HAL
open science

Les boîtes arc-en-ciel : une nouvelle technique pour visualiser les relations d'instanciation dans les ontologies et son application à la comparaison des propriétés des médicaments

Jean-Baptiste Lamy, Hélène Berthelot, Madeleine Favre

To cite this version:

Jean-Baptiste Lamy, Hélène Berthelot, Madeleine Favre. Les boîtes arc-en-ciel : une nouvelle technique pour visualiser les relations d'instanciation dans les ontologies et son application à la comparaison des propriétés des médicaments. Conférence Nationale en Intelligence Artificielle, Jun 2016, Clermont Ferrand, France. hal-01558578

HAL Id: hal-01558578

https://hal.science/hal-01558578v1

Submitted on 8 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les boîtes arc-en-ciel : une nouvelle technique pour visualiser les relations d'instanciation dans les ontologies et son application à la comparaison des propriétés des médicaments

Jean-Baptiste Lamy[1] Hélène Berthelot[1] Madeleine Favre[2,3]

[1] LIMICS Université Paris 13 - Sorbonne Paris Cité, 93017 Bobigny, INSERM UMRS 1142
UPMC Université Paris 6, Sorbonne Universités, Paris, France
[2] Université Paris Descartes, Faculté de Médecine, Département de Médecine Générale, Paris, France
[3] Société de Formation Thérapeutique du Généraliste (SFTG), Paris, France

jean-baptiste.lamy@univ-paris13.fr

Résumé

Dans les ontologies, les relations hiérarchiques de subsomption (est-un) et d'instanciation (est-l'instance-de) jouent un rôle important et sont notamment utilisées pour visualiser les ontologies sous forme de graphe. Cependant ces graphes deviennent vite illisibles lorsque les relations sont complexes et font intervenir de l'héritage multiple. C'est le cas dans l'exemple simple des acides aminés (20 acides aminés et 10 classes), mais aussi dans des applications plus complexes telles que la comparaison des propriétés des médicaments (2-10 médicaments de même indication et 10-100 classes correspondant à une propriété commune, contre-indication ou effet indésirable).
Dans cet article, nous présentons les boîtes arc-en-ciel, une nouvelle technique de visualisation adaptée à la présentation des relations d'instanciation complexes et multiples dans des ontologies de faible taille. Nous décrivons l'application de cette technique aux deux exemples précédents. Nous présentons aussi une expérimentation utilisateur sur l'application médicale, montrant que les médecins accèdent plus rapidement à l'information avec les boîtes arc-en-ciel qu'avec une approche « classique » à base de tableaux.

Mots Clef

Visualisation d'ontologie, Visualisation de connaissance, Médicaments, Expérimentation utilisateur

Abstract

In ontologies, hierarchical relations of subsumption (is-a) and instantiation (is-instance-of) play an important role, in particular for visualizing the ontology, generally using a graph. However, these graphs become hard-to-read when the relations are numerous and complex, and when multiple inheritance is involved. This is the case in the simple example of amino-acids (20 amino-acids and 10 classes), but also in more complex applications such as the comparison of drug properties (2-10 drugs with the same indication and 10-100 classes corresponding to a common property, contraindication or adverse effect).
In this paper, we present rainbow boxes, a novel visualization technique adapted to the presentation of complex and numerous instantiation relations in small ontologies. We describe the application of this technique to the two previous examples. We also present a user study in the medical application, showing that physicians find information faster with rainbow boxes than with a «classic» table approach.

Keywords

Ontology visualization, Knowledge visualization, Drugs, User study

1 Introduction

Les ontologies sont de plus en plus employées dans le monde de l'informatique. Dans les ontologies, les relations hiérarchiques de subsomption (*est-un*) et d'instanciation (*est-l'instance-de*) jouent un rôle important. Ces relations sont notamment utilisées pour organiser graphiquement et visualiser les classes et les instances d'une ontologie, généralement à l'aide d'arbres ou de graphes. Cependant les graphes deviennent vite illisibles lorsque les relations sont nombreuses et complexes, et font intervenir de l'héritage ou de l'instanciation multiple.
Un exemple est celui des acides aminés. Il existe 20 acides aminés (Alanine, Proline... que l'on peut abréger en code de 3 lettres : Ala, Pro...). Plusieurs acides aminés ont des propriétés physiques ou chimiques communes, telles que leur *petite* taille ou la présence d'un cycle *aromatique*. Une dizaine de propriétés sont traditionnellement considérées et visualisées dans un diagramme de Venn (Figure 1). Ces connaissances peuvent être formalisées dans une ontologie

comprenant 20 individus (les acides aminés) et 10 classes (regroupant les acides aminés partageant une propriété, par exemple la classe des *petits* acides aminés ou la classe des acides aminés *aromatiques*). Les classes ne sont pas indépendantes les unes des autres : par exemple, il est évident que les acides aminés *minuscules* sont aussi *petits* (subsomption), et les acides aminés *aromatiques* ne peuvent pas être *petits* (disjonction, car le cycle aromatique est volumineux). À cause du nombre important de relations et de l'instanciation multiple, la visualisation de l'ontologie des acides aminés conduit au graphe de la Figure 2, qui est encore moins lisible que le diagramme de Venn, déjà complexe.

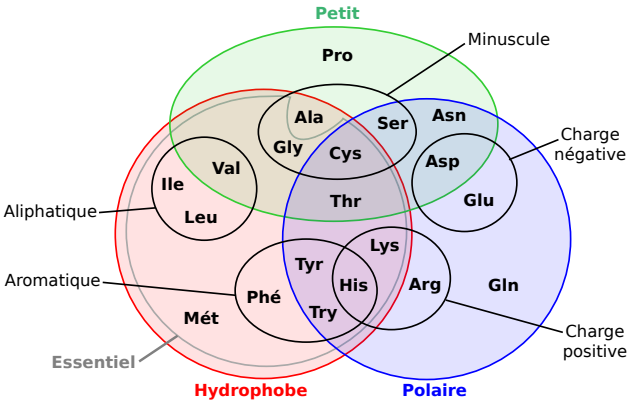

FIGURE 1 – Diagramme de Venn montrant les propriétés des acides aminés.

