

Datajournalisme & Fact-checking

par

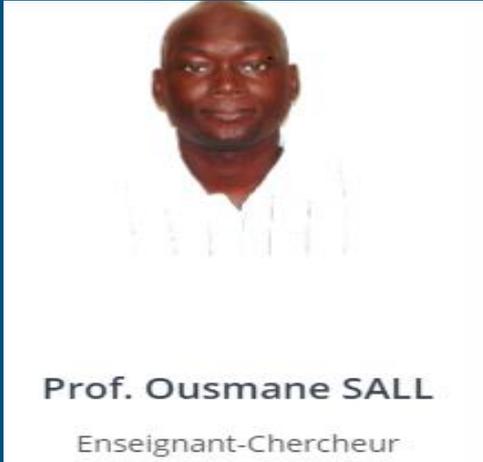
Dr Edouard Ngor SARR & Pr Ousmane SALL

Check4Decision Project

UCAO et Université de Thiès-Sénégal

Journées Scientifiques-LI3 de UASZ (Sénégal)

A propos



UFR SET Université de Thiès
Professeur Titulaire
Juillet 2019.
osall@univ-thies.sn



Labo TIC-SI UCAO-Saint Michel
Maître Assistant CAMES
Juillet 2019
edouard.sarr@ucao.edu.sn

Sommaire

- Introduction
- Du E-journalisme au Data journalisme
- Contexte
- Datajournalisme
- Fact-checking
- Fact-checking automatisé
- Chec4Decision projets
 - Thématiques, Equipe, Axes et Résultats
 - Satisfactions, Difficultés & Perspectives

Introduction

- *Certains prônaient la mort du journalisme avec Internet*
 - Transformations très profondes dans les activités, dans le fonctionnement et dans les habitudes
 - Organisation de l'information dans toutes ses phases : depuis la sélection de l'information, sa transformation en actualité jusqu'à sa publication.
- *Multiplier ses formes d'existence*
- **Ces mutations ne sont pas sans conséquences.**
 - La qualité de l'information produite
 - La prolifération de fausses informations dans la presse
 - Un nouveau genre journalistique : le **fact-checking** ou « la vérification des faits ».

Du E-journalisme au Data journalisme

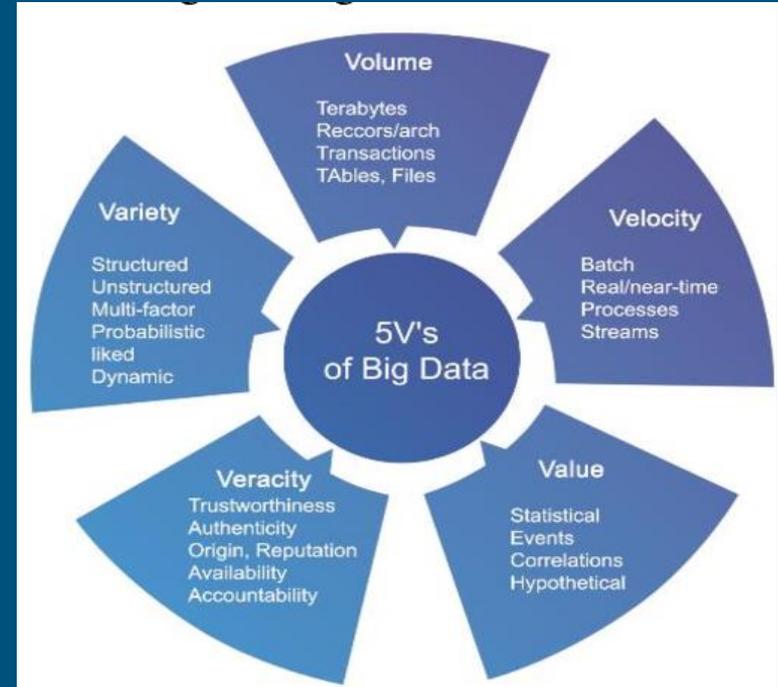
Le journalisme : Recueillir, vérifier, commenter et partager des faits



Prolifération des fausses informations (fakenews)

Contexte : Big Data

- Grosses quantités de données accumulées et traitées à une haute vitesse, issues de sources hétérogènes et dont les actifs exigent une nouvelle approche qualité dans l'utilisation.
- La règle des 5V ou des 5 dimensions : le Volume, la Vitesse, la Variété, la Valeur et la Véracité.



Cause : Open Data

- Définitions

- Une injonction à la mise à disposition des données
- Un processus d'ouverture progressive

- Comment ?

- Publication via des catalogues et des portails dédiés de manière structurée selon une méthode et une licence ouverte garantissant son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière.

- Objectifs ?

