

Directives importantes pour la conduite des projets

Les projets sont distribués et commentés durant le cours du 29 janvier et la constitution des groupes avec les choix de projet commence à cette date. Les choix définitifs des groupes devront être rendus lors du TD du 30 janvier. N'oubliez pas d'indiquer aussi *qui est le modérateur du groupe*.

Les projets seront à rendre par mail, *a priori* pour le 23 avril, à Gilles.Bernot@univ-cotedazur.fr sous forme d'un *fichier PDF*. Le sujet du mail devra être « projet BD GB4 ».

Le cours du 29 janvier fournira les éléments de gestion de tels projets, ce sont des techniques standard de développement de produits technologiques, adaptées aux produits logiciels, mais qui s'appliquent en fait à toute activité d'ingénierie. Les groupes devront se réunir en dehors des horaires d'enseignement pour avancer le projet.

Les ambiguïtés sur les besoins des utilisateurs et sur les informations nécessaires dans la base de données devront être entièrement inventoriées pour le TD du 19 février. Le 19 février, l'enseignant répondra aux questions *préalablement réfléchies dans chacun des groupes* afin de trancher les choix possibles : il jouera donc le rôle du « *client* » qui précise ses besoins (mais seulement en réponse aux questions qu'on lui pose) et avec lequel les concepteurs de la base de données *négoient* d'éventuels assouplissements (y compris du calendrier). Plus aucune négociation, ni de calendrier, ni sur les besoins des utilisateurs, ne sera possible ultérieurement.

Le 13 mars, chaque groupe aura conçu un *schéma de base de données* en prenant soin d'*éviter toutes redondances* qui conduiraient à devoir gérer la cohérence des données lors des mises à jour. Le groupe aura également réfléchi aux *tests fonctionnels* qui seront utiles pour valider le projet. L'enseignant jouera alors le rôle de « *conseiller technique* » sur le schéma de la base.

Le cours et les TD vous donneront les bases pour écrire ensuite vos requêtes en SQL afin de réaliser les fonctionnalités demandées. N'hésitez pas à consulter internet pour vous inspirer mais il est très important que vous compreniez suffisamment ce que vous programmez pour expliquer en détail le fonctionnement de chaque requête dans le document rendu. Les sites <https://sql.sh> ou <https://www.w3schools.com/sql> sont des références très utiles.

Il vous est demandé de rendre à la fin du projet les éléments suivants :

1. une reformulation *complète* de la description du projet qui supprime toutes les ambiguïtés du présent document et définit exhaustivement les services rendus par la base après négociations avec le client (*≈ spécification des besoins complète*),
2. le schéma de la base de données, exprimé d'une manière graphique de votre choix et commenté. Le commentaire expliquera et justifiera les choix effectués (*≈ spécification détaillée*),
3. l'expression en SQL des requêtes sur la base de données, qui sont utiles pour rendre les services sus-mentionnés, avec des explications bien structurées de chaque requête (*≈ description du codage*),

Important : chaque membre du groupe devra prendre en charge l'explication détaillée du fonctionnement d'une des requêtes SQL et le document rendu devra indiquer *qui* a rédigé cette explication. S'il y a moins de requêtes que de membres du groupe, alors les requêtes les plus longues pourront être prises en charge par un binôme. S'il y a plus de requêtes que de membres du groupe alors les plus courtes seront prises en charge par la totalité du groupe, ou bien regroupées et prises en charge ensemble par un des membre du groupe.

4. la description du jeu de données utilisé pour tester les différentes requêtes sous des conditions typiques. La taille de ce jeu de données peut être faible mais on justifiera soigneusement le choix de ces données pour la **complétude** des tests (= *validation*),
5. une analyse critique du déroulement de votre projet : découpage et répartition du travail en équipe, difficultés rencontrées et réflexion sur des améliorations de l'organisation des tâches au sein du groupe de travail... mais aussi les aspects bien réussis et les éléments clefs de cette réussite. Bref, comment vous y prendriez-vous « *si c'était à refaire* » ? Il ne s'agit *pas* de proposer des améliorations du produit logiciel en soi.

