

Analyse Formelle de concepts

Dans ces travaux, nous utiliserons :

- RCA-COGUI que vous trouverez à l'URL <http://www.lirmm.fr/cogui/> et dont une documentation est fournie sur Moodle.
- ou le logiciel **RCAexplore** plus ancien et avec moins de fonctionnalités que vous trouverez à l'URL <http://dataqual.engees.unistra.fr/logiciels/rcaExplore>, pour construire des treillis, des AOC-posets, ou des treillis Iceberg (documentation succincte en annexe à ce document)
- Open office, Libre office ou Excel : Il est en effet recommandé de saisir vos données brutes dans une feuille de calcul afin de les manipuler avec facilité puis de les transformer dans les formats d'entrée des logiciels cibles.

Vous rédigerez un compte-rendu d'expérience incluant toutes vos données et tous vos résultats (il peut être rendu par groupes de 1 à 3 étudiants).

Les étapes du travail consistent à :

- Comprendre les données du fichier déposé sur Moodle (Auteur Dame Samb sous licence Creative Common). Ces données ne doivent pas être diffusées. Elles décrivent des méthodes de fouille de données en flux. Si vous êtes seul, utilisez le premier tableau seulement. Si vous êtes 2, utilisez les deux premiers tableaux. Si vous êtes 3, utilisez les 3 tableaux.
- Identifier les attributs binaires pertinents pour décrire ces méthodes à partir des tableaux du document.
- Identifier des hiérarchies de spécialisation sur les attributs lorsqu'ils s'y prêtent.
- Produire les tableaux binaires décrivant les méthodes.
- Calculer des structures conceptuelles (treillis, AOC-poset, AC-poset, Iceberg).

Dans l'analyse des structures calculées, nous demandons de répondre aux questions suivantes :

1. Comparez les différentes structures (taille, concepts, etc.).
2. Extrayez des grandes catégories et sous-catégories de méthodes de fouilles de données qui apparaissent dans les structures conceptuelles.
3. Extrayez plusieurs règles d'implication.
4. Observez des caractéristiques qui ne sont jamais présentes ensemble.

1 Annexe : Installation et utilisation de RCAexplore ; pour RCA-COGUI voir documentation sur Moodle

Récupérez et renommez le binaire en `rcaexplore.jar`.

La documentation vous montrera comment lancer l'outil et l'utiliser avec une interface graphique et ci-dessous vous avez les principales commandes pour l'utiliser en ligne de commandes.

Comme il s'agit d'un prototype de recherche, comme pour les autres outils, faites régulièrement des sauvegardes dans différents fichiers pour sécuriser vos données.

1.1 L'éditeur de famille de contextes

On le lance en ligne de commande ainsi :

```
java -jar rcaexplore.jar editor &
```

Il permet de créer des familles de contextes ; il charge et sauve des fichiers dans un format texte lisible et éditables facilement, le format RCFT (extension des fichiers `.rcft`).

1.2 Le générateur de treillis de concepts, AOC-posets, et treillis Iceberg

Vous trouverez à la fin du document des exemples de fichiers d'entrée.

On lance le générateur en ligne de commande par :

```
java -jar rcaexplore.jar explogui <fichier.rcft> <dossier de sortie>
```

pour lancer RCA exploratoire sur le fichier `rcft` avec une interface graphique

En choisissant les options correspondantes au démarrage, l'exploration produit le fichier `result.xml` et divers autres fichiers qui peuvent vous intéresser. Elle produit toujours un fichier `.dot` contenant le treillis, un fichier `trace.csv` qui contient les options de configuration de chaque étape, et un fichier `latticebuilder.sh` qui sert à générer des pdf pour visualiser les treillis.

Observez la configuration courante, qui est celle de l'étape 0, par `display configuration`. Vous pourrez la changer à chaque étape lorsque vous serez en mode manuel, en choisissant au début les contextes objet-attribut, puis ensuite ces mêmes contextes ainsi que les contextes objet-objet (qui apparaissent plus tard).

On peut créer (`choose construction algorithm`) des treillis (FCA) des sous-hiérarchies de Galois (GSH, encore appelés AOC-posets, par l'algorithme ares) et des treillis Iceberg. `icebergXX` correspond au treillis de concepts dont on n'a conservé que les concepts dont la taille de l'extension est supérieure ou égale à XX% du nombre total d'objets. On garde aussi le concept bottom.

`continue` permet d'aller d'une étape à la suivante en changeant la configuration si vous restez en mode `manual`. Sinon, toute la construction (par `auto`) se fait avec la configuration de départ.

À partir de l'étape 1, en mode manuel, vous pouvez choisir les contextes mais aussi les opérateurs de scaling qui vous intéressent.

En cours de route, vous pouvez examiner les treillis construits (dans le répertoire `<dossier de sortie>`).

Les fichiers `stepi.dot` et `stepi-j.dot` vous présentent les structures construites à chaque étape de l'exploration par RCA. Ils peuvent être transformés en fichiers SVG, PDF, PNG, PS ou autres formats grâce à la commande `dot` en ligne de commande (logiciel GraphViz). Le fichier `latticebuilder.sh` contient les commandes pour créer ces fichiers sans écrire vous-même toutes les commandes `dot` nécessaires.

Un exemple de commande est donné ci-dessous si vous voulez le faire vous-même (transformation en fichier pdf) :

```
dot -Tpdf stepi-j.dot -o stepi-j.pdf
```

A la fin, `result.xml` correspond à l'ensemble des structures conceptuelles générées pour toutes les étapes du processus, et une option de visualisation vous montre les structures directement. Il peut

être analysé automatique pour tirer différents résultats a posteriori.

Pour résumer, le mode automatique permet de lancer toutes les étapes de RCA. Le mode manuel permet d'ajouter et de retirer à chaque étape des contextes et des relations en surveillant que c'est réalisé de manière cohérente. Il permet aussi de modifier à chaque étape les choix d'opérateurs de scaling et de structure (treillis, treillis Iceberg ou AOC-poset).

1.3 Le navigateur de concepts

Il permet de naviguer de concept à concept. On le lance en ligne de commande uniquement :
`java -jar rcaexplore.jar browser <fichier.xml>`

Pour voir le contenu d'un concept il faut choisir l'étape, le contexte et taper le nom du concept puis faire update. Les noms des concepts d'un contexte sélectionné sont affichés avec (entre parenthèses) leur support (taille de l'extension). Il est possible de faire un filtrage dans cette liste pour ne voir que les concepts dont le support est supérieur à une valeur donnée. On peut voir à partir d'un concept ses parents, ses enfants, son extension, son intension simplifiée et son intension. Si on sélectionne un élément de l'intension simplifiée et un élément de l'intension on peut voir juste au-dessous une représentation de la règle d'implication correspondante.