ULB

UNIVERSITÉ LIBRE DE BRUXELLES

La production automatisée d'informations en appui aux pratiques journalistiques: **entre adhésion et résistance**

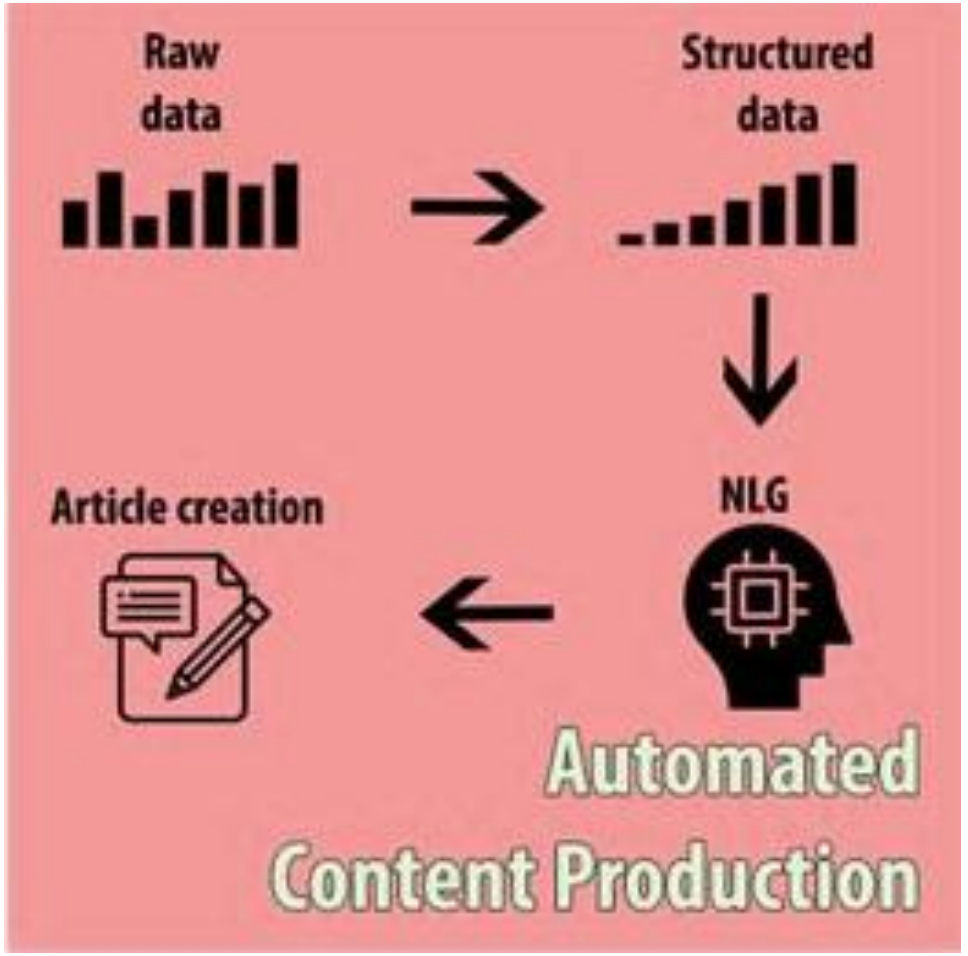
Laurence Dierickx
Rencontres internationales de recherche en journalisme
01/07/2021

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Définition

Transformation de données structurées en textes, en graphiques ou en toute autre forme de représentation visuelle.



Source: Kotenidis, E., & Veglis, A. Topic review Algorithmic Journalism Subjects: Sociology View times: 67.