L'usage de l'IA n'est pas interdit mais le périmètre décrit ci-dessous doit être scrupuleusement respecté :

1. La reformulation de la description du projet doit être faite par vous mais l'IA peut être utilisée pour améliorer le style de votre texte.
2. Le schéma de la base de données doit être entièrement fait par vous à partir de votre analyse des besoins.
3. Une première version de l'expression en SQL des requêtes peut être faite par IA *mais* vous devez alors comparer plusieurs versions, *les comprendre* et choisir votre solution finale en ayant parfaitement compris ce que font chacune des instructions des requêtes.
4. Le choix des jeux de données et la justification de complétude doivent être faits par vous mais l'IA peut être utilisée pour améliorer le style de votre texte.
5. L'analyse critique du déroulement de votre projet doit être faite par vous mais l'IA peut être utilisée pour améliorer le style de votre texte.

Enfin *vous devrez décrire brièvement à la fin de votre document toutes les aides de l'IA que vous avez utilisées.*

Vous effectuerez une implémentation de ce schéma de base de données dans un environnement de bases de données relationnelles de votre choix mais il est *inutile* d'en rendre une sauvegarde : indiquez seulement quels sont les tests qui ont éventuellement échoué... par conséquent la justification sus-mentionnée de *complétude des tests* est cruciale.

La note finale sera principalement fondée sur la qualité globale du dossier rendu, donc par défaut commune à tous les membres du groupe, l'objectif étant de préserver la solidarité du groupe. Une variation individuelle mineure de la note sera fondée sur la pédagogie des explications des requêtes SQL faites par chaque membre du groupe.

Tous les projets ont été calibrés pour être *a priori* de difficulté identique. Par conséquent le choix du projet n'a pas d'influence sur la complexité de réalisation. En revanche, une négociation bien menée avec le client peut avoir une *influence notable sur la simplicité/facilité du projet*, de même qu'un bon usage du conseiller technique en anticipant les questions, courtes et précises, qu'on souhaite lui poser sur le schéma de la base.

Description des projets

Projet n°1

Etude de réseaux neuronaux

Une équipe de recherche sur les réseaux neuronaux vient de concevoir un équipement permettant de mesurer simultanément l'activité de plusieurs neurones dans le cerveau. On peut ainsi calculer la fréquence de production de « spikes » de chaque neurone mesuré (évaluée sur une durée de l'ordre de la seconde) et l'on peut effectuer ces mesures simultanées de fréquences un petit nombre de fois par minute. En revanche, l'équipement ne permet pas de mesurer si deux spikes individuels sont émis exactement en même temps par deux neurones différents.

Une expérience consiste à assigner une activité bien définie à un sujet durant quelques minutes et à observer les fréquences successives de spikes de chaque neurone mesuré tout au long de l'expérience.

L'équipe vous demande de mettre en place une base de données pour collecter les résultats des diverses expériences. On devra entre autres gérer :

- un nom pour chaque électrode (par hypothèse il y a une électrode par neurone mesuré) avec la région du cerveau où elle se trouve,
- pour chaque expérience stockée, la succession des mesures de fréquences de chaque capteur, l'activité imposée, la date et l'heure de l'expérience,
- les activités possibles sont en nombre très limité car le protocole est parfaitement défini et contrôlé pour chacune d'elles. En particulier, s'il y a des changements d'environnement du sujet ou une succession de gestes à effectuer durant une expérience donnée, ils sont parfaitement fixés à un temps précis de l'expérience et la durée de l'activité/expérience est fixée avec précision.

