

Network analysis for information retrieval

Julien Vercin

Master MALIA / MIASHS

2023-2024

Outline

- Motivation
 - ubiquity of information networks
 - applications (in particular to IR)
 - importance of indexing
- Representation of documents
 - sparse representations
 - dense representations
 - topic models
- Network analysis
 - spectral clustering, modularity
 - representation learning for graphs
- Analyzing information networks
 - Graph Neural Networks

(part 1/4)

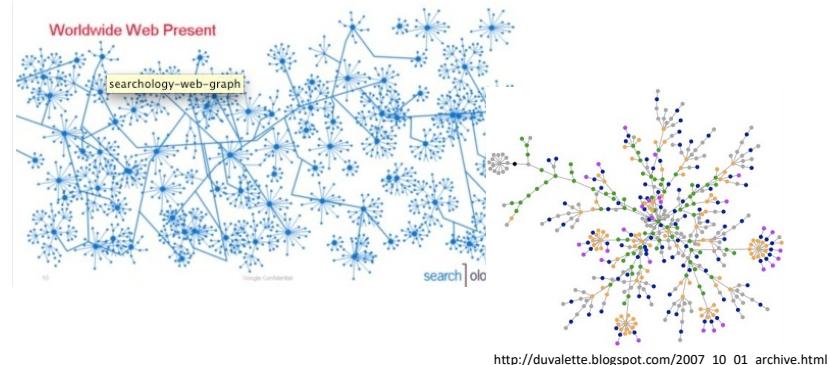
Motivation

Network analysis for information retrieval, M2 MALIA-MIASHS, Julien Vercin

Ubiquity of information networks

- Information networks = networks with information flows
- Information is usually coded as string
- More formally:
 - graph $G = (V, E)$
 - weighted $G = (V, E, W)$
 - Attributes/features associated to E and/or V
- We can model more complex IN (e.g., multiplex networks, heterogeneous networks)

Web graph(s)



Tous Actualités Images Vidéos Livres Plus Outils de recherche

Environ 142 000 000 résultats (0,29 secondes)

Images correspondant à climate change

Signaler des images inappropriées

Plus d'images pour climate change

NASA: Climate Change and Global Warming
climate.nasa.gov Traduire cette page
Vital Signs of the Planet: Global Climate Change and Global Warming. Current news and data streams about global warming and climate change from NASA.
Evidence - Scientific consensus - Causes - Effects

Climate change - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Climate_change Traduire cette page
Climate change is a change in the statistical distribution of weather patterns when that change lasts for an extended period of time (i.e., decades to millions of ...
Global warming - Scientific opinion on climate ... - Effects of global warming on ...

What is Climate Change? What Causes Global Warming?
www.takepart.com/flashcards/what-is-climate-change Traduire cette page
Climate change, also called global warming, refers to the rise in average surface temperatures on Earth. An overwhelming scientific consensus maintains that ...

Home | Climate Change | US EPA
https://www3.epa.gov/climatechange Traduire cette page
Official government site provides comprehensive information on the issue of climate change and global warming including climate change science, U.S. climate ...

6

Blogs, forums, chats

23/04/2009

Avarice du Charles-de-Gaulle : pourquoi c'est compliqué

Depuis les semaines, le Charles-de-Gaulle est à quai, à cause d'un problème mécanique, comme nous le révèlent alors une commission d'enquête, réunissant la Marine nationale et l'industriel DCNS, a été mise en place pour comprendre pourquoi le système de propulsion du porte-avions n'a pas pu être mis en marche. Pour l'instant, les experts n'ont pas de réponse et le ministère de la Défense a reconnu que les travaux pourraient durer "de quatre à six mois". A l'heure d'écouter les détails, les experts ont également évoqué toutes les possibles causes avec ce qu'ils ont de mieux à dire de celles-ci.

L'appareil propulsif d'un bateau de guerre est en effet un système très complexe et il faut faire un long cours de mécanique pour réussir à comprendre pourquoi il est plus difficile de faire fonctionner le Charles-de-Gaulle que sa voisine...

Contrairement à un navire civil, même de grande taille, le système de propulsion d'un bateau de guerre prend en compte des spécificités initiales. Il doit notamment être compact et durable.

