Introduction

Cette formation de trois jours propose un enseignement approfondi des fondamentaux de l'Exploration Avancée de Données sur l'environnement KNIME®. La formation est destinée aux personnes ayant déjà une première connaissance de KNIME® et une expérience préalable dans l'exploration de données.

Objectif

Le thème principal abordé au travers du cours est la mise en place de protocoles plus complexes par l'apprentissage de concepts de programmation évolués tels que l'encapsulation de protocoles en composants, l'utilisation de variables, la notion de streaming et l'utilisation de différentes variantes de boucles dans l'outil KNIME®. La formation aborde dans un deuxième temps l'apprentissage de nœuds d'exploration de données plus évolués pour la transformation complexe de données par des opérations telles que la création et la gestion de collections et le regroupement et dégroupement de données, entre autres.

A qui s'adresse cette formation ?

Cette formation s'adresse aux scientifiques qui souhaitent acquérir une connaissance plus approfondie de l'exploration de données sur l'environnement KNIME® avec de nombreux exemples pratiques pour illustrer la théorie.

Quels sont les prérequis ?

•Avoir suivi la formation Pikaïros « Introduction à l'Exploration de Données à travers KNIME® » (IEDK) ou avoir une première expérience prouvée (certification KNIME L1: https://www.knime.com/certification-program) dans l'utilisation de l'environnement KNIME® et l'exploration de données.

•Comprendre l'anglais informatique élémentaire (outil en langue anglaise).

Modalités, délais d'accès et adaptation aux personnes en situation d'handicap

Cette formation est entièrement basée sur le logiciel pédagogique KNIME avec la mise en place de workflows comme méthode d'apprentissage. Une évaluation continue par QCM est effectuée sur les exercices proposés lors de la formation. La formation est proposée uniquement en présentiel en intra-entreprise. Adaptation possible des temps et mode de formation aux personnes en situation d'handicap, en lien avec l'entreprise. Délai d'accès maximum d'un mois entre la commande et le début de la formation.

Programme completFirst day

Premier jour

Programmation avancée dans l'environnement d'exploration de données KNIME

Méta-Nœuds et Composants

- Définition et utilité des Méta-Nœuds et Composants.
- Configuration extérieure des Composants.
- Programmation à l'intérieur des Composants.
- Définition, configuration et exécution de nœuds en mode Streaming.

Utilisation de variables dans les Protocoles KNIME

- Définition et utilité des variables.
- Initialisation des variables par des Widgets. Initialisation programmatique des variables, modification, transfert de données entre variables et tableaux, et vice-versa.
- Paramétrage de nœuds par des variables.
- Découverte des nœuds les plus utiles pour l'utilisation des variables dans un protocole.

Utilisation du contrôle de flux dans la programmation d'un protocole

- Définition et utilité du contrôle de flux, CASE ou IF Switches.
- Découverte des nœuds les plus utiles pour le contrôle du flux dans un protocole.

Mise en application : Réorganisation d'un protocole complexe à l'aide de variables et méta-nœuds

- Installation, étude et compréhension d'un protocole complexe avec lecture, transformation et écriture de données.
- Transformation du protocole avec paramétrisation par des variables.

Deuxième jour

Notion de boucle et variantes élémentaires

- Définition et utilité des boucles.
- Contextes propices et à éviter lors de l'utilisation de boucles dans l'environnement KNIME.
- Boucles élémentaires dans KNIME (Count Loup, Generic Loop).

Boucles spécifiques à l'itération sur lignes et colonnes

- Boucles pour l'itération sur les lignes d'un tableau (Row to Variable, Chunk Loop, Interval Loop, Group Loop).
- Boucles pour l'itération sur les colonnes d'un tableau (Column List Loop).
- Boucles avec génération de variables par itération (Row To Variable).

Utilisation avancée des boucles dans l'environnement KNIME

- Boucles récursives.
- Boucles avec contrôle et correction d'erreurs (Try & Catch).

Mise en application : Programmation de boucles

- Mise en place d'un modèle prédictif simple.
- Apprentissage, test et optimisation d'un modèle prédictif à l'aide de boucles.
- Sélection de portions homogènes dans le protocole et encapsulation de ces portions en composants de façon à améliorer la modularité du protocole, sa lisibilité et faciliter la réutilisation de sous-protocoles.

Troisième jour

Découverte de nœuds avancées pour l'optimisation des protocoles KNIME :

- Extraction et réutilisation des méta-informations des protocoles KNIME
- Informations sur l'environnement KNIME.
- Informations génériques sur les tableaux.
- Contrôle du temps d'exécution des protocoles KNIME et mise en cache des tableaux.

Manipulation avancée de tableaux et de données

- Conversion de colonnes multiples en colonne nominale et vice-versa
- Agrégation (C. Aggregator), combinaison (C. Combiner) et fusion (C. Merger) de colonnes.
- Manipulation de collections : création (Collection Creator) et division de collections (Collection Splitter).

Manipulation avancée de données

- Création de vecteurs de bits (Create Bit Vector) et extracteur de vecteurs de bits (Expand Bit Vector).

Découverte de nouveaux nœuds KNIME d'utilité générale

- Nœuds de lecture : List Files, Read Images, Table Reader, Model Reader, Excel Reader.
- Nœuds d'écriture : Write Images, Table Writer, Model Writer, Excel Writer & Appender.
- Nœuds génériques : RowID et Constant value Column.

Mise en application : Lecture, transformation et sauvegarde de données issues de différentes sources de données à l'aide des nœuds étudiés lors de la formation.