

# Cartographie de portails Wikipédia

## Compte rendu de la réunion du 6 février

### Attendu

#### 0. Choix des métadonnées stockées à définir

- Métadonnées de la page wikipédia
- Nombre de pages homologues (en langue étrangère)
- Longueur de la page (nombre de caractères)
- Nombre de visites ?

#### 1. Viewer

**Choix du Viewer** : soit celui que va nous envoyer E. Fritsch, soit Google Earth (avec du KML)

**Affichage dans le viewer** : carte + timeline

On clique sur un marqueur posé sur la carte ou sur la timeline, une infobulle s'ouvre avec les informations et le lien wikipédia de la page.

Tests unitaires du Viewer en attendant la mise en place de l'aspirateur.

#### 2. Aspirateur en back office

**Langage** : PHP pur/PHP-Javascript client/Node

Pour PHP, passer par un proxy ou en file

**Lecture de chaîne** : - Expression régulière

L'aspirateur doit être récursif : Pour chaque lien, on vérifie que la page appartient au thème (nom du portail) et si oui, on traite la page et chaque lien dans cette page.

Si le thème n'a pas de portail correspondant, il faut trouver une condition d'appartenance dans les informations données dans le cadre à droite de la page.

Il faut encapsuler les tâches.

Les informations qui ne sont pas géolocalisées sont à garder pour la timeline

#### 3. Personnalisation

L'utilisateur doit pouvoir cocher les événements qu'il veut voir apparaître sur la carte et sur la timeline.

Timeline : échelle en fonction du thème (Par ex, pour la 1<sup>re</sup> Guerre Mondiale, durée de la bataille en jours ; pour la Guerre de Trente ans, étendue de la bataille qui correspond à son importance (voir alors les critères pour déterminer l'importance d'une bataille: nombre de morts,...))

#### 4. Si temps restant : descendre en granularité dans l'aspirateur

Récupérer les informations sur des événements qui n'ont pas de page Wikipédia mais seulement un paragraphe dans une page en rapport avec le thème (Par exemple, dans la page sur la bataille de Verdun, paragraphe sur les combats autour de Douaumont).

### Autres

**Contraintes** : le responsable des tests unitaires doit être différent du responsable développement

**Besoins matériels** : voir pour le serveur et la base de données

Jusqu'au 25 mars : 3 cycles de 2 séances

=> **Réunion le 18 février**