Visitor has requested translation from language: French

Stephen says:

Hello Marie. My name is Stephen how can I help you?
(Bonjour Marie. Mon nom est Stephen comment osez je vous aide ?)

Marie On www.whoson.com says:

I need price [J'ai besoin de prix]
Link Sent: http://www.whoson.com/installableorder.aspx
Stephen says:
Ok. For prices please go to http://www.whoson.com/installableorder.aspx
(Ok. Pour des prix veuillez voir à http://www.whoson.com/installableorder.aspx)

THE NEW REDDIT JOURNAL OF SCIENCE

4389 hot new rising controversial top

Humans have triggered the last 16 record-breaking hot years experienced on Earth (up to 2014), with the new research tracing our impact on the global climate as far back as 1937. The findings suggest that without human-induced climate change, recent hot summers and years would not have occurred. + 1999cog 15 hours ago by dreepooole 3720 comments share

Top 200 Comments show 500

sorted by: best (suggested) ▾

[+] old-table 665 points 13 hours ago
So what can we actually do to combat this? Aside from colonizing space and getting humans off this planet?
permalink

[+] XiiCuded 1957 points 13 hours ago
Switch to nuclear energy.
edit: thanks for the gold nuclear energy fw
permalink parent

[+] Mr_Industrial 839 points 13 hours ago
Good luck convincing several million people that nuclear energy is safer than most other forms of energy. It's not about the facts, it's about perception of the facts.
permalink parent

[+] climbreer 828 points 12 hours ago
You don't have to. The public rarely has input into power plant construction etc. Once they're up and running no-one cares about it anymore.
If you ask people if they'd like a change, 90% will say no, 95% if you say it might involve danger. If you make the change and ask how happy people are most are just as happy.
permalink parent

[+] Mr_Industrial 158 points 12 hours ago
This is a good point. The thing you have to remember though is that the people in charge who have the power to decide what type of

sign in  subscribe search

dating more International

the guardian

UK world sport football opinion culture business lifestyle fashion environment tech travel

headlines

Now 4°C Lyon 12:00 15:00 18:00 21:00 9°C 13°C 9°C 6°C Weather of climate emergency after surface temperatures 1.3°C warmer than average temperature for the month

Great Barrier Reef Severe coral bleaching worsens

Climate change / February breaks global temperature records by 'shocking' amount

Worries of climate emergency after surface temperatures 1.3°C warmer than average temperature for the month

Japan US sailor arrested in Okinawa on suspicion of rape

US elections 2016 Clinton and Sanders attack 'pathological liar' Trump

German elections Anti-refugee AfD party makes dramatic gains

Thailand Eight die in bank after chemical fire extinguisher leak

Ivory Coast Gunmen open fire on tourist resort, killing 16

Brazil More than a million protest over 'horror' government

Egypt Justice minister sacked for saying he would arrest prophet Muhammad

Turkey President vows to defeat terror as bomb kills 34 in Ankara

United Arab Emirates Plane reported missing in Yemen

+ More headlines

highlights

100 Best Nonfiction Books of All Time No. 7 - *The Right*

amazon.fr

Toutes nos boutiques

Amazon.fr Ventes Flash Meilleures ventes Offres recommandées Nos idées cadeaux Services Amazon Amazon Assistant

Star Wars : Battlefront - édition limitée > Commentaires client

Commentaires client

3,2 sur 5 étoiles

5 étoiles	17
4 étoiles	14
3 étoiles	7
2 étoiles	8
1 étoile	13

Evaluez cet article Écrire un commentaire

Hidden for obvious reasons

Meilleur commentaire positif

Voir les 31 commentaires positifs

★★★★★ Pas parfait mais un Star Wars

Par Client d'Amazon le 21 décembre 2015

Le titre pourrait être plus riche en terme de contenu, surtout en solo qui fait seulement guise d'introduction aux bases, mais l'immersion est tellement réussie que les fans de l'univers Star Wars seront conquis.

L'ambiance sonore et visuelle est magistrale, et incarner un stormtrooper en pleine bataille d'Endor ou sur Hoth est un réel plaisir !

A éviter si vous ne jouez pas en ligne.