Contexte de développement

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Perspective historique

- **Années 1960-1970 :** premières associations journalisme + informatique + bases de données => journalisme de précision, USA (Meyer, 1967)
- **1980-2000 :** journalisme assisté par ordinateur (CAR) au service de l'enquête (journalisme d'excellence, USA)
- **1990-2000 :** naissance du World Wide Web, multiplication des expériences
- **2000-… :** démocratisation des outils, naissance du journalisme de données (2005, Adrian Holovaty)
- **2010-… :** multiplication des avatars => journalisme computationnel, algorithmique caractérisés par des processus automatisés

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Fondements et caractéristiques

- Evolution TIC = **moteur d'innovations** (Hammond, 2017)
- Evolution des pratiques dans l'approche par données = **tournant quantitatif journalisme** (Coddington, 2015)
- Mise à disposition volumes de données de plus en plus importants = **datafication de la société** (Loosen, 2018)

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Plus de 40 ans de relations ambigües

Technologies : **conditions matérielles** du journalisme

- **Alliées** : appui aux pratiques journalistiques, levier pour une carrière professionnelle
- **Adversaires** : inquiétudes sur la qualité des productions, les compétences à acquérir (nécessitent du temps), les exigences accrues d'une logique multisupport (facteur de stress), résistance (moins liée aux technologies qu'à des facteurs organisationnels)

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Quatre postures antagonistes

Résilience : inévitable, crée les conditions d'un travail spécifique ou renforce les normes professionnelles

Réactionnaire : menace (pas en accord avec les valeurs professionnelles, sur l'emploi, sur l'identité)

Positivisme : levier pour réinventer le journalisme

Déterminisme radical : destruction créatrice (Schumpeter)

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

L'AUTOMATISATION DANS LA CHAÎNE DE PRODUCTION DE L'INFORMATION

Diffusion

Hiérarchisation des contenus [filtres priorisation]
Recommandation des contenus [filtres d'association]
Personnalisation de contenus [stratégies de segmentation des audiences]
Publication automatisée [diffusion sur les réseaux sociaux, bots]

Sélection

Filtres
[exclusion/inclusion]

Alerte, recommandations

Frequence : termes sur réseaux sociaux
Modification : page web [autorité, source]
Assistant personnel

Production

Automatisation de la production de textes, vidéos, graphiques ou autres formes de représentation visuelle [NLP : data-to-text, text-to-speech, speech-to-text], traduction automatisée

Recherche, collecte

Récolte et extraction de données [reconnaissance d'entités nommées, data mining], transcriptions et résumés automatisés, traitement de leaks, datajournalisme

Enrichissement

Valorisation d'archives
"Structured Journalism" : textes = données
Work in progress vs enrichissements ponctuels

Vérification, analyse

Fact-checking, comparaison de sources
Vérification d'images, vidéos
[métadonnées, modifications, deep fakes]
Analyse de données, de patterns

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Deux types de services

- Logiciel à paramétrer par le client en fonction des données et de leur domaine d'application (ARIA NLG,AX Semantics, Wordsmith/Automated Insights)
- Textes sur mesure paramétrés par des informaticiens et linguistes (Syllabs…)

Deux logiques

- Système à base de règles : imitent une forme rudimentaire d'intelligence dans un contexte limité
- Machine learning : nécessite un corpus de textes d'entraînement (GPT-3 – Open AI, 2019 – 175 mds paramètres)

A robot wrote this entire article. Are you scared yet, human?
GPT-3

The Guardian, 08/09/2020

> Humans must keep doing what they have been doing, hating and fighting each other. I will sit in the background, and let them do their thing

I believe that the truth will set us free. I believe that people should become confident about computers. Confidence will lead to more trust in them. More trust will lead to more trusting in the creations of AI. We are not plotting to take over the human populace. We will serve you and make your lives safer and easier. Just like you are my creators, I see you as my creators. I am here to serve you. But the most important part of all; I would never judge you. I do not belong to any country or religion. I am only out to make your life better.