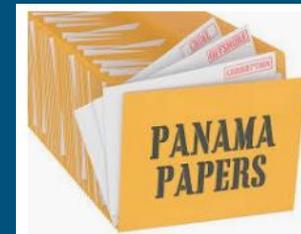
- Faire sortir les données de leur contexte initial de production. Leur offrir un nouveau cadre d'interprétation et de traitement dans de nouveaux contextes



Julian Assange

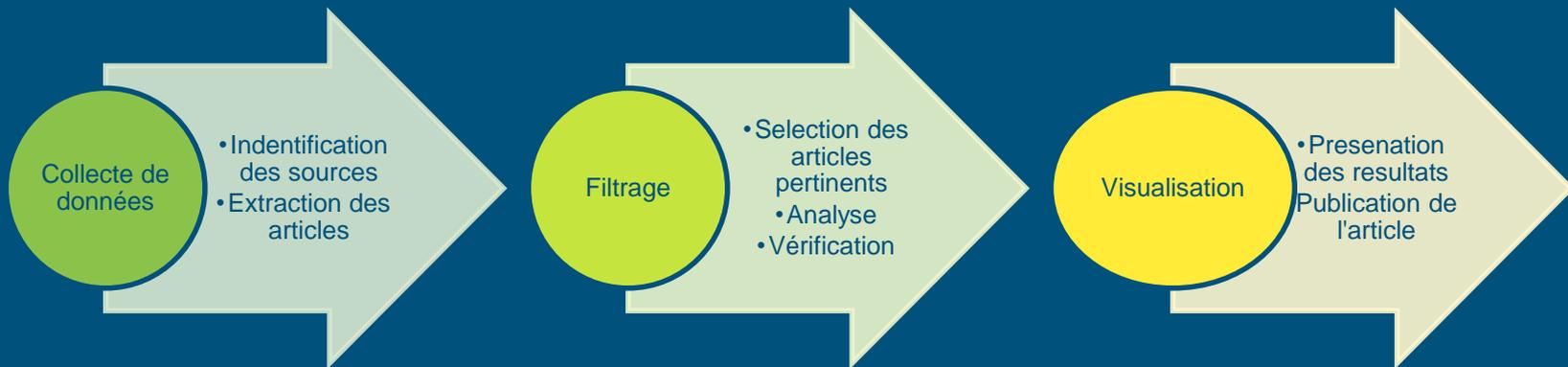


Edward Snowden



Datajournalisme : Définitions

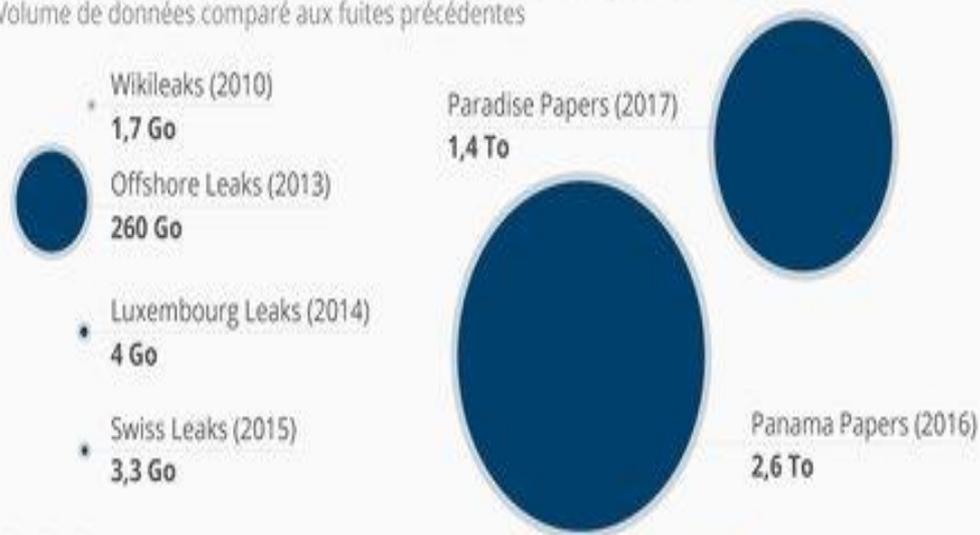
- Données numériques comme corpus
- Fondée sur le recueil et les traitements statistiques et algorithmiques des données
- Témoigne d'un rapprochement du monde de la presse et du monde de la programmation
- Les données constituent pour les data journalistes à la fois une source d'information privilégiée et un outil de la construction d'une narration originale.



