on formal contexts `remedies` and `ingredients`.

Relational Concept Analysis [9] is an extension of Formal Concept Analysis [8] which considers relational data, formalized within a Relational Context Family (RCF). An RCF is a pair $(\mathbf{K}, \mathbf{R})$ where $\mathbf{K}$ is a set of object-attribute contexts[1] (each context corresponding to an object category) – and $\mathbf{R}$ is a set of object-object/relational contexts (relations between objects of the same or various categories). For illustration, Table 1 (right-hand-side) introduces a relational context `composition` linking the objects of contexts `remedies` and `ingredients`.

The principle idea of RCA consists in integrating object-object relations as new attributes (called *relational attributes*) in the object-attribute contexts of $\mathbf{K}$ thanks to scaling quantifiers. It produces a set of concept lattices in an iterative way – one lattice per object category – interconnected through relational attributes. This set of lattices is called a Concept Lattice Family (CLF). The concepts in a given lattice group objects according to the shared attributes and to

[1] We use the term 'formal context' for FCA, and 'object-attribute context' for RCA.

Table 1: RCF example about remedies and their ingredients.

object-attribute contexts

remedies	pill	potion
remedy1	×	
remedy2	×	×
remedy3		×

ingredients	fruit	seed	bark
cinnamon			×
seed_celery		×	
acorn	×	×	

object-object contexts

composition	cinnamon	seed_celery	acorn
remedy1			×
remedy2	×	×	
remedy3	×		×

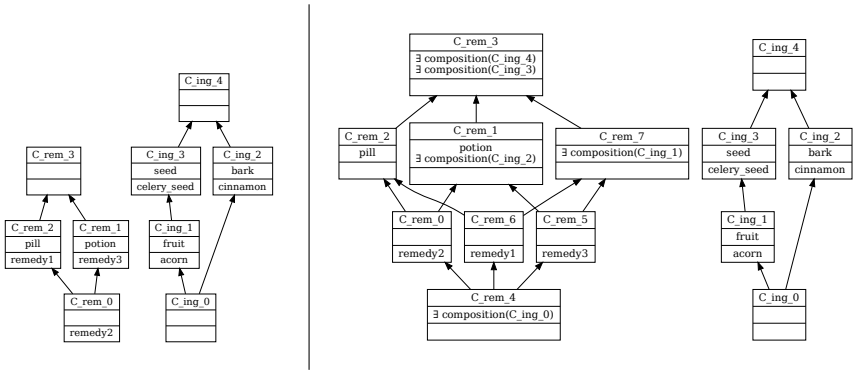

Fig. 1: RCA results on the RCF of Table 1 – initialisation (left) and at the end of RCA process (right).

the shared connections they have with objects of another category. RCA process is as follows. First, lattices are built on the two formal contexts, `remedies` and `ingredients` (Fig. 1, left-hand-side). Then the `remedies` context is extended by relational attributes linking `remedies` objects to `ingredients` concepts, based on the `composition` relational context as shown in Fig. 1 (right-hand-side). The concept `C_rem_1` with $Extent(\texttt{C\_rem\_1}) = \{\texttt{remedy2}, \texttt{remedy3}\}$, has an attribute `potion` and a relational attribute $\exists\, \texttt{composition}(\texttt{C\_ing\_2})$ in its intent. This means that at least one ingredient of $Extent(\texttt{C\_ing\_2})$ is a constituent of each remedy `remedy2`, `remedy3`.

In this work, we use the existential ($\exists$) and universal strict ($\exists\forall$) scaling quantifiers and the corresponding scaling operations are defined as follows [3]. $K = (O, A, I)$ and $K_r = (O_r, A_r, I_r)$ are two object-attribute contexts, r is a relation where $dom(r) = O$, and $ran(r) = O_r$; $\mathcal{C}_r$ is the concept set built on K_r. The image set of $o \in O$ is denoted by $r(o) = \{o_2 \in O_r | (o, o_2) \in r\}$.

Definition 1 (Existential Scaling). *For $o \in O$ and $C_i \in \mathcal{C}_r$, if $r(o) \cap Extent(C_i) \neq \emptyset$, then the relational attribute $\exists r(C_i)$ is added to the attribute set of o.*

Definition 2 (Universal strict Scaling). *For $o \in O$ and $C_i \in \mathcal{C}_r$, if $r(o) \neq \emptyset$ and $r(o) \subseteq Extent(C_i)$, then $\exists\forall r(C_i)$ is added to the attribute set of o.*

A generality relation can be established between quantifiers, e.g. $\exists\forall \preceq \exists$ which means that the concept introducing an attribute $\exists r(C)$ includes the extent of the concept introducing $\exists\forall r(C)$ [3]. More details about relational attributes can be found in [9, 3].

3 Exploiting Old Arabic Pharmacopoeias

3.1 Text Processing

Our main corpus is made up of the annotations of the complete remedies of a pharmacopoeia: Oliver Kahl's English translation of the work " Dispensatory in the Recension of the 'Aḍudī Hospital" written by Sābūr ibn Sahl in the 9th century [11]. The dispensatory, attributed to Sābūr ibn Sahl, a prominent Persian Christian physician and pharmacologist operating at the Academy of Gondishapur before moving to Baghdad, offers a unique glimpse into the pharmacological practices of the time. Sābūr ibn Sahl's work, especially through its recension under the auspices of the 'Aḍudī Hospital, underscores a systematic approach to drug composition and therapeutic applications. We chose this manuscript for its well-structured style of writing, as the text has also been edited by the translator-historian who is the chooser of the authoritative copy of the old manuscript. The choice of a well-structured corpus was made because it enhances the accuracy of data extraction, as relevant entities and relationships are more easily identified. A clear and organised manuscript reduces the complexity of preprocessing steps, saving time and resources in the overall work.