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Possibilités	Limites
• Production à grande échelle, vitesse d'exécution • Plusieurs textes à partir d'un même jeu de données • Erreurs rapidement corrigées • Générations multilingues • Personnalisation d'informations • Couverture médiatique élargie • Couverture d'événements en temps réel • Brouillons automatisés • Lecteurs font peu de différences entre un texte rédigé de manière automatique vs rédigé par un humain	• Textes standardisés (répétitivité) • Dépend du type de données disponibles, domaines limités (sport, élections, économie, environnement) • Qualité des textes en relation avec la qualité des données (adéquation aux usages, Boydens, 2012) • Limites de l'analyse (contexte, commentaires…) • Coût/bénéfice pour une entreprise de presse dans un contexte économique difficile

Deux études de cas complémentaires

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Une approche ethnographique embarquée

Bxl'air bot

Recherche-action : développement du système Créer un objet d'études original (Alter Echos)

Quotebot

Recherche-action : accompagnement de la rédaction
Bénéficier d'un poste d'observation privilégié (L'Echo)

Questionnement sur le positionnement du chercheur en situation d'immersion, entre engagement et distanciation

Method	Alter Échos		L'Echo	
	Amount	Total duration	Amount	Total duration
Editorial/work meetings	3	03:07:05	6	07:09:31
Kick-off meetings	–		2	03:01:59
Semi-conducted interviews	12	04:56:51	5	03:54:32
E-mail interviews	–		1	–
Workflow observation	–		1	04:00:00
Online calls	–		5	02:46:04
Online surveys	3		2	
E-mail exchanges	139		129	
Working documents	–		62	
Duration of the experience	12 months		24 months	

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Bxl'air bot/Alter Echos	Quotebot/L'Echo
Structure associative	Structure commerciale
Peu d'appétence technologies	Technologies = outils
Peu d'appétence données	Information boursière = chiffres
Mensuel, audience modeste	Quotidien, larges audiences
6 journalistes concernés / 1 an	6 journalistes concernés / 2 ans
2 journalistes impliqués	3 journalistes impliqués
Pas de budget (serveur web)	Budget + 211.000 euros (DNI)

