

Détection de Désinformation : Vérification Automatique des Faits

Auteur : Jérémie BRESSAND

Parcours : DECOL

Lieu : LIRMM, CNRS.

Encadrants :

- Konstantin TODOROV, LIRMM, Montpellier.
- Mathieu ROCHE, CIRAD, Montpellier.
- Stefan DIETZE, L3S, Leibniz University Hanover.



Plan

1. Introduction
2. Etat de l'art
3. Définition du problème
4. Les données
5. Vue générale de notre approche
6. Découverte d'information complémentaire
7. Prédiction de la véracité d'une information
8. Evaluation
9. Conclusion

1 Introduction

Réseaux sociaux



Désinformation



Donald J. Trump ✓

@realDonaldTrump



The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.

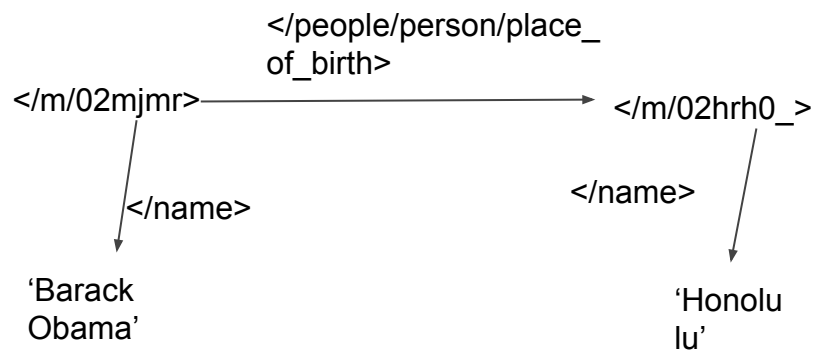
21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



2 Etat de l'art

Approches avec bases de connaissances (Xin Luna Dong, 2014), (Ran Yu, 2018)



Efficacité, transparence

Domaine d'application limité

Approches avec données non structurées (Naeemul Hassan, 2015), (William Yang Wang, 2017), (Kashyap Popat, 2017), (Xuezhi Wang, 2018), (Soroush Vosoughi, 2018)



Domaine d'application vaste



Efficacité ?
Transparence ?

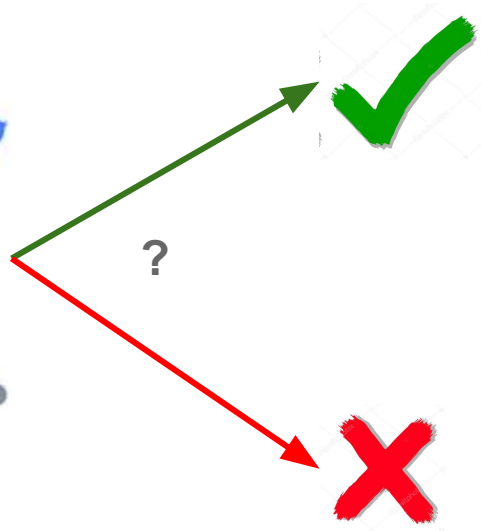
3 Définition du problème



The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.

21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



VRAI

FAUX

Problème n°1 : découverte d'information complémentaire permettant de caractériser la véracité d'une information



The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.

21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



Over the past 50 years, **the average global temperature has increased** at the fastest rate in recorded history.

<https://www.nrdc.org/stories/global-warming-101>

Problème n°2 : prédiction de la véracité d'une information en utilisant de l'information complémentaire

Over the past 50 years, the
**average global temperature has
increased** at the fastest rate in
recorded history.

<https://www.nrdc.org/stories/global-warming-101>



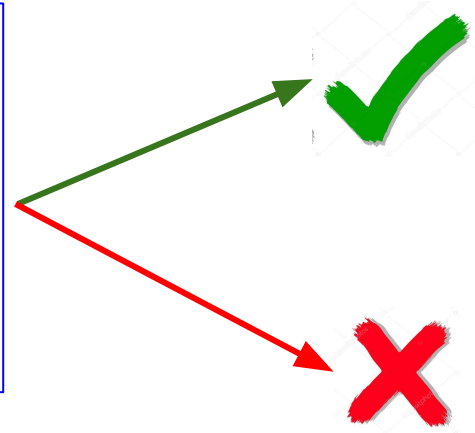
Donald J. Trump ✓
@realDonaldTrump



The concept of global warming was created by and for the
Chinese in order to make U.S. manufacturing non-competitive.