The first half of the document contains the text in its original Arabic whereas the second half is the English Translation by Oliver Kahl, the latter half being the focus of our work. The corpus is divided into chapters based on drug categories or therapeutic applications, such as pastilles, lohochs, beverages, oils, cataplasms, enemas, powders, and collyria. Each entry within the chapters provides information on the preparation, dosages, and intended therapeutic use of the compounds. In total, the corpus describes 292 remedies encompassing a wide array of substances and ingredients from various geographical origins including vegetable, animal, mineral, and, occasionally, human substances. Of those 292 remedies, 287 contain plant-based substances that are analysed in the following.

Within each chapter, individual entries provide detailed information for each described remedy. These entries typically include:

- Name and Description: the name of the drug and a brief description of its intended use or therapeutic properties and the symptoms or pathologies it aims to treat.
- Ingredients: a list of components used in the preparation, often with precise quantities or proportions. This includes a diverse array of substances.
- Preparation instructions: steps or actions to be taken for preparing the pharmaceutical compound.

The prescription of the pomegranate flower pastille which is useful for (the treatment of) abrasion, haemorrhage, dysentery, and bloody expectoration. Cassia, Armenian bole, and gum-arabic four dirham of each ; gumsenegal, stalkless roses, and pomegranate flowers eight dirham of each; tragacanth one and a half dirham. All (this) is pounded, kneaded with the cooked water of fresh pomegranate flowers or rose-water, dried, (and) a potion (may be made by using) two dirham from it.

Fig. 2: An excerpt from the pharmacopoeia (remedy 17) [11].

– Application and Dosage: guidelines for the administration of the medicine as well as dosages.

Figure 2 showcases an excerpt from the translated manuscript, describing a remedy denoted by "pomegranate flower pastille" and useful for various symptoms. The remedy is described as a recipe, with a list of ingredients (e.g. Cassia, tragacanth, etc.) and preparation (kneaded, dried) and application (potion) methods. Dirham is a weight measure. Most of the ingredients are plants or plant parts.

The corpus' tokenisation was performed using NLTK [1] to prepare the corpus for annotation. The annotation was performed manually by a computational linguist part-time over a period of one month, then reviewed in depth by an expert historian of medieval Arabian medicine, benefiting from the original available Arabic text to increase the understanding of entities and their annotation. The corpus is made up of 36,961 tokens, which were annotated with custom labels. To carry out the annotation, 4 types of labels were used:

– Type: the form of the remedy (pastille, pill, etc.);
– Sym : a symptom of a disease;
– Ing: a used ingredient;
– Org: a mentioned organ.

This resource can serve as the basis for the Named Entity Recognition task in order to analyse other books, as described by [7].

3.2 A Database on Old Remedies and their Ingredients

Textual entities from the corpus have been represented in a graph-oriented database (using Neo4j[2]). Remedies, symptoms, ingredients, etc. and their relations are formalised with respect to the model presented in Fig. 3.

The model is structured around the `CONTAINS` relationship (2890 instances) between the `Remedy` nodes (292 inst.) and `Ingredient` nodes (986 inst.) of which 716 are plant-based ingredients. This relationship is endowed with two attributes: the original name of the ingredient in the manuscript, as well as a specified geographic origin in the name where applicable, such as *Antioch* for the ingredient *Antioch scammony* (scammony of Antioch). Ingredients can be an entire plant

[2] https://neo4j.com/fr/

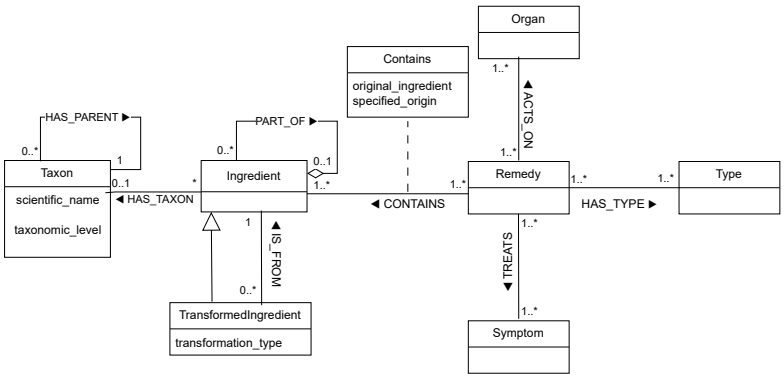

Fig. 3: The database model represented by a UML Diagram

or a part of a plant: `Ingredient` node is then linked by the `PART_OF` relationship (265 inst.) to another `Ingredient` node (plant or part of plant). An `Ingredient` node is also linked to a `Taxon` node (268 inst.) through the `HAS_TAXON` relationship (461 inst.) which indicates the scientific name of the species. Moreover, each `Taxon` node has a `HAS_PARENT` relationship that further provides the parent of the species to the respective family in the plant kingdom, making up 84 distinct families. Finally, our model takes into account the transformations of ingredients in remedies. Out of the 2890 instances of the `CONTAINS` relationship, 572 of them link the remedies to the transformed ingredients (`TransformedIngredient` node, 266 inst.). The `IS_FROM` relationship (265 inst.) allows linking the transformed ingredients to the original ingredient, with the `transformation_type` attribute allowing for the categorization of the transformation.