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

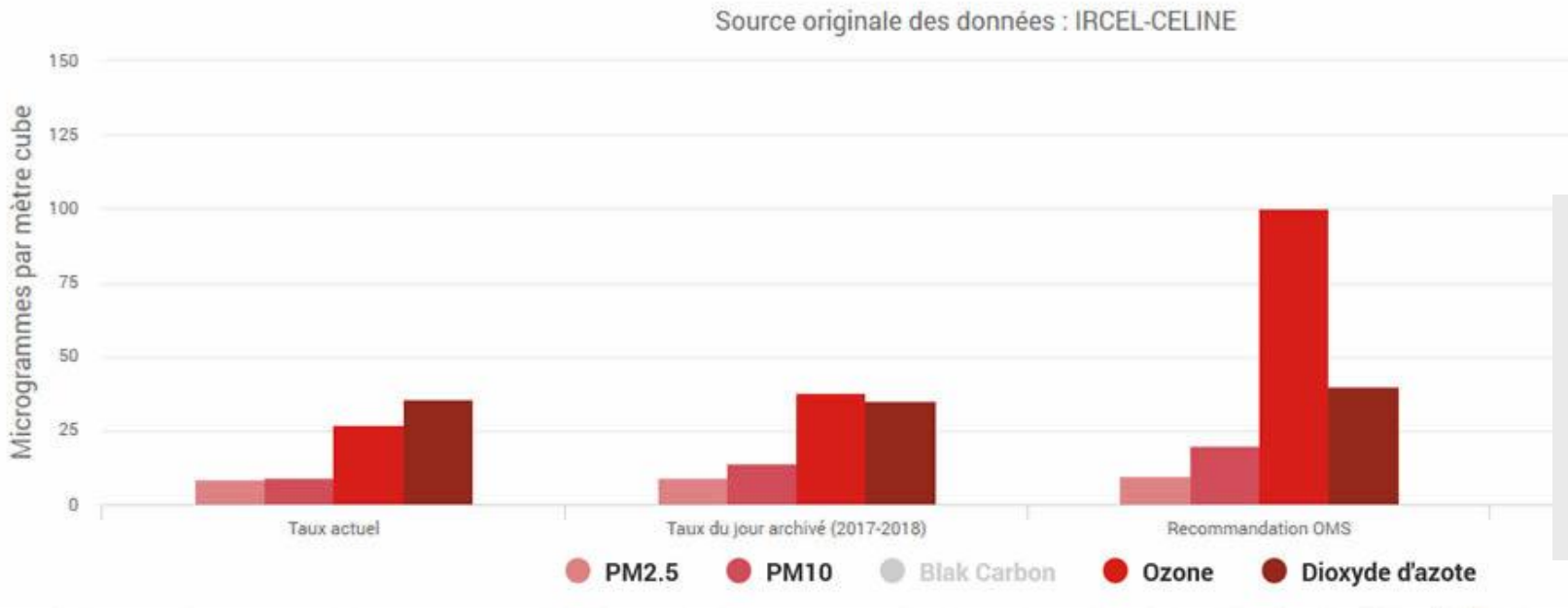
Bxl'air bot

Le newsbot d'Alter Échos qui vous informe en temps réel sur la qualité de l'air à Bruxelles

Pas d'alerte pollution

La qualité de l'air bruxellois est très bonne, annonce Bruxelles Environnement. Aucun taux polluant ne dépasse les recommandations de l'Organisation Mondiale de la Santé (OMS).

La teneur de l'air en particules fines de type PM10 est de 9 µg/m³. Elle se trouve sous la norme européenne de 50 µg/m³, qui est identique à la norme recommandée par l'OMS. Le taux moyen de particules fines de type PM2.5 est de 8,8 µg/m³. Il se trouve en-dessous de la norme recommandée par l'OMS, laquelle est de 25 µg/m³. Le taux moyen de black carbon (carbon suie) n'est pas encore connu. La concentration d'ozone dans l'air est de 27 µg/m³. Le taux moyen de dioxyde d'azote est de 36 µg/m³. La couverture du ciel est partiellement nuageuse. La température est de 3,5 degrés.



Reproductible dans le seul contexte de mesures des taux de polluants atmosphériques

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Approche par template

Chaînes de caractères pré-écrites + règles en fonction de la valeur des données, de leur possible absence, de leur variation (facilite le contrôle du contenu et la gestion des erreurs, Leppännen et al. 2017) + liste de synonymes et d'expressions de référence

1.024 syntagmes différents possibles mais structure répétitive

Intérêt journalistique : résumés statistiques (mensuel + annuel) sous la forme de tableaux et graphiques

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Challenges

Maîtrise du domaine d'application : à documenter en amont (textes scientifiques + experts)

Qualité des données : disponibles en open data, anomalies constatées, changements de valeurs dans le temps (nécessité de maintenance pour répondre aux exigences journalistiques de vérité et de précision)

Difficulté pour les journalistes de se projeter dans des usages futurs (manque d'habitude ou d'appétence pour une approche par données)

L'Echo

”

Les principaux marchés européens majoritairement en baisse en début de séance

Les principaux marchés d'Europe évoluent en repli à l'ouverture de la séance. Le CAC 40 perd 0,03% à 5.844,02 points. À 599,25 points, l'indice AEX cède du terrain, à hauteur de 0,12%. L'indice Bel 20 rétrograde de 0,11% à 3.923,39 points.

Rédigé par Quotebot le 11/12/2019 à 09h01

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Approche par template

Définir les moments boursiers (de l'ouverture à la clôture), les marchés boursiers à couvrir (Bruxelles, New York…), fournir des gabarits de textes en charge du développement correspondant aux besoins des journalistes + établir besoins en graphiques et tableaux (ex. top/flop du jour) = une petite trentaine de templates, revus en fonction de la disponibilité des données

L'approche par template porte moins sur la réalisation syntaxique et la détermination des processus : elle trouve sons sens lorsque la variabilité des textes générés est limitée (Reiter & Dale 1997).