Meilleur commentaire critique

Voir les 28 commentaires critiques

7 sur 7 personnes ont trouvé cela utile

★★★★☆ Déçu

Par julien le 6 décembre 2015

Pas de campagne, juste un multi joueur qui se rattrape par de super graphisme mais sa ne suffit pas... Et bien évidemment le reste sera en DLC ce qui fera grimper le jeu à environ 130€ (édition deluxe) donc pas pour moi...

Accueil Notifications Messages #malavita

#malavita

Top Direct Comptes Photos Vidéos Autres options

Suggestions - Actualiser - Tout afficher

Kaplan International @k... Suivre Sponsored

Aras BOzkurt @arasbozkurt Suivre

Stéphane Pouilly @spouilly... Suivre

Trouver des amis

Tendances - Modifier

#EnVolure Sponsored par Allianz France

#ASSEPG #JacquelineSauvage #Camping #TheVoice #Maaviv Milan Bordeaux Ronaldo Florian Thauvin Benoit Violier

11 nouveaux résultats

Marion @Marion_LaJct 2 min Trop cool ce film de LucBesson #Malavita #TF1

Gabi @gabi_01_01 2 min Et putain... #Malavita

Mouna Camara @mouna_camara 2 min Très bon film #Malavita

Stephanie Lij @SLijouren 2 min Après #Malavita place à #LOLUSA

Black Mamba @Smallikawakey 2 min Bon ce film était bof. Rien d'exceptionnel, des petits passages marrant. Dommage avec un tel casting... #Malavita

Ree @ReeRae 2 min Film d'action américain en normandie 😎 #Malavita

9

Articles scientifiques

Un cadre pour la représentation et l'analyse de débats sur le Web

Anna Stavrianou*, Julien Velcin**, Jean-Hugues Chauchat***
 ERIC Labeyrie - Université Lumière Lyon 2, Université de Lyon,
 5 avenue Pierre Mendès France 69676 Bron Cedex, France
 ** anna.stavrianou@univ-lyon2.fr
 *** julien.velcin@univ-lyon2.fr
 ... jean-hugues.chauchat@univ-lyon2.fr

Résumé. Les débats en ligne sont souvent modélisés par des réseaux sociaux d'utilisateurs représentés sous forme de graphes, chaque noeud correspondant à l'un des participants du débat. Ici, nous proposons une nouvelle modélisation des débats en ligne qui prend en compte la nature des messages échangés et chaque arc entre un message à celui auquel il répond ; ces arcs peuvent être caractérisés par les mots-clés résumant la relation entre les messages connectés. Cette modélisation permet une meilleure représentation des débats en ligne et facilite leur analyse. Nous comparons les deux représentations, graphes représentant les utilisateurs et graphes représentant les messages, puis nous analysons la connaissance qui peut être extraite à partir de chacun d'eux. Nos expériences sur des débats réels valident le modèle proposé et montrent les informations complémentaires qui sont apportées par le graphe des messages.

1 Introduction

Le développement de Web2.0 a provoqué la création d'un grand nombre de blogs, de forums et de sites sociaux. Ces derniers sont utilisés par les organisations publiques pour l'industrie ou le secteur commercial. Les discussions en ligne contiennent les intérêts et les avis des internautes, des critiques de produits, la présentation d'idées politiques, etc. En conséquence analyser ces données devient une problématique statutaire.

The online discussions are often modeled as a social network of users and they are represented by a graph where each node denotes a participant of the discussion. In this paper, we propose a new framework for the representation and analysis of online debates. In this framework, each node corresponds to a message of the discussion and each edge is labeled with the message the specific node replies to. The edges can be weighted by the keywords that characterize the connected messages. This model allows a better representation of the content of the discussion and facilitates its classification or identification of discussion chains. We compare the two representations and based on the modeling of the graph, we show the different information that can be extracted from them. Our experiments with real data validate the proposed framework and show additional information that can be extracted from a message-based graph.

Summary

The online discussions are often modeled as a social network of users and they are represented by a graph where each node denotes a participant of the discussion. In this paper, we propose a new framework for the representation and analysis of online debates. In this framework, each node corresponds to a message of the discussion and each edge is labeled with the message the specific node replies to. The edges can be weighted by the keywords that characterize the connected messages. This model allows a better representation of the content of the discussion and facilitates its classification or identification of discussion chains. We compare the two representations and based on the modeling of the graph, we show the different information that can be extracted from them. Our experiments with real data validate the proposed framework and show additional information that can be extracted from a message-based graph.