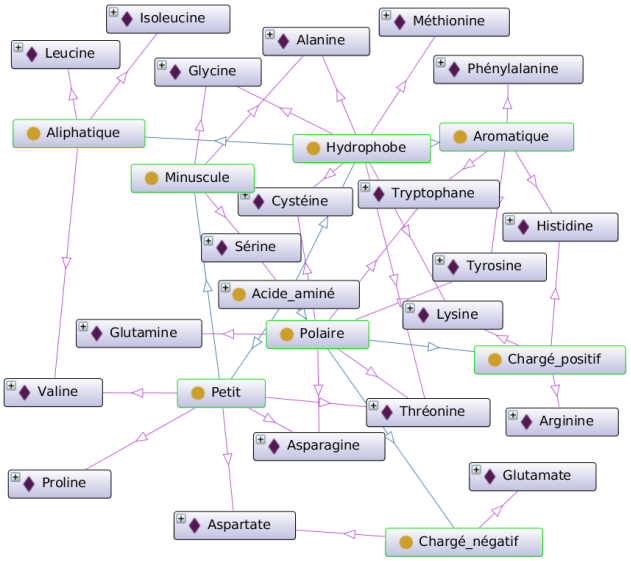

FIGURE 2 – Graphe représentant l'ontologie des acides aminés (généré avec OntoGraph).

Un autre exemple plus complexe concerne la comparaison des propriétés de médicaments similaires. Les propriétés des médicaments telles que les indications, contre-indications, interactions et effets indésirables sont listées dans des textes officiels appelés Résumés des Caractéristiques Produits (RCP), et ensuite rassemblées dans des bases de données. Parmi les médicaments partageant une même indication, plusieurs classes peuvent être définies en fonction de leurs propriétés, par exemple la classe des médicaments *contre-indiqués en cas d'hypertension* ou la classe des médicaments *pouvant provoquer des nausées* (effet indésirable). Un même médicament peut bien sûr appartenir à plusieurs de ces classes. La présentation visuelle des relations d'instanciation entre ces classes et les médicaments permettrait aux médecins d'avoir une vue d'ensemble sur les médicaments disponibles pour une indication donnée. Ce travail a été réalisé dans le cadre du projet VIIIP (Visualisation Intégrée de l'Information sur l'Innovation Pharmaceutique) qui vise à permettre la comparaison visuelle des propriétés des médicaments, notamment pour les nouveaux médicaments.

Dans cet article, afin de faciliter la visualisation, nous proposons de nous limiter aux seules relations d'instanciation. Nous présentons pour cela les *boîtes arc-en-ciel* (*rainbox boxes* en anglais), une nouvelle technique de visualisation que nous avons mise au point et qui permet la représentation de relations d'instanciation complexes et multiples pour des ontologies de faible taille (environ 2-25 individus et 5-100 classes).

L'article est organisé de la manière suivante. La section 2 présente un état de l'art sur la visualisation des ontologies. La section 3 décrit les boîtes arc-en-ciel en s'appuyant sur l'exemple simple des acides aminés. La section 4 décrit l'utilisation des boîtes arc-en-ciel dans un domaine plus complexe, la comparaison des propriétés des médicaments (contre-indications et effets indésirables). La section 5 présente une expérimentation utilisateur sur cette application médicale. Enfin, la section 6 discute les principaux résultats et propose des perspectives.

2 État de l'art

La visualisation des relations d'instanciation dans les ontologies s'apparente à la visualisation d'*ensembles non disjoints* (ou ensembles chevauchant, *overlapping sets* en anglais) : les individus d'une ontologie peuvent être considérés comme des éléments et les classes comme des ensembles incluant tout ou partie de ces éléments. Les ensembles ainsi constitués se chevauchent, c'est-à-dire qu'un élément peut appartenir à plusieurs ensembles (en cas d'instanciation multiple). La visualisation d'ensembles non disjoints a donné lieu à de nombreux travaux dont B. Alsakallah *et al.* ont effectué une revue récente [2]. La grande majorité des travaux s'inspire des diagrammes d'Euler ou de Venn, ou bien s'intéresse à la représentation de jeux de données volumineux.

En revanche, peu de travaux ont cherché à visualiser de la connaissance plutôt que des données, c'est-à-dire des volumes relativement faibles mais complexes et pour lesquels il est souhaitable de visualiser la totalité des éléments simultanément. La visualisation des ontologies et des hiérar-

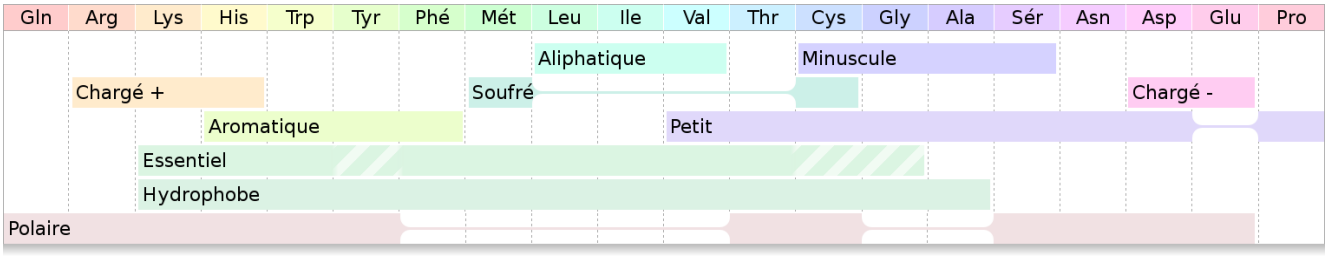

FIGURE 3 – Boîtes arc-en-ciel montrant 10 propriétés des 20 acides aminés.

chies de subsomption a fait l'objet de deux revues [9, 6]. La plupart des approches présentent les relations hiérarchiques sous forme de graphes, deux cas particuliers bien connus étant les arbres et les listes indentées. Ces méthodes ont par exemple été appliquées à la visualisation d'ontologies en anatomie [5] et de terminologies médicales [1]. Les diagrammes de Venn et les *treemaps* ont aussi été utilisés, par exemple en bioinformatique [3], et ont donné lieu à des variantes telles que Jambalaya [16] et CropCircle [19]. Cependant, une étude [10] comparant quatre modes de visualisation a montré que la liste arborescente indentée (présente de base dans l'éditeur Protégé), bien que très simple, donnait les meilleurs résultats en termes de temps de réponse. Il reste donc certainement des améliorations possibles dans ce domaine.

