Master 2 Sciences et Numérique pour la Santé parcours Bioinformatique, Connaissances, Données

Rapport de stage

Analyse formelle de données pour la sélection de plantes à effet pesticide pour la santé animale et végétale dans les pays du Sud

---

Auteur :
Amirouche OUZERDINE

Dirigé par :
Marianne Huchard (LIRMM)
Pierre Martin (CIRAD)

Suivi par :
Pierre Silvie (IRD/CIRAD)

Tutrice :
Anna-Sophie FISTON-LAVIER

19 août 2018

Table des matières

1 Introduction **1**
- 1.1 Contexte . . . 1
- 1.2 Problématique . . . 2
- 1.3 État de l'art . . . 3

2 Les trois approches de navigation **3**
- 2.1 Navigation dans un seul treillis . . . 3
  - 2.1.1 Analyse formelle de concepts : principe de base . . . 4
  - 2.1.2 Exemple de navigation dans un treillis . . . 4
  - 2.1.3 Inconvénients . . . 5
- 2.2 Navigation entre les treillis . . . 5
  - 2.2.1 Analyse relationnelle de concepts : principe de base . . . 5
  - 2.2.2 Exemple de navigation entre les treillis . . . 6
  - 2.2.3 Inconvénients . . . 7
- 2.3 Navigation sur demande : nouvelle approche de navigation . . . 8
  - 2.3.1 Algorithme *On-demand* . . . 8
  - 2.3.2 RCAExplore . . . 9
  - 2.3.3 Implémentation . . . 9
  - 2.3.4 Exemple de navigation incrémentale . . . 10

3 Analyses de données et résultats **11**
- 3.1 Les données d'analyse . . . 12
- 3.2 Avantages de l'analyse relationnelle de concepts . . . 12
  - 3.2.1 Hiérarchisation du résultat . . . 13
  - 3.2.2 Extraction de règles d'implication . . . 14
  - 3.2.3 Création d'hypothèses à partir des connaissances . . . 15
- 3.3 Scénarios *Knomana* . . . 15
- 3.4 Résultats d'analyse . . . 16
- 3.5 Difficultés liées à l'analyse . . . 22

4 Conclusion et Perspectives **23**

A Analyses et résultats **25**
- A.1 Avantages de l'analyse relationnelle de concepts . . . 25
- A.2 Résultats d'analyse . . . 26

1 Introduction

Tous les jours, une grande quantité de données est générée dans différents domaines d'application (banque, écologie, santé, commerce,etc.). Cette masse de données représente une importante mine d'informations qui peut être exploitée par les acteurs du domaine pour répondre à différents objectifs : accroître le chiffre d'affaire, étudier les besoins récurrents des clients et ainsi leur proposer de nouveaux services plus adaptés, relancer la recherche et inférer de nouvelles connaissances en médecine ou encore prendre les précautions nécessaires pour préserver ou améliorer l'état d'un environnement en écologie.

Quel que soit le domaine d'application, les données recueillies sont très souvent hétérogènes et difficilement exploitables à l'état brut. Il est donc indispensable de mobiliser des méthodes d'analyse afin de les traiter, de les structurer et de restituer le résultat à l'utilisateur de manière à ce qu'il puisse naviguer facilement dans les données, c'est-à-dire, démarrer d'une information/connaissance et se déplacer dans les données pour en découvrir de nouvelles.

Dans la suite de cette section, nous présentons le contexte du travail de recherche et la problématique afférente. Nous présentons dans la section 2 les approches de navigation dans les données dans le cadre des méthodes d'analyse que nous appliquons, ainsi qu'une nouvelle approche de navigation développée dans ce projet. Nous présentons dans la section 3 les données sur lesquelles portera l'analyse, les avantages de la méthode d'analyse utilisée ainsi que deux études de cas.

1.1 Contexte

Dans certains pays de l'Afrique francophone, les agriculteurs et les éleveurs souhaitent réduire l'usage des pesticides chimiques de synthèse pour protéger leurs cultures agricoles et les animaux élevés (bovins, etc.) en faveur de produits moins agressifs, à savoir, les plantes à effet pesticide. Le projet *Knomana* [1] (*Knowledge management on pesticides plants in Africa*) a été initié afin de répondre à ce besoin.

Knomana est financé par le métaprogramme Glofoods INRA [2] - CIRAD [3]. Son objectif est de réaliser une première analyse des plantes à effet pesticide et de leurs usages présentés dans les revues scientifiques.
Dans un premier temps, une base de connaissances nommée PPAf (Plantes Pesticides d'Afrique francophone) a été développée sous forme d'un tableau (figure 1) où chaque ligne représente une connaissance. Chaque connaissance décrit un usage de plante au moyen de 35 types d'informations. On y trouve différentes catégories :

— plante à effet pesticide : décrite par le nom d'espèce, le genre, la famille, la partie de la plante utilisée pour le traitement, la modalité d'application (huile essentielle, extrait,etc.), la méthode d'extraction des principes actifs (hydrodistillation, éthanol, eau, etc.), le pays où se trouve la plante.

— organisme ciblé par la plante à effet pesticide : décrit par sa taxonomie à savoir le nom d'espèce, le genre, la famille, l'ordre, la classe et le règne.

— organisme à protéger : décrit par le nom, le genre et la famille.

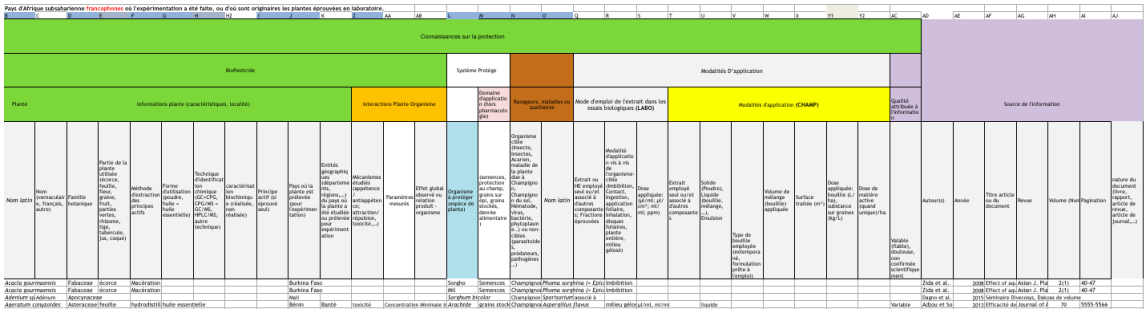
1. https ://www.cirad.fr/actualites/toutes-les-actualites/articles/2017/science/recenser-les-plantes-naturellement-pesticides-en-afrique-knomana

2. http ://www.glofoods.inra.fr/

3. Centre de coopération internationale en recherche agronomique pour le développement : centre de recherche français qui répond, avec les pays du Sud, aux enjeux internationaux de l'agriculture et du développement. https ://www.cirad.fr/

— source d'information : décrit par le titre de l'article/la revue, le nom de l'auteur, l'année de publication et la fiabilité de l'information.

Dans un second temps, PPAf doit être analysée de manière à extraire et inférer de nouvelles connaissances susceptibles d'être exploitées sur le terrain ou des connaissances faisant émerger des activités de recherches.



| Pays d'Afrique subsaharienne francophones où l'expérimentation a été faite, ou d'où sont originaires les plantes découvertes en laboratoire. | | |

Aussi dans ce rapport, nous nous intéressons à l'AFC et à son extension l'ARC, aux résultats qu'elles produisent et à la manière de naviguer dans ce résultat afin d'en extraire de la connaissance.

1.3 État de l'art

Dans une étude bibliographique qui porte sur la navigation conceptuelle et l'exploration des données [12], différents outils de navigation dédiés aux méthodes d'analyses formelle et relationnelle de concepts ont été recensés.

La plupart des outils présentent les résultats sous forme de graphe, à l'exemple de *LatViz* [4], *RCAExplore* [5] ou *Galicia* [6]. Bien que cette présentation sous forme de graphe permette d'avoir une vue d'ensemble sur le résultat, elle n'est cependant pas intuitive pour des utilisateurs peu habitués aux méthodes d'analyses formelle et relationnelle de concepts. On peut également noter que plus le volume des données à analyser est important, plus la taille des graphes générés est importante, et plus ils deviennent difficiles à visualiser et donc à analyser même pour un utilisateur aguerri des méthodes d'analyse.

Tous les outils de navigation recensés [12] ont un point commun, ils présentent les résultats d'analyse dans leur totalité et en une seule fois à l'utilisateur. Ces résultats étant très souvent volumineux, l'utilisateur s'y noie très rapidement et n'en tire aucune connaissance. Aucun de ces outils ne propose encore une navigation progressive et à la demande, c'est-à-dire permettant d'accompagner pas-à-pas l'utilisateur dans sa navigation, de lui donner le choix pour orienter sa navigation ou de l'arrêter quand il le souhaite. C'est cet aspect-là que nous allons traiter dans ce rapport, à savoir proposer une nouvelle approche de navigation incrémentale, l'implémenter dans l'outil *RCAExplore* et l'appliquer par la suite sur les données du projet *Knomana* pour en extraire de la connaissance.

2 Les trois approches de navigation

Nous présentons dans cette section trois approches pour naviguer dans les données. Nous rappelons dans les sections 2.1 et 2.2 le principe de base des méthodes d'analyses formelle et relationnelle de concepts, la manière de naviguer dans les données avec chacune des méthodes ainsi que leurs inconvénients. Nous présentons dans la section 2.3 une nouvelle approche dite "sur demande" pour naviguer dans les données, développée et implémentée dans le cadre de ce travail de recherche.

2.1 Navigation dans un seul treillis

Nous rappelons dans cette section le principe de base de l'analyse formelle de concepts, nous expliquons ensuite par un exemple comment naviguer dans un treillis, et présentons les limites de cette méthode.

2.1.1 Analyse formelle de concepts : principe de base

L'analyse formelle de concepts (AFC) est une méthode mathématique basée sur la théorie des treillis [5]. À partir d'un tableau binaire elle regroupe différents objets en fonction de leurs attributs communs, et organise les différents groupes obtenus dans une structure hiérarchique où les groupes les plus spécifiques se trouvent en bas de la structure

4. https ://latviz.loria.fr/
5. http ://dolques.free.fr/rcaexplore/
6. http ://www.iro.umontreal.ca/ galicia/

hiérarchique et les groupes les plus généraux en haut. Pour cette méthode, le tableau binaire est appelé "*contexte formel*", un groupe obtenu "*concept*" et la structure hiérarchique qui contient tous les groupes "*treillis de concepts*".

plant	lauraceae	asteraceae	aromatic	comestible	toxic	evergreen	contraceptive	antidysenteric	applicOil	applicEssentialOil	applicExtract
cinnamomumZeylanicum	x		x	x		x			x		
chromolaenaOdorata		x	x		x	x				x	x
aspiliaAfricana		x	x		x	x	x			x	x
ageratumConyzoides		x	x		x	x		x		x	x

pest	attacksPeanuts	attacksCheese	attacksRice	aspergillus
aspergillusFlavus		x		x
aspergillusParasiticus	x			x
aspergillusOchraceus	x			x
aspergillusTamarii		x	x	x

TABLE 1 – Contextes formels Plant (à gauche) et Pest (à droite).