Zhang, J., M. Ackerman, and L. Adamic (2007). Expertise networks in online communities: Structure and algorithms. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, 23–32. ACM.

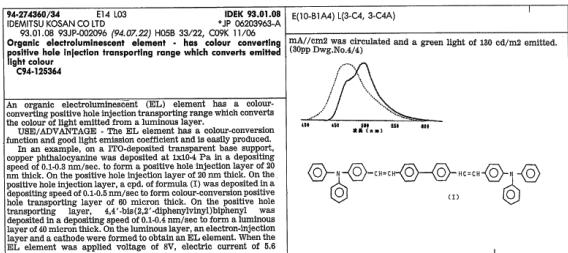
Zhou, D., I. Cohen, P. Zha, and C.-L. Giles (2007). Discovering temporal communities from social network documents. In *ICDM: International Conference on Data Mining*, pp. 745–750. IEEE Computer Society.

11

10

12

Brevets



13

Applications (in particular to IR)

- Search engines
- Personal assistants
- Opinion mining / e-reputation
- Community detection
- Multi-document summarization

Some applications

Network analysis for information retrieval, M2 MALIA-MIASHS, Julien Velcin

Applications

- moteurs de recherche à mots-clefs
- systèmes de Question-Réponse



Watson gagne le Jeopardy! en 2011

Google

text mining

Environs 4 060 000 résultats (15,1 secondes)

Fautes de typages - Wikipedia

Cette page a été générée par notre moteur de recherche sous l'anglicisme text mining. C'est un ensemble de termes interconnectés consistant à extraire des connaissances ...

Wikipedia.org/w/index.php?title=Fautes_de_typer&oldid=5900000

YAHOO!

FRANCE

text mining

Environs 4 060 000 résultats (15,1 secondes)

Fautes de typages - Wikipedia

Cette page a été générée par notre moteur de recherche sous l'anglicisme text mining. C'est un ensemble de termes interconnectés consistant à extraire des connaissances ...

Wikipedia.org/w/index.php?title=Fautes_de_typer&oldid=5900000

Introduction au Text mining - Christian Faust

30 mai 2007 - Les outils de text mining ont pour vocation d'automatiser la structuration des données textuelles et leur analyse. ...

Wikipedia.org/w/index.php?title=Introduction_au_text_mining&oldid=5900000

Les outils de Text Mining - Les critères de choix

19 mai 2007 - Les outils de text mining sont utilisés dans diverses applications : ...

Wikipedia.org/w/index.php?title=Les_outils_de_Text_Mining&oldid=5900000

Solutions de Data Mining et Text Mining - SAS

SAS a été l'un des premiers éditeurs à s'intéresser au text mining, c'est-à-dire à l'analyse de documents et à l'exploration et à l'exploitation de connaissances dans les documents écrits. ...

Wikipedia.org/w/index.php?title=Solutions_de_Data_Mining_et_Text_Mining&oldid=5900000

Zoom_moteur de recherche sémantique et text mining

Un moteur de recherche sémantique est une technique de recherche qui utilise des algorithmes de langage naturel, outil de text mining et de web mining. ...

Wikipedia.org/w/index.php?title=Zoom_moteur_de_recherche_sémantique_et_text_mining&oldid=5900000

Quel est l'âge d'Emmanuel Macron ?

Tous Actualités Vidéos Images Shopping Plus Paramètres Outils

Environs 27800000 résultats (0,59 secondes)

Emmanuel Macron / Âge

41 ans

21 décembre 1977

Emmanuel Macron

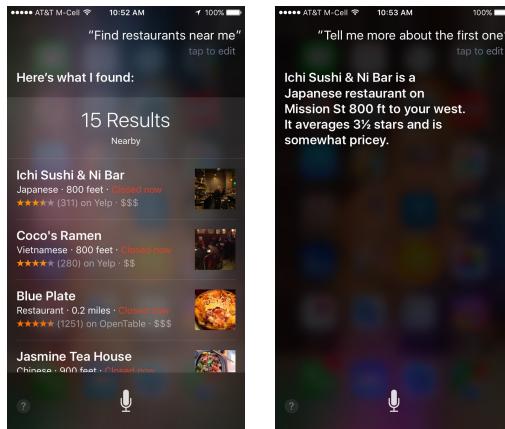
Président de la République française

Emmanuel Macron (né le 21 décembre 1977 à Amiens), est un haut fonctionnaire et homme d'État français. Il est président de la République française depuis le 14 mai 2017. Sorti de l'École nationale d'administration en 2004, il est inspecteur des finances. Wikipédia

