

Interface de chargement, de contrôle et d'analyse statistique des données pour la constitution d'un graphe de flux

Résumé du sujet

DCbrain est le « Waze » des réseaux physiques : notre logiciel donne du sens aux milliards de mesures issues du nombre toujours grandissant de capteurs *IoT* répartis dans les réseaux industriels de nos clients, en utilisant les technologies du *Big Data*, du *Machine Learning* et des Graphes.

L'arrivée massive de capteurs industriels permet une approche par la donnée de nombreux problèmes de l'industrie, à condition de savoir tirer parti de cette masse de données semi-structurées (séries temporelles + logs divers).

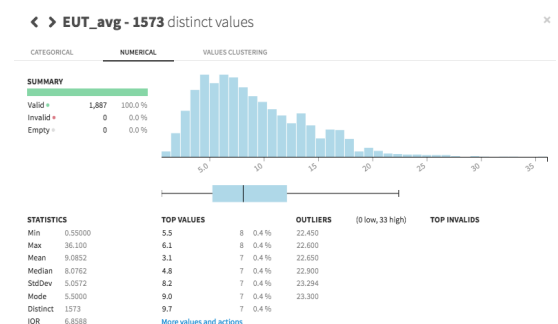
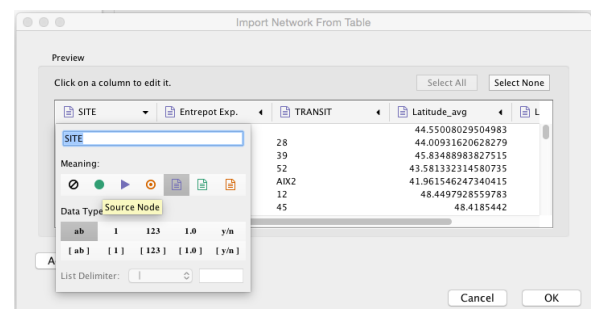
Nous utilisons aujourd'hui différentes techniques de *Machine Learning* classiques et de *Graph mining* afin d'aider nos clients créer de la valeur à partir des données collectées sur leurs réseaux. Afin de charger plus simplement les données dans notre moteur d'analyse, il est nécessaire de les contrôler et de les analyser.

L'objectif de ce sujet est le développement d'une interface de chargement, de contrôle et d'analyse statistique des données pour la constitution d'un graphe de flux.

Travail à réaliser :

1. Une bibliographie sur l'analyse descriptive des données
2. Un état de l'art des solutions et bibliothèques OpenSource utilisables
3. Une Interface de chargement d'un fichier csv
4. Un mécanisme de détection du type de colonne
5. L'assignation d'un rôle à chaque colonne

6. Une statistique descriptive de chaque colonne à la demande
7. Un chargement via des API



Calendrier

Remise du cahier des charges	Mercredi 15 mars 2017
Présentation orale	Mercredi 22 mars 2017
Remise des spécifications	Mardi 18 avril 2017
Remise du compte rendu	Vendredi 26 mai 2017
Soutenance et démonstration	Vendredi 2 juin 2017

Le cahier des charges ainsi que le compte rendu devront obligatoirement être tapés en LATEX.