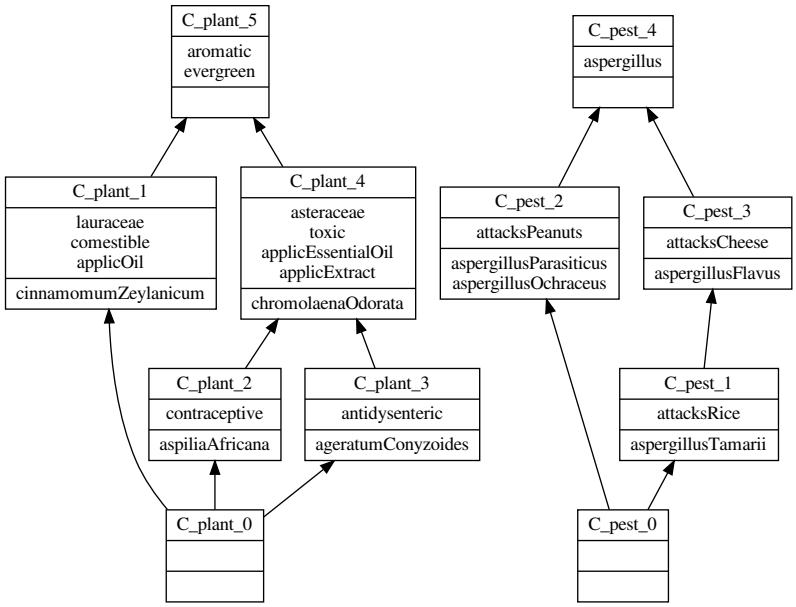


FIGURE 2 – Treillis de concepts plant (à gauche) et pest (à droite), obtenus avec l'analyse formelle de concepts.

2.1.2 Exemple de navigation dans un treillis

L'AFC calcule un treillis pour chaque contexte formel. Dans le cas de PPAf, nous appliquons l'AFC calculant un treillis pour chaque catégorie d'information. Nous expliquons ici comment naviguer dans un treillis en prenant pour exemple celui des plantes (à gauche de la figure 2).

Dans le treillis *plant* (à gauche de la figure 2) le concept *C_plant_2* indique que la plante *Aspilia africana* possède l'attribut *contraceptive*. La flèche sortante de ce concept indique que la plante *Aspilia africana* possède aussi les attributs de tous les concepts au-dessus qui sont directement ou indirectement liés au concept *C_plant_2*, c'est-à-dire les

attributs *asteraceae, toxic, applicEssentialOil, applicExtract* du concept *C_plant_4* et les attributs *aromatic, evergreen* du concept *C_plant_5*.
Le concept *C_plant_4* indique que les plantes *Chromolaena odorata, Aspilia africana* et *Ageratum conyzoides* possèdent les attributs *asteraceae, toxic, applicEssentialOil, applicExtract, aromatic* et *evergreen*. En effet, bien que les plantes *Aspilia africana* et *Ageratum conyzoides* ne soient pas explicitement mentionnées dans le concept *C_plant_4*, les deux flèches entrantes vers *C_plant_4* indiquent que ce concept possède aussi les objets de tous les concepts en-dessous qui lui sont directement ou indirectement liés, c'est-à-dire les objets *Aspilia africana* du concept *C_plant_2*, *Ageratum conyzoides* du concept *C_plant_3* et les objets du concept *C_plant_0*, bien que ce dernier n'en possède pas. La flèche sortante de *C_plant_4* indique que tous les objets de ce concept possèdent également les attributs *aromatic et evergreen* du concept *C_plant_5* bien qu'ils ne soient pas mentionnés explicitement dans *C_plant_4*.
Le concept *C_plant_0* est le groupe de plantes qui possèdent tous les attributs à la fois. Cependant dans notre exemple, il n'existe pas de plante qui possède tous les attributs en même temps.

2.1.3 Inconvénients

Un des inconvénients de l'AFC est le nombre de concepts générés dans chaque treillis. Pour un contexte formel de n lignes et m colonnes, le nombre maximum de concepts générés est de $2^{min(n,m)}$. Donc plus le nombre de concepts est important, plus il est difficile pour un utilisateur de naviguer dans le treillis.

Un autre inconvénient de l'AFC est de ne pas permettre de combiner des contextes formels. Ce faisant, l'AFC ne permet pas la navigation entre les treillis. Par exemple, on ne peut pas savoir ici quelle plante utiliser pour lutter contre un ravageur. Pour répondre à cette question, on peut utiliser l'analyse relationnelle de concepts (ARC) présentée dans la section 2.2.

2.2 Navigation entre les treillis

Nous rappelons dans cette section le principe de base de l'analyse relationnelle de concepts et son avantage par rapport à l'analyse formelle de concepts. Nous expliquons ensuite par un exemple comment naviguer entre les treillis issus de cette méthode et présentons à la fin l'inconvénient de cette méthode.

2.2.1 Analyse relationnelle de concepts : principe de base

L'analyse relationnelle de concepts (ARC) [7] est une extension de l'analyse formelle de concepts (AFC). En sus aux calculs permis par l'AFC, elle permet de traiter les relations entre contextes formels. Ce faisant, l'ARC permet de répondre à des questions du type "quelle plante utiliser pour lutter contre un ravageur ?". Pour ce faire, elle nécessite la définition d'un contexte relationnel pour chaque paire de contextes formels à relier, un contexte relationnel correspondant à un tableau binaire qui met en relation les objets de deux contextes formels.

Par exemple, aux contextes formels *plant* et *pest* de la table 1, nous ajoutons le contexte relationnel présenté en table 2 qui met en relation des plantes qui peuvent servir de traitement pour lutter contre des ravageurs. Les plantes représentent la source de la relation et les ravageurs la cible de la relation.

Le contexte relationnel *treat* permet en utilisant le treillis de concepts *pest* (figure 2 à droite) d'étendre le contexte formel *plant* (table 1 à gauche) avec de nouveaux attributs

treat	aspergillusFlavus	aspergillusParasiticus	aspergillusOchraceus	aspergillusTamarii
cinnamomumZeylanicum	x			x
chromolaenaOdorata			x	
aspiliaAfricana		x		
ageratumConyzoides	x			

TABLE 2 – Contexte relationnel -Treats- qui décrit qu'une plante (indiquée sur une ligne) traite un ou plusieurs ravageurs (les colonnes).

dits *attributs relationnels* (un attribut relationnel est noté : ∃ *treat (Nom du concept)*). Dans notre exemple, ces attributs relationnels représentent les concepts (groupes) de ravageurs avec lesquels une plante peut être en relation par la relation *treat*. Par exemple ∃ *treat (C_pest_3)* indique qu'une plante traite au moins un ravageur du groupe de ravageurs du concept *C_pest_3*, c'est-à-dire *Aspergillus flavus* ou *Aspergillus tamarii*.
De manière générale, chaque plante du contexte formel *plant* (table 1 à gauche) est mise en relation avec un attribut relationnel, si cette plante traite au moins un ravageur du concept constituant l'attribut relationnel. La table 3 et la figure 3 représentent respectivement le contexte *plant* étendu par la relation *treat*, et les treillis résultants de l'ARC. Le treillis *pest* reste inchangé car il n'existe pas de relation (contexte relationnel) dont l'ensemble des ravageurs soit la source pour étendre son contexte formel.

Plant	lauraceae	asteraceae	aromatic	comestible	toxic	evergreen	contraceptive	antidysenteric	applicOil	applicEssentialOil	applicExtract	∃ *Treats* (c_Pest_0)	∃ *Treats* (c_Pest_1)	∃ *Treats* (c_Pest_2)	∃ *Treats* (c_Pest_3)	∃ *Treats* (c_Pest_4)
cinnamomumZeylanicum	x		x	x		x			x				x		x	x
chromolaenaOdorata		x	x		x	x				x	x			x		x
aspiliaAfricana		x	x		x	x	x			x	x			x		x
ageratumConyzoides		x	x		x	x		x		x	x				x	x

TABLE 3 – Contexte Plant étendu par la relation *Treats*.

2.2.2 Exemple de navigation entre les treillis

On navigue dans le treillis étendu *plant* (figure 3 à gauche) de la manière présentée dans la section 2.1.2. Cependant l'avantage de celui-ci est de posséder les attributs relationnels qui permettent de naviguer vers le treillis des ravageurs. Par exemple le concept *C_plant_6* possède l'attribut relationnel ∃ *treat (C_pest_2)*. Ce dernier indique que les plantes du concept *C_plant_6*, c'est-à-dire *Chromolaena odorata* et *Aspilia africana* traitent au moins un ravageur du groupe de ravageurs *C_pest_2* c'est-à-dire *Aspergillus parasiticus* ou *Aspergillus ochraceus* et que ce groupe de ravageurs attaque les *Peanuts* " *attackPeanuts* ".

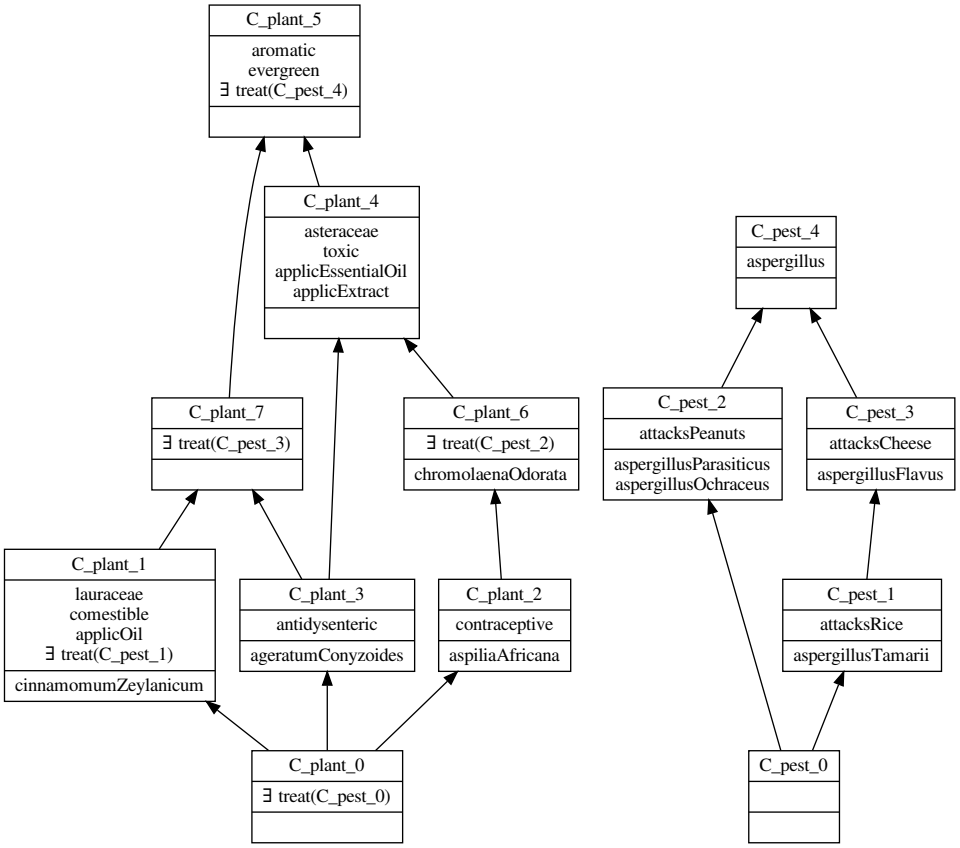


FIGURE 3 – Treillis de concepts *plant* étendu (à gauche) et *pest* (à droite), issus de l'analyse relationnelle de concepts.

Remarque : Il existe dans l'ARC des paramètres de quantification de la force de la relation entre les objets. Cette quantification permet par exemple de chercher une plante qui traite au moins un ravageur d'un groupe de ravageurs ou une plante qui ne traite que les ravageurs d'un groupe particulier de ravageurs. Ces paramètres dits *quantifieurs de relation* ne seront pas traités dans ce rapport, et sont étudiés en détails dans [2].

