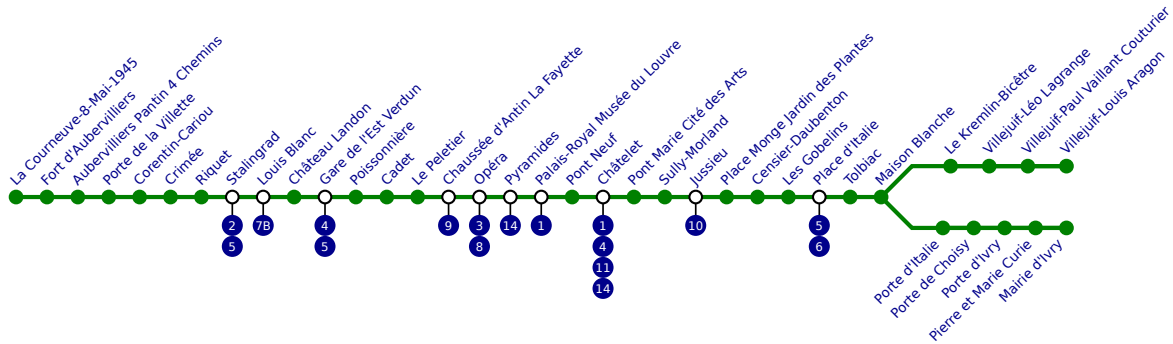


Projet de XML



Le but de ce projet est de convertir par programme, au format XML, un ensemble de données brutes décrivant la structure du métro parisien, puis d'extraire de ces données converties, à l'aide de XSLT, des cartes des différentes lignes du métro – sur le modèle des cartes horizontales affichées dans les rames.

I) Conversion des données CSV en XML

Le point de départ de ce projet est un fichier CSV fourni avec cet énoncé (**base_ratp.csv**). Vous pouvez l'ouvrir avec un éditeur de texte, mais il est beaucoup plus facile de visualiser son contenu en l'important dans un tableur tel que **soffice** (le séparateur est le point-virgule).

a) Données brutes

Le fichier CSV décrit tous les trajets du métro de longueur maximale effectuable sans changer de train. Chaque trajet part donc de l'un des terminus d'une ligne, et se poursuit de manière directe jusqu'à un autre terminus de cette ligne. Par exemple, à la ligne "La Courneuve - Mairie d'Ivry / Villejuif - Louis Aragon" représentée ci-dessus sont associés quatre trajets :

- Depuis "La Courneuve" jusqu'à "Mairie d'Ivry" ; le trajet inverse.
- Depuis "La Courneuve" jusqu'à "Villejuif - Louis Aragon" ; le trajet inverse.

Chaque ligne du fichier CSV autre que la première décrit un unique arrêt sur l'un de ces trajets. Les informations sur un arrêt, nommées dans la toute première ligne du fichier CSV, sont les suivantes :

1. **station_id** : l'identifiant de la station de l'arrêt.
2. **station_name** : le nom de cette station.
3. **station_desc** : l'adresse de cette station.
4. **station_lat**, **station_lon** : la latitude et la longitude de cette station.
5. **stop_sequence** : l'index de l'arrêt dans le trajet (à partir de 1).
6. **route_id** : l'identifiant du trajet.
7. **service_id** : l'identifiant de la ligne de métro du trajet.
8. **direction_id** : la direction du trajet (0 ou 1 pour aller ou retour).
9. **service_short_id** : le "numéro" de la ligne de métro du trajet (10, 7B,...).
10. **long_name_first**, **long_name_last** : les deux noms formant le nom complet de la ligne.

Noter que les "directions" des trajets sur une même ligne de métro sont tous spécifiées relativement à l'un de ses terminus, choisi de manière à ce que chacun des trajets partant de ce terminus soit considéré comme un "aller", chacun des trajets se finissant sur ce terminus soit considéré comme un "retour".

b) Choix d'un format XML et conversion des données brutes

À partir de ces données brutes, très redondantes et mal organisées, vous devrez :

1. Trouver un format XML permettant de représenter de manière optimale la structure du réseau du métro parisien. Votre format doit éviter au maximum toute duplication d'information, et ne doit contenir que des informations indispensables (c'est-à-dire ne pouvant pas être déduite d'autres informations présentes).
2. Écrire dans le langage de votre choix (C, Java, OCaml, Python, ...) un programme prenant en entrée le nom du fichier CSV, et produisant un fichier XML dans le format choisi. Certaines de ces informations peuvent être ignorées, mais les valeurs d'identifiants ne doivent pas être modifiées dans le XML produit.
3. Écrire un Schéma XML (fichier `xsd`) validant votre format. Le XML devra en particulier respecter les contraintes suivantes (elles-mêmes respectées par le fichier CSV) :
 - (a) aucun partage d'identifiant, de nom, d'adresse, ou d'emplacement (couple formé par la latitude et la longitude) entre deux stations distinctes,
 - (b) aucun partage d'identifiant, de numéro ou de nom complet entre deux lignes de métro distinctes.