Epoque : Brigitte Macron (m. 2007) Tendances

16

Personal assistants



source : cours de D. Jurafsky : <https://web.stanford.edu/~jurafsky/slp3/>

17

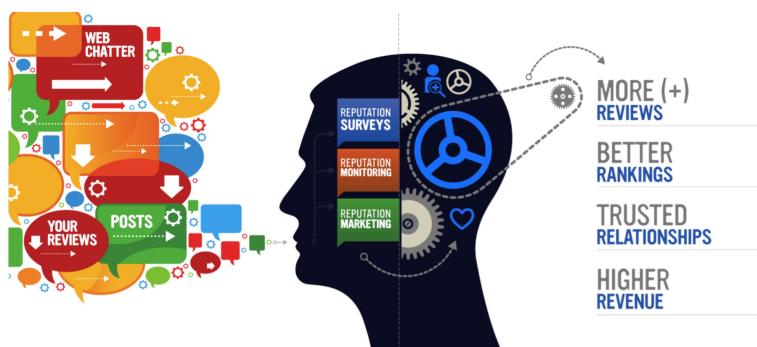
Opinion mining



source : <http://politoscope.org/>

18

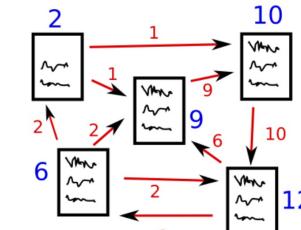
E-reputation management



source : <https://www.mibwebtech.com>

19

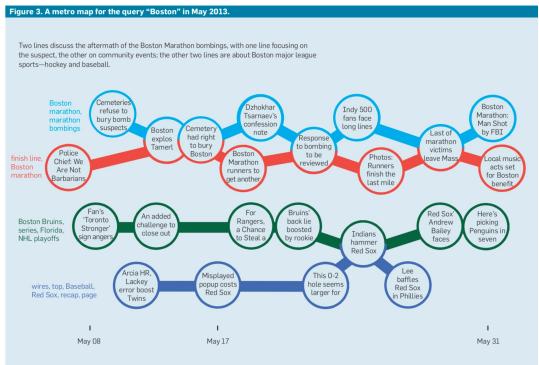
Multi-document summarization



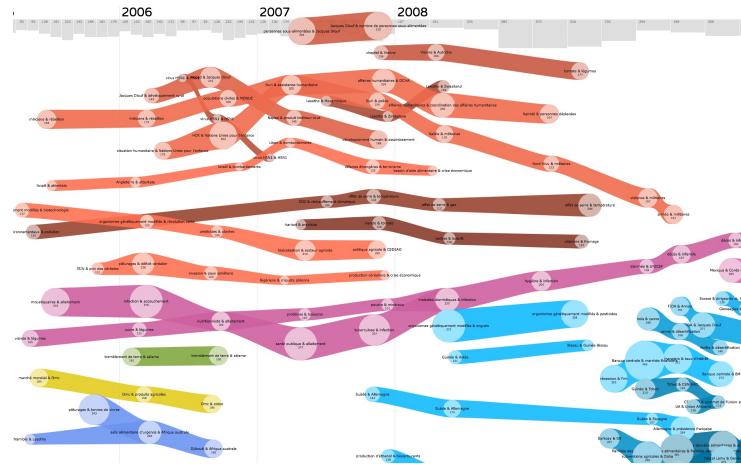
Source: <http://www.scottbot.net/HIAL/>

20

Metromaps (Shahaf et al., 2015)