2.2.3 Inconvénients

Tout comme l'AFC, l'ARC présente l'inconvénient de générer un nombre important de concepts dans le cas d'un jeu de données conséquent. De plus, le fait de traiter les relations avec l'ARC génère un nombre de concepts beaucoup plus important que celui généré avec l'AFC. Plus il y a de relations, plus la taille des treillis devient importante et plus la navigation entre les treillis devient compliquée pour un utilisateur. Une solution alternative serait de pouvoir naviguer au sein d'un treillis et entre les treillis à la demande, c'est-à-dire de visualiser exclusivement les concepts voisins au concept d'intérêt et d'afficher les voisins au gré des déplacements. Nous présentons dans la section 2.3 une nouvelle approche de navigation qui permet à un utilisateur de naviguer de façon progressive et à la demande, à la fois dans un treillis et entre les treillis.

2.3 Navigation sur demande : nouvelle approche de navigation

Nous proposons dans cette section une nouvelle approche de navigation dans les treillis. Cette approche est basée sur l'algorithme *On-demand* [1], qui permet à l'utilisateur de ne construire que la partie de treillis qu'il veut visualiser. Nous présentons d'abord comment l'algorithme *On-demand* est mis en place sur une étape, ensuite sur plusieurs étapes de la navigation. Nous présentons par la suite l'outil *RCAExplore* dans lequel l'algorithme a été implémenté et expliquons comment cet algorithme lui a été intégré.

2.3.1 Algorithme *On-demand*

Nous avons vu dans la section 2.2.3 que l'ARC présentait l'inconvénient de générer des treillis volumineux pour un jeu de données de taille importante. Il devient donc très compliqué pour un utilisateur de naviguer dans les treillis. Ceci a pour conséquence d'augmenter le risque d'erreur, par exemple l'utilisateur risque de naviguer vers le mauvais concept (à cause du nombre important de liens entre les concepts), ou encore risque de ne pas voir certains concepts à partir desquels on pourrait extraire des connaissances intéressantes. De plus, pour répondre à une question ciblée il n'est parfois pas nécessaire de construire un treillis en entier, mais une construction partielle peut suffire pour y répondre. C'est pour pallier à ces contraintes que l'algorithme *On-demand* [7] [1] a été considéré.

L'algorithme *On-demand* met l'accent sur l'interaction avec l'utilisateur par laquelle ce dernier peut orienter la navigation comme il le souhaite. C'est un processus itératif qui se décompose en deux étapes : (1) sélection des paramètres d'entrée et (2) calcul du voisinage d'un concept.

1. **Sélection des paramètres d'entrée :** À la première étape l'utilisateur sélectionne un contexte formel, un ensemble de relations et un ensemble d'attributs autour desquels il veut démarrer la navigation. L'algorithme étend le contexte formel avec l'ensemble de relations sélectionné, suivant le processus ARC expliqué dans la section 2.2.1.
2. **Calcul du voisinage d'un concept :** L'algorithme calcule d'abord le concept $\mathcal{C}$ autour de l'ensemble d'attributs sélectionné à l'étape (1). Il calcule ensuite le voisinage de ce concept, c'est-à-dire tous les concepts qui lui sont immédiatement supérieurs et inférieurs. Les attributs relationnels associés au concept $\mathcal{C}$ sélectionné permettent de se déplacer vers les concepts des treillis voisins. Les concepts directement supérieurs (resp. inférieurs) constituent la couverture supérieure (resp. inférieure). Les concepts connectés par des attributs relationnels constituent la couverture relationnelle. Le concept $\mathcal{C}$ et ses couvertures forment un treillis partiel mettant ainsi en évidence les attributs sélectionnés et ce qu'il y a autour.

L'algorithme reprend à l'étape (1) en proposant à l'utilisateur (s'il souhaite continuer la navigation) de sélectionner un concept (parmi ceux calculés dans la précédente itération) pour continuer à naviguer dans le même treillis ou un attribut relationnel pour basculer vers un treillis voisin, ainsi que de choisir les relations qui l'intéressent. L'étape (2) est ensuite relancée pour le calcul du voisinage de ce nouveau concept sélectionné.

L'algorithme *On-demand* a été implémenté et intégré à l'outil *RCAExplore* que nous présentons dans la section 2.3.2.

7. Dans ce rapport nous éxpliquons seulement le fonctionnement de l'algorithme *On-demand* sans détaillé l'aspect algorithmiques. L'aspect algorithmique est présenté en détail dans [1]

2.3.2 RCAExplore

RCAExplore[8] est un outil qui implémente les méthodes AFC et ARC. Il fournit à l'utilisateur une grande flexibilité quant à la manipulation de son jeu de données pendant tout le processus d'analyse. Par exemple, à chaque étape de l'ARC, l'utilisateur peut sélectionner les contextes formels et les relations qui l'intéressent, traiter chaque relation de façon indépendante et choisir pour chaque contexte formel un algorithme particulier[9] pour la construction du treillis. RCAExplore offre aussi la possibilité de visualiser tous les treillis construits pendant tout le processus ARC.

2.3.3 Implémentation

Pour implémenter l'algorithme *On-demand* dans *RCAExplore* il a d'abord fallu étudier de façon minutieuse la structure existante. Cette structure développée en Java et réalisée sous forme de classes et de paquetages, est volumineuse[10], et très complexe de part l'enchevêtrement des fonctions. *RCAExplore* implémente différents algorithmes de construction de treillis à l'aide de plusieurs fonctions. Nous avons réutilisé certaines de ces fonctions pour intégrer l'algorithme *On-demand*, nous avons donc dû les modifier tout en prenant garde à ne pas altérer leurs résultats attendus dans les algorithmes qui les utilisent. Ce travail a demandé d'expliciter l'architecture logicielle de l'application afin d'identifier de manière précise où ces fonctions intervenaient, de comprendre leur fonctionnement et leurs interactions avec le reste du logiciel tout en veillant à ne pas modifier leur spécifications fonctionnelles externes.

L'implémentation de l'algorithme *On-demand* a nécessité aussi l'étude d'autres algorithmes pour calculer des traverses minimales [13] [10] [11] utiles pour le calcul de la couverture inférieure d'un concept. Nous avons retenu l'algorithme MTMiner pour notre implémentation [10], car il présentait un compromis entre la rapidité et la facilité d'implémentation.

L'algorithme *On-demand* permet à l'utilisateur de choisir l'ensemble de relations qu'il veut étudier à chaque étape de la navigation. Cependant pour cette implémentation nous avons préféré ignorer cette option et étendre le contexte formel sélectionné comme entrée pour la navigation (étape 1) avec toutes ses relations. Nous avons fait ce choix parce que :

— Il arrive souvent qu'un ensemble de relations forme un cycle. Traiter à chaque étape seulement un sous-ensemble de relations de ce cycle complexifie les calculs dans l'algorithme.

— Cela rend l'algorithme facile et rapide à implémenter par rapport au temps imparti pour cette tâche.

— Nous voulions mettre en pratique une première version de l'algorithme pour la tester et l'améliorer ensuite.

2.3.4 Exemple de navigation incrémentale

L'exemple illustré dans cette section est un résultat de l'utilisation de l'algorithme *On-demand* appliqué sur les contextes formels des tables 1 et 2. Dans cet exemple, l'utilisateur souhaite d'abord naviguer dans le treillis des plantes, et voir ce qu'il y a autour de l'attribut *antidysenteric*. La figure 4 répond à la requête de l'utilisateur.

8. http ://dolques.free.fr/rcaexplore/

9. Il existe divers algorithmes pour construire les treillis. Ces algorithmes n'ont aucune influence sur la navigation dans les treillis, quel que soit l'approche de navigation adoptée (un seul treillis, entre treillis ou à la demande). Ce faisant, il n'a donc pas été jugé utile de présenter ces algorithmes dans ce rapport.

10. 23 paquetages et 124 classes.

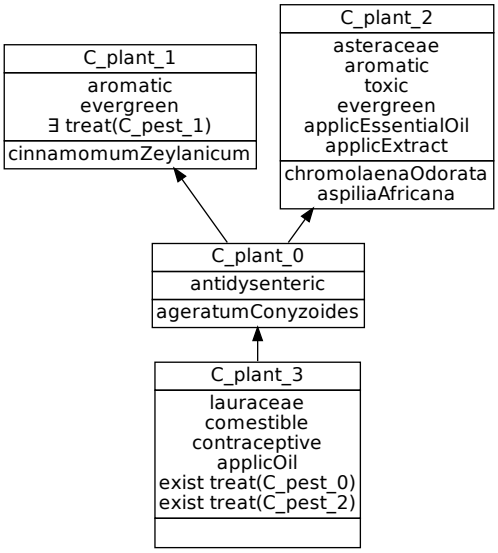


FIGURE 4 – Itération 1 de l'algorithme *On-demand* pour répondre à la question : qu'est ce qu'il y a autour de l'attribut *antidysenteric* dans le treillis *plant* ?

Le treillis de la figure 4 indique que le concept *C_plant_0* est le groupe de plantes qui possèdent la propriété *antidysenteric* en l'occurrence *Ageratum conyzoides* et que les plantes qui sont *antidysenteric* ont d'autres propriétés (toutes celles des concepts supérieurs). Le concept *C_plant_2* informe que les plantes *Ageratum conyzoides, Chromolaena odorata* et *Aspilia africana* ont en commun les attributs *asteraceae, aromatic, toxic, evergreen, applicEssentialOil, applicExtract*. Le concept *C_plant_1* informe via l'attribut relationnel ∃ *treat* (*C_pest_1*) que les plantes *Ageratum conyzoides* et *Cinnamomum zeylanicum* traitent au moins un ravageur du groupe de ravageurs du concept *C_pest_1*.

En sélectionnant l'attribut relationnel ∃ *treat* (*C_pest_1*), l'utilisateur poursuit sa navigation vers le treillis des ravageurs pour savoir quels sont les ravageurs que traitent les plantes *Ageratum conyzoides* et *Cinnamomum zeylanicum*. L'algorithme va donc calculer ce qu'il y a autour du concept *C_pest_1* comme le montre la figure 5.

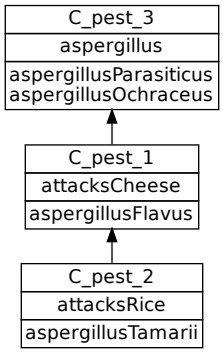


FIGURE 5 – Itération 2 de l'algorithme *On-demand* pour répondre à la question "quels ravageurs traitent les plantes *Ageratum conyzoides* et *Cinnamomum zeylanicum* ?".

Dans le treillis de la figure 5, le concept *C_pest_1* est le groupe de ravageurs qui attaquent le fromage (*attacksCheese*) en l'occurrence *Aspergillus flavus* et *Aspergillus tamarii*. La réponse à la question de l'utilisateur est donc : les plantes *Ageratum conyzoides* et *Cinnamomum zeylanicum* traitent au moins un des ravageurs *Aspergillus flavus* et/ou

Aspergillus tamarii, ces derniers attaquant le formage.

L'utilisateur peut aussi voir ce qu'il y a autour du concept *C_pest_1*. Par exemple, le concept *C_pest_2* informe que *Aspergillus tamarii* attaque aussi le riz (*attackRice*) et donc une plante qui le traite peut protéger des stocks de fromage et de riz. Le concept *C_pest_3* informe que les ravageurs *Aspergillus parasiticus, Aspergillus ochraceus, Ageratum conyzoides* et *Cinnamomum zeylanicum* sont de la famille *Aspergillus*.

Ainsi l'utilisateur a obtenu des réponses à ses questions sans construire tous les treillis et sans se noyer dans les données. Il peut à ce stade soit continuer sa navigation s'il cherche encore des réponses à certaines questions ou juste faire de la navigation, soit arrêter le processus de navigation.

3 Analyses de données et résultats

Pour faire les analyses, nous nous basons sur le schéma relationnel de la figure 6. Ce schéma représente une sous-partie du schéma relationnel de PPAf qui est plus complexe, mais il est suffisant pour apporter des éléments de réponse dans ce rapport. Il faut noter que le terme *pest* s'applique en réalité à un organisme ciblé dans les études, qui peut être un bioagresseur (insecte ravageur, champignon pathogène) ou un ennemi naturel de ce bioagresseur (insecte parasitoïde ou prédateur du ravageur). Par facilité de lecture le terme *pest* est adopté au lieu d'"organisme-cible".