Representation of Parts of Plants. Many ingredients in remedies are parts of plants, such as seeds, fruits, roots, and leaves. Each part of a plant has been modelled as a node in a hierarchical tree alongside the `Ingredient` label (not shown in the model for the sake of simplicity). The root of the hierarchical tree is the "whole plant" node, which is linked to the taxon node corresponding to the plant in our model. Figure 4 shows an extract of the database with ingredients being various plant parts. In order to make the parts of the used ingredient explicit and to make its reading easier in this paper, we have embedded the parts in the name of the ingredient. For instance, if the original ingredient is *citron peels*, it would be represented as "peel_fruit_citron tree". Here, three categories of plant parts are obvious: the peel, the fruit, and the tree (the whole plant).

Representation of Transformed Ingredients. A second inquiry addresses how transformed ingredients are represented, particularly when their chemical composition and medicinal properties have been modified from their original form. In the graph database, transformed ingredients are denoted by the node type `TransformedIngredient`. These nodes are connected to their unmodified counterparts through a `IS_FROM` relationship, which includes an attribute detailing the

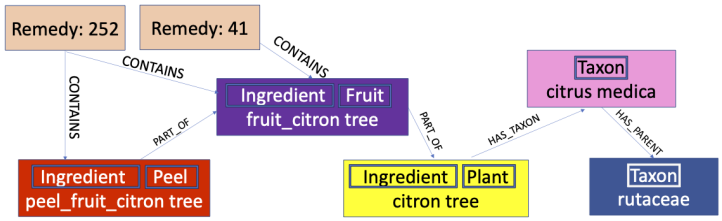

Fig. 4: Representation of remedies which contain ingredients that are parts of the citron tree.

transformation process, such as drying or grinding. The `TransformedIngredient` node is subsequently connected to the remedy by a `CONTAINS` relationship, inherited from `Ingredient` node. Figure 5 presents an example from the database. This setup effectively captures both the original and the transformed state of the ingredient, as well as the transformation details.

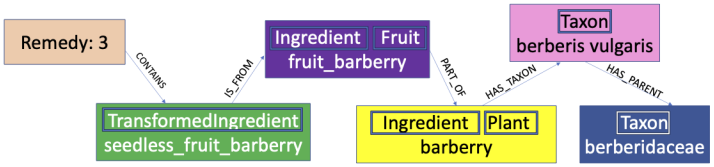

Fig. 5: Representation of a remedy which contains a transformed part of the barberry plant.

4 Querying Old Remedies with FCA and RCA

When looking at the pharmacopoeia books, biologists interested in the conception of new drugs ask themselves a number of questions, as for example: "which ingredients appear the most often?" (Q1); "which ingredients often appear together?" (Q2); "which remedies appear together in a specific form?" (Q3); "are there groups of ingredients that can be associated to groups of symptoms?" (Q4). To answer these questions, we first perform FCA on a formal context that describes remedies by ingredients, and we then perform RCA on a dataset about remedies, their ingredients and the symptoms they treat.

4.1 Analysing Ingredients Frequencies and Co-occurences with FCA

The data model we treat in this section is a formal context that describes remedies by their ingredients; it is made up of 287 remedies (rows) and 586 ingredients

(columns). Table 2 presents a small extract of this formal context. It is denoted by $\texttt{Dataset}_1$ in the following and will enable us to answer questions Q1 and Q2. The analysis carried out on $\texttt{Dataset}_1$ is in line with the work conducted in [2] namely the search of the frequent and co-occurring ingredients, which are the information of interest to biologists. In addition, grouping remedies according to ingredients they share will also provide information on remedies that have the most ingredients in common which may lead to questions about possible links between the symptoms treated by these remedies.

Table 2: A small extract from the `remedies-ingredients` context ($\texttt{Dataset}_1$).

`remedies/ingredients`	alhagi	asarabacca	barberry	barley	camphor tree	fruit_barberry	...
Remedy: 1	×			×			
Remedy: 2					×	×	
Remedy: 3	×		×		×	×	
Remedy: 4	×	×	×			×	
...							

The lattice obtained on $\texttt{Dataset}_1$ contains 1158 concepts (not counting the $\top$ and the $\bot$)[3]. A top-down traversal of this lattice allows an exploration of the most frequent to the least frequent ingredients as well as the associated remedy clusters. As we are interested in the most frequent ingredients (Q1), we focus on the most general concepts, i.e, concepts with the greatest extents. Table 3 summarizes some general concepts with their extent cardinality in brackets, and the detail of their intent (ingredients). The first, second and third rows of this table describe respectively concepts where intent has only one, two and three elements. For instance, `plant for wine (51)` represents the concept of remedies having `plant for wine` as intent and containing `51 remedies` in its extent. Table 3 thus reveals that `plant for wine` is the ingredient that appears the most often, followed by the ingredient `rose` appearing in `47 remedies`. Indeed, 51/287 remedies have wine or wine vinegar as ingredient, probably used as thinner, and generally without precision on the original plant, grapes or others. Besides, the ingredient `rose` comes in different forms, rose water (12), rose oil (23) or dry/stalkless rose: roses could have healing and soothing properties. In the same way, {`plant for wine, rose`} `(11)` materialises the concept of remedies having {`plant for wine, rose`} as intent and `11 remedies` in the extent. This table is just an excerpt of the information revealed by this analysis.