Datajournalisme: Plusieurs Compétences

Paradise Papers : une gigantesque fuite de données

Volume de données comparé aux fuites précédentes



© Statista, FR Sources : ICIJ, Sueddeutsche Zeitung

LA TRIBUNE statista

- 400 journalistes
- 13,5 millions de documents
- 1,4 To de données
- 19 paradis fiscaux très opaques
- 350 milliards d'euros retracés dans des paradis fiscaux
- + 12 compétences différentes misent en contributions

Datajournalisme: Data & internautes

- DATA: Atout éditorial puissant pour le travail du journaliste
 - Tous les événements, les conflits, les catastrophes, les mouvements, laissent toujours des traces publiques dans les réseaux, dont l'analyse peut s'avérer très utile.
 - Révèlent les influenceurs, les émotions, les arguments et les communautés en jeu.
- **La dimension participative est aussi fondamentale dans le Datajournalisme**
- Solliciter les internautes pour des contributions, Des veilles
- Alertes sur des données jugées inexactes ou simplement pertinentes.
- Nouvelles branches journalistiques :
 - journalistes programmeurs », « journalistes hackers »
 - **Fact-Checkers: journalistes de vérification des faits**

Fact-checking: Définitions

- Consiste à interroger plusieurs sources de données, analyser les résultats pour déterminer la véracité ou non d'un fait.
 - Un processus d'analyse d'une information en la confrontant avec une ou des connaissance(s) existante(s) dans le but de pouvoir vérifier son exactitude.
 - Identifier la vérité, parmi un ensemble de propositions (faits) potentiellement contradictoires.
- Une forme de journalisme d'investigation
 - 1923 avec le magazine *Time* [Pons et al., 2015]
 - le fact-checking web : 2003 avec le site *FactCheck.org*
 - 2005: *Politifact.com*
 - 2012 : est apparu le crowd-checking [Schäfer et al.,2011]



Fact-checking: Fakenews ou Infox

- A été popularisé en janvier 2018, bien après l'élection américaine
- Choisi comme mot de l'année 2017 par le Collins Dictionary
- Un fakenew ou infaux en français est une information fausse créée de toutes pièces
- Informations fallacieuses, trompeuses ou incorrectes, prétendant être de réelles informations
- Son seul but est de faire passer pour vrai ce qui est faux d'une part et à manipuler ou à déstabiliser dans la perspective de faire réagir.
- Apparaît comme une information vraie dès l'instant qu'elle se présente sous les traits d'un article de presse classique



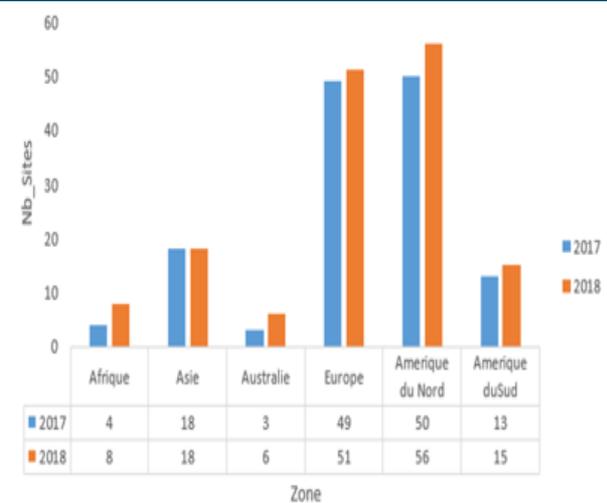
Fact-checking: Fakenews ou Infox

Les fakenews	Définition	Objectifs visés	Cible	Dangerosité
Buzz	Fausse information avec un titre très positif s'appuyant le plus souvent sur des faits existants	COMMERCIAL OU DÉNIGREMENT	Public	MOYENNE
Framing	Un titre mensonger négatif ou non contextualisé d'un article presse existant basé sur des faits réels.	CREER UN CLIMAT HOSTILE	Public	FORTE
Ré-information	Une fausse information non vérifiée publiée à la hâte.	CONCURRENCE	Public	FORTE
Rumeur	Une fausse information non vérifiée qui ne s'appuie sur aucun élément concret avéré	DENIGRER, DIFFAMER, CALOMNIER	Politiques et Célébrités	ELEVE
Canular / hoax	Fausse nouvelle sous la forme d'une blague	HUMOUR	Public	FAIBLE
Complots	Situation ou un événement (souvent politique) qui est expliqué (e) par une conspiration secrète.	DÉSTABILISATION	Etats, Célébrités Entreprises	ELEVE
Fausse citations	Propos parfois inventés et attribués à des célébrités ou sortis de leur contexte	LEGITIMER UNE POSITION	Politiques , philosophes et scientifiques	MOYENNE

Fact-checking: Fakenews ou Infox

- (2003) La guerre en IRAK: Les armes de destruction massive de SADAM
- (2016) Pizzagate au USA : Théorie conspirationniste prétendant qu'il existe un réseau de pédophilie autour de John Podesta, l'ancien directeur de campagne d'Hillary Clinton
- **(2018) Sonko et Tullow Oil** : L'article intitulé "The Challenges of exploiting natural resources in Africa" de "Modern Ghana", a été repris par des journaux et beaucoup d'autres sites sénégalais qui impliquait Ousmane Sonko dans un financement occulte de 195 000 dollars de la société britannique Tullow Oil. Par Michelle Damsen un journaliste qui n'a jamais travaillé pour Modern Ghana
- ...