Concernant la comparaison des propriétés de plusieurs médicaments, les seules approches de visualisation utilisées dans la littérature sont des tableaux ou des matrices. Wroe *et al.* [20] ont proposé DOPAMINE, un outil à base de matrices, principalement à des fins de rédactions et de vérifications des propriétés dans les bases de données. Iordatii *et al.* [8] ont proposé une approche à base de tableaux pour comparer les contre-indications et les effets indésirables d'un nouveau médicament à un médicament de référence. Les *Drug Fact Boxes* [17] offrent une information comparative sur le médicament, mais limitée à un faible nombre d'entre eux et destinée aux patients plutôt qu'aux médecins.

3 Les boîtes arc-en-ciel

3.1 Principe général

Les boîtes arc-en-ciel sont une nouvelle technique pour visualiser des relations d'instanciation ou des ensembles non disjoints. Dans les boîtes arc-en-ciel, chaque instance est représentée par une colonne et chaque classe par une boîte rectangulaire qui occupe les colonnes correspondant aux instances membres de la classe. Le nom de la classe est affiché à l'intérieur de la boîte. Lorsque cela est nécessaire, des « trous » sont présents dans les boîtes ; les instances sont ordonnées de sorte à limiter au maximum les trous, à l'aide d'un algorithme décrit en section 3.2. Les boîtes les plus larges sont placées en bas (la largeur étant calculée globalement, sans tenir compte des trous), ensuite les boîtes sont empilées comme les blocs d'un jeu de « Tétris » ; plusieurs boîtes peuvent donc être côte à côte du moment qu'elles n'occupent pas les mêmes colonnes. La Figure 3 présente l'ontologie des acides aminés avec des boîtes arc-en-ciel.

Par défaut, les boîtes arc-en-ciel n'utilisent pas la couleur pour apporter une information supplémentaire, mais plutôt pour faciliter la lecture et rendre la visualisation plus attractive. Chaque instance reçoit une couleur différente, du rouge (à gauche) au violet (à droite) en traversant tout le spectre selon l'ordre des colonnes. La couleur d'une boîte est la moyenne des couleurs (en coordonnées RVB) des instances de la classe. Par conséquent, les boîtes regroupant peu d'instances ont tendance à prendre la couleur de leurs instances tandis que les boîtes avec beaucoup d'instances tirent sur le gris.

Néanmoins les couleurs peuvent aussi être spécifiées manuellement pour chaque couple (classe, instance) afin d'apporter une information complémentaire (voir exemple sur les effets indésirables en section 4.2). Les boîtes peuvent aussi inclure des hachures pour indiquer une appartenance « partielle » d'une instance à une classe, par exemple pour les acides aminés semi-essentiels sur la Figure 3.

Les boîtes arc-en-ciel indiquent l'appartenance des instances aux différentes classes, mais permettent aussi de découvrir de nouvelles relations ou similarités. Sur la Figure 3, les relations de subsomption entre classes ne sont pas représentées explicitement mais peuvent néanmoins se deviner « par extension », par exemple les acides aminés *minuscules* sont tous *petits* puisque la boîte *petit* couvre la totalité des colonnes couvertes par la boîte *minuscule*. De même pour les disjonctions, par exemple il est facile de constater qu'aucun acide aminé n'est à la fois *petit* et *aromatique*. Par ailleurs, il est frappant de constater que la classe des acides aminés *essentiels* correspond quasiment à celle des *hydrophobes*.

3.2 Optimisation de l'ordre des colonnes

Pour tracer des boîtes arc-en-ciel avec le plus petit nombre possible de trous, l'ordre des colonnes (c'est-à-dire des instances) doit être optimisé. Le nombre d'ordres possibles est $n!$ (factoriel n) où n est le nombre d'instances. La complexité du problème est donc $\mathcal{O}(n!)$. Pour des valeurs de n relativement élevées (au-delà de 10), il n'est pas possible de tester tous les ordres possibles. Pour l'ontologie des acides aminés, il y a $20\,! \approx 2{,}43\text{x}10^{18}$ ordres possibles. Pour résoudre ce problème dans un temps satisfaisant, nous

Algorithme 1 Heuristique pour trouver les ordres candidats (quasi-)optimaux (*ordres_candidats*). *instances* est l'ensemble de toutes les instances et *classes* l'ensemble de toutes les classes.

premiers_éléments_candidats = { *i0* pour chaque instance *i0* dans *instances* qui appartient au plus grand nombre de classes dans *classes* }
ordres_en_construction = { tuple (*i0*) pour chaque instance *i0* dans *premiers_éléments_candidats* }
ordres_candidats = { } (ensemble vide)

Tant que *ordres_en_construction* $\neq$ { } :
  ordre = Prendre et enlever un tuple de *ordres_en_construction*
  restant = *instances* $\setminus$ *ordre*
  insertions_possible = { (*i*, "début") pour chaque *i* dans *restant* } $\bigcup$ { (*i*, "fin") pour chaque *i* dans *restant* }
  Pour chaque couple (*i*, *position*) dans *insertions_possible* avec le plus grand **score_d_insertion**(*i*, *position*, *ordre*) :
    Si *position* est "début" : *nouvel_ordre* = (*i*, *ordre*)
    Sinon si *position* est "fin" : *nouvel_ordre* = (*ordre*, *i*)
    Si *instances* $\setminus$ *nouvel_ordre* = { } : Ajouter *nouvel_ordre* dans *ordres_candidats*
    Sinon : Ajouter *nouvel_ordre* dans *ordres_en_construction*