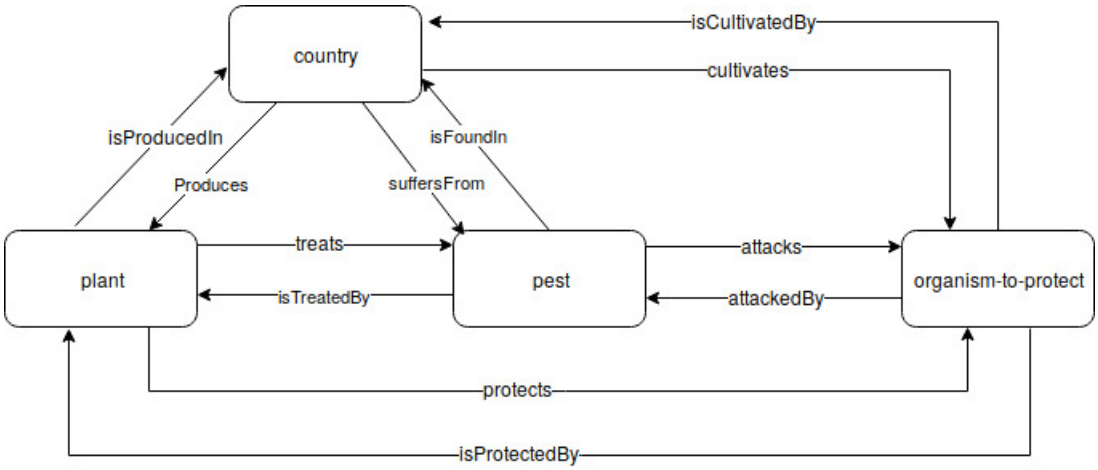


FIGURE 6 – Schéma relationnel utilisé pour répondre aux requêtes des utilisateurs.

Dans le cadre de l'analyse relationnelle de concepts, chaque rectangle de la figure 6 représente un contexte formel et chaque flèche représente une relation, donc un contexte relationnel. La double relation entre deux contextes formels nous permet en fonction de la question posée, de sélectionner le contexte formel de départ pour commencer la navigation. Par exemple, si on cherche les organismes-cibles (par exemple les insectes ravageurs) d'une plante à effet pesticide, on se positionne au départ sur le contexte formel *plant* pour aller vers le contexte formel *pest* en passant par la relation *treats*. Et si on cherche les plantes à effet pesticide qui traitent un ravageur, on démarre du contexte formel *pest* pour aller vers le contexte formel *plant* en passant par la relation *isTreatedBy*.

Dans cette section nous analysons les données de *Knomana* et présentons quelques résultats. Nous présentons tout d'abord dans la section 3.1 les données de PPAf sur lesquelles portera l'analyse. Dans la section 3.2 nous appliquons la méthode d'analyse relationnelle de concepts (ARC) et nous présentons les avantages qu'offre cette méthode. Nous présentons

ensuite dans la section 3.3 quelques sollicitations reçues de la part de personnes intéressées par l'usage des plantes pesticides et pour lesquelles nous appliquons l'analyse par navigation sur demande, dont les résultats sont présentés et discutés dans la section 3.4. Nous parlons dans la section 3.5 des difficultés auxquelles nous avons été confronté pendant l'analyse de données.

3.1 Les données d'analyse

Actuellement PPAf recense plus de 33000 connaissances sur l'usage de plantes à effet pesticide. Cependant, pour nos analyses, nous n'en utilisons qu'une partie composée de 5287 connaissances. Chaque connaissance est décrite au moyen de 35 types d'informations. On y trouve par exemple des informations sur la plante, l'organisme à protéger, l'organisme cible, les modalités d'application du traitement, les pays dans lesquels sont effectuées les expériences, les articles et les revues scientifiques à partir desquels sont extraites ces connaissances, etc. Dans le cadre de ce rapport nous nous somme intéressés à 13 informations relatives aux plantes à effet pesticide (espèce, genre, famille), aux organismes à protéger (espèce, genre, famille), aux organismes cibles (espèce, genre, famille, ordre, classe, règne) ainsi qu'aux noms des pays dans lesquels sont effectuées les expériences. Le but est de trouver des liens entre certaines espèces ou familles d'espèces (plante/ravageur, plante/plante) ainsi que le lieu où se trouvent ces espèces.

PPAf est alimentée par saisie manuelle, de ce fait on y trouve plusieurs erreurs (fautes de frappe, différentes langues utilisées -Français/Anglais-, des informations manquantes, une même information saisie plusieurs fois, etc.). Elle a donc été nettoyée [11] afin que les données que nous utilisons soient plus propres et ainsi éviter les erreurs dans les résultats.

3.2 Avantages de l'analyse relationnelle de concepts

Comme présenté dans la section 1.2, l'analyse formelle de concepts (AFC) couvre un large spectre d'analyse : classification, *clustering*, extraction de règles d'implication/association, hiérarchisation du résultat. L'analyse relationnelle de concepts (ARC) est une extension de l'AFC, elle offre donc ces mêmes possibilités, en plus de l'étude des relations entre les objets. Nous présentons dans cette section quelques-uns de ces avantages par un exemple d'analyse effectuée sur PPAf.

Nous utilisons pour l'exemple le sous-schéma relationnel de la figure 7 [12]. L'application de l'ARC sur ce schéma relationnel génère les treillis *plant*, *pest*, *country* et *organism-to-protect* avec respectivement 697, 265, 49 et 53 concepts [13]. La figure 8 présente des parties des treillis *plant* (haut à gauche), *organism-to-protect* (haut à droite) et *pest* [14] (en bas) (le treillis *country* n'est pas présenté ici parce qu'il n'est pas utilisé dans les explications).

3.2.1 Hiérarchisation du résultat

Nous avons vu dans les sections 2.1 et 2.2 que les méthodes d'analyse formelle et relationnelle de concepts créent des groupes particuliers appelés concepts. Ces concepts

11. Le nettoyage de PPAf est effectué par un autre stagiaire qui travaille sur le projet *Knomana*.
12. L'application de l'ARC sur tout le schéma de la figure 6 génère des treillis de taille importante qui sont difficiles à analyser. Nous utilisons pour nos explication une sous-partie du schéma de la figure 6
13. Les treillis étant de taille importante, nous n'en affichons que des parties pour les explications dans la suite du rapport.
14. Par souci de lisibilité, dans cette figure, certains objets sont retirés des concepts du treillis *pest*. Leur retrait n'affecte pas le résultat ni le raisonnement à tenir. L'annexe A.1 présente la partie du treillis *pest* avec l'ensemble de ses objets.

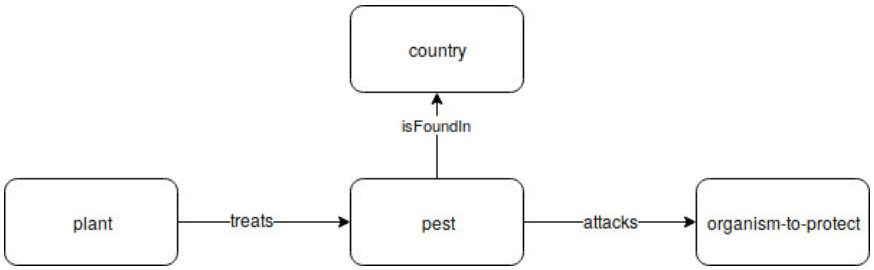


FIGURE 7 – Schéma relationnel utilisé pour l'analyse relationnelle de concepts.

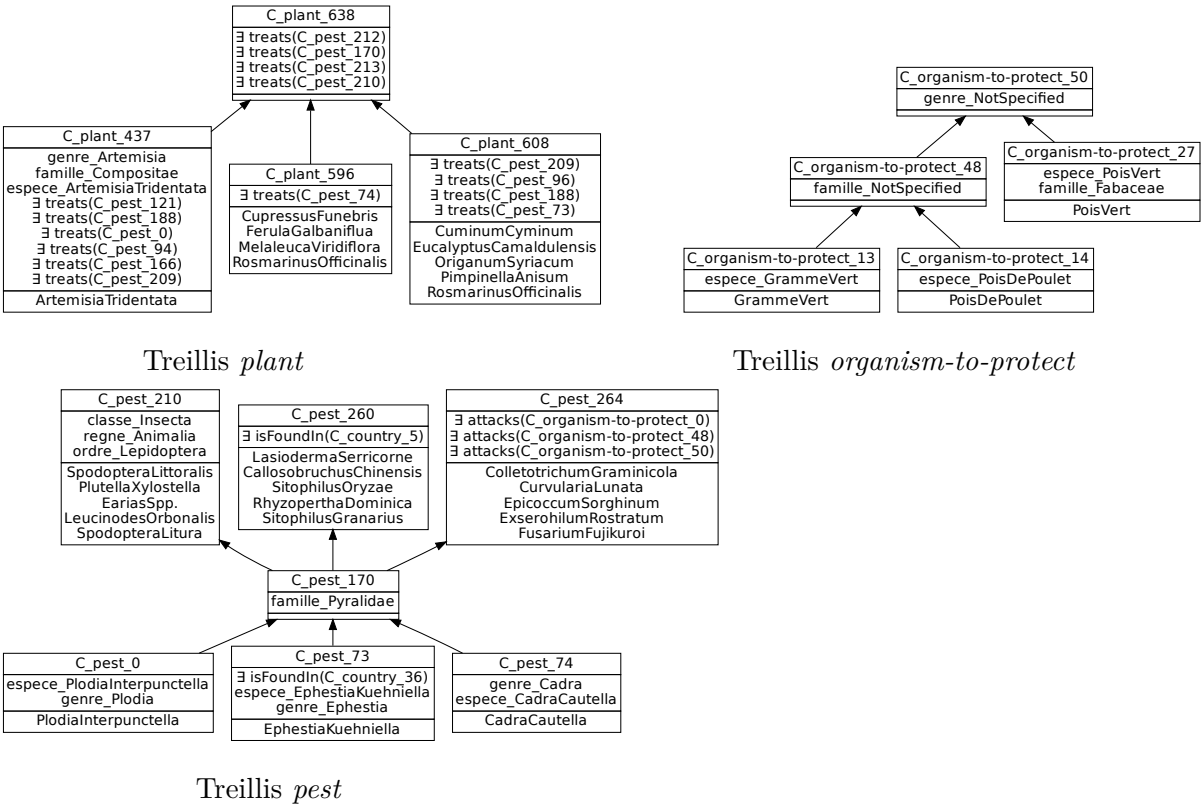


FIGURE 8 – Parties des treillis *plant* (haut à gauche), *organism-to-protect* (haut à droite) et *pest* (en bas).

sont organisés dans un treillis de façon hiérarchique où les concepts les plus spécifiques se trouvent en bas du treillis et les concepts les plus généraux en haut. Un concept spécifique apporte plus d'information sur les objets qu'il contient qu'un concept général.

Par exemple, la partie du treillis *plant* (figure 8) présente la hiérarchie des concepts. Le concept *C_plant_638* indique que l'ensemble des plantes des concepts *C_plant_437*, *C_plant_596* et *C_plant_608* ont en commun le fait qu'elles traitent au moins un ravageur de chaque concept des ravageurs *C_pest_170, C_pest_210, C_pest_212* et *C_pest_213*. Le concept *C_plant_638* est plus général que le concept *C_plant_437* qui est plus spécifique car il donne des informations supplémentaires sur la plante *Artemisia tridentata*, par exemple qu'elle est de la famille *Compositae* et qu'elle traite aussi d'autres ravageurs.