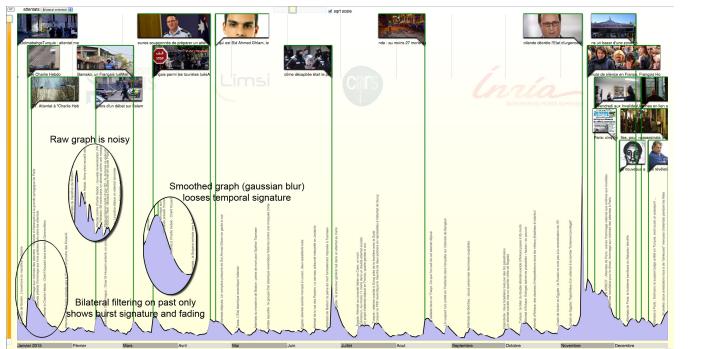
21



Project Pulseweb : <http://pulseweb.context.net>

22

Journalisme de données : exemple du projet Chronolines

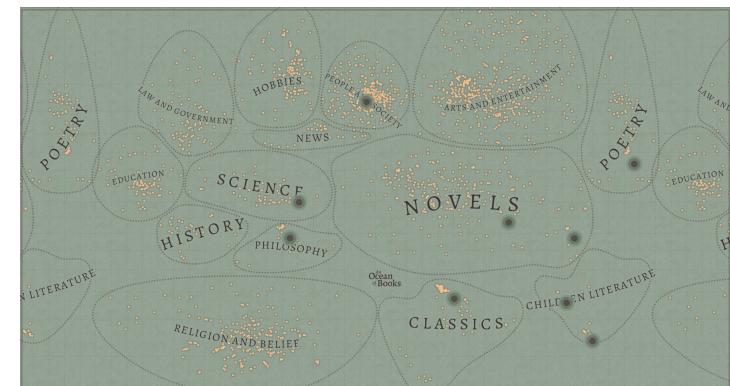


Kiem-Hieu Nguyen, Xavier Tannier, Véronique Moriceau. Ranking Multidocument Event Descriptions for Building Thematic Timelines. In *Proceedings of the 30th International Conference on Computational Linguistics (Coling 14)*. Dublin, Ireland, August 2014.

23

Projet LIFRANUM

- Projet ANR LIFRANUM MARGE, ERIC, BnF



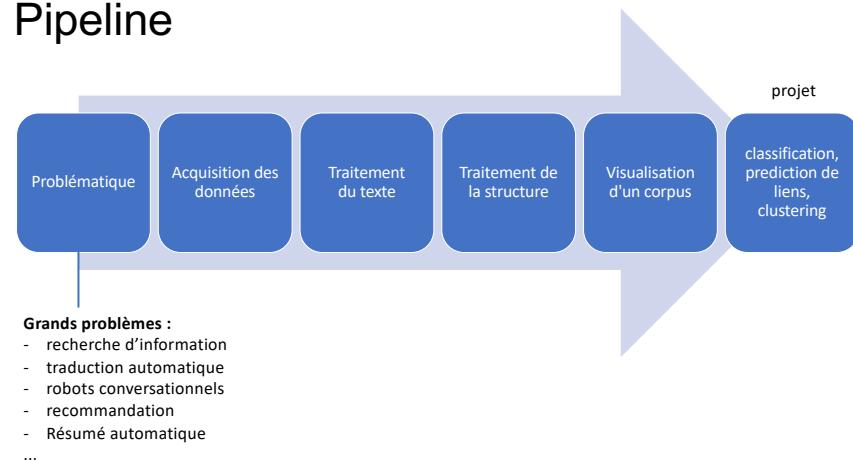
<https://artsexperiments.withgoogle.com/ocean-of-books>

24

Importance of indexing

- What we're looking for?
 - words (and n-grammes)
 - authors
 - sources
 - categories (predefined? discovered? both?)
- An heterogenous perspective : building/using (latent) spaces that link various objects (eg., documents and authors, topics and words, etc.)

