

RAPPORT DE GROUPE EN SCIENCES DES DONNÉES 4

Victor Hugo, Albert Enstein.



Département MIASHS, UFR 6 Informatique, Mathématique et Statistique
Université Paul Valéry, Montpellier 3

Mars 2023

SOUmis COMME CONTRIBUTION
POUR LE COURS SCIENCE DES DONNÉES 4

Déclaration de non plagiat

Nous déclarons que ce rapport est le fruit de notre seul travail, à part lorsque cela est indiqué explicitement.

Nous acceptons que la personne évaluant ce rapport puisse, pour les besoins de cette évaluation:

- la reproduire et en fournir une copie à un autre membre de l'université; et/ou,
- en communiquer une copie à un service en ligne de détection de plagiat (qui pourra en retenir une copie pour les besoins d'évaluation future).

Nous certifions que nous avons lu et compris les règles ci-dessus.

En signant cette déclaration, nous acceptons ce qui précède.

Signature: _____ Date: _____

Signature: _____ Date: _____

Remerciements

Nos plus sincères remerciements vont à nos encadrants pédagogiques pour les conseils avisés sur notre travail.

Nous remercions aussi ...

Résumé

L'image ci-dessous vous donne quelques conseils pour rédiger un bon résumé.

general background → Horizontal gene transfer (HGT), the acquisition of genetic material from non-parental lineages, is known to be important in bacterial evolution^{1,2}. In particular, HGT provides rapid access to genetic innovations, allowing traits such as virulence³, antibiotic resistance⁴ and xenobiotic metabolism⁵ to spread through the human microbiome. Recent anecdotal studies providing snapshots of active gene flow on the human body have highlighted the need to determine the frequency of such recent transfers and the forces that govern these events^{4,5}. Here we report the discovery and characterization of a vast, human-associated network of gene exchange, large enough to directly compare the principal forces shaping HGT. We show that this network of 10,770 unique, recently transferred (more than 99% nucleotide identity) genes found in 2,235 full bacterial genomes, is shaped principally by ecology rather than geography or phylogeny, with most gene exchange occurring between isolates from ecologically similar, but geographically separated, environments. For example, we observe 25-fold more HGT between human-associated bacteria than among ecologically diverse non-human isolates ($P = 3.0 \times 10^{-270}$). We show that within the human microbiome this ecological architecture continues across multiple spatial scales, functional classes and ecological niches with transfer further enriched among bacteria that inhabit the same body site, have the same oxygen tolerance or have the same ability to cause disease. This structure offers a window into the molecular traits that define ecological niches, insight that we use to uncover sources of antibiotic resistance and identify genes associated with the pathology of meningitis and other diseases.

specific background →

knowledge gap →

here we show... →

results with key, concrete values →

meaning of results →

All rights reserved by Nature.
Reproduced here for
educational purposes only.

The human body is a complex biological network comprising tan

Table des matières

Chapitre 1	Introduction	1
Chapitre 2	Gestion de projet	2
2.1	Introduction	2
2.2	Collaboration avec Git	2
2.3	Contribution	2
2.4	Planning	2
2.5	Conclusions	2
Chapitre 3	Base de données	3
3.1	Introduction	3
3.2	Traitements réalisés	3
3.3	Descriptif des tables	3
3.4	Modèle conceptuel des données	3
3.5	Modèle organisationnel des données	4
3.6	Requêtes réalisées	4
3.7	Conclusions	4
Chapitre 4	Description technique et fonctionnelle	5
4.1	Introduction	5
4.2	Description technique	5
4.3	Description fonctionnelle	5
4.4	Conclusions	5
Chapitre 5	Conclusions et perspectives	6
Annexes		7

CHAPITRE 1

Introduction

Vous devez introduire votre projet qui a sans doute évolué par rapport à la version de décembre. Quel est le nouveau périmètre du projet si celui-ci a évolué ?

A qui s'adresse t'il ?

En quoi votre projet est original et se démarque des sites déjà existants ? Comment pourriez vous en faire la publicité ?

Conclure l'introduction par l'url vers la vidéo de démonstration de votre projet ainsi que la vidéo de soutenance.

Donner le plan du mémoire.

CHAPITRE 2

Gestion de projet

2.1 Introduction

Vous introduirez ce chapitre et préciserez la méthodologie suivie. Vous donnerez le lien vers le site GitHub.

Donner le plan de la section.

2.2 Collaboration avec Git

Vous décrirez comment vous vous êtes servis de Git pour le développement collaboratif. Avez-vous utilisé des branches ? Comment avez-vous maîtrisé les conflits de merge ? Comment avez-vous alimenté votre fichier .gitignore et pourquoi ? Avez-vous profité de l'historique de commits (`git log`) pour récupérer du code à partir d'une ancienne version de votre projet ? etc.