Un autre exemple peut se voir dans la partie du treillis *pest* (figure 8). Les concepts *C_pest_0*, *C_pest_73* et *C_pest_74* sont des concepts spécifiques, ils décrivent un ra-

vageur en fonction de son genre et de son espèce. Le concept *C_pest_0* indique que le ravageur *Plodia interpunctella* appartient au genre *Plodia*, le concept *C_pest_73* indique que le ravageur *Ephestia kuehniella* appartient au genre *Ephestia* et le concept *C_pest_74* indique que le ravageur *Cadra cautella* appartient au genre *Cadra*. Le concept *C_pest_170* est un concept plus général car il ajoute une information commune à tous les ravageurs des concepts *C_pest_0*, *C_pest_73* et *C_pest_74* à savoir qu'ils font tous partie de la famille *Pyralidae*.

Cette notion de hiérarchisation de concepts est importante. En effet elle nous permet de très vite nous positionner dans les treillis selon les besoins, c'est-à-dire si nous cherchons des informations plus générales et qui sont communes à beaucoup d'objets, on se positionne en haut du treillis, et si nous cherchons des informations plus spécifiques qui concernent quelques objets, on se positionne en bas et on navigue dans le treillis.

3.2.2 Extraction de règles d'implication

L'extraction de règles d'implication contenant des informations sur les relations est une notion importante de l'analyse relationnelle de concepts. Une règle d'implication est toujours vraie, elle apporte une connaissance sur le lien existant entre les attributs.

Par exemple dans la partie du treillis *pest* (figure 8), le concept *C_pest_170* décrit le groupe de ravageurs qui sont de la famille *Pyralidae*. Le concept *C_pest_170* hérite de l'attribut relationnel $\exists$ *attacks (C_organism-to-protect_50)* du concept *C_pest_264*. Nous pouvons donc extraire une règle d'implication notée : *Pyralidae* $\longrightarrow \exists$ *attacks (C_organism-to-protect_50)*. Cette règle est interprétée comme suit : "à chaque fois qu'on observe un ravageur de la famille *Pyralidae*, on observe une attaque sur au moins un organisme à protéger du groupe d'organismes à protéger *C_organism-to-protect_50*". Le concept *C_organism-to-protect_50* du treillis *organism-to-protect* (figure 8) contient les organismes à protéger *Gramme Vert, Pois De Poulet* et *PoisVert*. En se basant sur ces observations, nous pouvons dire que : dans PPAf, les ravageurs faisant partie de la famille *Pyralidae* attaquent toujours un ou plusieurs des organismes à protéger suivants : *Gramme Vert, Pois De Poulet* et *PoisVert*.

Un autre exemple dans la partie du treillis *plant* (figure 8). Le concept *C_plant_437* décrit les plantes qui sont de la famille *Compositae*. Le concept *C_plant_437* hérite de l'attribut relationnel $\exists$ *treats (C_pest_170)*. Nous pouvons donc extraire une règle d'implication notée : *Compositae* $\longrightarrow \exists$ *treats (C_pest_170)*. Cette règle est interprétée comme suit : "à chaque fois qu'on observe une plante de la famille *Compositae*, on observe le fait qu'on puisse traiter au moins un ravageur du groupe de ravageurs *C_pest_170*". Nous avons vu que le concept *C_pest_170* décrit les ravageurs qui font partie de la famille *Pyralidae*, nous pouvons donc supposer qu'il y a une corrélation entre la famille de plantes *Compositae* et la famille de ravageurs *Pyralidae*.

3.2.3 Création d'hypothèses à partir des connaissances

L'analyse relationnelle de concepts nous permet de créer des hypothèses à partir des connaissances extraites. C'est-à-dire de raisonner sur la base des informations récemment obtenues pour en extraire des nouvelles.

Par exemple dans le treillis *plant*, le concept *C_pest_596* décrit les plantes à effet pesticide qui traitent au moins un ravageur du concept des ravageurs *C_pest_74* c'est-à-dire le ravageur *Cadra cautella*. Nous savons aussi à partir du treillis *plant* que le concept

C_pest_596 n'a aucun lien direct avec les concepts de ravageurs *C_pest_73* et *C_pest_0* qui contiennent respectivement les ravageurs *Ephestia kuehniella* et *Plodia interpunctella* qui font partie de la famille *Pyralidae*. Or le concept *C_plant_596* hérite de l'attribut relationnel ∃ *treats (C_pest_170)* du concept *C_plant_638*. Cela veut dire que le concept *C_plant_596* du treillis *plant* est en relation avec le concept *C_pest_170* du treillis *pest* qui décrit les ravageurs de la famille *Pyralidae*. À partir de ces observations, nous pouvons nous demander si les plantes à effet pesticide *Cupressus funebris, Ferula galbaniflua, Melaleuca viridiflora, Rosmarinus officinalis* du concept *C_plant_596* qui traitent le ravageur *Cadra cautella* du concept *C_pest_74*, peuvent aussi traiter d'autres ravageurs faisant partie de la même famille comme le ravageur *Plodia interpunctella* du concept *C_pest_0* et le ravageur *Ephestia kuehniella* du concept *C_pest_73*. Ce questionnement peut amener à développer de nouvelles expérimentations afin de confirmer ou d'infirmer cette hypothèse.

Un autre exemple de création d'hypothèse vient de ce que l'on observe dans le treillis *pest* dans lequel on voit le concept *C_pest_264* contient les ravageurs de la famille *Pyralidae* issus du concept *C_pest_170* en plus des ravageurs *Colletotrichum graminicola, Curvularia lunata, Epicoccum sorghinum, Exserohilum rostratum, Fusarium fujikuroi*. Cet ensemble de ravageurs a la particularité d'attaquer au moins un organisme à protéger des concepts d'organismes à protéger *C_organism-to-protect_48*, *C_organism-to-protect_50* et *C_organism-to-protect_0*. Nous avons vu que ces concepts d'organismes à protéger peuvent être protégés contre la famille de ravageurs *Pyralidae* par les plantes à effet pesticide du concept *C_plant_638*. On pourrait penser qu'il est donc possible d'utiliser ces mêmes plantes pour protéger ces organismes à protéger pour lutter contre les ravageurs *Colletotrichum graminicola, Curvularia lunata, Epicoccum sorghinum, Exserohilum rostratum, Fusarium fujikuroi* du concept *C_pest_264*.

Évidemment contrairement aux règles d'implication qui sont vraies, les connaissances extraites par création d'hypothèse ne le sont pas toujours. Mais elles ouvrent de nouvelles perspectives de recherches et potentiellement la mise en place de nouvelles solutions.

3.3 Scénarios *Knomana*

Le projet *Knomana* a éveillé un grand intérêt dans différentes communautés (chercheurs, agriculteurs, conseillers, membres d'ONG,etc.). Des sollicitations relatives au projet et des questions précises sur l'usage des plantes nous sont parvenues. Dans le cadre de ce rapport, deux études de cas ont été retenues auxquelles des éléments de réponse sont apportés dans la section 3.4 :

1. quelles sont les connaissances liées aux plantes *Azadirachta indica* et *Lantana camara* ? Dans quels pays ces espèces végétales se trouvent-elles ? Contre quels organismes sont-elles utilisées ? Si la plante n'est pas présente, quelle(s) autre(s) espèce(s) végétale(s) pourrai(en)t atteindre au moins un des organismes ciblés ?
2. quelle(s) plante(s) utiliser au Burkina Faso pour lutter contre le champignon *Fusarium oxysporum* ?

3.4 Résultats d'analyse

Nous présentons dans cette section les résultats d'analyses effectuées pour répondre aux questions formulées dans la section 3.3.

Où se trouvent *Azadirachta indica* et *Lantana camara* ?

Comme nous l'avons expliqué dans la section 2.3.2, l'outil d'analyse *RCAExplore* offre une grande flexibilité quant à la manipulation du jeu de données car on peut sélectionner seulement les contextes formels et relationnels qui nous intéressent pour l'analyse. Par exemple, pour répondre à la question : "dans quels pays se trouvent les plantes *Azadirachta indica* et *Lantana camara* ?", il n'est pas nécessaire d'appliquer l'analyse relationnelle de concepts (ARC) sur tout le schéma relationnel de la figure 6. Il suffit de sélectionner la sous-partie du schéma relationnel qui nous intéresse à savoir : les contextes formels *country*, *plant* et le contexte relationnel *isProducedIn* qui décrit qu'une plante se trouve dans un pays (figure 9).

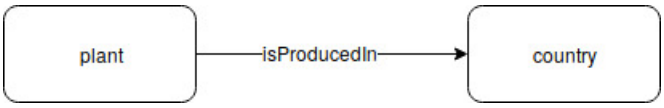


FIGURE 9 – Sous-schéma relationnel utilisé pour répondre à la question : dans quel pays se trouvent les plantes *Azadirachta indica* et *Lantana camara* ?

Une application directe de l'ARC comme expliqué dans la section 2.2 sur le schéma relationnel de la figure 9 génère les deux treillis *plant* et *country* avec respectivement 601 et 49 concepts. Il est difficile pour un utilisateur de naviguer dans le treillis des plantes et d'y chercher les plantes *Azadirachta indica* ou *Lantana camara* dans les 601 concepts. On utilise donc la navigation incrémentale proposée dans la section 2.3, où l'utilisateur indique en entrée le nom de la plante qu'il cherche, par exemple *Azadirachta indica*. On obtient les deux treillis *plant* et *country* avec respectivement 29 et 49 concepts. La figure 10 représente des parties des treillis *plant* (en haut) et *country* (en bas) [15] obtenus avec l'algorithme *On-demand* et dans lequel l'utilisateur sélectionne comme entrée la plante *Azadirachta indica*.

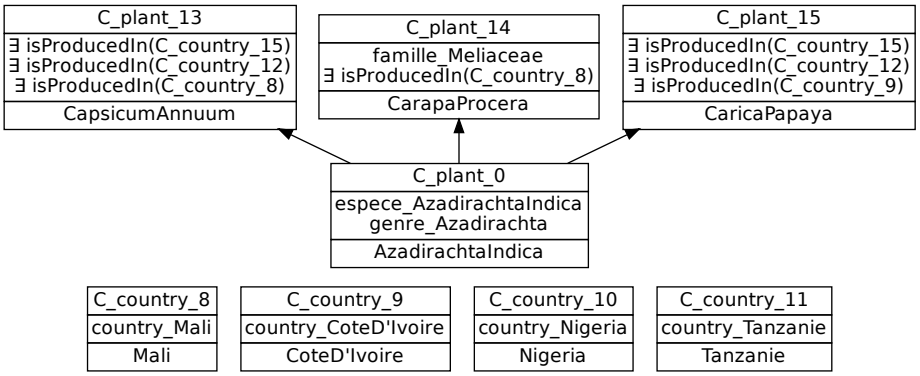


FIGURE 10 – Partie du treillis des plantes obtenu avec l'algorithme *On-demand* avec comme entrée la plante *Azadirachta indica* (en haut) et partie du treillis des pays (en bas).

On voit dans la partie du treillis des plantes (figure 10 - en haut) que c'est le concept *C_plant_0* qui nous donne des informations sur *Azadirachta indica*. Le concept *C_plant_0* hérite de l'attribut *Meliaceae* du concept *C_plant_14*. Cela nous dit que *Azadirachta indica* est une plante de la famille des *Meliaceae* et que *Azadirachta indica* et *Carapa procera* sont de la même famille botanique. Le concept *C_plant_0* hérite aussi de l'attribut relationnel $\exists\, isProducedIn(C_country_8)$ du concept *C_plant_14*, cela veut dire que la plante *Azadirachta indica* est présente dans au moins un des pays du concept *C_country_8*. Ce dernier contient uniquement le Mali, comme on peut le voir dans la

15. Bien que l'algorithme *On-demand* réduise la taille du treillis *plant* de 601 concepts à 29 concepts, ce treillis reste cependant illisible si on l'affiche en entier dans le rapport (voir annexe A.2).

partie du treillis des pays (figure 10 - en bas). En analysant tous les concepts de la couverture supérieure du concept *C_plant_0* on obtient les pays dans lesquels la plante *Azadirachta indica* est présente qui sont : Burkina Faso, Mali, Côte d'Ivoire, Tanzanie, Bénin, Cameroun, Ghana, Égypte, Togo, Kenya, Sénégal, Pakistan, Zimbabwe.
Pour avoir les pays (mentionnés dans PPAf) dans lesquels la plante *Lantana camara* est produite, on relance l'algorithme *On-demand* en prenant comme entrée la plante *Lantana camara* et on analyse les treillis de la même manière que pour la plante *Azadirachta indica*.