We have also identified the least frequent sets of ingredients, those that only appear (together) in the composition of a single remedy. Table 4 shows some of these ingredient sets (intent) with the associated remedy in columns (extent). Such remedies or ingredients also deserve further analysis, as they may reveal relevant information. For instance, `Remedy: 149` has 21 specific ingredients (out of 65) that do not appear in other remedies; this remedy is used for about 20 various symptoms, it seems like a panacea, the role of these ingredients in its

[3] In the rest of the paper, $\top$ and $\bot$ concepts are not considered in the analyses.

Table 3: Summary of the most general remedies concepts.

One ingredient in concept intent	- {plant for wine} (51) - {saffron} (33) -{seed_sesamum} (29) - {bark_cinnamon tree} (23)	- {rose} (47) - {indian spikenard} (32) - {plant of vinegar} (23) - {sap_tragacanth} (22	- {sap_acacia (34) - {mastic} (32) - {ginger} (23) -{fruit_olive tree} (21) ...
Two ingredients in concept intent	-{plant for wine, saffron} (16) -{sap_tragacanth, sap_acacia} (14) -{ginger, bark_cinnamon tree} (13) -{plant for wine, rose} (11) ...	- {saffron, indian spikenard} (14) -{plant for wine, indian spikenard} (13) - {mastic, indian spikenard} (12)	
Three ingredients in concept intent	-{ginger, long pepper, black pepper} (9) -{ginger, bark_cinnamon tree, long pepper} (8) -{saffron, ginger, bark_cinnamon tree} (7) -{clove, ginger, long pepper} (7) ...	- {saffron, plant for wine, bark_cinnamon tree } (9) - {indian spikenard, long pepper, black pepper} (8) -{plant for wine, myrrh, bark_cinnamon tree} (6)	

Table 4: A few sets of ingredients that only appear in a single remedy.

Remedy: 149	Remedy: 43	Remedy: 124
{cypress, dīqāmanūn iqrīṭī, faṭrāsāliyūn, fruit_Indian caraway, fruit_qardamānā, greek hypericum, hop marjoram, hyssop-water, hūfārīqūn, latex_qufiyun nārdīn iqlīṭī, sap_ūfāqiṣṭīdās, seed_rue, seed_babylonian garden peppercress seed_milk parsley, seed_white mustard, seed_wild celtic carrot, sweet flag ṭalāsfiyus, unspecified_roots water, usqūrdiyūn, valerian} (21)	{bark_pandanus, yellow sandalwoo root_fennel, stalk_fennel, kadar, pandanus, seed_fruit_pomegranate tree, unspecified_old white wine vinegar} (8)	{fruit_Syrian carob-tree, lote-tree, fruit_mulberry-tree, quincetree, stalk_service-tree, fruit_date-palm, unspecified_stalks} (7)

composition would be difficult to analyse. In the same way, `Remedy: 43` has 8 specific ingredients (out of 17) that do not appear in other remedies. The case of `Remedy: 124` is somehow special because it is made up of a total of 8 ingredients and 7 are specific to it. The only ingredient it shares in common with other remedies is `sandalwood` (actually a beverage made from sandalwood and used to dissolve the remedy).

As the concept lattice is very large, a good way to search for the co-occurring ingredients is to generate the base of implications. Table 5 summarizes the results of the implication rules obtained from $\texttt{Dataset}_1$, indicating the number of rules per support, and Table 6 shows some relevant implications rules. One rule appeared with a support of 9: `liquorice, sap_acacia` $\rightarrow$ `sap_tragacanth`. The three ingredients have various medicinal properties, the question is to explain how they combine in a remedy. Considering and exploring the sub-concepts of the concept that introduces `liquorice` and `sap_acacia` allows us to find out which ingredients other than `sap_tragacanth` also appear with these two ingredients. This analysis can then lead to a study of the respective roles of ingredients, to determine which ones have similar or complementary roles.

Table 5: Summary of implication rules.

Number of rules	Support
895	2
273	3
115	4
25	5
16	6
5	7
1	9

Table 6: Examples of implication rules obtained on $\texttt{Dataset}_1$.

Rule	Support
liquorice, sap_acacia → sap_tragacanth	**9**
black pepper, ginger, indian spikenard → long pepper	**7**
cassia, myrrh, plant for wine → saffron	**7**
bark_cinnamon tree, cassia → saffron	**6**
mace → clove	**6**
seed_quince → sap_acacia	**5**

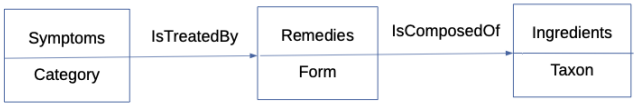

Fig. 6: Model for the Symptoms-Remedies-Ingredients RCF.