21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



VRAI

FAUX

4. Les données



DERRICK CROWE

Says each of the past three years "has been the hottest on record."

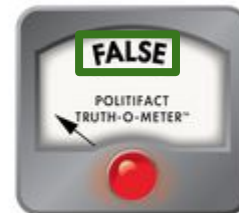
— *PolitiFact Texas* on Thursday, June 15th, 2017



AMERICANNEWS.COM

"BUSTED: NASA Accidentally destroys every climate change advocate by exposing massive truth behind Antarctica."

— *PunditFact* on Monday, November 20th, 2017



Assertion : une assertion est un texte contenant de l'information dont la véracité peut être vérifiée par un journaliste.



Véracité d'une assertion : l'assertion est qualifiée comme vraie ou fausse par un journaliste.

5. Position par rapport à l'état de l'art

(Xuezhi Wang, 2018)

Approches avec données non structurées (Naeemul Hassan, 2015), (William Yang Wang, 2017), (Kashyap Papat, 2017), (Xuezhi Wang, 2018), (Soroush Vosoughi, 2018)



Problèmes rencontrés

- Une majorité de documents ne discutant pas des mêmes sujets que les assertions associées.
 - “We also note that simply taking all related documents generated by the candidate generation phase achieves much lower accuracy/precision because **the majority of related documents are not relevant.**” (Xuezhi Wang, 2018).
- Beaucoup de documents à récupérer :
 - Temps de récupération des pages de résultats renvoyés par un moteur de recherche : 1 semaine.
 - Estimation du nombre de résultats par page : 10 résultats.
 - **Temps estimé de récupération des documents : 10 semaines.**

Solution



The Role of Twitter in the 2016 US Election

<https://books.google.fr/books?isbn=3319689819> - Traduire cette page

Christopher J. Galdieri, Jennifer C. Lucas, Tauna S. Sisco - 2017 - Political Science

Unconventional and yet not unsurprising, Mr. Trump's climate tweet content ... Trump tweeted on November 6, 2012: "The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive" ...


PAS
PERTINENT

Global Warming Facts, Definition, Causes and Effects | NRDC

<https://www.nrdc.org/stories/global-warming-101> ▼ Traduire cette page

11 mars 2016 - A: Here's a simple definition of global warming. (And yes, it's really happening.) Over the past 50 years, the average global temperature has ...

Are the Effects of Global - How You Can Stop Global - Amanda MacMillan


PERTINENT

Récupération
document →



Snippet : un texte correspondant à un résultat fourni par un moteur de recherche par rapport à une assertion.

6. Découverte d'information complémentaire

Lexique

- | | |
|----------------|----------------------|
| 1 : definition | 6 : average |
| 2 : global | 7 : temperature |
| 3 : warming | 8 : create, |
| 4 : years | 9 : u.s., |
| 5 : chinese | 10 : manufacturing |
| | 11 : non-competitive |



Donald J. Trump ✓
@realDonaldTrump

The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.

21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



[concept, global, warming, create, chinese, u.s., manufacturing, non-competitive]

Global Warming Facts, Definition, Causes and Effects | NRDC

<https://www.nrdc.org/stories/global-warming-101> ▼ Traduire cette page

11 mars 2016 - A: Here's a simple definition of **global warming**. (And yes, it's really happening.) Over the past 50 years, the average global temperature has ...

Are the Effects of Global · How You Can Stop Global · Amanda MacMillan

[definition, global, warming, years, average, global, temperature]

[1,
2,
3,
4,
6]

[2,
3,
5,
8,
9,
10,
11]

Techniques d'analyse
de texte (BOW, TF-IDF,
LSI, LDA, RP)

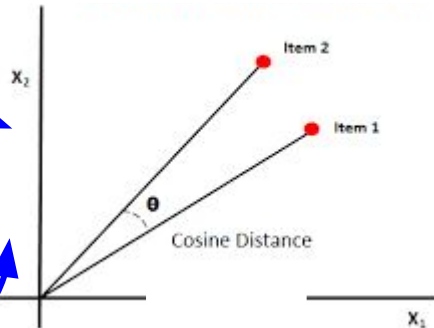
Représentation
de documents

[2,
3,
5,
8,
9,
10,
11]