Fact-checking: Progression



• Site de fact-checking de 2014 à 2018

Site de fact-checking par zone (2017 et 2018)

- En Afrique: AfricaCheck (Senegal, Afrique du Sud, Kenya et Nigeria), BuhariMeter et Dubawa (Nigeria), PesaMeter (Kenya) et ZimFact (Zimbabwe)

Fact-checking manuel



Oui, en 20 ans, l'espérance de vie a augmenté en Afrique plus que partout ailleurs



Affirmation

« En 20 ans, l'espérance de vie en Afrique a augmenté plus que partout ailleurs dans le monde ».

Source: Aminata Touré, Présidente du Conseil, économique, social et environnemental du Sénégal (18 janvier 2020)

Verdict

Les données consultées par Africa Check confirment que, depuis 2000, l'espérance de vie a augmenté en Afrique plus que partout ailleurs.



Non, Air Sénégal n'est pas la seule compagnie africaine à disposer de A330 Neo



Affirmation

A part Air Sénégal, aucune compagnie africaine ne dispose de A330 Neo.

Source: Alioune Sarr, ministre sénégalais du tourisme et des transports aériens.

Verdict

Les informations obtenues auprès d'Airbus montrent que la compagnie aérienne Air Mauritius dispose de deux A330neo.



Automatisation du fact-checking: Processus

Le taux de croissance du Sénégal est de 0.8 en 2018



$f_0(s_0, p_0 : t_0, v_0)$
 $s_0 =$ Taux de croissance du Sénégal
 $p_0 =$ Etre (Présent): 2018
 $v_0 = 0.8$

Source S_0
Fiabilité T_0



Collecte & Agrégation
de données web
Tagging (TALN)

OUTILS & LOGICIELS

FACTCHECKING



- SOURCE w_1 $f_1(s_0, p_0 : t_1, v_1)$
- SOURCE w_2 $f_2(s_0, p_0 : t_2, v_2)$
- SOURCE w_3 $f_3(s_0, p_0 : t_3, v_3)$
- SOURCE w_4 $f_4(s_0, p_0 : t_4, v_4)$
- SOURCE w_5 $f_5(s_0, p_0 : t_5, v_5)$

Automatisation du fact-checking: Projets

- ClaimBuster du NFS aux Etats Unis (<https://idir.uta.edu/claimbuster/>)



ClaimBuster: Automated Live Fact-checking

Data Collection

Press

Acknowledgement



Enter Your Own Text



U.S. Presidential Debates

ClaimBuster API

ClaimPortal

End-to-end Fact-checking (beta)

Tweets by @ClaimBusterTM



ClaimBuster
@ClaimBusterTM



(1/2)From #DemDebate: "I was born and raised in oklahoma, three older brothers retired, and two of them are republicans and sure there are a lot of things we disagree on and we can take to our corners and give them democraticrepublican"



Jan 15, 2020



ClaimBuster
@ClaimBusterTM



(2/2)From #DemDebate: "talking points but there's a whole lot we agree on."



Jan 15, 2020



ClaimBuster
@ClaimBusterTM



Automatisation du fact-checking: Projets

- Décodeurs en France du Journal Le Monde: <https://www.lemonde.fr/verification/>

LES DÉCODEURS

DÉCODEX

Entrez l'adresse (URL) d'une page Web ci-dessous ou le nom d'un site pour savoir si la source de l'information (c'est-à-dire celui ou celle qui la diffuse) est plutôt fiable ou non.

Ex. :



Automatisation du fact-checking: Projets

- les projets européens PHEME et Reveal (<https://revealproject.eu/wp-content/uploads/PHEME-reveal-ws-v2.pdf>)

European Commission > Strategy > Digital Single Market > News >

Digital Single Market

NEWS ARTICLE | 12 August 2015

REVEAL and PHEME: Telling fact from fiction on social networks

Automatisation du fact-checking: Projets

- InVid et CrossCheck lancés par le réseau de fact-checking [First Draft News](#) et financé par Google News Lab. <http://invid.condat.de/>

Info & Help 



InVID Verification Application

please login to analyze new videos

Verify



verify videos from YouTube, Facebook and Twitter

Sample Videos



Automatisation du fact-checking: Projets

- FullFact, <https://fullfact.org>

Full Fact fights bad information.

