

Collecte des Données (1/2)

Problèmes :

- Pas de corpus a priori
- Pas d'API
- Pas de structuration uniforme
- Pas d'encodage uniforme
- Pas d'unicité des références

Collecte des Données (2/2)

Démarche :

- Identification arbitraire d'un corpus avec temps d'arrêt
- Extraction récursive des liens avec élagage
- Extraction des contenus
- Extraction des textes propres
- Décodage
- Stopwords et lemmatisation

Exemple

```
>————>————>————>————><li>
><a href="
http://www.profmichelle.com/2017/01/nos-cahiers-de-noel.htm
l" title="Voilà le super travail de mes élèves de
sixième... ..en attendant Noël! Bravo!">
>————>
></a>
><p class="caption">
>————><a href="
http://www.profmichelle.com/2017/01/nos-cahiers-de-noel
.html" title="Voilà le super travail de mes élèves de
sixième... ..en attendant Noël! Bravo!">
```

Exemple

```
http://www.profmichelle.com/2017/01/nos-cahiers-de-noel.htm
```

```
1"
```

```
>
```

```
http://www.profmichelle.com/2017/01/nos-cahiers-de-noel  
.html
```

Exemple

```
"color:#000080;"><span lang="FR" style="background:white">Auschwitz</span>
"FR"><span style="background:white">.</span></span><span lang="FR"> L&#39;
&agrave; la 2<sup>e</sup> place du classement des 50 BD essentielles &eac
</i></span></span></span></span></p></div>
<div class="ob-section ob-section-link"><div class="ob-ctn ob-ctn--withIma
https://www.kizoa.it/Movie-Maker/d92418675kP302368765o111/la-shoa-et-la-ba
ob-pull-left ob-media-left" target="_blank"><a href="https://www.kizoa.it/Movie-Maker/d92418675kP302368765c
"ob-link" target="_blank">Movie Maker: La Shoa et la bande dessinée Maus</
non è stata attivata dalla persona che ha fatto questa creazione. Puoi ric
personalizzato. Questa funzionalità è disponibile ...</p><p class="ob-url"
https://www.kizoa.it/Movie-Maker/d92418675kP302368765o111/la-shoa-et-la-ba
https://www.kizoa.it/Movie-Maker/d92418675kP302368765o111/la-shoa-et-la-ba
<div class="ob-section ob-section-html"><p style="margin-bottc
sont les animaux qui apparaissent le plus fr&eacute;quemment dans les imag
```

Exemple

vocabulaire planche entiere Bande Dessinee compose cases vignettes indiquer changement lieu temps appeles elements narratifs phylactere prend forme bulle contient paroles forme nuage composition personnage holocauste Maus Spiegelman Maus bande dessinee ecrite par Art Spiegelman New York traitant persecutions juifs annees 1930 1940 Shoah univers animalier nazis representes chats Juifs souris Polonais oeuvre recu prix Pulitzer special 1992 traduite huit langues narratives pere Spiegelman racontant deportation recit relate par son pere raconte Vladek Spiegelman rescape juif camps nazis fils auteur de la bande dessinée pere portes Auschwitz trottoirs New York deroule temps annees 1940 survivre survivant Spiegelman rend appartement pere New York Regarde ensemble evenements vecus pere lors Seconde Guerre mondiale] premieres pages Spiegelman aborde theme majeur litterature Solidarite deportees pere jeune fils pleure amis partis amis amis argent sauver selectionnees Auschwitz album classe 2012]

Topic model (1/2)

évaluer
compétence

CAPES
formation

ministre
réforme

Sep
21

Foutage de gueule

Il paraît que le prof doc n'évalue pas les élèves.

Le prof doc passe un Capes ! Certificat d'Aptitude au
Professorat de l'Enseignement du Second degré ! Comment
peut-on enseigner sans évaluer ? Sans faire un suivi des
élèves ? Là, j'aimerais bien qu'on m'explique...

J'enseigne et j'évalue mes élèves. Je suis prof, c'est mon
boulot, faire passer des connaissances et m'assurer qu'elles
sont comprises, acquises.

On l'entend partout, notre ministre le dit partout,

chaque **topic** est une distribution de mots

Topic model (1/2)

évaluer
compétence

CAPES
formation

ministre
réforme

Sep
21

Foutage de gueule

Il paraît que le prof doc n'évalue pas les élèves.

Le prof doc passe un Capes ! Certificat d'Aptitude au
Professorat de l'Enseignement du Second degré ! Comment
peut-on enseigner sans évaluer ? Sans faire un suivi des
élèves ? Là, j'aimerais bien qu'on m'explique...