Fonction **score_d_insertion**(*instance*, *position*, *ordre*) :
  Si *position* est "début" : *voisin* = Premier élément de *ordre*
  Sinon si *position* est "fin" : *voisin* = Dernier élément de *ordre*
  score = 0
  Pour chaque classe *X* dans *classes*, tel que (*instance* $\sqsubseteq X$) et (*voisin* $\sqsubseteq X$) : *score* = *score* + 2
  Pour chaque classe *Y* dans *classes*, tel que (*instance* $\not\sqsubseteq Y$) et ($\nexists\, j \in ordre, j \sqsubseteq Y$) : *score* = *score* + 1
  Retourne *score*

proposons ici une heuristique pour obtenir un ordre (quasi) optimal dans un temps raisonnable. Cette heuristique génère dans un premier temps une liste réduite d'*ordres candidats* (204 dans l'exemple des acides aminés) et ensuite teste chacun de ces ordres candidats pour garder le meilleur (c'est-à-dire celui conduisant au moins de trous et aux trous les plus courts dans la visualisation).

L'Algorithme 1 génère la liste des ordres candidats. Les ordres candidats sont construits progressivement, en commençant par *i0*, l'instance qui appartient au plus grand nombre de classes et qui est donc la plus susceptible de créer des trous puisqu'elle fait intervenir le plus grand nombre de boîtes. Ensuite, les instances restantes sont ajoutées soit au début soit à la fin de l'ordre en cours de construction. À chaque étape, l'instance à ajouter et sa position (début ou fin) sont déterminées à l'aide d'un score d'insertion (l'insertion avec le meilleur score étant retenue). Ce score prend en compte (a) le nombre de classes auxquelles appartiennent à la fois l'instance à ajouter et sa voisine (c'est-à-dire la première ou la dernière instance actuelle de l'ordre en cours de construction, selon que l'on ajoute au début ou à la fin, respectivement) et (b) le nombre de classes auxquelles n'appartiennent ni l'instance à ajouter ni aucune des instances déjà présentes dans l'ordre en cours de construction. La partie (a) favorise l'ajout d'une instance qui appartient aux mêmes classes que sa future voisine, tandis que la partie (b) favorise l'ajout d'une instance qui n'appartient pas à des classes qui sont pour l'instant absentes de l'ordre en cours de construction, c'est-à-dire que (b) tend à « retarder » l'ajout de nouvelles boîtes dans la visualisation. Enfin, lorsque plusieurs *ex aequo* sont présents pour choisir *i0* ou pour choisir l'instance à ajouter ou sa position, toutes les possibilités sont considérées, ce qui conduit à la génération de plusieurs ordres candidats.

4 Application à la comparaison des propriétés des médicaments

Nous décrirons dans cette section l'application des boîtes arc-en-ciel à la comparaison des contre-indications et des effets indésirables des médicaments.

4.1 Comparaison des contre-indications

La Figure 4 présente sous forme de boîtes arc-en-ciel les contre-indications de 8 médicaments pour la dysfonction érectile. Les boîtes arc-en-ciel « basiques » décrites à la section 3 ont été enrichies avec : (a) des en-têtes indiquant les classes pharmaco-thérapeutiques, au-dessus des en-têtes de colonnes, tels que « Prostaglandine E1 » pour Vitaros®, (b) des hachures pour représenter les situations où le médicament est déconseillé, c'est-à-dire des contre-indications relatives et non absolues, telles que l'insuffisance rénale sévère pour Cialis®, (c) des *widgets*, pour ajouter des informations spécifiques sur un couple (médicament, contre-indication) donné, tels que l'étiquette « (non-controllée) » pour la contre-indication hypertension artérielle de Cialis® ou les barres colorées dans la boîte « Âge », (d) des icônes VCM à droite des boîtes.

Le langage iconique VCM (Visualisation des Concepts en Médecine) [12, 13, 14] propose des icônes pour représenter les principaux concepts médicaux, et notamment les conditions patients et les maladies. VCM comprend un ensemble de primitives graphiques (couleurs, formes et pictogrammes) et une grammaire graphique qui permet de combiner les primitives pour créer les icônes.

Les boîtes arc-en-ciel de la Figure 4 montrent les contre-indications de 8 médicaments, mais montrent aussi les similarités entre médicaments et en particulier les ressemblances qui existent au sein de chacune des deux

