

Bases de données avancées – Projet

LIFBDW2 – Automne 2018

Décembre 2018

1 Consignes

Ce projet à réaliser seul ou en binôme est à déposer sur TOMUSS (1 ou plusieurs fichiers dans une archive) avant le **21 décembre 2018, 23:59**. *Chaque entête de fichier doit être renseigné par le(s) nom(s) et numéro d'étudiant(s)*. Veuillez également à bien commenter chacune de vos réponses (i.e., des réponses sans aucune explication pourront être considérées comme invalides lors de la correction).

2 Contexte

Nous disposons de données sur le parlement européens (députés, votes, etc.) lors de la huitième législature que nous souhaitons importer dans une base de données à des fins d'analyse. Pour cela, nous aurons besoin de définir des contraintes, des fonctions et des déclencheurs.

3 Description des données

Les données sous format CSV sont accessibles à l'adresse suivante :
https://perso.liris.cnrs.fr/marc.plantevit/ENS/LIFBDW2/TP/EP_DATA.zip.

Plusieurs fichiers sont disponibles :

- **BALLOTS.csv**: la liste des scrutins qui ont été votés entre le 21 Octobre 2014 et le 04 Octobre 2018. Chaque scrutin est décrit par son identifiant unique (BALLOTID), sa date (VOTE_DATE) et sa commission (COMMITTEE).
- **TAGS.csv**: contient les thématiques que traitent les scrutins votés dans le parlement. Les tags sont organisés dans une hiérarchie arborescente, et chaque scrutin peut être annoté par un ou plusieurs tags (la correspondance entre les tags et les scrutins est décrite dans le fichier **BALLOTS_TO_TAGS**). La hiérarchie est exprimée sous forme d'arbre, et la relation d'ascendance est décrite dans l'identifiant (TAG_ID) des tags où chaque point annonce les fils d'un noeud. (Par exemple: *le tag '4.15.05' a*

comme tag parent, le tag '4.15', qui lui même, à comme tag parent le tag '4'). Chaque tag a un label décrit par l'attribut: TAG_LABEL.

- **BALLOTS_TO_TAGS.csv**: décrit pour chaque scrutin l'ensemble des thématiques qu'il traite. Il est à noter, que dans ce fichier, seuls les tags les plus précis sont donnés. Ainsi, un scrutin qui possède comme tag '4.15.05' est implicitement annotés par les tags '4.15' et '4'.
- **MEPS.csv**: décrit l'ensemble des députés de la huitième législature du parlement européen. Chaque député possède un identifiant (MEPID) et est décrit par son nom (NAME_FULL), son pays (COUNTRY), son parti politique (NATIONAL_PARTY), son groupe politique (GROUPE_ID) et son genre (GENDER). Il est à noter, que pour les données fournies, on ne garde que le parti courant.
- **OUTCOMES.csv**: Cette table fait le lien entre les scrutins (BALLOTID) et les députés (MEPID) en précisant pour chaque député le vote qu'il a exprimé pour chaque scrutin le cas échéant, sinon le député est dit absent. Les députés dans le parlement européen peuvent voter pour (FOR), contre (AGAINST) ou s'abstenir (ABSTAIN). Notons que le fichier peut contenir des erreurs de saisie (*exemple: OFR au lieu de FOR*).

4 Création des Tables et Contraintes

Créer les tables correspondantes aux fichiers CSV sources en déclarant pour chaque table, les contraintes nécessaires (clés, contraintes d'intégrité référentielles, check). Puis importer les données des fichiers. Alternativement vous pouvez ajouter les contraintes une fois les données insérées.

5 Requêtes SQL

1. Écrire une requête qui permet de retourner l'ensemble des députés appartenant à un parti politique.
2. Écrire une requête qui permet d'ordonner les parlementaires selon leur assiduité (pourcentage de scrutins dans lesquels le parlementaire a été présent).
3. Quelle est l'assiduité moyenne des membres du parlement?
4. Écrire une requête qui retourne pour chaque groupe, le nombre de membres.
5. Écrire une requête qui retourne pour chaque groupe, l'assiduité moyenne.
6. Écrire une requête qui retourne pour chaque groupe politique, le député qui est le plus en accord avec le groupe (sur un scrutin, un député est

considéré en accord avec son groupe si son vote est égal au vote majoritaire du groupe), afficher aussi le % d'accord du député.