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

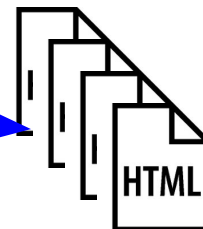
Représentation
d'assertions

[1,
2,
3,
4,
6]

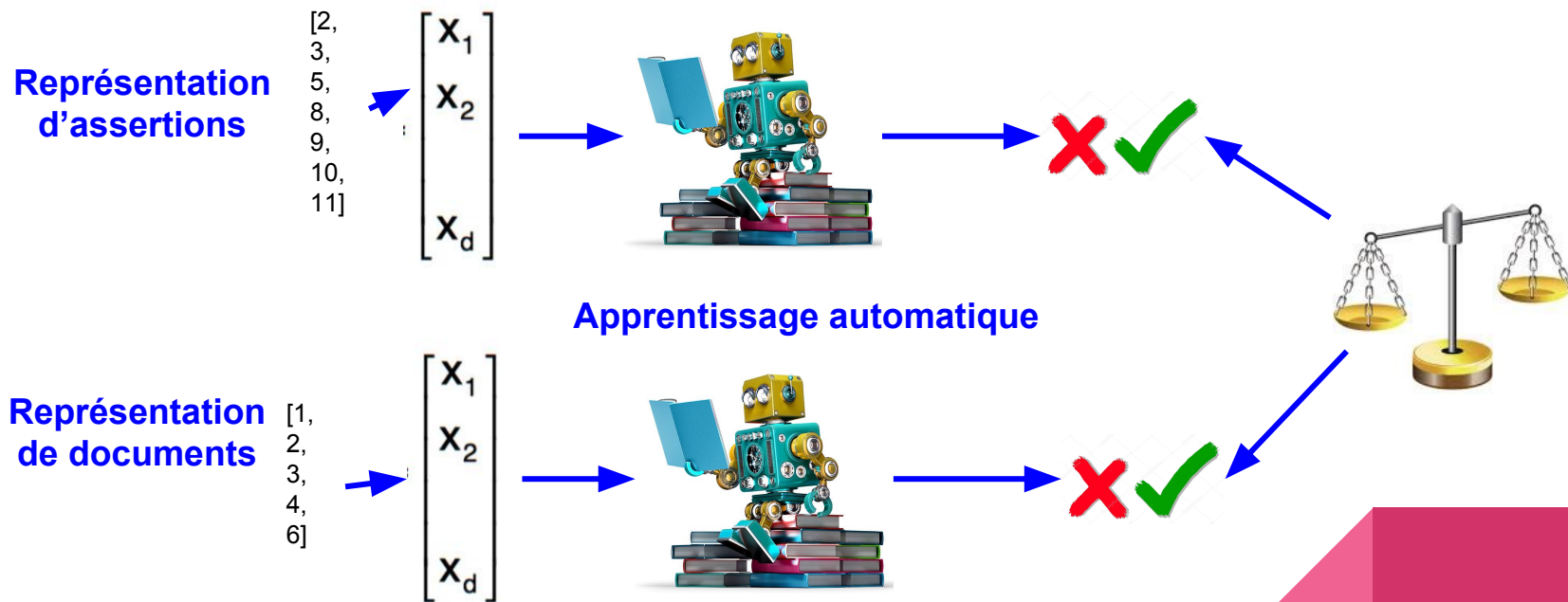
$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$