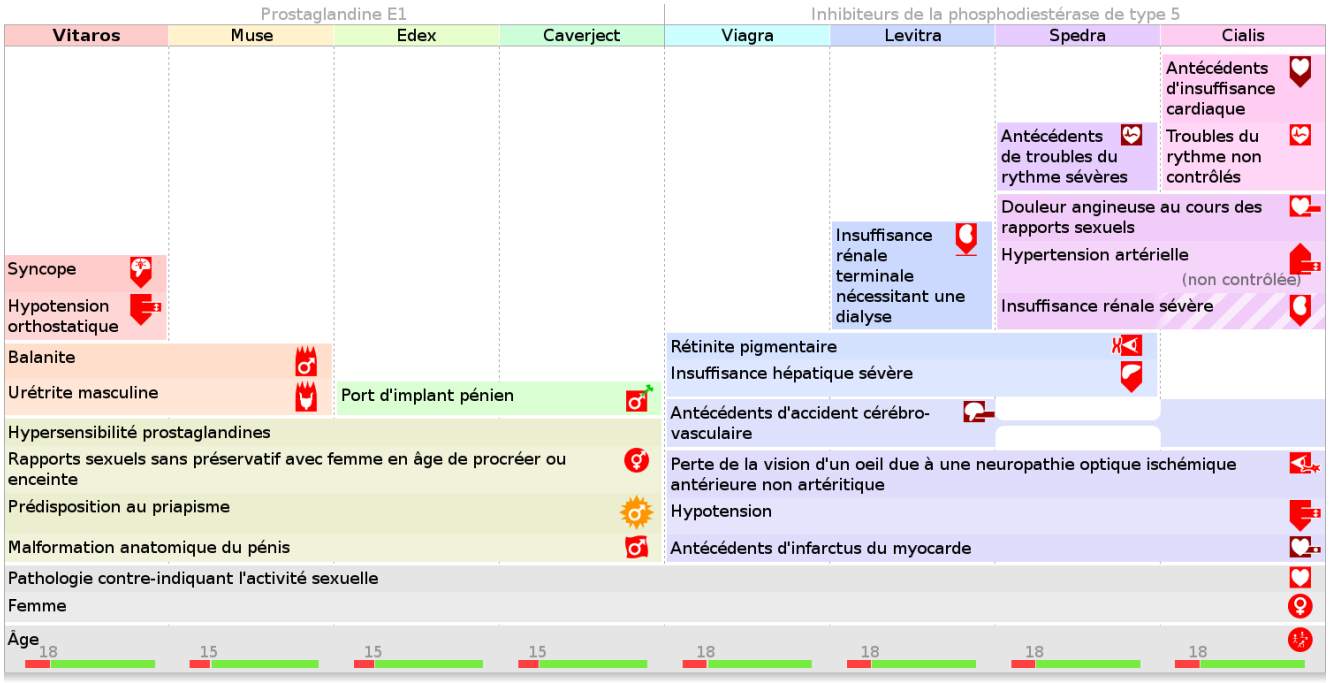

FIGURE 4 – Boîtes arc-en-ciel montrant les 79 contre-indications (26 distinctes) de 8 médicaments pour traiter la dysfonction érectile.

classes pharmaco-thérapeutiques. Par exemple, la contre-indication avec la prédisposition au priapisme est clairement une propriété commune à la classe des prostaglandines E1.

La visualisation a été rendue interactive : en cliquant sur les colonnes, l'utilisateur peut afficher une comparaison limitée aux deux médicaments choisis (Figure 5). Cette comparaison comprend trois parties : en haut à gauche, les contre-indications spécifiques du premier médicament, en haut à droite, celles du second médicament, et en bas, les contre-indications communes aux deux. Dans chaque partie, les contre-indications sont triées par organes.

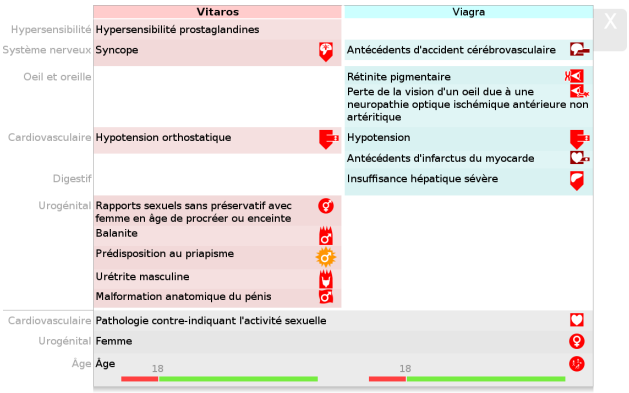

FIGURE 5 – Comparaison des contre-indications de deux médicaments (Vitaros® et Viagra®).

4.2 Comparaison des effets indésirables

La Figure 6 montre un exemple de boîtes arc-en-ciel pour présenter les effets indésirables des huit médicaments. Les effets indésirables sont plus complexes à visualiser que les contre-indications, car ils sont caractérisés par deux attributs additionnels : la gravité (attribut booléen) et la fréquence de l'effet (mesurée sur l'échelle standard de cinq valeurs : très rare, rare, peu fréquent, fréquent, très fréquent). Seuls les effets graves et/ou fréquents (ou très fréquents) ont été retenus, en effet les médecins ne sont pas intéressés *a priori* par les effets non-graves et peu fréquents, d'après un focus groupe réalisé au début du projet.

Les deux attributs ont été représentés à l'aide de couleurs personnalisées dans chaque cellule. La teinte indique la gravité (rouge : grave, orange : non grave) tandis que la saturation indique la fréquence (les couleurs plus saturées correspondent à des effets plus fréquents). Pour certaines boîtes, la couleur peut changer d'une colonne à l'autre (par exemple la boîte « Douleur du pénis » est moins saturée pour Vitaros® car cet effet est moins fréquent avec ce médicament qu'avec les autres prostaglandines E1).

La Figure 6 donne une vue globale des effets indésirables des huit médicaments et permet plusieurs constats tels que : Vitaros® a de nombreux effets indésirables non graves, Viagra® a de nombreux effets graves, la plupart étant communs avec Cialis®, ou Spedra® est le médicament qui présente le moins d'effets indésirables.