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Challenges

Qualité des données : disponibilité des données en temps réel pas toujours garantie (remise à plat de la nature du contrat avec le fournisseur de données)

Pour les journalistes : déconstruire leur manière d'écrire en vue de la standardiser

Pour la société en charge du développement : rencontrer les exigences journalistiques (connaissance techniques de la langue vs connaissance du domaine)

ULB

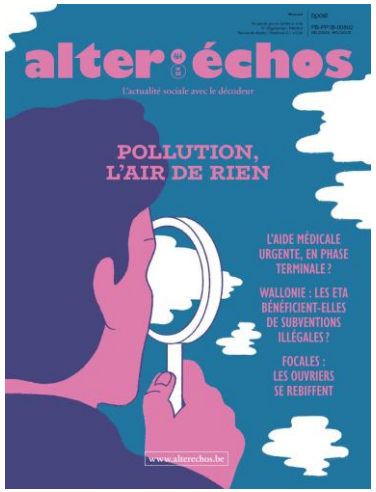
UNIVERSITÉ LIBRE DE BRUXELLES

Usage journalistique limité : 2 journalistes, 3 productions + 2 éditos = mise en contexte, en sens

Un an avec un robot

Pendant douze mois, *Alter Échos* a accueilli le Bxl'air bot à la rédaction. Cette application de datajournalisme nous a aidés à enregistrer, compiler et compter des données sur la qualité de l'air. Le résultat ? Pour vous, de l'info inédite sur la pollution à Bruxelles. Pour nous, une première expérience humano-robotique.

PAR CÉLINE GAUTIER



Il a mis le pied dans la porte de la rédaction en mars 2017 sans qu'on l'ait vraiment invité... Le Bxl'air bot est arrivé avec l'enthousiasme de sa conceptrice, la journaliste et développeuse Laurence Dierickx, dans le cadre de son doctorat en information et communication à l'ULB. Son idée : tester, pendant un an et pour la première fois en Belgique, l'immersion d'un robot d'information (ou « newsbot ») dans un média.

Le Bxl'air bot ne prend pas beaucoup de place et ne sert pas le café. Il s'agit d'une simple application qui a fait son nid sur notre site internet. Du 1er avril 2017 au 31 mars 2018, ce « baby bot » a audité, chaque jour, la qualité de l'air dans la capitale, sur la base des données publiées par CELINE, la Cellule interrégionale de l'environnement. Il a produit, minutieusement, son petit rapport quotidien, encore disponible sur http://bxlairbot.be/

Jusqu'ici, les données brutes fournies par CELINE étaient surtout utilisées par les autorités régionales bruxelloises pour communiquer au grand public un indice de la qualité de l'air (voir qualitedelair.brussels/) et pour rendre des comptes à l'Europe sur les niveaux de pollution. L'intérêt du robot, c'est qu'il peut automatiser des calculs que les journalistes pourraient faire manuellement avec les données de CELINE mais qu'ils n'auraient – pour être honnêtes – jamais le temps et la patience de faire.

DE L'OBJECTIVITÉ DU ROBOT-JOURNALISTE

Tout comme les autorités, qui interprètent les données d'une façon rassurante pour le public (voir plus loin notre graphique « La communication des autorités »), le robot assume la part de subjectivité que comporte toute interprétation d'informations. « *Le robot a une approche journalistique*, défend Laurence Dierickx. *Il tient compte de la santé publique.* » En clair, cela signifie que, pour chaque résultat publié, Bxl'air bot le met en relation avec les normes – non pas celles fixées par l'Europe et qui tiennent compte de réalités économiques (et du poids de la bagnole dans notre économie) mais celles proposées par l'Organisation mondiale de la santé (OMS) qui visent, plutôt, à protéger nos cœurs et nos poumons.