4.2 Linking Symptoms to Ingredients with RCA

To answer question Q4, we propose to use an RCF, called $\texttt{Dataset}_2$ and built according to the diagram depicted in Fig. 6, linking symptoms, remedies and ingredients. Remedies are described by their form, allowing to answer the question Q3. For this analysis, data are restricted to a subset of remedies treating *fever* symptoms, their ingredients and symptoms.

The RCF is defined by $(\mathbf{K}, \mathbf{R}) = (\{\texttt{symptoms, remedies, ingredients}\}, \{\texttt{isTreatedBy, isComposedOf}\})$: the `symptoms` context ($105\times9$) describes symptoms by their type (category) – these categories are based on expertise – the `remedies` context (26×13) describes remedies by their forms and the `ingredients` context (156×147) describes the ingredients by their taxons (species and family). The two relational contexts are as follows: the context `isTreatedBy` describes the fact that a *symptom is treated by a remedy* and the context `isComposedOf` describes the fact that *a remedy is composed of certain ingredients*.

As described in Sect. 2, RCA first calculates the concept lattices for each formal context before extending the object-attribute contexts with relational attributes, and updating the lattices. In the following, we first describe the concept lattices of each formal context of $\texttt{Dataset}_2$; then we apply two combinations of scaling quantifiers to update the concept lattice family: (1) the $\exists$ quantifier is used for both relations, (2) the $\exists\forall$ quantifier is used for relation `isTreatedBy`, while the $\exists$ quantifier is used for `isComposedOf`.

CLF Obtained on $\texttt{Dataset}_2$ at step 0. Lattices built on the initial formal contexts are briefly described by their number of concepts, and their most general concepts.

`Symptoms` lattice contains 9 concepts corresponding to the different expert categories, with the category `miscellaneous` grouping symptoms belonging to no category, so we focus on the other 8 categories during the analysis. These categories are ranked from the most frequent to the less frequent, specifying in brackets the number of symptoms: `fever` (19), `digestive` (12), `dermatological` (12), `neurological` (12), `hepatic` (6), `psychiatric` (5), `respiratory` (4), `hematological` (3). Names of symptoms are very diverse, e.g. for digestive symptoms: gastric debility, acid belching, pain in the belly, intestinal putridity ... that have to be interpreted.

`Remedies` lattice contains 13 concepts as presented in Fig. 7. `Pastille` is the most common form of remedy (`C_remedies_13`, with 11 remedies) followed by `potion` (`C_remedies_12`). This information may for example, lead to the question of whether the form of the remedy is related to the organ it treats. Besides,

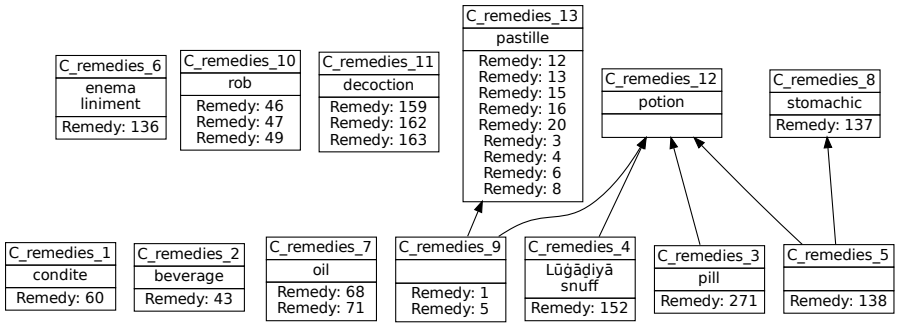

Fig. 7: `remedies` lattice from $\texttt{Dataset}_2$ (step 0) without $\top$ and $\bot$.

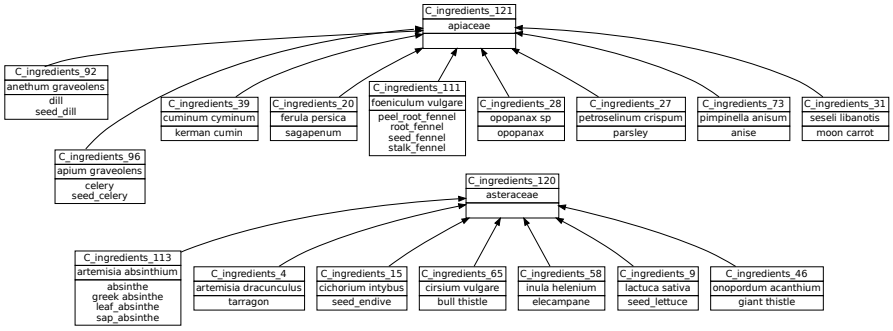

Fig. 8: Extract of the `ingredients` lattice from $\texttt{Dataset}_2$.

some remedies appear to have different forms such as {`Remedy: 1, Remedy: 5`} in `C_remedies_9` which have `pastille` and `potion` forms: actually in both remedies, a pastille is prepared and then dissolved before absorption to make a potion, highlighting how a medicine is preserved and administered.