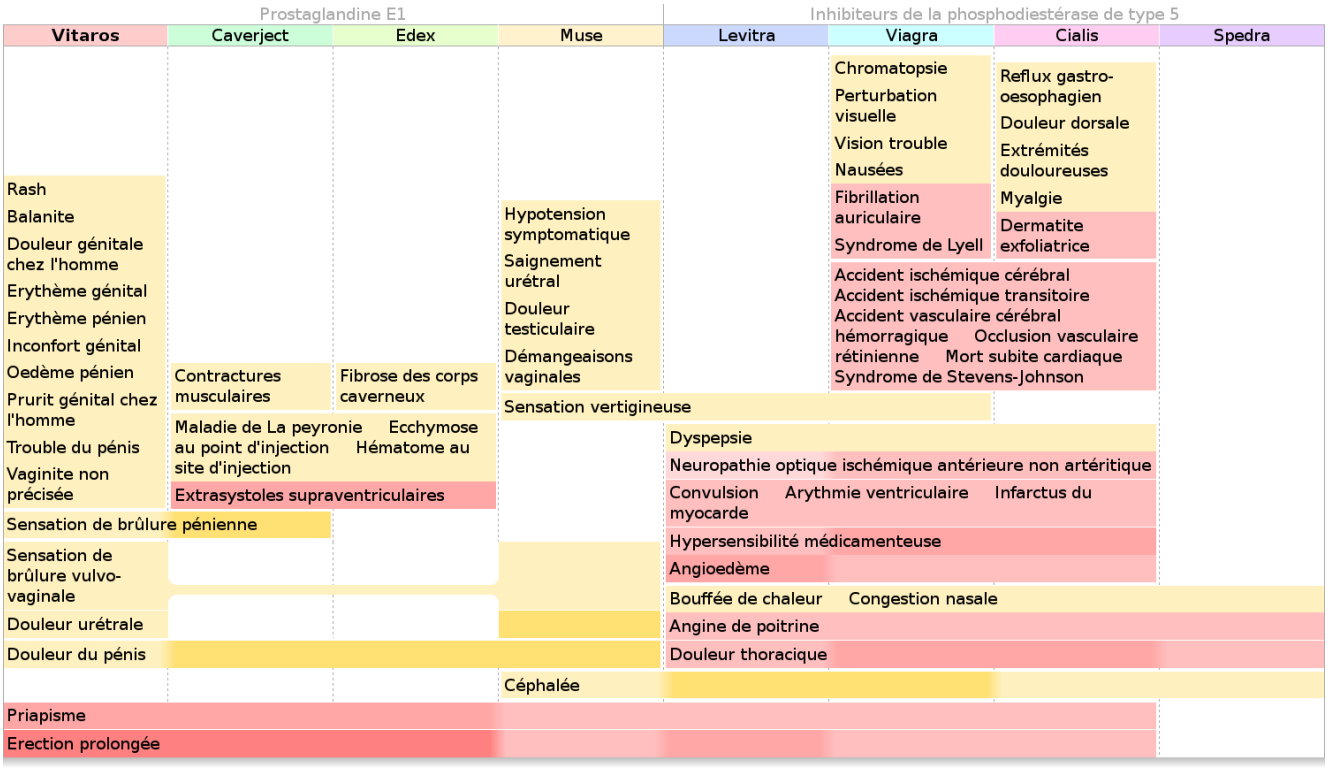

FIGURE 6 – Boîtes arc-en-ciel montrant les 116 effets indésirables (56 distincts) de 8 médicaments pour traiter la dysfonction érectile. Des couleurs personnalisées sont utilisées pour indiquer la gravité (teinte rouge ou orange) et la fréquence (saturation) des effets.

4.3 Implémentation

L'application présentée ici a été implémentée en Python 3, en deux parties : un module pour la génération des boîtes arc-en-ciel et un autre pour la gestion des propriétés des médicaments. Le premier module produit des pages HTML avec des feuilles de style CSS et du JavaScript. Le second permet soit la saisie manuelle des propriétés des médicaments soit leur extraction automatique à partir de la base de données sur le médicament Thériaque ou à partir des textes des RCP (pour les effets indésirables seulement). Les contre-indications ont été codées avec la CIM-10 (Classification Internationale des Maladies, 10ème version) et les effets indésirables avec MedDRA (*Medical Dictionary for Regulatory Activities*). PyMedTermino [15] a été utilisé pour gérer les deux terminologies. Les propriétés des médicaments ont été enregistrées dans une ontologie formelle en OWL/XML avec le module Python OwlReady [11].

La version complète du site web du projet VIIIP, visant à comparer les nouveaux médicaments aux anciens et portant sur quatre nouveaux médicaments, est disponible à l'adresse suivante : http://www.lesfleursdunormal.fr/static/viiip_proto/html.
Le code source du module Python permettant la génération des boîtes arc-en-ciel en HTML est disponible à l'adresse suivante : http://bitbucket.org/jibalamy/rainbowbox, sous licence libre (GNU LGPL v3).

5 Expérimentation utilisateur

5.1 Protocole

Les boîtes arc-en-ciel ont été évaluées lors d'une étude sur l'application aux propriétés des médicaments, en les comparant à des tableaux dynamiques (avec les médicaments en colonne et les contre-indications ou les effets indésirables en ligne, et des symboles colorés indiquant dans chaque case les contre-indications absolues ou relatives, ou la fréquence et la gravité des effets indésirables). Les tableaux étaient interactifs et permettaient d'afficher les propriétés d'un, de deux ou de tous les médicaments. Les tableaux ont été choisis comme référence lors de l'étude car ils paraissaient la manière la plus « évidente » de présenter les propriétés des médicaments aux auteurs mais aussi dans la littérature [20, 8].
L'étude a été réalisée selon un protocole en *cross-over*. 3 jeux de données sur les médicaments (dont celui sur la dysfonction érectile présenté plus haut) ont été extraits manuellement, chaque jeu comprenant 2 comparaisons (contre-indications et effets indésirables) et chaque comparaison a produit 2 documents (un avec les tableaux et l'autre avec les boîtes arc-en-ciel). Cela a conduit à 12 documents. 12 questions à choix fermés et uniques ont été écrites et validées par deux experts (2 questions pour chaque comparaison). Chaque question impliquait au moins deux médicaments et la réponse était présente dans

le document, par exemple : « Vitaros® est-il préférable à Edex® chez les patients porteurs d'un implant pénien ? (Oui/Non) » ou « Quel médicament présente le plus d'effets indésirables cardiaques graves ? (Vitaros®/Caverject®/..., 8 réponses possibles) ». Deux questions « d'échauffement » sur un 4ème jeu de données ont été ajoutées au début de l'évaluation, mais leurs résultats n'ont pas été pris en compte pour l'analyse.

