

## Extraction de la variabilité dans des familles de produits interconnectées *Travaux pratiques évalués*

Dans ces travaux, nous appliquerons les principes vus lors du cours sur la réingénierie de lignes de produits multiples (plusieurs dimensions, plusieurs préoccupations) par Analyse Relationnelle de Concepts. L'objectif sera d'extraire de possibles relations de variabilité entre caractéristiques d'une même dimension ou entre caractéristiques de différentes dimensions. Le résultat peut prendre la forme d'une liste de relations, mais il serait encore plus pertinent d'établir des *feature models* (par exemple, 1 par dimension et 1 central) connectés par des références (ou d'autres stratégies de modélisation à imaginer).

Dans ces travaux, vous étudierez des matrices de comparaison de produits (PCMs) afin d'en extraire de possibles relations de variabilité entre caractéristiques de la même PCM ou de différentes PCMs. Choisissez un exemple constitué de 3 à 4 matrices, un extrait des produits et un extrait des caractéristiques car il y a un travail manuel de nettoyage à effectuer. Les données proposées sont sur Moodle : <https://moodle.umontpellier.fr/course/view.php?id=27109>. Elles proviennent de wikipedia et peuvent vous avancer dans votre travail. Par exemple, les répertoires `Comparison_of_accounting_software` proviennent de la description des logiciels de comptabilité (*accounting software*) qui se situent par ailleurs sur cette page : [https://en.wikipedia.org/wiki/Comparison\\_of\\_accounting\\_software](https://en.wikipedia.org/wiki/Comparison_of_accounting_software). Cette description est effectuée la plupart du temps en plusieurs tableaux. Ici, elle contient des colonnes décrivant les licences, les système d'exploitation, les bases de données et les langages de programmation, qui sont décrits par d'autres répertoires provenant d'autres matrices de comparaison de produits, ce qui va vous permettre de travailler avec un modèle de données du même type que celui de la diapo 14 (que vous complèterez).

Vous travaillerez par groupes de 2 à 3 personnes. Rédigez un compte-rendu d'expérience en établissant les points permettant de restituer votre démarche le plus clairement possible. Vous rendrez ce compte-rendu et tous les fichiers et logiciels (ou liens vers les logiciels) permettant de reproduire votre travail. Ce qui sera évalué ne sera pas seulement un résultat mais le soin apporté à l'analyse et aux explications données.

Format pour rendre le travail : sur Moodle, dans le dépôt ouvert dans ce but (une seule personne du groupe dépose le travail du groupe et les noms de tous sont bien inscrits en clair dans les documents).

Vous suivrez les étapes classiques d'un **processus de fouille de données** que vous décrirez dans votre compte-rendu pour votre situation :

- jeu de données utilisé, sources correctement référencées
- sélection des données, vous n'utiliserez pas l'intégralité des données, donc vous expliquerez comment vous avez effectué la sélection
- nettoyage des données
- présentation du modèle de données choisi sous forme d'un modèle UML et en expliquant comment les contextes formels et relationnels se positionnent par rapport aux classes et associations de ce modèle UML
- extraction d'information (dans votre cas, des relations de variabilité et des feature models avec des références)
- évaluation, restitution, description de la connaissance extraite et discussion

Pour calculer les structures conceptuelles interconnectées, vous utiliserez l'outil FCA4J. La partie de la documentation dont vous avez besoin se trouve à l'adresse <https://www.lirmm.fr/fca4j/ImportwithFamily.html> pour les commandes et à l'adresse <https://www.lirmm.fr/fca4j/Downloads.html> pour les exemples. Lorsque vous les calculerez, vous obtiendrez aussi un fichier JSON qui peut être navigué avec RCAviz (<https://rcaviz.lirmm.fr/>).