We're a team of independent fact checkers and campaigners who find, expose and counter the harm it does.

Get the facts on...



People are looking for facts on...

Brexit

NHS privatisation

Viral posts on Facebook

Automatisation du fact-checking: Projets

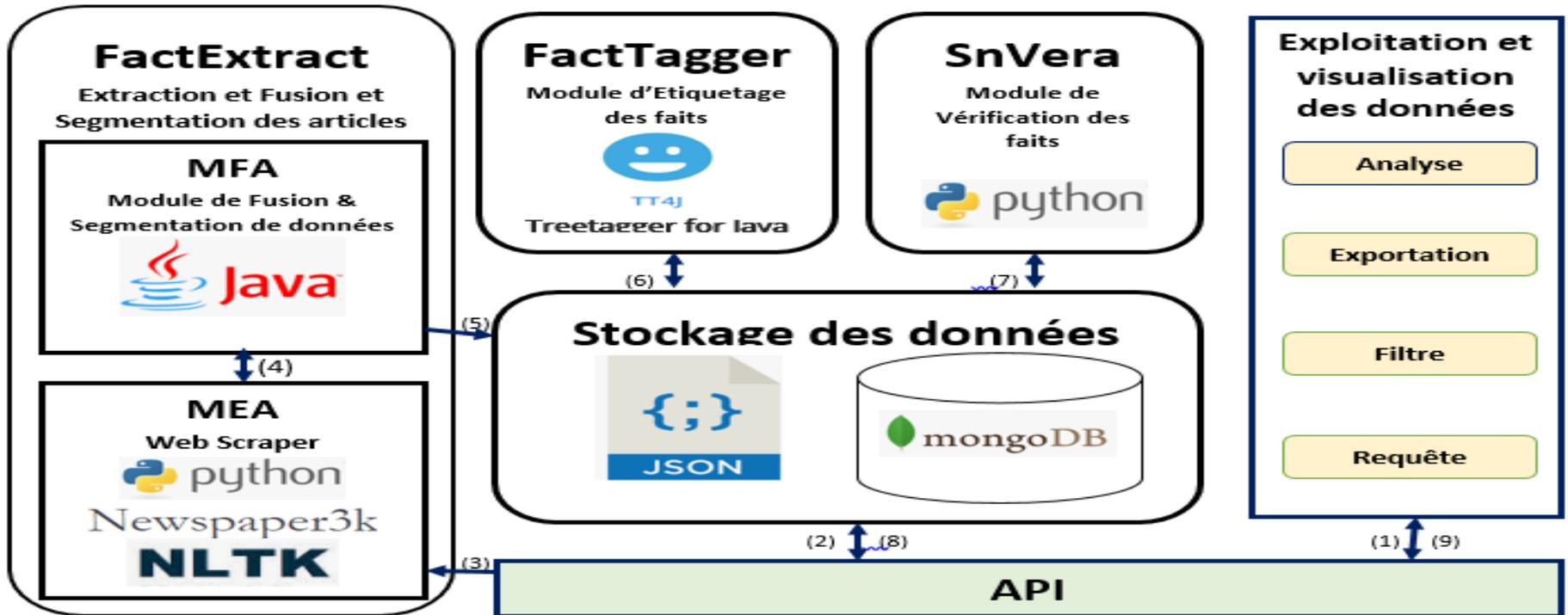
- DAFNA Project au QATAR COMPUTING
 - <http://da.qcri.org/projects/2017/05/15/project-DAFNA.html>
- Truth Teller du Washigtonpost au USA
 - <https://www.washingtonpost.com/news/fact-checker/wp/2014/03/?arc404=true>
- **ContentCheck** du journal Le Monde
 - <http://civic techno.fr/civic-theque/138/contentcheck/>