Des médecins ont été recrutés et brièvement formés aux deux modes de visualisation (15 minutes environ). Ils ont ensuite répondu sur ordinateurs aux questions, chaque question étant accompagnée d'un document. Les médecins étaient répartis en deux groupes ; pour une comparaison donnée, l'un des groupes avait la première question avec les tableaux et la seconde avec les boîtes arc-en-ciel, et l'autre l'inverse.

Durant le test, nous avons mesuré la justesse des réponses et le temps de réponse ; après, nous avons mesuré l'utilisabilité perçue à l'aide d'un questionnaire standard SUS (*System Usability Scale*) [4]. Les données ont été analysées avec le logiciel R version 3.2.3. Une discussion générale avec les médecins a ensuite été organisée.

5.2 Résultats

22 médecins ont été recrutés (12 hommes, 10 femmes, âge moyen : 54,6 ans). Ils ont été grandement intéressés à l'idée de comparer les propriétés des médicaments. 9 médecins ont préféré les tableaux dynamiques et 7 les boîtes arc-en-ciel (6 n'ont pas exprimé de préférence). Le score SUS était de 72,0 pour les tableaux contre 66,9 pour les boîtes arc-en-ciel, la différence n'est pas significative (p = 0,25, test de Wilcoxon).

99 réponses sur 132 (75%) étaient corrects avec les tableaux contre 103 sur 132 (78%) avec les boîtes arc-en-ciel, la différence n'est pas significative (p = 0,66, test du Chi-2 corrigé). Le temps de réponse moyen était de 39,0 secondes avec les tableaux contre 29,7 secondes avec les boîtes arc-en-ciel, la différence est significative au risque $\alpha = 5\%$ (p = 0,011, test T de Student sur *log*(temps de réponse) pour normaliser les distributions).

Lors de la discussion, les médecins nous ont dit qu'ils avaient plus de difficultés à utiliser les boîtes arc-en-ciel. Ils ont aussi trouvé que les deux modes de visualisation leur permettraient de répondre a des questions différentes dans leur pratique quotidienne : les tableaux seraient plus appropriés en consultation lorsque le médecin a un patient précis en tête, tandis que les boîtes arc-en-ciel seraient plus adaptées pour la formation, lorsque le médecin a besoin d'une vue d'ensemble des propriétés des médicaments.

6 Discussion

6.1 Boîtes arc-en-ciel

Les boîtes arc-en-ciel permettent de visualiser les relations d'instanciation entre instances et classes (*est-instance-de*) dans une ontologie, tout en restant lisibles lorsque ces relations sont nombreuses et complexes. En revanche, les relations de subsomption entre classes (*est-un*) ne sont pas représentées de manière explicite. Par exemple, il est facile de constater sur la Figure 3 que tous les acides aminés *aliphatiques* sont aussi *essentiels*, cependant les boîtes arc-en-ciel n'indiquent pas si ce fait est purement fortuit ou bien s'il existe une relation de subsomption entre la classe des acides aminés *aliphatiques* et celle des *essentiels*.

En l'état, les boîtes arc-en-ciel peuvent être utiles pour découvrir des similarités entre instances ou entre classes, et faire des hypothèses sur des relations de subsomption. Elles pourraient cependant être améliorées en prenant en compte les relations de subsomption entre classes. L'héritage simple entre classes pourrait par exemple être représenté en incluant les boîtes des classes plus spécifiques dans les boîtes des classes plus générales (par exemple la boîte des acides aminés *minuscules* pourrait être incluse dans celle des *petits*). De même, la visualisation des disjonctions entre classes pourrait être facilitée si l'on plaçait les classes disjointes sur la même ligne. Ces connaissances ontologiques pourraient ainsi intervenir dans l'heuristique.

Pour l'optimisation de l'ordre des colonnes, nous avons testé d'autres approches qui n'ont pas conduit à des résultats aussi bons que l'heuristique présentée ici. Nous avons essayé l'*analyse en composante principale* (ACP) après avoir transformé l'ontologie en un espace à n dimensions (où n est le nombre de classes) dans lequel chaque instance est un point dont les coordonnées sont 1 (si l'instance est membre de la classe associée à la dimension) ou 0 (sinon). L'ACP permet ensuite de calculer l'axe permettant le mieux de séparer les points ; les instances sont alors placées dans l'ordre où elles figurent sur cet axe. Nous avons aussi essayé des méthodes de construction d'arbre (*unweighted pair group method with arithmetic mean*, UPGMA, et *neighbor-joining*, NJ). Ces méthodes permettent de regrouper les instances similaires de manière hiérarchique, cependant elles ne définissent pas un ordre à proprement parler, en particulier les fils de chaque nœud de l'arbre ne sont pas ordonnés. Il faut alors tester tous les ordres possibles satisfaisant l'arbre, et cela conduit à tester un nombre d'ordres bien plus grand que l'heuristique que nous avons proposée (pour les acides aminés, 524 288 pour UPGMA et 786 432 pour NJ). Les algorithmes de classification double (*biclustering*) pourraient aussi être envisagés, avec néanmoins le même problème : l'objectif est ici de trouver un ordre optimal et non pas un ensemble de classes (*clusters*).

6.2 Application à la comparaison des propriétés des médicaments

L'expérimentation utilisateur a montré que les médecins avaient une légère préférence pour les tableaux, mais qu'ils trouvaient l'information significativement plus rapidement avec les boîtes arc-en-ciel. Ces résultats ne sont pas contradictoires : en effet le score SUS mesure l'*utilisabilité perçue* mais celle-ci n'est pas nécessairement corrélée à l'utilisabilité réelle. Temps de réponse, justesse des réponses et satisfaction de l'utilisateur sont trois paramètres à mesu-

rer lors des expérimentations utilisateurs et ceux-ci doivent être considérés *a priori* comme non corrélés [7].
Les difficultés rencontrées par les médecins avec les boîtes arc-en-ciel peuvent s'expliquer par le fait que cette visualisation est nouvelle pour eux, alors qu'ils sont habitués aux tableaux, et par l'âge élevé d'un certain nombre des évaluateurs (certains mentionnant eux-mêmes ce point dans les questionnaires). Les boîtes arc-en-ciel pourraient être utiles pour la formation initiale et continue, par exemple pour les enseignements de pharmacologie ou l'information sur les nouveaux médicaments.