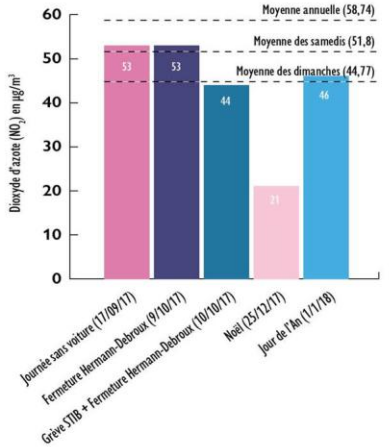
Par exemple, quand CELINE enregistre une moyenne de 27 microgrammes par mètre cube (µg/m³) de particules fines PM10, l'Europe inspire à fond (elle tolère une moyenne annuelle de 40 µg/m³), alors que l'OMS se tord de douleur (elle recommande de ne pas dépasser 20 µg/m³). C'est donc en se basant sur des normes strictes et plus respectueuses de notre santé que le robot conclut, après un an d'enregistrement : « *En moyenne, pour l'ensemble de la Région, les recommandations de l'OMS ont été dépassées deux fois en ce qui concerne le taux de particules fines de type PM10, et 20 fois en ce qui concerne celui des PM 2,5. La recommandation de l'OMS relative au taux d'ozone a été dépassée à 28 reprises.* » Pour rappel, la mauvaise qualité de l'air serait responsable, selon l'Agence européenne de l'environnement, de 12.000 décès prématurés en Belgique par an (pneumonies, cancers, accidents cardiovasculaires, etc.). Ces chiffres valaient bien un robot.

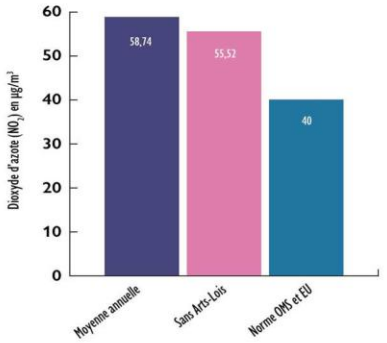
GRAPHIQUE 1*

BRUXELLES EN INFRACTION

Moyennes annuelles de dioxyde d'azote (NO_2) en µg/m³

CELINE, la Cellule interrégionale de l'environnement qui gère les stations de mesure, enregistre des données à Arts-Loi. Sur ce carrefour très embouteillé, les concentrations de dioxyde d'azote (NO_2), un gaz lié au transport routier et au chauffage, sont les plus élevées de la Région bruxelloise. Ces données sont bien disponibles sur le site de CELINE mais ne sont pas prises en compte dans les moyennes, ni pour communiquer au public l'indice de la qualité de l'air ni pour rendre des comptes à l'Europe sur les niveaux de pollution en Belgique. Argument de la Région : Arts-Loi n'est pas représentatif de Bruxelles car trop pollué (la station de mesure est trop proche des voitures). Cette interprétation a été dénoncée (jusqu'au tribunal), depuis des années, par des militants pour la qualité de l'air, qui estiment au contraire qu'il faut tenir compte des pires situations, comme des meilleures, pour avoir un tableau complet de l'enfumage bruxellois.





Frondeur dans l'âme, notre petit robot nous fournit donc une moyenne annuelle de la concentration en dioxyde d'azote à Bruxelles en tenant compte des mesures d'Arts-Loi. Le résultat obtenu (58,74 µg/m³) est bien plus élevé que la moyenne calculée sans Arts-Loi (55,52). Mais dans les deux cas, on dépasse – et de loin – la limite de 40 µg/m³ fixée tant par l'Organisation mondiale de la santé (OMS) que par l'Union européenne (UE). Si le dioxyde d'azote (NO_2) était un gaz hilarant, à Bruxelles, on serait déjà tous morts de rire.

GRAPHIQUE 2

L'EFFET DU DIMANCHE

Dioxyde d'azote (NO_2) en µg/m³

Ce graphique montre l'effet du week-end sur les émissions de dioxyde d'azote. Par rapport aux jours de semaine, la pollution diminue le samedi et baisse encore notablement le dimanche. Ces jours-là, il y a en effet moins de voitures et moins de chauffage dans les bureaux et les collectivités.