< seuil

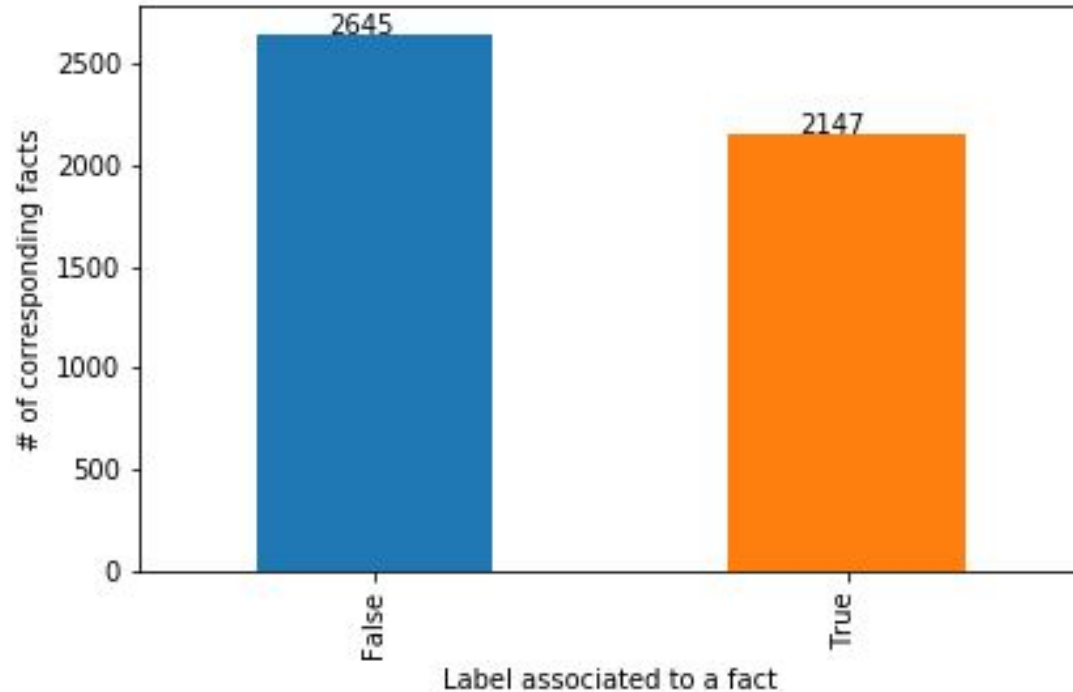
>= seuil



7. Prédiction de la véracité d'une information



8. Expérimentations



- Sélection d'un modèle pour la classification d'assertion :
 - Nombre de mots :
 - Modèles baselines : BOW et TF-IDF
 - Modèles plus complexes : LSI, LDA et RP.
 - Nombre de thématiques :
- Sélection d'un modèle pour la classification de documents :
 - Nombre de mots :
 - Modèles baselines : BOW et TF-IDF
 - Modèles plus complexes : LSI, LDA et RP
 - Nombre de thématiques :
- Protocole d'évaluation :
 - classifieur Logistic Regression, 10-fold cross validation
 - grid-search, score F-mesure
- Objectif : réduction du nombre de modèles

- Comparaison des performances entre le modèle pour les assertions et le modèle pour les documents
 - Classifieur SVM
 - 10-fold cross-validation
 - Optimisation grid-search pour les hyperparamètres
 - Score F-mesure

9.Evaluation

Modèles baselines

	90	100	110	120	130	140	150	160	170	180
BOW	0.61	0.57	0.57	0.56	0.58	0.58	0.58	0.60	0.61	0.60
TF-IDF	0.62	0.60	0.60	0.59	0.60	0.59	0.58	0.61	0.60	0.60

Modèles plus complexes

		90	100	110	120	130	140	150	160	170	180
LDA	10	0.59	0.67	0.64	0.57	0.55	0.59	0.53	0.58	0.51	0.59
	20	0.62	0.62	0.58	0.62	0.55	0.53	0.60	0.59	0.61	0.60
	30	0.62	0.65	0.62	0.55	0.50	0.59	0.61	0.59	0.63	0.65
	40	0.56	0.62	0.52	0.59	0.62	0.53	0.52	0.54	0.61	0.61
	50	0.64	0.52	0.62	0.53	0.58	0.53	0.49	0.51	0.54	0.57
LSI	10	0.64	0.61	0.61	0.60	0.61	0.61	0.61	0.64	0.61	0.62
	20	0.65	0.63	0.64	0.64	0.64	0.64	0.64	0.62	0.63	0.61
	30	0.63	0.63	0.65	0.62	0.61	0.63	0.63	0.62	0.60	0.62
	40	0.63	0.62	0.63	0.63	0.62	0.63	0.62	0.60	0.59	0.60
	50	0.63	0.62	0.64	0.62	0.63	0.61	0.60	0.59	0.59	0.59
RP	10	0.58	0.55	0.55	0.57	0.50	0.51	0.52	0.52	0.60	0.50
	20	0.58	0.50	0.59	0.56	0.53	0.61	0.58	0.60	0.60	0.56
	30	0.53	0.56	0.54	0.63	0.55	0.50	0.55	0.52	0.56	0.53
	40	0.61	0.44	0.59	0.57	0.55	0.58	0.54	0.57	0.55	0.50
	50	0.57	0.59	0.57	0.66	0.58	0.60	0.56	0.53	0.62	0.54