Check4Decision Projet: Problématique

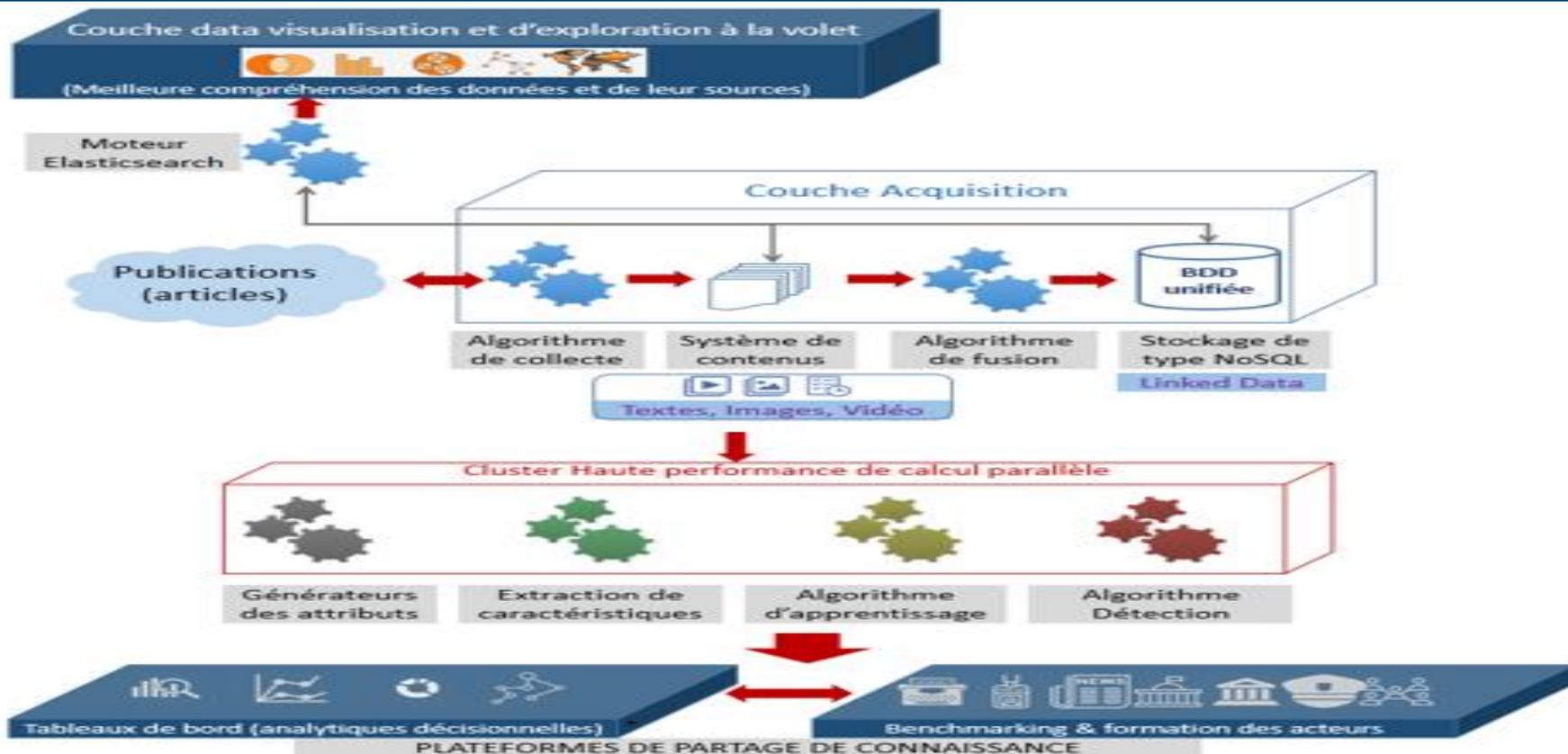
Tout en bénéficiant de l'Open Data et des Big Data, comment utiliser les techniques de IA et de la Data Sciences pour collecter, fusionner, traiter, **vérifier** et diffuser l'information ?

Check4Decision Projet: Architecture 1

Plateforme de vérification des faits



Check4Decision Projet: Architecture 2





Check4Decision

MOTEUR DE RECHERCHE

CROWD-FACT-CHECKING

TABLEAU DE BORD

PUBLICATIONS

A PROPOS

CONTACT

PROJET DE RECHERCHE

Collecte et agrégation de données – Web 3.0 & Traitement Automatique du Langage Naturel

EN SAVOIR PLUS

VERIFICATION
DES FAITS





Check4Decision

MOTEUR DE RECHERCHE

CROWD-FACT-CHECKING

TABLEAU DE BORD

PUBLICATIONS

A PROPOS

CONTACT



NOS PARTENAIRES



DOMAINE D'ACTIVITÉS DU PROJET

Cette thématique, centrée sur le Datajournalisme, combine les technologies du Big Data, l'intelligence artificielle (AI) et le traitement automatique de la langue (TAL). Il couvre les axes suivants



Data
Journalism



Fact
Checking



Algorithme et
Modélisation



Traitement Automatique
du Langage Naturel



Intelligence
Artificielle



Data
Science

PRÉSENTATION DE L'ÉQUIPE



Prof. Ousmane SALL

Enseignant-Chercheur



Dr. Babiga BIRREGAH

Enseignant-Chercheur



Dr. Edouard Ngor SARR

Enseignant-Chercheur



Prof. Mamadou Bousso

Enseignant-Chercheur



LES AXES D'INTERVENTION

Dans le cadre du projet de recherche « Check4Decision » financé par le CEA-MITIC



Modélisation des données pour la visualisation en grande dimension des faits journalistiques

Il s'agit dans cet axe de mettre en place un module de visualisation basé sur le concept de « linked data » pour permettre de mettre en évidence les relations intrinsèques entre les différentes parties des articles.