La base de données devra être utilisable pour traiter en particulier les questions suivantes :

1. Mémoriser une nouvelle expérience
2. Lors d'une expérience donnée, quels sont les neurones ayant une activité similaire (i.e. des fréquences proches tout au long de l'expérience) ?
3. Les neurones ayant des activités similaires à un neurone donné sont-ils situés dans la même région du cerveau ?
4. Etant données deux expériences avec la même activité, quels sont les neurones ayant des activités similaires d'une expérience à l'autre ?
5. Quels neurones augmentent leur fréquence de spikes lors d'un changement d'environnement ou le début d'un geste ?
6. Lesquels anticipent l'activité d'un délai donné ?

D'autres questions pourront éventuellement compléter ce panel.

Projet n°2

Gestion d'articles de référence pour la formation

Le service de formation permanente d'un grand groupe pharmaceutique souhaite mettre en place une base de données qui aide leurs formateurs à trouver des articles scientifiques pertinents (et raisonnablement récents) pour préparer leurs cours. Chaque article est caractérisé par son titre, ses auteurs, l'année de parution, l'éditeur de l'article et enfin les références chez cet éditeur. Des informations complémentaires doivent être gérées afin de conseiller au mieux les formateurs :

- une liste de domaines (scientifiques ou non) dans lesquels l'article peut s'inscrire,
- une liste de mots-clefs qui précisent les sujets abordés,
- le ou les niveaux de généralité de l'article : est-ce un article encyclopédique ? un chapitre de livre pédagogique ? une revue scientifique qui fait le point sur un domaine ? ou bien un article spécialisé ?
- l'avis des formateurs ayant déjà exploité cette référence :
 - quel est le niveau pédagogique de l'article ?
 - peut-il être directement distribué ou conseillé aux personnels en formation ?
 - suffit-il à lui seul pour traiter complètement un domaine ?

La base de donnée a pour but de soulager les responsables du service de formation en donnant aux formateurs les informations les plus fréquemment utiles, et l'on doit traiter les questions suivantes :

1. Ajouter un nouvel article.
2. Etant donné un domaine à enseigner, quels sont les articles qui traitent complètement (respectivement partiellement) le domaine et sont les plus pédagogiques possibles ?
3. Etant donnée quelques mots-clefs, quels sont les articles qui les abordent tous ?
4. Etant donné un auteur, quels sont tous les articles auxquels il a participé ?
5. Etant donné un mot, quels sont tous les articles dont le titre contient ce mot ?
6. Etant donné un article, quels sont les articles qui partagent le plus de mots-clefs avec lui ?
7. Un formateur doit pouvoir ajouter (ou redonner) son avis sur les trois questions (niveau pédagogique, directement distribué ou non, complétude par rapport à un domaine).
8. Quels sont les articles unanimement considérés comme non pédagogiques ?

D'autres questions pourront éventuellement compléter ce panel.

Projet n°3

Base de médicaments

Un groupement de pharmacies souhaite informatiser le traitement des ordonnances et la gestion des stocks dans ses officines. Il vous demande de concevoir et réaliser une base de données apte à répondre à ce besoin. Votre architecture de base de donnée sera ensuite dupliquée dans chacune des officines.

Les informations que l'on doit pouvoir retrouver sur un médicament incluent :

- ses indications,
- sa ou ses molécule(s) active(s),
- la posologie associée,
- les contre-indications,
- ses interactions avec d'autres molécules ou médicaments à éviter,
- sa disponibilité dans l'officine concernée et le nombre de ventes durant les 12 derniers mois,
- son prix.

Les informations spécifiques à chaque médicament sont supposées connues à l'avance dans la base de donnée et leur mise à jour ne relève donc pas des officines, sauf les informations de disponibilité et de prix. Dans une officine, la base de données devra être utilisable pour traiter en particulier les questions suivantes :

1. modifier le prix d'un médicament
2. liste des 3 médicaments les plus vendus dans l'officine,
3. conformité d'une ordonnance (en regard d'une pathologie donnée ou des interactions médicamenteuses),
4. remplacement d'un médicament en rupture de stock par un autre en stock ayant la même molécule active,
5. prix total d'une ordonnance,
6. mise à jour du nombre de boîtes en stock d'un médicament donné, à l'issue d'une vente ou au contraire à l'issue d'un réapprovisionnement,
7. liste des médicaments à renouveler avant rupture de stock.

