

Extraction automatique du sens de documents textuels

Mathilde Da Cruz

Master 2 IASD - Intelligence Artificielle, Systèmes, données

21 Septembre 2022

- 1 Introduction
- 2 Inconvénients des méthodes classiques
- 3 Solution proposée
- 4 Conclusion

Introduction - Contexte

Une demande formulée par le bureau du dialogue social → tâche de veille sociale

- Par quel moyen et sous quelle forme extraire le sens de documents ?
- Comment évaluer ce travail ? En particulier, comment évaluer une synthèse de documents ?

Introduction - Intérêt du sujet

- De nombreuses utilisations ;
- Une tâche pas vraiment résolue ;
- Plusieurs éléments à définir, à discuter :


→ Qu'est ce qu'un résumé ? Comment extraire du sens ? Comment le transmettre efficacement ?

→ Qu'est ce qu'un bon résumé ? Comment mesurer l'efficacité d'un modèle d'extraction de sens ?

→ Légitimité des modèles SATO ?

Inconvénients des méthodes classiques - L'abstractif

- Abstractif VS Extractif : lequel est meilleur ?
- Des incohérences ...

Référence	Le Bayern-Munich a fait un score de 3-0 contre l'ASSE ce qui est tout à fait historique en 2012. C'est une année très importante pour ce club qui continue de faire des scores impressionnants, au détriment d'autres équipes.
Résumé 1	Le Bayern Munich a réalisé un triplé en 2012 et 2013 face à l'équipe de football de l'Olympique de Marseille. 
Résumé 2	Le club bavarois a remporté son premier titre de champion d'Allemagne en 2012, en battant l'ASSE (3-1).

Inconvénients des méthodes classiques - L'évaluation

$$ROUGE - N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (1)$$

Inconvénients des méthodes classiques - L'évaluation

	eval_A	eval_C	eval_E	eval_L	eval_M1	eval_M2	eval_Z
TFIDF	0.2	0.266667	0.6	0.375	0.25	0.5	0.266667
Comp. Vecteurs	0.1	0.133333	0.3	0.125	0	0.3	0.133333
TextRank PMI	0.2	0.4	0.3	0.25	0.375	0.2	0.133333
TextRank Embed.	0.1	0.133333	0.1	0.125	0.125	0.1	0

Figure – Résultats des évaluations de quatres méthodes d'extraction de mots clés différentes, avec le score $F1$, en comparant à différents ensemble de mots clés référence.

Inconvénients des méthodes classiques - L'évaluation

Référence A	"Une femme accouche d'un bébé de 6 kilos."
Résumé A1	"Un bébé accouche d'une femme de 6 kilos."
Résumé A2	"Une femme accouche de 6 kilos d'un bébé."
Résumé A3	"De kilos d'un femme bébé une accouche 6."
Référence B	"Le vieillard répond au policier."
Résumé B1	"Le policier répond au vieillard."

- Dépendance au résumé référence ;
- Informations contenues et compression ;
- Cohérence et grammaire ;
- Consistance.

Solution proposée - Vers de nouvelles méthodes d'évaluation

Evaluer les qualités réelles d'un résumé :

- La pertinence ;
- La consistance ;
- La fluidité ;
- La cohérence.

Solution proposée - Visualisation et mise en contexte



Les cotisations sociales

sa scilicet à quel ?

Les gouvernements successifs et le patronat n'ont eu de cesse de nous parler de coût du travail, de charges sociales... Ils tentent de faire croire qu'on y gagnerait à ne plus payer de cotisations sociales. Pourtant il y a regardé de plus près rien n'est plus faux. Les cotisations retraites, chômage, maladie... payées par l'employeur et/ou par les salariés (pour les cotisations retraite), c'est du salaire. Ce n'est pas le salaire qu'on touche à la fin du mois, mais c'est celui qu'on touche toute sa vie, quand on est malade qu'on est au chômage, en retraite...

