

## A LIRE AVANT DE RENSEIGNER LE DOCUMENT

L'objectif de ce document est de recueillir les actions qui ont été menées sur les données des cas d'études. La synthèse de ces documents servira à identifier les besoins en collecte et analyse de données pour la création d'indicateurs et à connaître les pratiques en analyse de données dans le cadre de Hubble. Cette synthèse permettra la création du modèle conceptuel pour la plateforme Hubble.

En fonction du travail de collecte et/ou d'analyse que vous avez fait, vous renseignerez uniquement les parties qui concernent votre travail. Nous vous demandons d'être le plus précis possible dans vos descriptions. Si des scripts pour la collecte ou l'analyse de données ont été conçus (e.g. script R) nous vous demandons de bien vouloir les inclure dans ce document.

Merci.

=====

Date de rédaction : 16/11/2015

Nom du rédacteur du document : Marie Lefevre

Spécialités : ~~Didactique, Sciences de l'éducation, Informatique, Analyse de données~~ (*razer les mentions inutiles*)

=====

Cas d'étude Hubble: LIF3

Scénario hubble : 1 (Qualité des questions de types (QCM) faites) et 2 (Identifier des 'types' d'apprenants)

Personnes impliquées pour la collecte et l'analyse : Marie Lefevre, Alexis Lebis, Nathalie Guin

Période de la collecte : chaque semestre d'automne et de printemps depuis septembre 2014

Periode de l'analyse : Avril 2015 - ...

### **Dispositif d'apprentissage (Etude de cas de Hubble)**

Type de dispositif : Plateforme d'exercices pour l'auto-évaluation

Finalité de l'apprentissage : Pas d'apprentissage, révision des acquis et entraînement sur les notions vues en amphi à l'université

Utilisation du dispositif et fonctionnalités : après chaque cours en amphi, des exercices sont mis à disposition des étudiants pour qu'ils s'auto-évaluent et qu'ils manipulent les notions vues en cours

Contexte de production de données : de septembre à janvier, puis de janvier à mai en 2014-15 et 2015-16.

Au besoin indiquer les différents moments de la production (savoir si des données ont été produites sur plusieurs années)

## Décrire en quelques mots la problématique posée :

Objectifs de l'analyse :

- Identifier les modèles de questions de QCM qui ne sont pas pertinents (trop faciles / difficiles).
- Identifier l'utilisation de la plateforme par les étudiants (abandon), les types d'apprenants et leur progression au cours du temps

## Description du stockage des données:

Plateformes/outils utilisés: Les traces sont stockées dans la base de données de la plateforme Claroline Connect et doivent être transférées vers KTBS

Points forts de ces plateformes	Points faibles
Les traces du plugin ASKER sont anonymisées car nous n'avons pas les infos sur les comptes de Claroline Connect	Le passage de la base de données vers KTBS s'avère très compliquées au vue de la complexité des tables dans la BD

## Production des données avant le traitement :

Décrire le processus de production des données brutes : un collecteur de traces est contenu dans la plateforme d'exercices.

Liste des variables initiales : Nom et Description

Contient les tables (entre autres) :

- Table **Claire\_exercise\_item**
  - Id : de l'exercice généré dans la BD
  - Exercise\_id : [Id de l'exo](#) pour la plateforme ASKER
  - Type : QCM, QROC, groupement... => pour QCM = multiple-choice
  - Content : JSON de la forme (pour les QCM) :

```
{“question”:"xxx",
“proposition”:[
  {“text”: “la proposition 1”, right:bool(right/false)},
  ...],
“origin-resource”:"n° de la ressource utilisée pour générer la question,
“item-type”:"mutiple-choice-question"}
```
- Table **Claire\_exercise\_attempt**
  - Id : de la tentative dans la BD
  - Exercise\_Id : [id de l'exercice](#) auquel est associée la tentative
  - User\_Id : [utilisateur](#) ayant fait la tentative
  - Test\_attempt\_id : NULL tout le temps...