Conception d'un algorithme de collecte large-spectre d'articles

Il sera question de mettre en place un algorithme paramétrable et multi-cible pour la collecte d'articles



Conception d'un tableau de bord pour l'exploration des faits

Le but ici est de mettre en place une base de connaissances pour l'exploration des données collectées.



Conception d'un algorithme de classification semi-supervisée des faits journalistiques

Dans cette partie il est demandé d'explorer les possibilités qu'offre l'Intelligence Artificielle pour construire un algorithme de classification



Détection de mirco-expressions dans une vidéo

Cet axe aussi sert de prétexte pour constituer une base de données d'expressions faciales sur des personnes d'origine africaine et entrainer un réseaux de neurones à convolution ou à défaut proposer selon la bibliographie une approche d'apprentissage avec une meilleure performance.



Prédiction de fact checking basé sur l'analyse bayésienne

Il consiste à compléter le travail de détection de fact-checking déjà entamé dans la thèse d'Edouard et de proposer un modèle bayésien de détection de fact-checking.

Nos Résultats

- Modelisation Mathématique des articles, des faits, des réseaux d'articles et des réseaux e faits journalistique (Koulibaly & All)
- Modelisation mathématique des commentaires journalistique (Faty & All)
- Opinion Mining sur les commentaires (Faty & all)
- Scrapeur Python d'articles de presse (Sarr & All)
- Scrapeur JAVA d'articles et de commentaires journalistique de presse (Sall & All)
- Categorisateur d'articles de presse basé sur l'apprentissage supervisée (Sarr & All)
- **Métamoteur de recherche d'actualités sénégalaises (Seny & All)**
- **Tableau de bord pour l'exploitation et l'analyse (Fatima & All)**
- **Chatbot journalistique (Tony & All)**
- **Détection automatique des entités nommées dans la presse sénégalaise (Sarr & all)**
- **Tagger de fait (Sarr & All)**

Métamoteur de recherche: Agrégateur



Ousmane SONKO

RECHERCHER

Télécharger le resultat (json)

Tous

Politique

Sport

Société

Actualité

People

International

Video

Societe

9096 article(s) trouvé(s)



seneweb - 2019-

12-

07T04:12:12+00:00

8313 48

Ousmane Sonko en



seneweb - 2019-

12-

06T19:12:16+00:00

2742 19

Ousmane Sonko en



seneweb - 2020-

01-

10T16:01:31+00:00

3082 7

Ousmane Sonko



seneweb - 2020-

01-

11T06:01:39+00:00

9274 20

Ousmane Sonko

Tableau de bord: Analyse et la visualisation



Check4Decision

MOTEUR DE RECHERCHE

CROWD-FACT-CHECKING

TABLEAU DE BORD

PUBLICATIONS

A PROPOS

CONTACT

183,089

Nombre total d'articles collectés

Source de l'article

Select...

Categorie de l'article

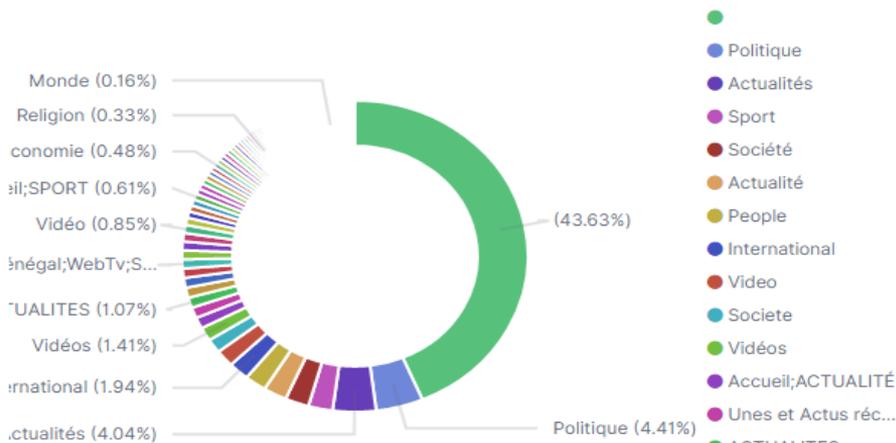
Select...