Sélection d'un modèle
pour la classification
d'assertion

Modèles baselines

	2K	4K	6K	8K	10K	12K	14K	16K	18K
BOW	0.53	0.48	0.48	0.48	0.47	0.46	0.48	0.47	0.48
TF-IDF	0.57	0.50	0.53	0.46	0.53	0.46	0.60	0.47	0.47

Modèles plus complexes

		2K	4K	6K	8K	10K	12K	14K	16K	18K
LDA	100	0.64	0.57	0.54	0.55	0.57	0.48	0.55	0.49	0.46
	140	0.55	0.47	0.49	0.47	0.53	0.58	0.44	0.56	0.50
	180	0.60	0.48	0.53	0.56	0.48	0.53	0.57	0.47	0.58
	220	0.54	0.55	0.53	0.62	0.60	0.58	0.53	0.58	0.57
	260	0.49	0.60	0.46	0.47	0.58	0.52	0.55	0.47	0.52
	300	0.46	0.52	0.54	0.58	0.66	0.55	0.53	0.47	0.47
LSI	100	0.51	0.53	0.60	0.49	0.52	0.54	0.52	0.55	0.53
	140	0.53	0.52	0.52	0.55	0.53	0.54	0.49	0.51	0.49
	180	0.51	0.53	0.66	0.59	0.51	0.54	0.55	0.50	0.52
	220	0.51	0.63	0.59	0.53	0.51	0.54	0.53	0.50	0.48
	260	0.52	0.50	0.51	0.46	0.54	0.60	0.52	0.53	0.48
	300	0.51	0.53	0.53	0.46	0.54	0.54	0.53	0.50	0.52
RP	100	0.51	0.54	0.53	0.54	0.53	0.51	0.50	0.54	0.52
	140	0.51	0.54	0.50	0.51	0.54	0.50	0.48	0.51	0.51
	180	0.50	0.49	0.52	0.53	0.46	0.53	0.52	0.55	0.52
	220	0.51	0.47	0.50	0.54	0.50	0.54	0.44	0.51	0.52
	260	0.53	0.46	0.48	0.53	0.56	0.53	0.50	0.54	0.50
	300	0.52	0.46	0.53	0.48	0.52	0.47	0.46	0.54	0.51

Sélection d'un modèle
pour la classification de
documents

Comparaison des performances entre le modèle pour les assertions et le modèle pour les documents

Score f-mesure pour classification d'assertions	Score f-mesure pour classification de documents
0.64	0.58

Pourquoi de tels résultats ?

Stratégie de recherche d'information



The concept of global warming was created by and for the chinese in order to mak



Donald J. Trump ✓
@realDonaldTrump



The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.

21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



Beijing to Donald Trump: Climate change is not a Chinese hoax

<https://www.telegraph.co.uk/news/2016/11/17/beijing-to-trump...> ▼

The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive. — Donald J. Trump (@realDonaldTrump) 6 ...

Author: Neil Connor Location: Beijing

China on Donald Trump's Claim It Invented Global Warming: Nah

<https://gizmodo.com/china-on-donald-trumps-claim-it-invented...>

"The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive ... Chinese in order to make U.S ...

Global Warming Facts, Definition, Causes and Effects | NRDC

<https://www.nrdc.org/stories/global-warming-101> ▼

Q: What is **global warming**? A: Here's a simple definition of **global warming**. (And yes, it's really happening.) Over the past 50 years, the average **global** temperature ...

Global Warming: 2018 Articles, Facts, Causes & Effects

<https://www.livescience.com/topics/global-warming> ▼

13/06/2018 · The latest articles and facts about **global warming** and its causes, plus a look at the effects of climate change: rising sea level and severe weather.



global warming



Représentation sémantique de l'information



The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.