- Created\_at : date de la forme 20xx-xx-xx hh:mm:ss
- Position : NULL tout le temps...
- Table **Claire\_exercice\_answer** (format pour les QCM) :
  - Id : de la réponse dans la DB
  - Item\_Id : n° de la question dans le QCM
  - Attempt\_Id : n° de la tentative pour l'item
  - Content : JSON de la forme {"content": [0,1,0,...]}
  - Mark : score en pourcentage

Plateformes/outils utilisés:

Points forts	Points faibles

### Description des pré-traitements:

Objectifs des pré-traitements

Décrire le processus de pré-traitement:

Pour l'automne :

Entrée :

- created-at (date)

Action :

- supprimer les données avant le 8/9 et après le 6/01

Résultat :

- mêmes infos dans les traces sur une période plus courte

**REMARQUE : en rouge les paramètres à faire varier pour réexploiter au printemps...**

Entrée :

- résultat précédent

Action :

- pour chaque apprenant
- séparer les tentatives par semaine à partir du 8/9
- compter le nombre de tentatives par semaine

Résultat :

- autant de listes que d'apprenants avec autant d'entiers que de semaine entre le 8/9 et le 6/01, ici 18 semaines

Entrée :

- résultat précédent

Action :

- fusionner les semaines 8 (du 27/10) et 9 (du 3/11) car vacances scolaires
- fusionner les semaines entre le 15/12 et le 6/01 car vacances scolaires

Résultat :

- autant de listes que d'apprenants avec 15 semaines donc 15 entiers

=> fin de la préparation "temporelle" des données

Entrée :

- résultat précédent

Action :

- pour chaque nombre de tentative, si nombre de tentatives >1, remplacer l'entier par 1, sinon par 0

Résultat :

- autant de listes que d'apprenants avec 15 entiers valant 0 ou 1

=> fin de la normalisation, début de l'analyse des types d'apprenants

Plateformes/outils utilisés: pas mis en oeuvre techniquement

Points forts	Points faibles

**Description des analyses :** (Faire une description de chacune des analyses conduites)

Liste des variables : Nom et Description

résultat précédent : pour 1 user, 15 semaines représentées par des 0 ou 1

Liste des méthodes mise en œuvre :

#### CAS 1

Entrée :

- résultat précédent : pour 1 user, 15 semaines

Action :

- faire la somme de toutes les semaines pour 1 utilisateur
- si somme > 10 alors type régulier = vrai, sinon faux

Résultat :

- pour chaque apprenant un booléen indiquant le type "régulier"

#### CAS 2

Entrée :

- résultat précédent : pour 1 user, 15 semaines

Action :

- faire la somme de S1 à S7 pour 1 utilisateur => valeur SommeDebut
- faire la somme de S8 à S15 pour 1 utilisateur => valeur SommeFin
- si SommeDebut > 5 et SommeFin < 2 alors type au-début-pas-apres = vrai, sinon faux

Résultat :

- pour chaque apprenant un booléen indiquant le type "au-début-pas-apres"

#### CAS 3

Entrée :

- résultat précédent : pour 1 user, 15 semaines

Action :

- faire la somme de toutes les semaines pour 1 utilisateur
- si somme <5 et valeur de S9=1 alors type avant-tp-note = vrai, sinon faux

Résultat :

- pour chaque apprenant un booléen indiquant le type "avant-tp-note"

#### CAS 4

Entrée :

- résultat précédent : pour 1 user, 15 semaines

Action :

- faire la somme de toutes les semaines pour 1 utilisateur
- si somme <5 et valeur de S15=1 alors type avant-exam = vrai, sinon faux

Résultat :

- pour chaque apprenant un booléen indiquant le type "avant-exam"

#### CAS 5 : type "occasionnel" : 1 semaine sur 2 (ou presque...)

Mode opératoire technique, logiciels utilisés

Scripts produits pour l'analyse des données (*exemple script R*)

Résultats obtenus: Pas encore mis en oeuvre

Points forts des analyses	Points faibles des analyses

#### Description des données produites au cours du traitement

Objectif de la création de ces nouvelles données :

Mode de calcul de ces variables :

#### Description des nouvelles variables : Nom et description

Nom	description

### **Description des Itérations**

Pourquoi le processus d'analyse a été reproduit ?

Points forts des itérations	Points faibles des itérations