Pipeline

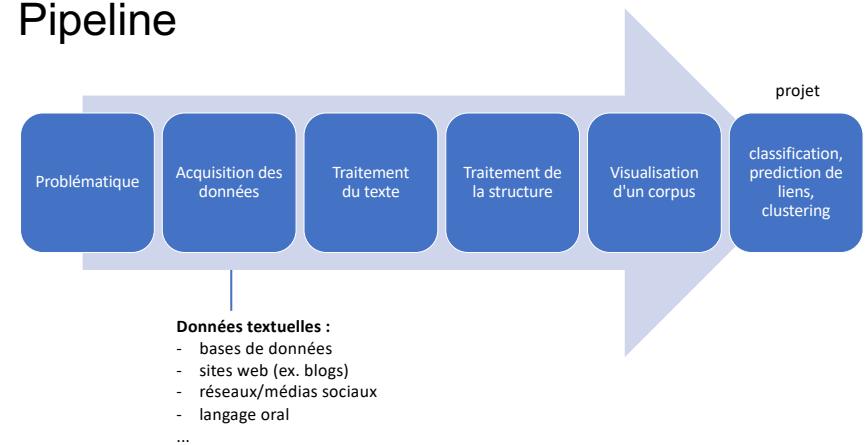


27

Suggested pipeline

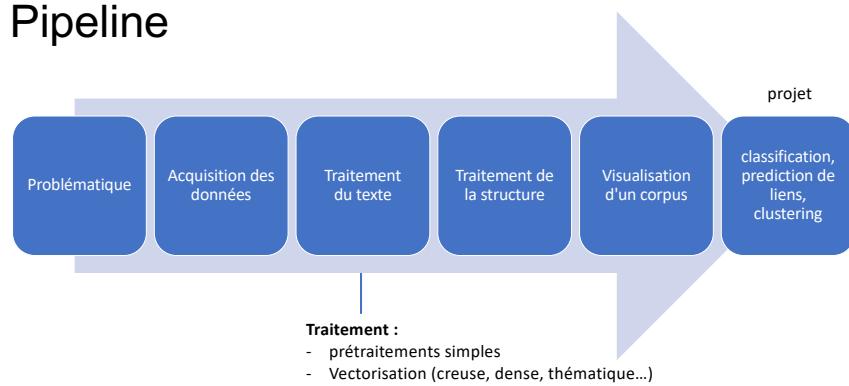
Network analysis for information retrieval, M2 MALIA-MIASHS, Julien Velcin

Pipeline

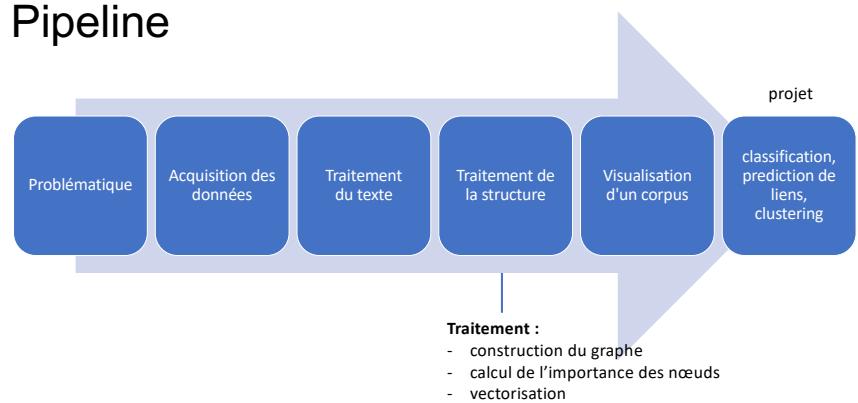


28

Pipeline



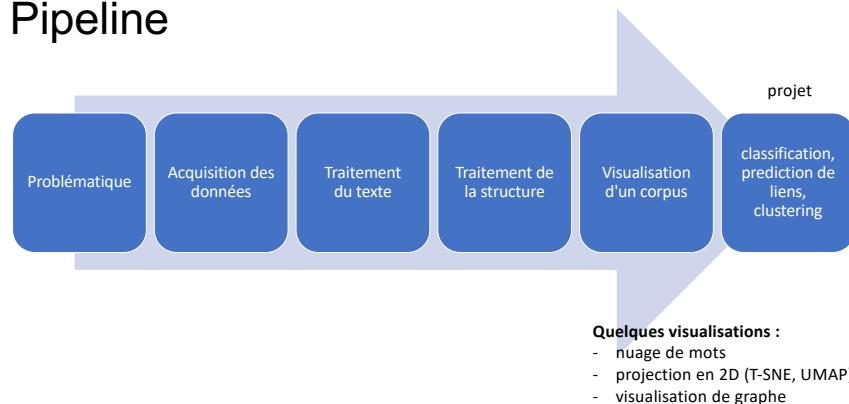
Pipeline



29

30

Pipeline



31