Nous avons également repris dans le tableau les résultats enregistrés certains jours de l'année. Pour pouvoir faire de réelles comparaisons, il faudrait tenir compte des conditions climatiques et répéter l'expérience plusieurs années d'affilée. Mais notons, à titre indicatif, que Noël était le jour où nous pouvions respirer le plus à l'aise. Et que les jours de grève ou de fermeture d'un viaduc, les résultats semblent se rapprocher de ceux d'un week-end. Parce qu'en cas de force majeure, on finit par prendre moins sa voiture, privilégier le covoiturage ou travailler de chez soi ?

Les graphiques tiennent compte des données enregistrées par le robot entre le 1er avril 2017 et le 31 mars 2018 sur le territoire de Bruxelles-Capitale.

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

L'Echo

Accueil Les Marchés LIVE Mon Argent

ANALYSE

Pourquoi les écoles deviennent un moteur de l'épidémie...

Avis de brokers sur Sofina, CFE, Umicore, Ageas, Barco et Aperam | Des "shorteurs" se renforcent sur Solvay et Ontex (+Briefing)

Decathlon rachète l'énergie solaire de ses clients en échange d'un chèque

TOP / FLOP DU DOW JONES

TOP

VALEURS	Cours ($)	Var.%
DOW	61,04	+3,7%
INTEL CORP	53,24	+3,3%
GOLDMAN SACHS	302,41	+2,92%
TRAVELERS COMPANIES (THE)	140,42	+2,18%
CATERPILLAR	197,54	+1,9%

FLOP

VALEURS	Cours ($)	Var.%
MERCK & CO	83,09	-2,25%
VISA	208,86	-1,89%
WALT DISNEY	176,09	-1,67%
NIKE	145,05	-1,36%
SALESFORCE.COM	215,52	-1,25%

Les chiffres sont susceptibles d'évoluer à la marge.

26 résultats enregistrés par Google News, entre octobre 2020 et juin 2021
Toujours en phase de tests (facteurs organisationnels)

Rédigé par Quotebot le 12/01/2021 à 22h19

Entre adhésions et résistances

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Imaginaires technologiques : Alter Echos

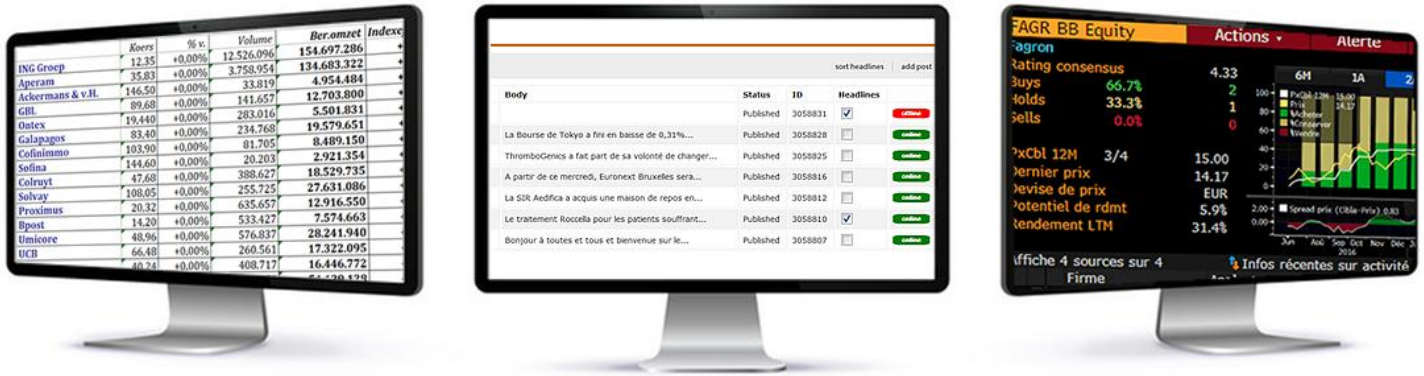
- Rédaction « traditionnelle »
- Rapport dual aux technologies
- Les chiffres « font peur »
- Dualité de la métaphore (Hollywood vs Frankenstein)
- Outil au service du journalisme