21:15 - 6 nov. 2012

♡ 66,7 k 💬 116 k personnes parlent à ce sujet



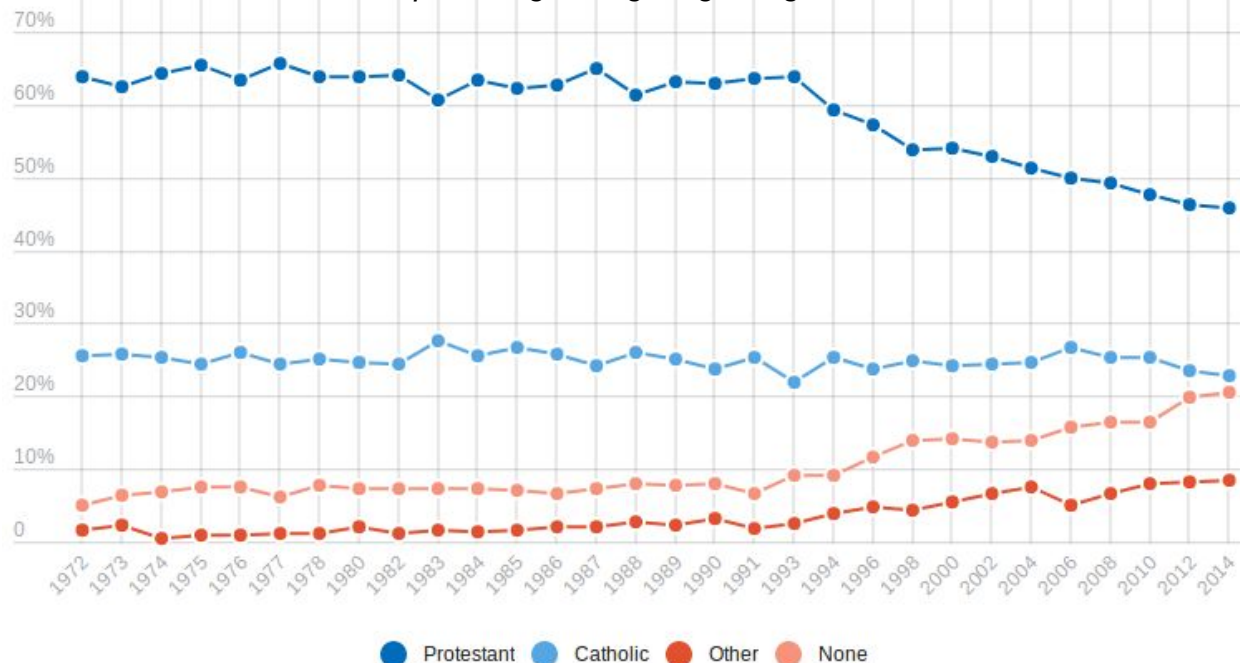
Over the past 50 years, **the average global temperature has increased** at the fastest rate in recorded history.



<https://www.nrdc.org/stories/global-warming-101>

Analyse temporelle de l'information

<http://www.politifact.com/truth-o-meter/statements/2018/may/16/mike-pence/mike-pence-right-religion-gaining-new-life-us/>



MIKE PENCE

"Religion in America isn't receding. It's just the opposite. Faith is gaining new life across America every day."

— *PolitiFact National* on Wednesday, May 16th, 2018

7 Conclusion

- **Ce que j'ai appris :**
 - Domaine de recherche compliqué, nécessité d'une représentation de l'information de plus haut niveau (complexité sémantique, temporelle) et de stratégies plus complexes de recherche d'information.
- **Ma contribution :**
 - Identifier des limites de certaines approches de vérification automatique des faits, notamment l'aspect temporelle de l'information.
- **Perspectives :**
 - Comparer les performances de notre approche avec les approches précédentes.
 - Collecter des données à partir d'autre sites web de vérification des faits.
 - Trouver de nouveaux descripteurs (aspects temporel et sémantique).
 - Trouver des nouvelles stratégies de recherche d'information.

Bibliographie

- [1] Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wiko Horn, Ni Lao, Kevin Murphy, Thomas Strohman, Shaohua Sun, Wei Zhang : “Knowledge Vault : A Web-Scale Approach to Probabilistic Knowledge Fusion.” paru dans les actes de la 20eme conférence internationale ACM SIGKDD sur la découverte de connaissances et l’exploration des données, pages 601-610, ACM, 2014
- [2] Ran Yu, Ujwal Gadiraju, Besnik Fetahu, Olivier Lehmberg, Dominique Ritze et Stefan Dietze. “KnowMore - Knowledge Base Augmentation with Structured Web Markup.”, paru dans le journal Semantic Web, 2018.

