

## TP1 : Lecture, édition et validement de fichier XML

### Exemple d'application: le flux RSS

L'objectif de ce TP est de vous familiariser avec la syntaxe du langage XML à travers un exemple d'application concret : le flux RSS. A la fin de ce TP vous serez capable de :

- Représenter le document selon une structure en arbre.
- Créer la DTD pour définir la grammaire du document XML.
- Créer un document XML en utilisant le jeu de données réel, et tester s'il est bien formé et valide.

#### Rappel

- Un document XML bien formé (well formed) est un document XML syntaxiquement correct: conforme à la norme XML qui spécifie Règles de nommage des balises, et qui respecte l'ordre d'ouverture et de fermeture de celles-ci. Exemples:
  - Pas d'espaces, pas d'apostrophe, pas de «/»
  - Premier caractère alphabétique ou «\_»
  - Noms sensibles aux majuscules-minuscules.
  - Les noms ne peuvent pas débuter par un nombre ou un caractère de ponctuation.
- Un document XML est valide est un document qui suit une définition de document. C'est-à-dire s'il est conforme à un schéma XML (DTD ou XSD).

Un flux RSS est représenté par un fichier XML dont le contenu est produit de façon automatique en fonction des mises à jour d'un site web. En utilisant un agrégateur de flux RSS, vous restez informé des dernières actualités de vos sites web favoris. L'objectif est alors de créer une application permettant de regrouper plusieurs flux RSS dans un seul flux fédérateur.

Flux à utiliser :

<http://hightech.bfmtv.com/rss/smartphone/>

<https://news.google.com/news/rss/headlines/section/topic/TECHNOLOGY?ned=us&gl=US&hl=en>

[http://www.lemonde.fr/technologies/rss\\_full.xml](http://www.lemonde.fr/technologies/rss_full.xml)

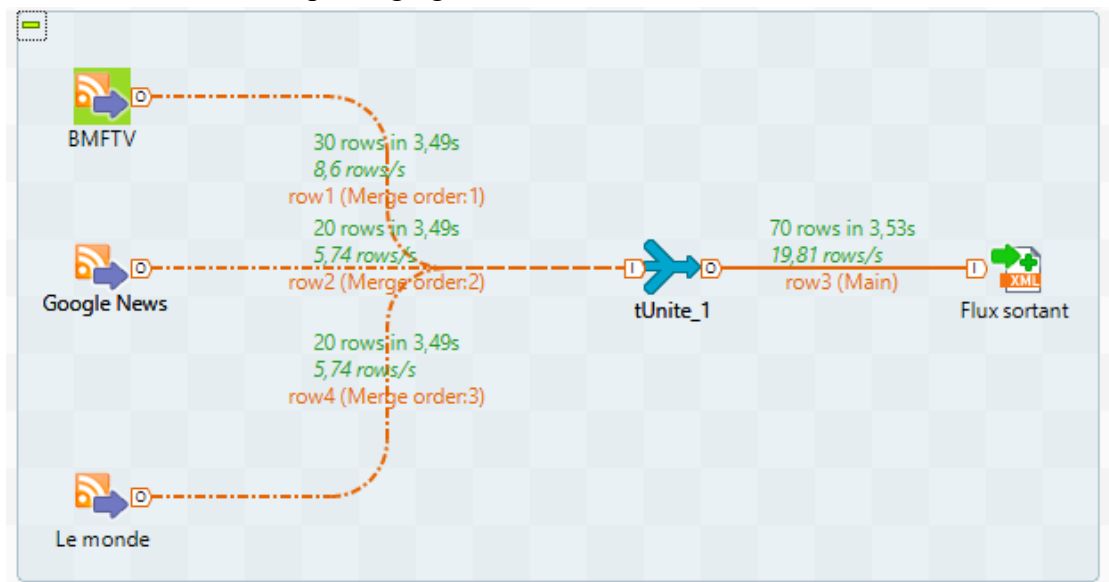
Travail demandé :

Partie 1 :

1. Représenter le flux RSS sous forme d'un arbre.
2. Créer une DTD pouvant valider ce document.
3. Vérifier à l'aide d'un parseur que votre document XML est bien formulé et valide.

Partie 2 :

1. Utiliser l'ETL **Talend Data Intégration** : c'est un outil qui vous permettra d'agréger plusieurs flux RSS en un seul.
2. Créer un nouveau Projet -> puis créer un nouveau Job.
3. Déposer l'élément tRssInput et spécifier l'url du flux.
4. Répéter l'étape 3 pour tous les flux rss entrant.
5. Déposer un élément de type tFileOuputXml.
6. Utiliser l'élément tUnit pour agréger les différents flux.



7. Exécuter le job talend.
8. Exporter le job sous forme d'une application que vous lancerez de façon régulière (Chaque jour par exemple).

**Pour aller loin:**

- Générer un flux rss de sortie (tRssOutput) que vous mettrez dans votre site web comme flux fédérateurs de plusieurs sources. Pour cela, il faut chercher un moyen pour le déposer de façon périodique dans un serveur.
- Exporter les données dans une base de données en vue de leur traitement par une application.