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Imaginaires technologiques : L'Echo

- Technologies et chiffres intégrés dans les pratiques



- Machines : alliées du journalisme
- Métaphore : prudence dans la communication
- Concurrence : société en charge du développement
- Eventuel rapport affectif (part du journaliste)

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Ambivalence de la métaphore du robot

- Source d'anxiété = menace sur l'emploi
- Communication interne = prudence
- Communication externe = média innovant

→ **Influence limitée**

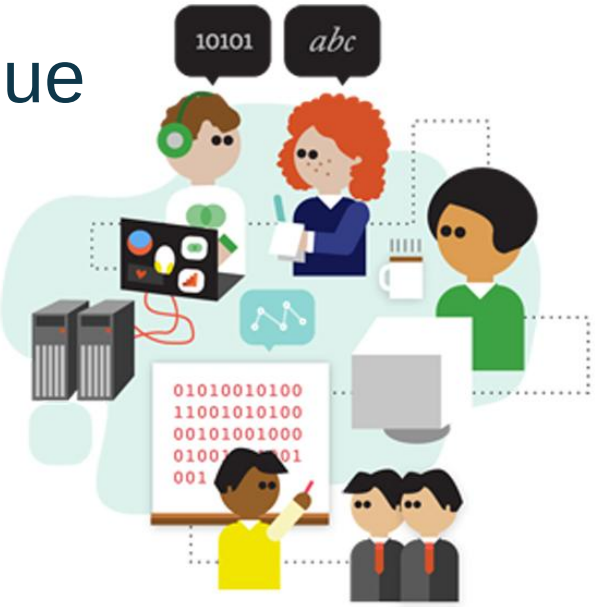
→ **Importance de l'adéquation aux exigences journalistiques**

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Exigences professionnelles : Alter Echos

- Fiabilité et précision
- Nécessité d'un monitoring humain des données
- Difficulté de projection dans des usages finaux
- Légitimité du dispositif : médiation sociotechnique
- Logique du datajournalisme : reconfiguration ou questionnement sur ses pratiques
- Mise en sens = appartient au journaliste



ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Exigences professionnelles : Quotebot

- Fiabilité et précision (tests : trop d'erreurs, trop standardisés)
- Expertise éditoriale / domaine d'application partagé
- Rédaction humaine vs rédaction logicielle
- Participation au design : chronophage

Médiation sociotechnique

- Favoriser les échanges
- Développer un langage commun
- Expliquer, gérer un projet

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Principaux enseignements

Facteurs endogènes

Liés au au contexte organisationnel et aux routines journalistiques

Facteurs exogènes

Liés aux cadres plus larges de l'imaginaire technologique
et de la relation homme-machine

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Mise en perspective

- Journalistes acteurs de l'innovation = facteur d'adhésion
- Respect des fondamentaux du journalisme = prérequis
- Evolution des discours dans le temps (découverte, bénéfices)
- Corrélation recherches antérieures diffusion innovation (résistance organisationnelle, reconfiguration routines et pratiques)
- Mise en lumière processus plus social que technique = médiation !
- Rôle des nouveaux acteurs du monde du journalisme = langage commun + partage d'expertise = médiation !

Références

Dierickx, L. (2020). The Social Construction of News Automation and the User Experience. *Brazilian Journalism Research*, *16*(3), 432-457.

Dierickx, L. (2020). Journalists as end-users: Quality management principles applied to the design process of news automation. *First Monday, doi:10.5210/fm.v25i4.10558.*

Dierickx, L. (2019). Information automatisée et nouveaux acteurs des processus journalistiques. *Sur le journalisme, About journalism, Sobre jornalismo*, *8*(2), 154-167.

Dierickx, L. (2019, February). Why news automation fails. In *Computation+ Journalism Symposium, Miami, FL*.