
Règles

Après lecture de quelques fichiers XML, je suis parvenu à en saisir quelques récurrences (liste non exhaustive) :

- « Poisoning perpetrated against [victims] »
- « Assassinated [victims] »
- « demembred [victims] »
- « (the) murder of [victims] »
- « (was) killed by [murder] »

De ce fait, on peut en conclure à un premier traitement :

- Transformer le mot en **lexème** (ex : Assassinated → Assassinat)
- Comparer le lexème obtenu avec une liste du champ lexical du **meurtre** (ex : Assassinat , Kill , Poison , etc.)
- Si oui, alors on peut s'intéresser à ce mot (on va le nommer « **cible** »)

Un deuxième traitement consisterait à analyser les mots qui se trouvent autour de la « cible » :

- **Après** un « by », on sait qu'il y aura le nom du meurtrier (le nom du fichier est le nom de ce meurtrier)
- **Après** un « of », on sait que cela concerne les victimes :
 - Cela peut être un **chiffre**, en nombre ou en lettre (cela ne nous intéresse pas car il n'y aura pas de noms propres).
 - Cela peut être un **mot-clé** : en Xml, ça ressemble à [[mot-clé]] : si le mot-clé peut être un nom commun (comme le mot prince), on peut tomber sur un nom propre (qui nous intéresse, car ça va être celui de la victime)
- Au lieu de « of », on peut tomber sur un mot de « viser » (ex : perpetrated **against**)
- Lorsqu'il y a un pronom comme « him » par exemple, il faudrait pouvoir se souvenir des noms propres enregistrés et prendre soit le dernier ou les derniers (si on a « them »). Si tout se passe bien, le nom en rapport à un pronom n'est jamais trop loin (ex : il y a une phrase avant).