II) Conversion des données XML en SVG

Dans la seconde partie du projet, votre travail consistera à écrire un fichier XSLT prenant en argument l'identifiant d'une ligne et le fichier XML extrait par votre programme, et produisant un fichier SVG affichant le plan de cette ligne sur le modèle de la figure de la page précédente : noms des stations, représentation graphique des connexions entre stations adjacentes et des bifurcations, indication des correspondances¹. Voici deux autres exemples de plans produits, tout comme le plan de la page précédente, par un unique fichier XML et un unique fichier XSLT, sans aucune modification de ces fichiers. Toutes les figures de cet énoncé sont à la même échelle :



1. Vous disposez d'une certaine marge de liberté pour le design et l'esthétique de ce plan, pourvu qu'il ne s'écarte pas trop de ce modèle. Un critère d'appréciation pourrait être celui-ci : à peut-être quelques informations mineures manquantes près, votre plan devrait pouvoir remplacer celui affiché dans une rame sans provoquer la colère immédiate des usagers.

Un fichier d'exemple SVG minimal, suffisant pour construire de telles cartes, est fourni avec cet énoncé (**sample.svg**) (<line>, <circle> et <text> suffisent pour tout le projet, <path> permet d'abrégier certains tracés, et rien ne vous empêche d'utiliser d'autres éléments).

III) Extensions

De nombreuses extensions sont envisageables : plan global avec les stations à leurs latitudes et longitudes, recherche et affichage de trajets, etc. La seule contrainte est d'effectuer ces traitements en XSLT.

IV) Modalités

a) Groupes

Le projet doit être impérativement réalisé en binôme. Toutefois, en cas d'impossibilité majeure de satisfaire cette contrainte (groupe en nombre impair, conditions d'études particulières, etc.), nous vous invitons à en discuter le plus tôt possible avec votre enseignant de cours magistral - aucun travail non réalisé en binôme ne sera accepté sans son accord explicite. La répartition des tâches au sein d'un groupe doit être raisonnablement équilibrée. En cas de déséquilibre avéré, les notes finales pourront être individualisées.

b) Individualité de chaque projet

De manière évidente, votre code doit être strictement personnel. Tout partage de code entre groupes est évidemment interdit (nous ferons tourner sur les rendus un détecteur de plagiat très efficace, insensible aux changements de noms ou aux modifications de l'ordre des différents éléments du code)². Il relève de votre responsabilité de faire en sorte que votre code reste inaccessible aux autres groupes : par exemple, si vous vous servez d'un dépôt `git`, ce dépôt doit être privé.

c) Forme du rendu

Les dates de rendu et de soutenance seront précisées ultérieurement. Votre rendu consistera en :

- Tous les fichiers autres que le fichier CSV de votre projet (le code de votre extracteur, le `xml` produit, les fichiers `xsd` et `xsl`).
- Un fichier texte nommé `README` contenant vos noms et numéros d'étudiants, et précisant la manière d'utiliser vos fichiers.
- Un rapport de quelques pages au format PDF décrivant votre projet, et expliquant et justifiant les choix de conception ou d'implémentation,

Tous ces éléments seront placés dans une unique archive compressée en `.tar.gz`. L'archive devra s'appeler `nom1-nom2.tar.gz`, et s'extraire dans un répertoire `nom1-nom2/`, où `nom1` et `nom2` sont les noms des deux personnes constituant le groupe. Par exemple, si vous vous appelez Denis Diderot et René Descartes, votre archive devra s'appeler `diderot-descartes.tar.gz` et s'extraire dans un répertoire `diderot-descartes/`.

2. La soutenance de projet est un examen comme un autre. Le plagiat en projet constitue une fraude aux examens passible de lourdes sanctions disciplinaires – ce cas s'est produit plus d'une fois dans cette UFR.