[3] Naeemul Hassan, Bill Adair, James T. Hamilton, Chengkai Li, Mark Tremayne, Jun Yang, Cong Yu. “The Quest to Automate Fact-Checking”, paru dans les actes du Computation + Journalism Symposium, 2015.

[4] William Yang Wang, “‘Liar, Liar Pants on Fire’: A New Benchmark Dataset for Fake News Detection’, paru dans les actes du rendez-vous annuel de l’Association for Computational Linguistics, pages 422-426, 2017.

[5] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen et Gerhard Weikum. “Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media.”, paru dans les actes de la 26ème International World Wide Web Companion, pages 1003-1012, 2017.

[6] Xuezhi Wang, Cong Yu, Simon Baumgartner et Flip Korn. “Relevant Document Discovery for Fact-Checking Articles.”, paru dans Companion Proceedings of the Web Conference, pages 525-533, 2018.

[7] Soroush Vosoughi, Deb Roy and Sinan Aral. “The spread of true and false news online.”, paru dans Science 359, pages 1146-1151, 2018.

Annexe A.1: Résultats découverte de snippets pertinents

Table 6.2: Relevant snippet discovery metric scores for topic modelling representations with associated sizes of dictionaries from 10 000 to 35 000 and associated number of topics from 100 to 300.

	LSI					LDA					RP				
	100	150	200	250	300	100	150	200	250	300	100	150	200	250	200
10K	0	1	1	0	0	0	5	4	2	6	3	3	5	2	6
15K	1	1	1	1	0	0	4	3	5	3	2	3	4	2	6
20K	1	1	1	1	0	6	2	1	1	1	2	3	1	3	4
25K	0	1	1	1	0	0	3	5	1	5	4	4	4	4	3
30K	0	1	1	1	1	3	0	7	1	5	6	3	4	1	2
35K	0	1	1	1	0	2	2	2	6	1	4	2	3	5	1

Annexe A.2 : Choix du seuil pour découverte de snippets pertinents

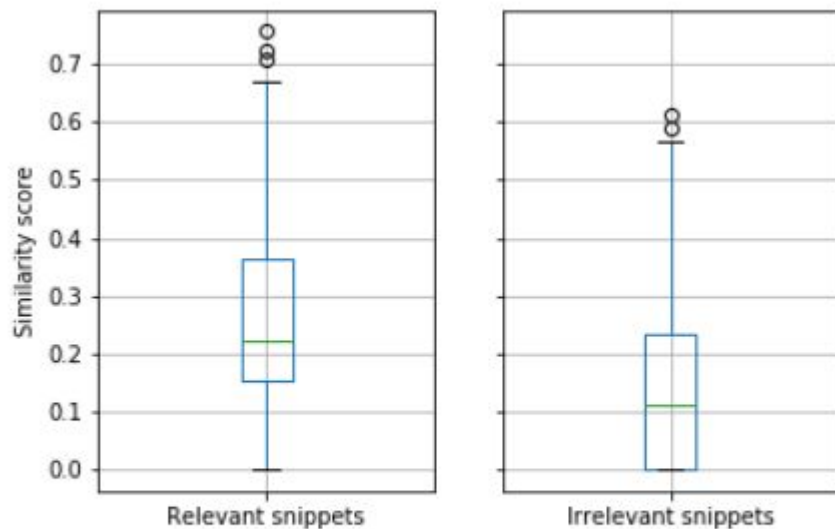


Figure 6.1: Distribution of similarity scores among relevant and irrelevant snippets with a dictionary size of 30K and 200 topics with LDA topic modeling.

Annexe B.1 : Résultats baselines classification assertions

Table 6.3: Results from binary classification experiment using BOW and TF-IDF representations of facts with size of dictionaries from 100 to 300 words.

	90	100	110	120	130	140	150	160	170	180
BOW	0.61	0.57	0.57	0.56	0.58	0.58	0.58	0.60	0.61	0.60
TF-IDF	0.62	0.60	0.60	0.59	0.60	0.59	0.58	0.61	0.60	0.60

Annexe B.2 : Résultats nouveaux descripteurs classification assertions

Table 6.5: F-measure scores from binary classification experiment using LSI, LDA and RP representations of facts with size of dictionaries from 90 to 180 words and number of topics from 10 to 50.