`Ingredients` lattice contains 121 concepts grouping ingredients by their `family` and `genre species`. Concepts of ingredients grouped by `family` are interesting since plants of the same family may share properties. Figure 8 shows an extract from the `ingredients` lattice which highlights the two most numerous ingredient families: `apiaceae` with 14 ingredients (`C_ingredients_121`) and `asteraceae` with 10 ingredients (`C_ingredients_120`). We also have the families `rosaceae`, `fabaceae`, and `lamiaceae`, each associated with 9 ingredients. Concepts of ingredients grouped by `genre species` may be too specific, but they also reveal the usage of various parts of the same plant, e.g., `C_ingredients_111`, *Foeniculum vulgare* shows that roots, seeds and stalks of `fennel` are used as ingredients in the remedies. Note that `ingredients` lattice remains unchanged along the RCA process.

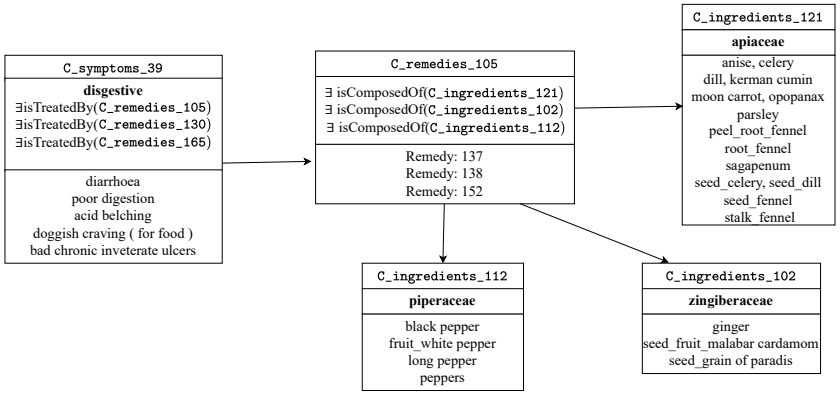

Fig. 9: An extract from RCA results on $\texttt{Dataset}_2$ with ∃ / ∃ quantifiers.

Existential Quantifier. The existential quantifier is applied on both relations to update `symptoms` and `remedies` lattices. Here we focus on `symptoms` or `remedies` concepts with a small number of elements in the extent.

A concept `C_symptoms_i` in the `symptoms` lattice has a relational attribute ∃`isTreatedBy(C_remedies_j)` pointing to a concept `C_remedies_j` of lattice `remedies` if, for each `symptom x` ∈ *Extent*(`C_symptoms_i`), there exists a `remedy y` ∈ *Extent*(`C_remedies_j`) that treats `symptom x`. Likewise, when a concept `C_remedies_j` has ∃`isComposedOf(C_ingredients_p)` as a relational attribute, pointing to a concept of lattice `ingredients`, then, for each `remedy y` ∈ *Extent*(`C_remedies_j`), there exists an `ingredient z` ∈ *Extent*(`C_ingredients_p`) that composes `remedy y`. This yields information of the form : `for each x` ∈ *Extent*(`C_symptoms_i`) there exists a `remedy y` that treats `x` and there exists an `ingredient z` that composes `y`, thus linking symptoms and ingredients.

Concept lattices `symptoms` and `remedies` contain respectively 837 and 168 concepts at the end of RCA execution. Therefore, to facilitate the analysis, an Iceberg lattice (4% threshold) [21] has been built on the `symptoms` context and is exploited in the following. The resulting Concept Lattice Family contains sequences of information in accordance with the diagram shown in Fig. 6, the relations being existentially quantified. Figure 9 shows a set of connected concepts extracted from these results. It presents symptoms (extent of `C_symptoms_39`) that are treated by at least one remedy of `C_remedies_105` (the most specific `remedies` concept) composed of at least one ingredient from the `apiaceae` (`C_ingredients_121`), `zingiberaceae` (`C_ingredients_102`) and `piperaceae` (`C_ingredients_112`) families. This extract suggests that ingredients from these families are useful to treat symptoms of `C_symptoms_39`. Actually peper and cardamom, e.g., are known for their effects on digestive troubles.

As seen before, `remedies` concepts also group remedies according to their form (Fig. 7). In particular, concept `C_remedies_9` gathers two remedies that have both `pastille` and `potion` forms. Figure 10 shows an extract of the neigh-

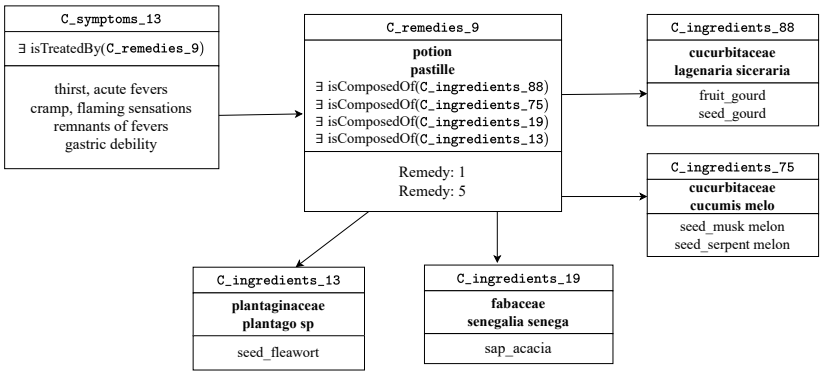

Fig. 10: An extract of information on `potion` and `pastilles` remedies.

bourhood of this concept. It includes the concept `C_symptom_13`, that gathers a set of symptoms treated by at least one remedy of `C_remedies_9`. Besides, remedies of `C_remedies_9` share some ingredients in common e.g. `seed_fleawort` and `sap_acacia` from `plantaginaceae` and `fabaceae` families respectively; and some ingredients such as seeds of melon and fruit or seed of gourd, that can thus be associated to the treatment of symptoms like thirst, cramp and fever.