7 Conclusion

Dans cet article, nous avons présenté les boîtes arc-en-ciel, une nouvelle technique pour la visualisation des relations hiérarchiques dans les ontologies. Cette technique est particulièrement intéressante pour mettre en évidence les similarités entre instances et entre classes, et pour découvrir des relations de subsomption. Nous avons montré comment elle pouvait être appliquée à l'exemple des acides aminés et à la comparaison des propriétés des médicaments. Ensuite, nous avons décrit une expérimentation utilisateur qui a montré que les boîtes arc-en-ciel permettaient aux médecins de répondre plus rapidement à des questions précises qu'une approche à base de tableaux ; cependant davantage de formation est sans doute nécessaire avant de pouvoir utiliser efficacement ce nouvel outil.
Les boîtes arc-en-ciel pourraient être appliquées dans bien d'autres domaines, tels que la bioinformatique (par exemple pour la visualisation de *Gene Ontology*), les résultats des moteurs de recherche, les classifications et terminologies ou l'analyse et l'appréciation des bases de connaissances [18]...

Remerciements

Ce travail a été financé par l'ANSM (Agence Nationale de Sécurité du Médicament et des produits de santé) au travers du projet de recherche VIIIP (AAP-2012-013).

Références

[1] I Alecu, C Bousquet, P Degoulet, and M C Jaulent. PharmARTS : terminology web services for drug safety data coding and retrieval. *Stud Health Technol Inform*, 129(Pt 1) :699–704, 2007.

[2] B Alsallakh, L Micallef, W Aigner, H Hauser, S Miksch, and P Rodgers. Visualizing Sets and Set-typed Data : State-of-the-Art and Future Challenges. In *Eurographics Conference on Visualization (EuroVis)*, 2014.

[3] E H Baehrecke, N Dang, K Babaria, and B Shneiderman. Visualization and analysis of microarray and gene ontology data with treemaps. *BMC bioinformatics*, 5 :84, 2004.

[4] A Bangor, P Kortum, and J Miller. Determining what individual SUS score mean : adding an adjective rating scale. *J Usability Stud*, 4 :114–123, 2009.

[5] A S Dadzie and A Burger. Providing visualisation support for the analysis of anatomy ontology data. *BMC Bioinformatics*, 6(74), 2005.

[6] M Dudás, O Zamazal, and V Svátek. Roadmapping and navigating in the ontology visualization landscape, 2014.

[7] E Frøkjær, M Hertzum, and K Hornbæk. Measuring Usability : Are Effectiveness, Efficiency, and Satisfaction Really Correlated ? In *Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems*, volume 345-352, The Hague, The Netherlands, 2000.

[8] M Iordatii, A Venot, and C Duclos. Design and evaluation of a software for the objective and easy-to-read presentation of new drug properties to physicians. *BMC medical informatics and decision making*, 15 :42, 2015.

[9] A Katifori and C Halatsis. Ontology visualization methods - A survey. *ACM Computing Surveys*, 39(4) :10, 2007.

[10] A Katifori, E Torou, C Vassilakis, G Lepouras, and C Halatsis. Selected results of a comparative study of four ontology visualization methods for information retrieval tasks. In *Research Challenges in Information Science*, volume 133-140, 2008.

[11] J B Lamy and H Berthelot. Ontopy : programmation orientée ontologie en Python. In *Actes du congrès d'Ingénierie des Connaissances (IC2015)*, Rennes, France, 2015.

[12] J B Lamy, C Duclos, A Bar-Hen, P Ouvrard, and A Venot. An iconic language for the graphical representation of medical concepts. *BMC Medical Informatics and Decision Making*, 8 :16, 2008.

[13] J B Lamy, C Duclos, and A Venot. De l'analyse d'un corpus de texte à la conception d'une interface graphique facilitant l'accès aux connaissances sur le médicament. In *Actes des 20es Journées Francophones d'Ingénierie des Connaissances*, volume 265-276, Hammamet, Tunisie, 2009. PUG.

[14] J B Lamy, L F Soualmia, A Venot, and C Duclos. Validation de la sémantique d'un langage iconique médical à l'aide d'une ontologie : méthodes et applications. In *Actes du congrès d'Ingénierie des Connaissances (IC2014)*, Clermont-Ferrant, France, 2014.

[15] J B Lamy, A Venot, and C Duclos. PyMedTermino : an open-source generic API for advanced terminology services. *Stud Health Technol Inform*, 210 :924–928, 2015.

[16] R Lintern and M A Storey. Jambalaya Express : Ontology visualization-on-demand. In *International Protégé Conference*, 2005.

[17] L M Schwartz, S Woloshin, and H G Welch. The Drug Facts Box : Improving the communication of prescription drug information. *Ann Intern Med.*, 150 :516–27, 2009.

[18] F Vexler, A Berger, J P Cotton, and A Belloni. Eléments d'appréciation et d'analyse d'une base de connaissance : l'expérience industrielle d'Ardans. In *Atelier aide à la décision à tous les étages (AIDE), conférence extraction et gestion des connaissances (EGC)*, volume 59-72, 2013.

[19] T D Wang and B Parsia. CropCircles : Topology sensitive visualization of OWL class hierarchies. In *International Semantic Web Conference (ISWC)*, 2006.

[20] C Wroe, W Solomon, A Rector, and J Rogers. DOPAMINE : a tool for visualizing clinical properties of generic drugs. In *Proceedings of the Fifth Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP)*, volume 61-65, 2000.