La réponse à la question : où se trouvent les plantes *Azadirachta indica* et *Lantana camara* ? est résumée dans la table 4 :

Plante	**Pays dans lesquels la plante se trouve**
Azadirachta indica	Burkina Faso, Mali, Côte d'Ivoire, Tanzanie, Bénin, Cameroun, Ghana, Égypte, Togo, Kenya, Sénégal, Pakistan, Zimbabwe
Lantana camara	Burkina Faso, Nigeria, Bénin, Kenya, Algérie

TABLE 4 – Réponse à la question : où se trouvent les plantes *Azadirachta indica* et *Lantana camara* ?

Contre quoi les plantes *Azadirachta indica* et *Lantana camara* sont-elles utilisées ? Et si la plante n'est pas présente, quelle(s) autre(s) espèce(s) végétale(s) pourrai(en)t atteindre au moins un des organismes ciblés ?

Pour répondre à ces questions nous utilisons le schéma relationnel de la figure 11 :

FIGURE 11 – Sous-schéma relationnel utilisé pour répondre à la question : que traitent les plantes *Azadirachta indica* et *Lantana camara* ?

Pour savoir quels ravageurs traitent les plantes *Azadirachta indica* et *Lantana camara*, il nous suffit d'étudier la relation *treats*. Après application de l'algorithme *On-demand* avec comme entrée la plante *Azadirachta indica*, nous obtenons le treillis des ravageurs dont une partie est représentée dans la figure 12 en bas, ainsi que le treillis des plantes qui contient le concept qui introduit la plante *Azadirachta indica*, sa couverture supérieure qui est composée de 40 concepts et sa couverture inférieure composée de 1 concept. Cependant nous nous intéressons tout d'abord au concept introducteur *C_plant_0* représenté dans la figure 12 (en haut).

Le concept *C_plant_0* possède plusieurs attributs relationnels dont l'attribut relationnel $\exists treatsC_Pest_73$. Cela veut dire que la plante *Azadirachta indica* traite au moins un ravageur du groupe de ravageurs *C_Pest_73*. En naviguant vers le treillis des ravageurs à travers cet attribut relationnel, on trouve que le concept *C_Pest_73* contient uniquement l'espèce *Diaeretiella rapae*. Si on continue la navigation à partir du concept *C_Pest_73* via la relation *attacks*, on peut savoir avec quelle(s) plante(s) l'espèce *Diaeretiella rapae* a une interaction. La figure 13 montre une partie du treillis des organismes cibles obtenue en continuant la navigation à partir du concept *C_Pest_73* (en haut) et une partie du treillis des organismes à protéger (en bas).

On voit dans la figure 13 que le concept *C_Pest_73* hérite de l'attribut relationnel $\exists\ attacks\ C_organism-to-protect_8$ du concept *C_pest_174*. Cela veut dire que

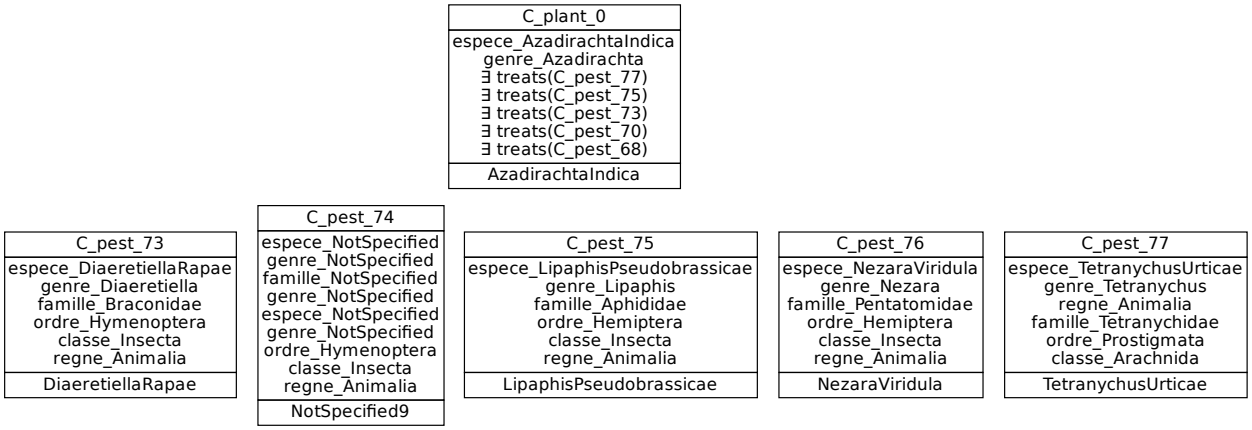


FIGURE 12 – Concept introducteur de la plante *Azadirachta indica* calculé avec l'algorithme *On-demand* (en haut). Partie du treillis des ravageurs (en bas).

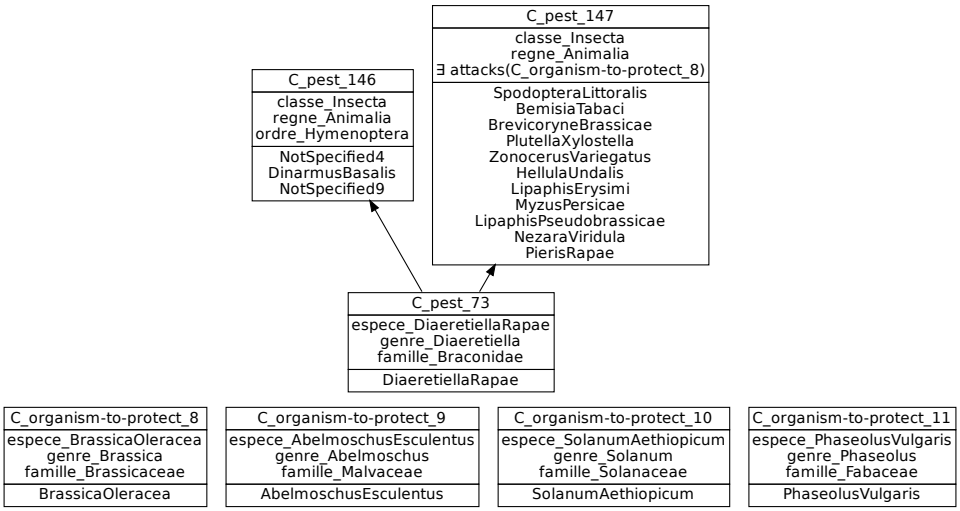


FIGURE 13 – Partie du treillis des ravageurs obtenue en continuant la navigation à partir du concept *C_Pest_73* de la figure 12 (en haut). Partie du treillis des organismes à protéger (en bas).

l'organisme *Diaeretiella rapae* du concept *C_Pest_73* est relié à au moins un organisme du groupe des organismes à protéger *C_organism-to-protect_8*. En navigant à travers cet attribut relationnel vers le treillis des organismes à protéger, on trouve que le concept *C_organism-to-protect_8* contient uniquement l'organisme à protéger *Brassica oleracea* (le chou).

La connaissance qu'on peut extraire à travers cette navigation (figures 12 et 13) est la suivante : actuellement, dans la base de connaissances PPAf, *Azadirachta indica* est la seule plante qui présente un lien entre la plante *Brassica oleracea* et *Diaeretiella rapae*.

En suivant le même raisonnement sur tous les attributs relationnels du concept *C_plant_0* (figure 12), on trouve les ravageurs qui sont uniquement traités par *Azadirachta indica* ainsi que les organismes qu'ils attaquent. En reprenant le même processus et le même raisonnement pour la plante *Lantana camara*, on ne trouve aucun ravageur qui ne soit traité que par la plante *Lantana camara*. La table 5 présente les ravageurs qui ne peuvent être traités

que par *Azadirachta indica* pour protéger les organismes que ces ravageurs attaquent.

Ravageurs traités seulement par l'*Azadirachta indica*	**Organisme attaqué par le ravageur**
Diaeretiella rapae, Lipaphis erysimi, Lipaphis pseudobrassicae	Brassica oleracea
Tetranychus urticae	Solanum lycopersicum
Poophilus costalis	Sorghum bicolor

TABLE 5 – Table des ravageurs uniquement traités que par *Azadirachta indica* et plantes attaquées par ces ravageurs.

On s'intéresse maintenant à la couverture supérieure du concept *C_plant_0*, pour traiter la question : quelles plantes utiliser à la place de *Azadirachta indica* ou de *Lantana camara* dans le cas où celles-ci ne sont pas présentes pour traiter au moins une de leurs cibles. La couverture supérieure du concept *C_plant_0* est composée de 40 concepts, nous en présentons seulement une partie dans la figure 14 (en haut).

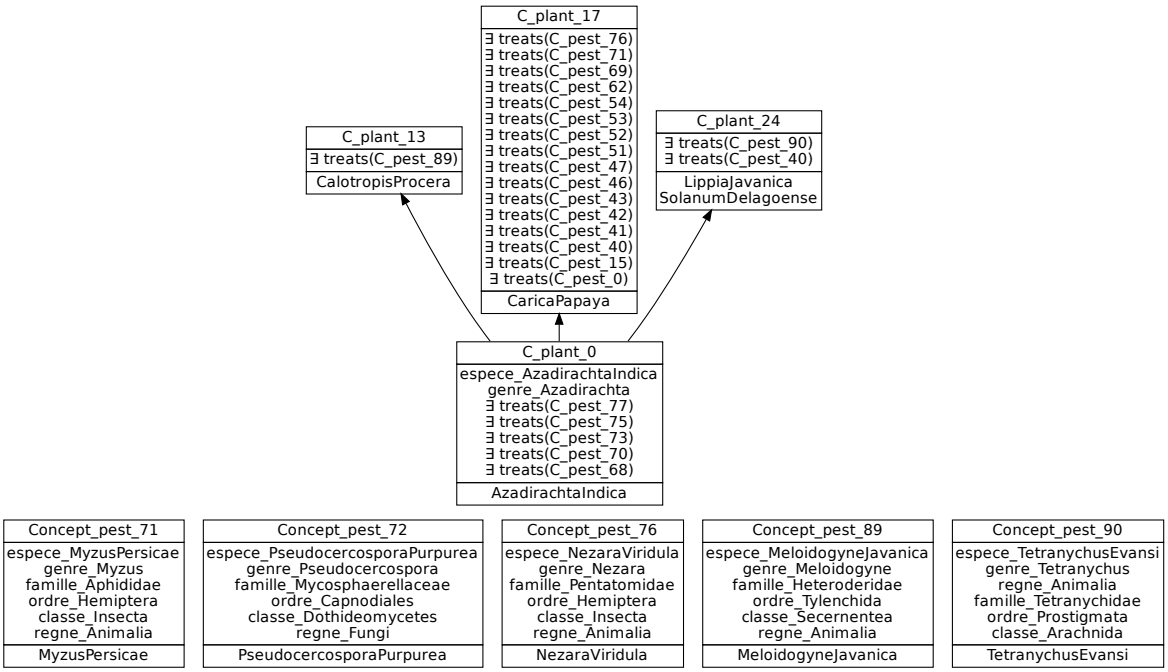


FIGURE 14 – Partie de la couverture supérieure du concept *C_plant_0* (en haut). Partie du treillis des ravageurs (en bas).