Existential and Universal Strict Quantifiers. The $\exists\forall$ quantifier is applied to the relation `isTreatedBy` and the $\exists$ quantifier on the relation `isComposedOf` to update `symptoms` and `remedies` lattices. As above, we focus on `symptoms` and `remedies` concepts with small extents. As previously, The `symptoms` lattice is built with the Iceberg algorithm.

A concept `C_symptoms_i` in the `symptoms` lattice has a relational attribute $\exists\forall$ `isTreatedBy(C_remedies_j)` in its intent if, for each `symptom x` $\in$ *Extent* `(C_symptoms_i)`, each `remedy y` that treats `symptom x` belongs to the extent of `C_remedies_j` [3]. If `C_remedies_j` has $\exists$ `isComposedOf(C_ingredients_p)` as relational attribute, this combination of quantifiers yields information of the form: for each `x` $\in$ *Extent*`(C_symptoms_i)`, all `remedy y` that treat `x` are in *Extent*`(C_remedies_j)` and there exists an `ingredient z` $\in$ *Extent*`(C_ingredients_p)` that composes `y`. Thus `z` is useful to treat `x`.

The `symptoms` concept shown on Fig. 11 has a smaller extent than the one shown in Fig. 9 while the relational attributes point to the same concepts of `remedies` lattice. Indeed, with the $\exists\forall$ quantifier on the relation `isTreatedBy`, all remedies that treat the symptoms of `C_symptoms_31` are in the extent of `C_remedies_105` (and of super-concepts). This explain why the `diarrhoea` symptom is absent from `C_symptoms_31` extent, this symptom being treated by `Remedy 152` and an other remedy not belonging to these `remedies` concepts. Finally, this concept gives a more precise information than the one built with the $\exists$ quantifier. It allows to conclude that there are atmost three remedies that treat these

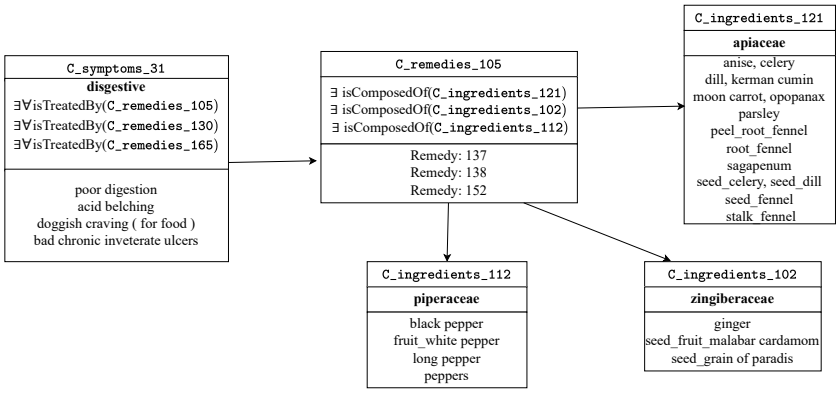

Fig. 11: An extract from RCA results on $\texttt{Dataset}_2$ with $\exists\forall$ / $\exists$ quantifiers.

four digestive symptoms and that their ingredients belong to a few plant families. This result can lead to the study of these specific remedies as well as the ingredients they share in common for the treatment of digestive symptoms.

5 Related Work

FCA-based models have been widely used for data mining or knowledge discovery purposes in various domains such as software engineering, web-documents analysis, text mining, biology or medicine [19]. In the medical domain, FCA has been mainly used for analysing data collected from physician or hospital networks. In [20] a Health Record System is analysed by means of FCA, investigating symptoms and diseases for enhancing medical diagnoses. In [22], FCA is used for discovering potential association between drugs and adverse effects from pharmacovigilance data. FCA and pattern structures have been used to analyse patient pathways from a French healthcare dataset on cancer [4]. Closer to our work, the authors of [14] have used FCA for characterising and clarifying syndromes in traditional china medicine, based on syndrome factors (disease sites, e.g. lung, head, and disease or pathological causes, e.g. cold, dampness).

RCA has been used for analysing data in various domains. It has been applied in software engineering in order to solve different problems pertaining to UML models [6], or for refactoring [16]. RCA has also been used in environmental applications, e.g. to discover patterns from monitoring data on a river network [17]. RCA was also used for Information Retrieval, e.g., for querying collections of legal documents connected through cross references [15]. Furthermore, RCA has been used for analysing data collected from contemporary texts about plant health in Sub-Saharan Africa. Data represent plants that grow or are used, and how they are used, in various countries to treat animal or plant diseases or pests [12]. The idea is to find plant-based extracts that can serve as alternative to synthetic pesticides and antimicrobials.

Other approaches have been used to explore old pharmacopoeias. The work in [5] is based on the objective of finding co-occurring ingredients in remedies extracted from a 15th century Middle English medical text, focusing on microbial infections. It uses community detection in a network of ingredients. Using FCA and RCA as we propose here can reveal more complex relations between ingredients, their characteristics or origin, and the targeted symptoms, thus guiding biologists in their search for ingredients with specific properties.