2.3 Contribution

Vous complèterez 2 tableaux :

- un tableau avec la contribution de chacun par tache est obligatoire : par exemple tache 1 : base de données réalisée par l'étudiant X à 30% et Y à 70%.
- un tableau avec la contribution de chacun par page est obligatoire : par exemple Page d'accueil : réalisée par l'étudiant X à 30% et Y à 70%.

Vous expliciterez les différence observable entre la contribution affichée et ce qui est visible sur Github. Commenter le nombre de commit par utilisateur, les dates... .

2.4 Planning

Le planning de janvier n'est sûrement pas celui qui a été suivi ce semestre. Indiquer les deux (l'ancien et le planning mis à jour).

Il faut donc expliquer pourquoi certaines taches ont été plus rapides ou lentes que prévu. Comment vous vous êtes adaptés ?

2.5 Conclusions

CHAPITRE 3

Base de données

3.1 Introduction

Vous introduirez ici le chapitre et donnerez le ou les url des sites sur lesquels vous avez récupéré des données.

Dans les jeux de données récupérés, expliquer pourquoi vous avez sélectionné certaines tables. Vous devez indiquer comment les données sont stockées (le format), la taille des fichiers, le nombre, ...

Donner le plan de la section.

3.2 Traitements réalisés

Vous décrirez ici les étapes réalisées pour pré-traiter vos données. Expliquer quels choix vous avez réalisés pour filtrer les lignes et les colonnes (éventuellement réduire le périmètre du projet) et décrire les critères de sélection (e.g., ne garder que 5 colonnes sur les 15, ne garder que les lignes qui correspondent à une ville en particulier, ...).

Préciser les nettoyages réalisés avant l'import comme l'uniformisation des valeurs des champs (e.g., Mr, M., Monsieur, ...) ou le remplissage des valeurs manquantes par une valeur moyenne ...

3.3 Descriptif des tables

Pour chaque table conservée, préciser le nombre de lignes et de colonnes après filtrage, lister les colonnes et donner pour chacune le type, la signification du champ et des caractéristiques (unique, clés, valeur manquante, ...) en remplissant le tableau ci-dessous.

Table 3.1: Nom de la table (nombre de lignes × nombre de colonnes)

Nom colonne	Type	Signification	Caractéristique

3.4 Modèle conceptuel des données

Pour le MCD, inclure une image réalisée avec le logiciel Mocodo [<https://www.mocodo.net/>] telle que celle visible sur la Figure 3.1 ci-dessous :

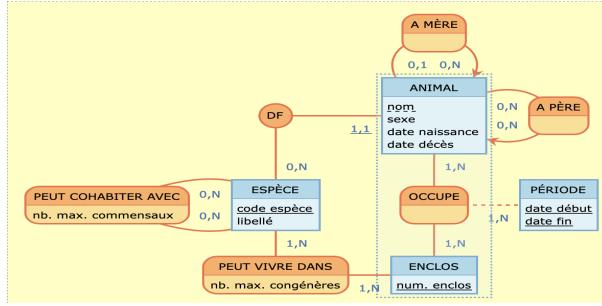


Figure 3.1: Relations.

3.5 Modèle organisationnel des données

Pour le MOD, inclure une image réalisée avec le designer de phpmyadmin

3.6 Requêtes réalisées

Pour les requêtes importantes dans votre interface, les exprimer en langage naturel puis en SQL. Puis donner le résultat obtenu (ou un extrait) et expliquer ce résultat.

3.7 Conclusions

Préciser la durée de la mise en place de la BD et les principales difficultés rencontrées.

CHAPITRE 4

Description technique et fonctionnelle

4.1 Introduction

Introduire le chapitre et donner le plan de la section.

4.2 Description technique

Vous listerez les technologies utilisées (Git, Php, Javascript, Mamp...)

 Vous inclurez un schéma de l'architecture de votre application

4.3 Description fonctionnelle

Par écran,

- quelles sont les fonctionnalités, que peut faire l'utilisateur ?
- quelle gestion des erreurs avez vous prévu : par exemple, sur un écran de connexion, on peut vérifier que un email contient un '@', qu'un mot de passe n'est pas trop facile..

Ne pas hésiter à lister les fonctionnalités que vous n'avez pas eu le temps d'implémenter pour montrer que vous y avez quand même pensé. Idem pour la gestion des erreurs.

4.4 Conclusions

Conclure et lister les difficultés rencontrées.

CHAPITRE 5

Conclusions et perspectives

Résumer vos contributions puis donner des perspectives à vos travaux. C est pour nous la partie importante : s'il l'on vous prenait en stage pour 2 mois sur votre sujet, auriez vous des idées et seriez vous autonomes pour améliorer votre application. Combien de temps auriez vous besoin pour ajouter ces fonctionnalités ?

On attend de vous deux types de perspectives : des perspectives à court terme pour améliorer rapidement votre approche et des perspectives à plus long terme qu'elles soient liées à la science des données ou au domaine métier pour lequel vous avez travaillé.

Annexes