On voit dans la figure 14 (en haut) que le concept *C_plant_24* possède les attributs relationnels ∃ *treat C_pest_90* et ∃ *treat C_pest_40*. Cela nous informe que les objets de ce concept à savoir les plantes *Lippia javanica, Solanum delagoense* et *Azadirachta indica* ont en commun le fait qu'elles traitent au moins un ravageur des groupes de ravageurs *C_pest_90* et *C_pest_40*. En poursuivant la navigation vers le concept *C_pest_90* du treillis des ravageurs via l'attribut relationnel ∃ *treat (C_pest_90)*, on trouve que ce concept contient le ravageur *Tetranychus evansi* (figure 14 (en bas)). En continuant la navigation incrémentale à partir du concept *C_pest_90*, on obtient les parties des treillis des ravageurs et des organismes protégés présentées dans la figure 15. Le concept *C_pest_90* de la figure 15 (en haut), hérite de l'attribut relationnel ∃ *attacks (C_organism-to-protect_6)* du concept *C_pest_146*, cela veut dire que le ravageur *Tetranychus evansi* attaque au

moins un organisme du groupe des organismes à protéger *C_organism-to-protect_6* à savoir la plante *Solanum lycopersicum*.

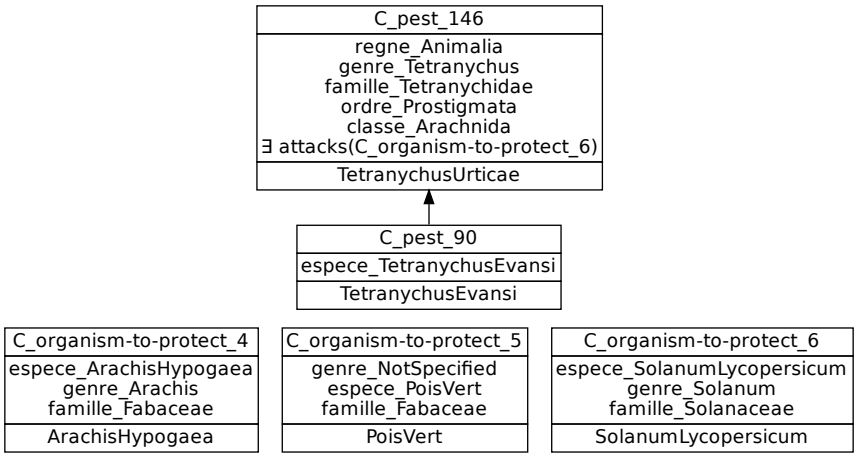


FIGURE 15 – Partie du treillis des ravageurs obtenue en continuant la navigation à partir du concept *C_Pest_90* de la figure 14 (en haut). Partie du treillis des organismes à protéger (en bas).

On peut en conclure à partir des figures 14 et 15 que pour protéger la plante *Solanum lycopersicum* contre le ravageur *Tetranychus evansi* on peut utiliser une des plantes suivantes : *Azadirachta indica, Lippia javanica* ou *Solanum delagoense*.

Les tables 6 et 7 présentent respectivement quelques résultats sur les plantes pouvant être utilisées à la place de *Azadirachta indica* et *Lantana camara* pour traiter un ravageur.

Plantes	**Ravageurs à traiter**
Carica papaya	Myzus persicae, Nezara viridula
Calotropis procera	Meloidogyne javanica
Hyptis suaveolens, Khaya senegalensis	Cryptophlebia leucotreta, Pectinophora gossypiella
Citrus sinensis, Moringa oleifera, Tithonia diversifolia	Pseudocercospora purpurea
Capsicum Spp, Chenopodium opulifolium, Ocimum gratissimum, Tephrosia vogelii	Bruchus rufimanus, Sitophilus spp, Tribolium spp

TABLE 6 – Plantes pouvant être utilisées à la place de *Azadirachta indica*.

Quelle(s) plante(s) utiliser au Burkina Faso pour lutter contre le champignon *Fusarium oxysporum* ?

Pour répondre à cette question, nous utilisons le sous-schéma relationnel présenté dans la figure 16. Nous partons des ravageurs en sélectionnant le ravageur *Fusarium oxysporum* comme entrée, pour trouver via la relation *isTreatedBy* quelles sont les plantes qui le traitent. Nous cherchons ensuite dans quels pays ces plantes sont produites en utilisant la relation *isProducedIn*. Si ces plantes sont produites au Burkina Faso, on suppose alors qu'elles peuvent être utilisées comme traitement pour lutter contre *Fusarium oxysporum* au Burkina Faso.

On voit dans la figure 17 (en haut) que le concept *C_pest_0* hérite de l'attribut relationnel $\exists$ *isTreatedBy* (*C_plant_*211) du concept *C_pest_27*. Cela veut dire que le ravageur *Fusarium oxysporum* du concept *C_pest_0* peut être traité par au moins une plante du concept *C_plant_211* (figure 17 en bas), c'est-à-dire la plante *Hyptis suaveolens*.

Plantes	Ravageurs à traiter
Hyptis spicigera	Colletotrichum graminicola
Hyptis spicigera, Ocimum americanum	Curvularia lunata, Exserohilum rostratum
Hyptis spicigera, Ocimum americanum, Europhobia hirta	Epicoccum sorghinum
Dysphania ambrosioides, Acorus calamus, Carum carvi, Tanacetum cinerariifolium	Sitophilus granarius
Ocimum canum, Ocimum gratissimum, Europhobia hirta, Azadirachta indica, Tephrosia vogelii	Fusarium oxysporum

TABLE 7 – Plantes pouvant être utilisées à la place du *Lantana camara*.

FIGURE 16 – Sous-schéma relationnel utilisé pour répondre à la question : quelle(s) plante(s) utiliser au Burkina Faso pour lutter contre le ravageur *Fusarium oxysporum* ?

Pour savoir si la plante *Hyptis suaveolens* se trouve au Burkina Faso, nous continuons donc la navigation à partir du concept *C_plant_211* vers le treillis des pays en utilisant la relation *isProducedIn*. Nous obtenons des treillis dont une partie est présentée par la figure 18.

On voit dans la figure 18 (en haut) que le concept *C_plant_211* hérite de l'attribut relationnel $\exists$ $isProducedIn$ ($C_country_1$). Cela veut dire que la plante *Hyptis suaveolens* se trouve au moins dans un des pays du concept *C_country_1*, c'est-à-dire le Burkina Faso (figure 18 en bas).
À partir des navigations des figures 17 et 18, on peut donc émettre l'hypothèse que la plante *Hyptis suaveolens* pourra être utilisée comme traitement pour lutter contre le ravageur *Fusarium oxysporum* au Burkina Faso.

En naviguant à travers tous les attributs relationnels du concept *C_pest_0* (figure 17 en haut), nous avons trouvé dans PPAf les plantes qui traitent le ravageur *Fusarium oxysporum* et qui sont produites au Burkina Faso. Ces plantes sont [16] : *Hyptis suaveolens, Lantana camara, Ocimum canum, Euphorbia hirta, Azadirachta indica.*

3.5 Difficultés liées à l'analyse

L'étape d'analyse était a été complexe à plus d'un titre, à commencer par la modélisation de données. En effet, à l'heure actuelle, certaines informations restent ambiguës et demandent des précisions complémentaires. Par exemple, nous ne savons pas si l'information sur le pays représente le pays qui cultive la plante à effet pesticide, ou s'il s'agit du pays dans lequel nous avons fait l'expérimentation sur le ravageur. Nous avons donc considéré pour notre analyse que la plante à effet pesticide est cultivée dans le même pays que celui où se déroule l'expérimentation sur le ravageur.
Une autre difficulté liée à l'analyse est la visualisation des treillis. Ces derniers sont de taille très importante et il est difficile de les analyser. Les connaissances sont tellement dispersées dans les treillis qu'il est très difficile de les appréhender sans un système de visualisation pour faciliter la navigation dans les treillis et l'analyse. Bien que la navigation sur demande réduise considérablement la taille des treillis, cela reste insuffisant. Nous

16. Les parties des treillis qui couvrent l'ensemble de ces plantes ne sont pas mises dans ce rapport à cause de leur grande taille.

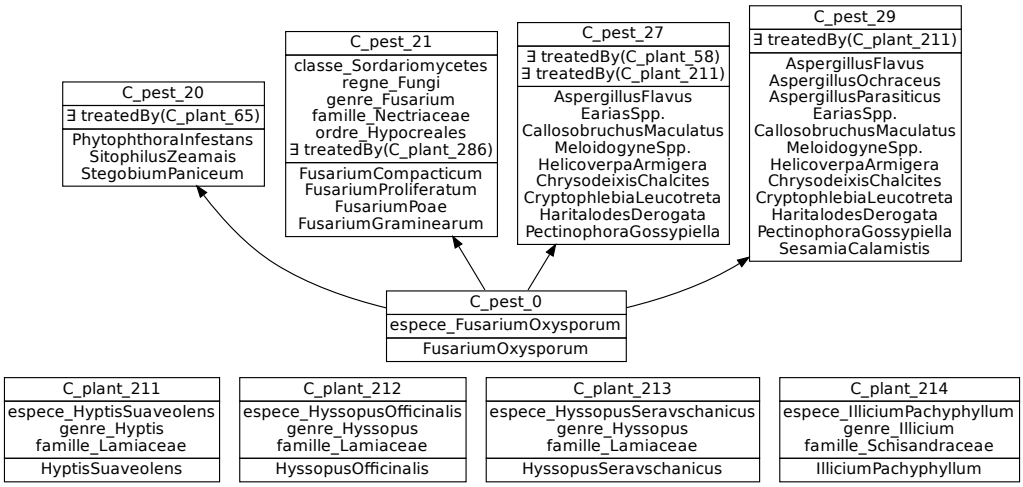


FIGURE 17 – Partie de la couverture supérieure du concept *C_pest_0* qui introduit la ravageur *Fusarium oxysporum* (en haut). Partie du treillis des plantes (en bas).

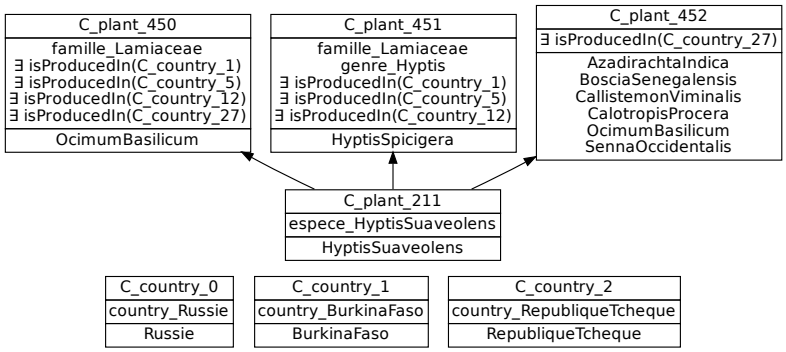


FIGURE 18 – Partie du treillis des plantes obtenue en continuant la navigation à partir du concept *C_plant_211* de la figure 17 (en haut). Partie du treillis des pays (en bas).

avons donc pensé à développer des scripts pour extraire les connaissances qui nous intéressent, mais les personnes intéressées par le projet *Knomana* et la réalité du terrain nous ont confronté à des questions diverses et variées auxquelles il était impossible de répondre par de simple scripts mais plutôt par la mise en place d'un système de requêtes complet dédié au treillis de concepts.