D'autres questions pourront éventuellement compléter ce panel.

Projet n°4

Analyses par calculs enzymatiques

Une start-up est spécialisée dans la détection rapide de métabolites dans un échantillon de sang, à des fins de diagnostic. Pour ce faire, elle crée un réseau enzymatique artificiel dans une solution de sa composition. Après dépôt d'une goutte de sang dans cette solution, la présence ou l'absence des métabolites recherchés va activer le réseau artificiel qui produira de nouveaux métabolites générant une coloration. Cette start-up a dégagé un grand ensemble d'enzymes avec lesquelles elle peut mettre en place ces « calculs enzymatiques ».

Par exemple, pour détecter si un fluide contient *à la fois* les métabolites a et b , il suffit de sélectionner une enzyme E prenant a et b pour substrat et produisant un métabolite c produisant une coloration (ou précurseur d'une telle coloration). Cette enzyme est une *porte AND*.

Si l'on veut détecter si *l'un au moins* des métabolites a ou b est présent, alors on sélectionne deux enzymes : E_1 qui prend a comme substrat et produit c (avec ses conséquences sur la coloration), et E_2 qui prend b comme substrat et produit le même c . En plaçant E_1 et E_2 dans la solution on crée une *porte OR*.

Pour détecter l'absence de a , on place dans la solution une enzyme E ayant pour substrat un métabolite e et pour produit c , mais telle que a est un inhibiteur enzymatique de E . On place également un stock de e dans la solution : si a est absent, c est produit : c'est une *porte NOT*. Enfin des enzymes jouent le rôle de *portes traductrices*, transformant un substrat a en un produit a' plus facile à intégrer dans une porte *AND*, *OR* ou *NOT* existante.

Le nombre d'enzymes pleinement étudiées par la start-up est important et ils disposent ainsi d'une grande collection de portes *AND*, *OR*, *NOT* ou *traduction*, d'où la nécessité de votre base de données. Cette base devra en particulier contenir :

- tous les liens enzyme-substrat connus entre les éléments de la base
- tous les liens enzyme-produit connus entre les éléments de la base
- toutes les réalisations de portes jugées fiables par la start-up avec les enzymes et métabolites de la base (ces portes sont nécessairement en accord avec les liens enzyme-métabolite précédents)
- et toutes les solutions en cours de conception.

La start-up veut utiliser la base de donnée afin d'assister la construction des réseaux enzymatiques à partir de formules logiques. Par exemple si une maladie est diagnostiquée par la présence de x et de y et l'absence de z , les experts de la start-up voudront construire la formule $(x \text{ AND } y) \text{ AND } \text{NOT}(z)$ et donc trouver deux portes *AND* et une porte de *NOT*. . . mais il peut être nécessaire de traduire x , y et z au préalable pour pouvoir utiliser les portes de la base, et d'avoir des traductions intermédiaires entre plusieurs portes. Ainsi la base devra répondre en particulier aux questions suivantes :

1. quelles sont toutes les portes *AND* telles qu'un métabolite donné en est un substrat ? et même question avec les portes *OR*, *NOT* et *traductions*,
2. inversement quelles sont toutes les portes d'un type donné qui produisent un métabolite donné ?
3. étant donné un ensemble de portes déjà utilisées dans une solution, quelles sont les portes qui n'interagissent pas avec elles ? En effet, si un métabolite d'une des portes utilisée a de l'affinité avec une enzyme d'une autre porte non prévue dans les calculs, cela créerait des « court-circuits ».

D'autres questions pourront éventuellement compléter ce panel.