		90	100	110	120	130	140	150	160	170	180
LDA	10	0.59	0.67	0.64	0.57	0.55	0.59	0.53	0.58	0.51	0.59
	20	0.62	0.62	0.58	0.62	0.55	0.53	0.60	0.59	0.61	0.60
	30	0.62	0.65	0.62	0.55	0.50	0.59	0.61	0.59	0.63	0.65
	40	0.56	0.62	0.52	0.59	0.62	0.53	0.52	0.54	0.61	0.61
	50	0.64	0.52	0.62	0.53	0.58	0.53	0.49	0.51	0.54	0.57
LSI	10	0.64	0.61	0.61	0.60	0.61	0.61	0.61	0.64	0.61	0.62
	20	0.65	0.63	0.64	0.64	0.64	0.64	0.64	0.62	0.63	0.61
	30	0.63	0.63	0.65	0.62	0.61	0.63	0.63	0.62	0.60	0.62
	40	0.63	0.62	0.63	0.63	0.62	0.63	0.62	0.60	0.59	0.60
	50	0.63	0.62	0.64	0.62	0.63	0.61	0.60	0.59	0.59	0.59
RP	10	0.58	0.55	0.55	0.57	0.50	0.51	0.52	0.52	0.60	0.50
	20	0.58	0.50	0.59	0.56	0.53	0.61	0.58	0.60	0.60	0.56
	30	0.53	0.56	0.54	0.63	0.55	0.50	0.55	0.52	0.56	0.53
	40	0.61	0.44	0.59	0.57	0.55	0.58	0.54	0.57	0.55	0.50
	50	0.57	0.59	0.57	0.66	0.58	0.60	0.56	0.53	0.62	0.54

Annexe B.3 : Résultats baselines classification documents

Table 6.4: F-measure scores from binary classification experiment using BOW and TF-IDF representations of aggregated relevant documents with size of dictionaries from 2000 to 18000 words.

	2K	4K	6K	8K	10K	12K	14K	16K	18K
BOW	0.53	0.48	0.48	0.48	0.47	0.46	0.48	0.47	0.48
TF-IDF	0.57	0.50	0.53	0.46	0.53	0.46	0.60	0.47	0.47

Annexe B.4 : Résultats nouveaux descripteurs classification documents

Table 6.6: F-measure scores from binary classification experiment using LSI, LDA and RP representations of documents with size of dictionaries from 2000 to 18000 words and number of topics from 100 to 300.

		2K	4K	6K	8K	10K	12K	14K	16K	18K
LDA	100	0.64	0.57	0.54	0.55	0.57	0.48	0.55	0.49	0.46
	140	0.55	0.47	0.49	0.47	0.53	0.58	0.44	0.56	0.50
	180	0.60	0.48	0.53	0.56	0.48	0.53	0.57	0.47	0.58
	220	0.54	0.55	0.53	0.62	0.60	0.58	0.53	0.58	0.57
	260	0.49	0.60	0.46	0.47	0.58	0.52	0.55	0.47	0.52
	300	0.46	0.52	0.54	0.58	0.66	0.55	0.53	0.47	0.47
LSI	100	0.51	0.53	0.60	0.49	0.52	0.54	0.52	0.55	0.53
	140	0.53	0.52	0.52	0.55	0.53	0.54	0.49	0.51	0.49
	180	0.51	0.53	0.66	0.59	0.51	0.54	0.55	0.50	0.52
	220	0.51	0.63	0.59	0.53	0.51	0.54	0.53	0.50	0.48
	260	0.52	0.50	0.51	0.46	0.54	0.60	0.52	0.53	0.48
	300	0.51	0.53	0.53	0.46	0.54	0.54	0.53	0.50	0.52
RP	100	0.51	0.54	0.53	0.54	0.53	0.51	0.50	0.54	0.52
	140	0.51	0.54	0.50	0.51	0.54	0.50	0.48	0.51	0.51
	180	0.50	0.49	0.52	0.53	0.46	0.53	0.52	0.55	0.52
	220	0.51	0.47	0.50	0.54	0.50	0.54	0.44	0.51	0.52
	260	0.53	0.46	0.48	0.53	0.56	0.53	0.50	0.54	0.50
	300	0.52	0.46	0.53	0.48	0.52	0.47	0.46	0.54	0.51