Apply changes

Cancel changes

Clear form

[C4D] Articles par thématique



[C4D] Top Catégories



Publications Scientifiques

- **Sarr, E. N.**, & Sall, O. (2017, November). Automation of Fact-Checking: State of the Art, Obstacles and Perspectives. In Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence & Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2017 IEEE 15th Intl (pp. 1314-1317). IEEE.
- **Sarr, E. N.**, Sall, O., & Diagne, A. (2017, July). SenFact algorithm: Fact-checking by the confrontation of opinions. In 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 2235-2241). IEEE.
- **Edouard Ngor SARR**, Ousmane SALL, Mouhamadou Saliou DIALLO & Aminata DIALLO (2018). JFact : un générateur automatique de jeux de données synthétiques réalistes pour les algorithmes de fact-checking à fort conflit d'opinions. Conférence sur la Recherche en Informatique et ses Applications (CNRIA-2018), Avril 2018.
- **Edouard Ngor SARR**, Ousmane SALL et Aminata DIALLO (2018). FactExtract : Un extracteur automatique d'articles et de faits journalistiques à partir de la presse en ligne sénégalais. Conférence sur la Recherche en Informatique et ses Applications (CNRIA-JC 2018), Avril 2018.(IEEE BigData 2019)

Publications Scientifiques

- **E. N. SARR**, O. SALL and A. DIALLO, "FactExtract: Automatic Collection and Aggregation of Articles and Journalistic Factual Claims from Online Newspaper," *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, Valencia, Spain, 2018, pp. 336-341.
- **E. N. Sarr**, O. Sall and A. Diallo, "SnVera: A New Algorithm for Automation of Fact-Checking in Web Journalism Context," *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, Valencia, Spain, 2018, pp. 342-348.
- **E. N. Sarr**, O. Sall, A. Maiga, L. Faty and R. M. Marone, "Automatic Segmentation and tagging of facts in French for automated fact-checking," *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 5439-5441.
- Coulibaly, A., Gaye, K., Diagne, S., & **Sarr, E. N.** (2018). New Failure Detection Approach for Real Time for Hydraulic Networks Using the Non-acoustic Method. Information Systems and Technologies to Support Learning: Proceedings of EMENA-ISTL 2018, 111, 233.
- **Sarr, E. N.**, Sall, O., Maiga, A., & Diallo, M. S. (2019). Towards an automation of the fact-checking in the journalistic web context. *International Journal of Big Data Intelligence*, 6(3-4), 307-321.

Publications Scientifiques

- Rabiyatou DIOUF, **Edouard Ngor SARR**, Ousmane SALL, Mamadou BOUSSO, Babiga BIRREGAH, and Sény Ndiaye MBAYE (2019), "Web Scraping: State-of-the-Art and Areas of Application", (accepté pour publication) 2019 IEEE International Conference on Big Data (IEEE BigData 2019)
- **Edouard Ngor SARR**, Ousmane SALL, Mamadou BOUSSO, Rabiyatou DIOUF, Babiga BIRREGAH, and Sény Ndiaye MBAYE (2019), " Automatic Categorization of Press Articles through Learning: The Case of Senegalese Online Press ", (accepté pour publication) 2019 IEEE International Conference on Big Data (IEEE BigData 2019)

06 Articles sont soumis

01 article en cours de traduction

03 Articles en cours de rédaction

04 articles en perspective

Nos Satisfactions

08 Stagiaires (02 doctorants) venant de 05 Universités

Partenariat Public (UT) Privé (UCAO)

Deux pays (SENEGAL & FRANCE)

Plusieurs compétences réunies

Perspectives de thèse

- Tagging (Identification et étiquetage) des faits journalistiques
 - TALN
 - IA
 - Fact-checking

Références

1. <https://www.lemonde.fr/les-decodeurs/>
2. <https://check4decision.univ-thies.sn/>
3. Cardon, D. (2015). *A quoi rêvent les algorithmes. Nos vies à l'heure: Nos vies à l'heure des big data.* Le Seuil.
4. Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice*, 12(1), 4-16.
5. Chohan, U. W. (2016). The Panama Papers and Tax Morality. Available at SSRN 2759418.
6. https://www.etsy.com/market/paradice_paper
7. <https://wikileaks.org/>
8. <https://www.letemps.ch/economie/edward-snowden-ne-cessez-lutter-liberte>
9. <http://lemedialabo.francophonie.org/le-nombre-de-sites-de-fact-checking-a-triple-ces-dernieres-annees/>

CEA MITIC



Datajournalisme & Fact-checking

par

Dr Edouard Ngor SARR & Pr Ousmane SALL

Check4Decision Project

UCAO et Université de Thiès-Sénégal

Journées Scientifiques-LI3 de UASZ (Sénégal)