6 Procédures et fonctions

1. Écrire la fonction PL/pgSQL `PARENTOF(tag1,tag2)` qui permet de vérifier si un tag (`tag1`) est un parent directe dans l'hierarchie des tags d'un autre tag (`tag2`).
2. Écrire la fonction PL/pgSQL `ANCESTOR(tag1,tag2)` qui permet de vérifier si un tag (`tag1`) est un ancêtre d'un autre tag (`tag2`).
3. Écrire la fonction PL/pgSQL `TAGGED(ballot,tag)` qui permet de vérifier si un Scrutin (`ballot`) est annotée par un tag (`tag`) en utilisant la fonction (`ANCESTOR`).
4. Écrire la fonction PL/pgSQL `MAJORITY_VOTE_PARTY(ballot,national_party)` qui permet de calculer le vote majoritaire d'un parti `national_party` pour un scrutin `ballot`. La fonction retournera les valeurs entières suivantes:
 - 1: si le vote majoritaire est 'FOR'
 - 2: si le vote majoritaire est 'AGAINST'
 - 3: si le vote majoritaire est 'ABSTAIN'
 - 4: si 'For' et 'AGAINST' sont les deux votes majoritaires.
 - 5: si 'For' et 'ABSTAIN' sont les deux votes majoritaires.
 - 6: si 'AGAINST' et 'ABSTAIN' sont les deux votes majoritaires.
 - 7: s'il y a égalité entre les trois votes possibles.
 - 0: Aucun député du parti donné paramètre n'a été présent dans le scrutin demandé.
5. Écrire la fonction PL/pgSQL `MAJORITY_VOTE_GROUPE(ballot,EU_group)` qui calcule de la même manière que la fonction précédente, le vote majoritaire d'un groupe politique `EU_group` pour un scrutin `ballot`.
6. Écrire la fonction PL/pgSQL `MAJORITY_VOTE_COUNTRY(ballot,country)` qui calcule de la même manière que la fonction précédente, le vote majoritaire d'un pays `country` pour un scrutin `ballot`.
7. Écrire la fonction `SIMILARITY_NATIONAL_PARTY(ballot,party1,party2)` qui permet de calculer la similarité entre deux partis nationaux sur la base d'un scrutin donnée. La fonction retournera 1, si les deux groupes comparés ont voté exactement de la même manière, 0 sinon.
8. Écrire les deux fonctions `SIMILARITY_GROUP` (resp. `SIMILARITY_COUNTRY`) qui calcule la similarité de votes majoritaires entre deux groupes politiques (resp. deux pays).

9. Écrire une fonction PL/pgSQL `USUAL_SIMILARITY_PARTY(party1, party2)` qui permet de calculer le pourcentage de scrutins sur lesquels deux parties politiques sont d'accord. Écrire également les deux fonctions `USUAL_SIMILARITY_GROUP` et `USUAL_SIMILARITY_COUNTRY` qui calculent respectivement l'accord usuel entre deux groupes politique et l'accord usuel entre deux pays.
10. Écrire une fonction PL/pgSQL `CONTEXTUAL_SIMILARITY_PARTY(tag, party1, party2)` qui calcule la similarité entre les votes majoritaires des deux parties confrontés sur la base des scrutins affiliés à un tag donné. Écrire également les deux fonctions `CONTEXTUAL_SIMILARITY_GROUP` et `CONTEXTUAL_SIMILARITY_COUNTRY`.
11. En utilisant les fonctions précédentes, écrire une procédure PL/pgSQL `STATISTICS_MEPS` qui permet de créer la table `STATISTIQUE` qui, pour chaque député, donne les informations suivantes:
 - nombre de scurtins auxquels le député a participé.
 - pourcentage de scrutins sur lesquels le député en question en accord avec la majorité de son parti politique.
 - pourcentage de scrutins sur lesquels le député en question en accord avec la majorité de son groupe politique.
 - pourcentage de scrutins sur lesquels le député en question en accord avec la majorité de son pays.

7 Triggers

1. Écrire le Trigger `TRIGGER_ADD_TAG(tag)` qui est interdit l'insertion d'un nouveau tag dans la table `TAGS` dans le cas où son parent directe n'est pas présent dans la table.
2. Écrire le Trigger `TRIGGER_ADD_OUTCOME(tag)` qui met à jour la table `STATISTICS_MEPS` à chaque ajout d'un nouveau vote dans la table `OUTCOMES`.

8 Fact Checking

1. *“L'opposition, c'est nous !”* Mettre en oeuvre un moyen de mesurer ces propos pour un parti/groupe.
2. *“Nous sommes unis!”* Comment évaluer cette affirmation pour un groupe/parti/pays?
3. *“Un groupe/parti, une idéologie ?”* Y a t-il des individus qui ont des comportements plus similaires à de personnes hors de son groupe/parti?
4. *“les votes sont purement idéologiques, pas nationalistes!”* Exhiber des sujets de votes où les députés d'un pays sont exceptionnellement en accord.