J'enseigne et j'évalue mes élèves. Je suis prof, c'est mon
boulot, faire passer des connaissances et m'assurer qu'elles
sont comprises, acquises.

On l'entend partout, notre ministre le dit partout,

chaque **topic** est une distribution de mots

chaque **document** est un mélange de quelques
topics

Topic model (1/2)

évaluer
compétence

CAPES
formation

ministre
réforme

Sep
21

Foutage de gueule

Il paraît que le prof doc n'évalue pas les élèves.

Le prof doc passe un Capes ! Certificat d'Aptitude au
Professorat de l'Enseignement du Second degré ! Comment
peut-on enseigner sans évaluer ? Sans faire un suivi des
élèves ? Là, j'aimerais bien qu'on m'explique...

J'enseigne et j'évalue mes élèves. Je suis prof, c'est mon
boulot, faire passer des connaissances et m'assurer qu'elles
sont comprises, acquises.

On l'entend partout, notre ministre le dit partout,

chaque **topic** est une distribution de mots

chaque **document** est un mélange de quelques
topics

chaque **mot** est tiré au sort dans un topic

Topic Model (2/2)



Sep
21

Foutage de gueule

Il paraît que le prof doc n'évalue pas les élèves.

Le prof doc passe un Capes ! Certificat d'Aptitude au Professorat de l'Enseignement du Second degré ! Comment peut-on enseigner sans évaluer ? Sans faire un suivi des élèves ? Là, j'aimerais bien qu'on m'explique...

J'enseigne et j'évalue mes élèves. Je suis prof, c'est mon boulot, faire passer des connaissances et m'assurer qu'elles sont comprises, acquises.

On l'entend partout, notre ministre le dit partout,

Dans la réalité, on observe les documents

Topic Model (2/2)



Sep
21

Foutage de gueule

Il paraît que le prof doc n'évalue pas les élèves.

Le prof doc passe un Capes ! Certificat d'Aptitude au Professorat de l'Enseignement du Second degré ! Comment peut-on enseigner sans évaluer ? Sans faire un suivi des élèves ? Là, j'aimerais bien qu'on m'explique...

J'enseigne et j'évalue mes élèves. Je suis prof, c'est mon boulot, faire passer des connaissances et m'assurer qu'elles sont comprises, acquises.

On l'entend partout, notre ministre le dit partout,

Dans la réalité, on observe les documents

Tout le reste constitue des **variables cachées**

Nous cherchons à les retrouver, en inversant le processus génératif

Exemple

6→liberte etat societe homme droit politique droi
economie selon monde pouvoir liberalisme justice se
7→triangle loys puzzle rectangle puzzles eleves z
maths dernier mathix emes reponses exemple isocèle
8→documents republique document trace travail cor
place entreprise etats france problematique composi
9→jeux console jouer generation playstation stree
attends dois couleurs nouvelle premier semble live
10→eleves histoire premiere travail etaient ulyse
terre matiere niveau sujets possible princesse troi
11→opinion verite lire challenge opinions albums l
roman envie suite delivrer chapitre extraits person
12→ecole eleves dalton classe acte parents educati

Exemple

```
le-data\bigbag\1-laviemoderne-1.txt 0 0 association 73
le-data\bigbag\1-laviemoderne-1.txt 1 1 populaire 84
le-data\bigbag\1-laviemoderne-1.txt 2 2 victime 5
le-data\bigbag\1-laviemoderne-1.txt 3 3 numerisme 82
le-data\bigbag\1-laviemoderne-1.txt 4 4 etude 82
le-data\bigbag\1-laviemoderne-1.txt 5 5 instructive 1
le-data\bigbag\1-laviemoderne-1.txt 6 6 enseignements 71
le-data\bigbag\1-laviemoderne-1.txt 7 7 soient 57
le-data\bigbag\1-laviemoderne-1.txt 8 8 scientifiques 1
le-data\bigbag\1-laviemoderne-1.txt 9 9 nouveaux 82
le-data\bigbag\1-laviemoderne-1.txt 10 10 vient 5
le-data\bigbag\1-laviemoderne-1.txt 11 11 publier 84
le-data\bigbag\1-laviemoderne-1.txt 12 12 afev 82
le-data\bigbag\1-laviemoderne-1.txt 13 13 numerique 82
le-data\bigbag\1-laviemoderne-1.txt 14 14 collegiens 1
```

Limites et espoirs

- Hypothèse forte du modèle
- Gérer l'hétérogénéité (sur le fond) du corpus
- Démarche en échantillonnage/test pour dépasser l'exploratoire.
- Vers une constitution itérative du corpus