4 Conclusion et Perspectives

Dans ce rapport nous avons présenté dans la section 1 le projet *Knomana* dont le but est de recenser et d'étudier l'usage des plantes à effet pesticide. Nous avons aussi présenté une des problématiques qui lui est liée et sur laquelle portent nos travaux, à savoir l'analyse et la navigation dans les données pour en extraire de la connaissances. Nous avons présenté dans la section 2 les méthodes d'analyse formelle et relationnelle de concepts utilisées pour l'analyse ainsi que les deux approches existantes pour naviguer (dans un treillis, entre treillis) dans les données. Nous avons aussi proposé et implémenté une nouvelle approche de navigation sur demande qui accompagne pas-à-pas l'utilisateur dans sa navigation dans les données. Dans la section 3 nous avons présenté dans un premier temps quelques résultats obtenus par l'analyse relationnelle de concepts sur la base de connaissances PPAf, et présenté quelques avantages qu'offre cette méthode d'analyse. Dans un deuxième temps, nous avons apporté des éléments de réponse à des sollicitations reçues

de la part de personnes intéressées par l'usage des plantes à effet pesticide.

Dans nos prochains travaux, nous intégrerons pour l'analyse l'ensemble des données et des relations qui existent dans PPAf. Nous optimiserons et étendrons l'approche de navigation sur demande de manière à ce qu'elle soit plus flexible et plus intuitive pour l'utilisateur. Nous essayerons de proposer et de développer une stratégie générale d'extraction de connaissances ainsi qu'un système de requêtes dédié au treillis de concepts pour l'extraction de connaissances. De plus, nous étudierons comment mettre en place un système de visualisation pour la présentation des résultats qui demeure un vrai challenge quand nous travaillons avec des treillis de taille importante.

Références

[1] Alexandre Bazin, Jessie Carbonnel, Marianne Huchard, and Giacomo Kahn. On-demand relational concept analysis. *CoRR*, abs/1803.07847, 2018.

[2] Agnès Braud, Xavier Dolques, Marianne Huchard, and Florence Le Ber. Generalization effect of quantifiers in a classification based on relational concept analysis. *Knowledge-Based Systems*, 2018.

[3] Sébastien Ferré. A proposal for extending formal concept analysis to knowledge graphs. In *ICFCA*, volume 9113 of *Lecture Notes in Computer Science*, pages 271–286. Springer, 2015.

[4] Sébastien Ferré and Peggy Cellier. Graph-fca in practice. In *ICCS*, volume 9717 of *Lecture Notes in Computer Science*, pages 107–121. Springer, 2016.

[5] Bernhard Ganter and Rudolf Wille. *Formal concept analysis - mathematical foundations*. Springer, 1999.

[6] Gillian J. Greene, Marvin Esterhuizen, and Bernd Fischer. Visualizing and exploring software version control repositories using interactive tag clouds over formal concept lattices. *Information & Software Technology*, 87 :223–241, 2017.

[7] Mohamed Rouane Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Relational concept analysis : mining concept lattices from multi-relational data. *Ann. Math. Artif. Intell.*, 67(1) :81–108, 2013.

[8] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining Concepts and Techniques*, volume 673, chapter 8, pages 330–350. Morgan Kaufmann, 3rd edition, 2011.

[9] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining Concepts and Techniques*, volume 673, chapter 1, pages 243–271. Morgan Kaufmann, 3rd edition, 2011.

[10] Céline Hébert, Alain Bretto, and Bruno Crémilleux. A data mining formalization to improve hypergraph minimal transversal computation. *Fundam. Inform.*, 80(4) :415–433, 2007.

[11] Mohamed Nidhal Jelassi. *Representation and applications of hypergraph minimal transversals, Chapter 2, Pages 19-21.* Theses, Université Jean Monnet - Saint-Etienne, December 2014.

[12] Priscilla Keip. Navigation conceptuelle et exploration pour la sélection de plantes pesticides et antibiotiques en santés animale et végétale dans les pays du sud. Master's thesis, 2018. Université de Montpellier.

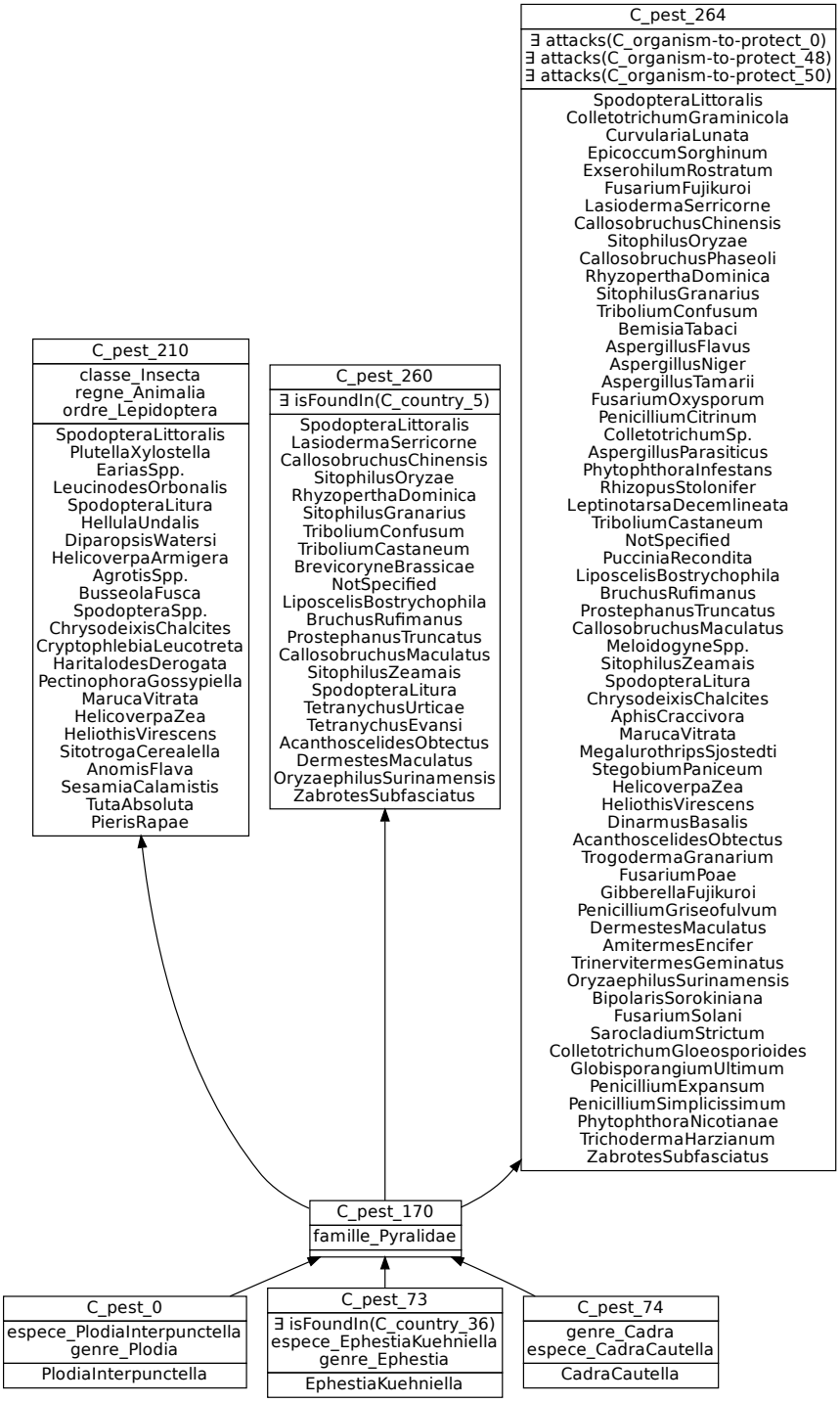
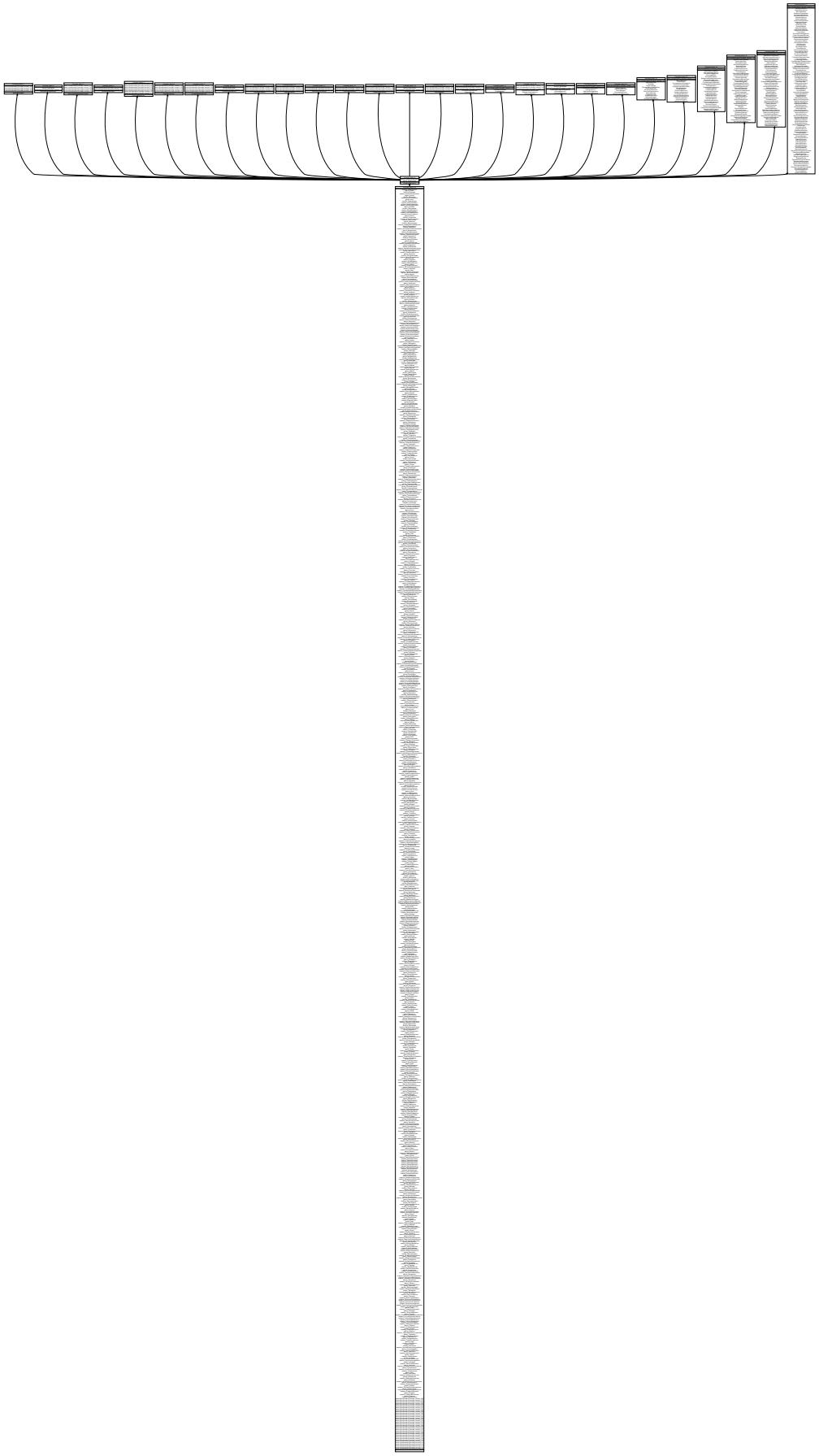
[13] Christine Largeron, Sadok Ben Yahia, and Nidhal Jelassi. Local-Generator : "diviser pour régner" pour l'extraction des traverses minimales d'un hypergraphe. In *Actes de la conférence Extraction et gestion des connaissances (EGC 2014)*, pages 245–256, Rennes, France, January 2014.

[14] Amirouche Ouzerdine. Étude de méthodes d'analyse de données et d'extraction de connaissances pour l'usage des plantes pesticides -projet knomana-. Master's thesis, 2018. Université de Montpellier.

Annexes

A Analyses et résultats

A.1 Avantages de l'analyse relationnelle de concepts



A.2 Résultats d'analyse