6 Conclusion and Future Work

This work aims to explore old pharmacopoeias to find active ingredients that can be useful to design new drugs. A database of old remedies has been built from a 9th century Arabian manuscript. The database describes remedies, their ingredients, mainly plants, the symptoms they can treat and the concerned organs. FCA and RCA have been used to analyse this database, focusing on remedies that treat fever symptoms, and their associated ingredients. The analysis was performed at different levels: FCA allowed to search for co-occurring ingredients, and ingredients that appear the most often; RCA was used to reveal links between symptoms and ingredients. We have presented a few results, that may be of interest for biologists searching for ingredients with unknown or forgotten relevant properties.

In the future, this analysis will be extended and deepened, and other datasets from the database will be studied, e.g. symptoms that can be linked to bacterial infections and their remedies. However, since the obtained lattices are large, the selection of concepts could be guided by the use of interestingness measures, e.g. concept stability [13]. Besides, other data models and other RCA quantifiers, such as the universal-percent scaling quantifier, $\exists\forall_{\geq n}$ [9, 3], can be used to refine the analysis. Finally, visualisation and exploration techniques, to facilitate the analysis of the results by biologists would be very useful.

References

1. Bird, S., Klein, E., Loper, E.: Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc." (2009)
2. Braud, A., Dolques, X., Fechter, P., Lachiche, N., Le Ber, F., Pitchon, V.: Analyzing the composition of remedies in ancient pharmacopeias with FCA. In: RealDataFCA'2021 (2021)
3. Braud, A., Dolques, X., Huchard, M., Le Ber, F.: Generalization effect of quantifiers in a classification based on relational concept analysis. Knowl.-Based Syst. **160**, 119–135 (2018)
4. Buzmakov, A., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A., Raïssi, C.: On mining complex sequential data by means of FCA and pattern structures. Int. J. General Syst. **45**(2), 135–159 (2016)
5. Connelly, E., del Genio, C.I., Harrison, F.: Data mining a medieval medical text reveals patterns in ingredient choice that reflect biological activity against infectious agents. mBio **11**(1) (2020)

6. Dolques, X., Huchard, M., Nebut, C., Reitz, P.: Fixing generalization defects in UML use case diagrams. Fundam. Informaticae **115**(4), 327–356 (2012)
7. El Haff, K., Antoun, W., Le Ber, F., Pitchon, V.: Reconnaissance des entités nommées pour l'analyse des pharmacopées médiévales. In: EGC 2023. RNTI, vol. E-39, p. 329–336 (Jan 2023)
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer (1999)
9. Hacene, M.R., Huchard, M., Napoli, A., Valtchev, P.: Relational concept analysis: mining concept lattices from multi-relational data. Ann. Math. Artif. Intell. **67**(1), 81–108 (2013)
10. Hajar, R.: The Air of History Part III: The Golden Age in Arab Islamic Medicine An Introduction. Heart Views **14**(1), 43–46 (2013)
11. Kahl, O.: Sābūr Ibn Sahl's Dispensatory in the Recension of the ʻAḍudī Hospital. BRILL (2009), arabic edition and English translation
12. Keip, P., Gutierrez, A., Huchard, M., Le Ber, F., Sarter, S., Silvie, P., Martin, P.: Effects of Input Data Formalisation in Relational Concept Analysis for a Data Model with a Ternary Relation. In: ICFCA 2019. pp. 191–207. Springer (2019)
13. Kuznetsov, S.O.: On stability of a formal concept. Ann. Math. Artif. Intell. **49**(1), 101–115 (2007)
14. Liu, X., Hong, W., Song, J., Zhang, T.: Using formal concept analysis to visualize relationships of syndromes in traditional chinese medicine. In: Medical Biometrics: ICMB 2010. pp. 315–324. Springer (2010)
15. Mimouni, N., Fernández, M., Nazarenko, A., Bourcier, D., Salotti, S.: A relational approach for information retrieval on XML legal sources. In: Int. Conf. on Artificial Intelligence and Law (ICAIL). pp. 212–216. ACM (2013)
16. Moha, N., Hacene, A.R., Valtchev, P., Guéhéneuc, Y.: Refactorings of design defects using relational concept analysis. In: 6th ICFCA. pp. 289–304. Springer (2008)
17. Nica, C., Braud, A., Le Ber, F.: Exploring heterogeneous sequential data on river networks with relational concept analysis. In: ICCS 2018. Springer (2018)
18. Ouzerdine, A., Braud, A., Dolques, X., Huchard, M., Le Ber, F.: Adjusting the exploration flow in relational concept analysis – an experience on a watercourse quality dataset. In: Adv. in Knowl. Dis. and Manag. pp. 175–198. Springer (2022)
19. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: A survey on applications. Expert Syst. with Appl. **40**(16), 6538–6560 (2013)
20. Săcărea, C., Şotropa, D., Troancă, D.: Using analogical complexes to improve human reasoning and decision making in electronic health record systems. In: ICCS 2018. pp. 9–23. Springer (2018)
21. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with Titanic. Data & Knowl. Eng. **42**(2), 189–222 (2002)
22. Villerd, J., Toussaint, Y., Lillo-Le Louët, A.: Adverse drug reaction mining in pharmacovigilance data using formal concept analysis. In: Machine Learning and Knowledge Discovery in Databases. pp. 386–401. Springer (2010)