Toutes ces estimations qu'il s'agissait sans les remplacez, c'est du fric en monnaie pour la sécurité sociale, c'est ce qui fait qu'on balance les allocations chômage, le remboursement des notes et les retraites. Mais ça n'est pas perdu pour tout le monde, le patronat en fait son affaire en réduisant considérablement la part qu'il doit verser pour chaque salarié.

Et les impôts ça sert à quoi ?

On nous fait un sacré

va prendre charge les augmentations de pouvoir d'achat que les patrons ne veulent pas payer (à peine le gouvernement a-t-il évoqué la possibilité d'augmenter 10 % du CICE que le Medef est immédiatement remonté au créneau), mais cette généralité a son côté parce que le gouvernement ne veut pas se laisser aller à des mesures de dégrèvement. Alors comment dans un tel contexte développer la proximité des services publics réclamée dans cette mobilisation ? **Une autre répartition des rôles** : ça va consister d'abord dans le rapport entre les salaires et les patronnes, par plus de salaire, moins de profits et de dividendes, et ça va consister ensuite au niveau de l'Etat par un système fiscal juste, par des politiques de redistribution, par des politiques de soutien à la petite entreprise, par des politiques de soutien à la petite entreprise et un développement des services publics au bénéfice de tous et toutes.

Et 0 pour les fonctionnaires ?

Totalement « inutilisés » du dispositif, ceux-ci, notamment celles et ceux dont les salaires sont égaux ou inférieurs au SMIC, se sentent donc à l'aise pour participer de leurs traitements, en aucune forme de flat et d'aise (sans pour les fonctionnaires de la police à ce jour). Ils et elles doivent continuer à supporter le poids de leurs salaires et subir les conséquences des baisses de recettes publiques pour financer les annonces de Macron. A savoir de nouvelles suppressions d'effectifs, des coupes dans les moyens et des suppressions de services publics pouvant en grande partie à l'origine de la colère des gilets jaunes. C'est en patissant et aux actionnaires de payer les annonces de Macron ! L'augmentation des salaires doit se faire pour toutes et tous !

Et 0 pour les chômeur-euses et les minimas sociaux ?

Pour Solidaires, ce sont les patrons qui licencient, ce ne sont pas les chômeurs-crisés qui sont responsables

Aucune mesure de revalorisation des minima sociaux n'a été annoncée par le gouvernement, rien pour le RSA, rien pour l'allocation aux adultes handicapés, rien pour le minimum vieillesse, rien pour les APL.

- Compense le manque de contexte ;
- Facilite l'évaluation manuelle.

Solution proposée - Visualisation et mise en contexte



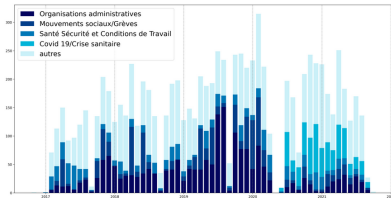
Tippex

SYNDICAT	PERIMETRE	THEME	DATE DE DEBUT	DATE DE FIN	POURCENTAGE DE PHRASES
Choose an option	Choose an option	Choose an option	2016/10/06	2021/09/01	20
<button>Appliquer les filtres</button>					

Le filtre ressort 8822 tract(s).



Nombre de tracts publiés par mois



Conclusion - Limites et évolution

- Synthèse abstractive ;
- Ré-entraîner un modèle sur nos données ;
- Utiliser la mise en page et les images ;
- Synthèse multi-documents ;
- Synthèse orientée selon un sujet ;
- Détection de nouveautés ;
- Mettre en place les méthodes d'évaluation proposées.

Conclusion

- Prendre en compte les spécificités d'un problème réel ;
- Mettre en balance l'efficacité, la complexité, et les ressources nécessaires ;
- Savoir remettre en question l'état de l'art ;
- Ne pas négliger l'évaluation.

Merci pour votre attention

