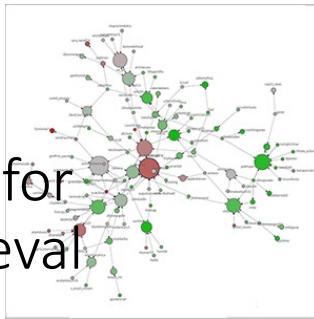


# Network analysis for Information Retrieval

Julien Velcin

Master MALIA / MIASHS

2024-2025



(part 1/4)

1

## 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

2

## 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)

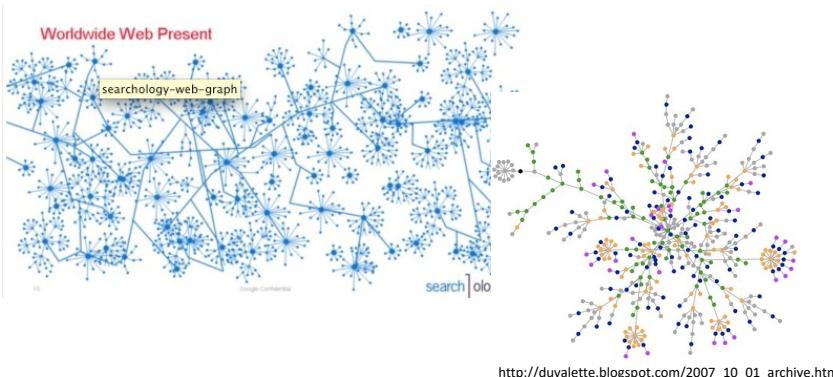
3

4

## Motivation

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

## Web graph(s)



5

Google climate change

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/flashtards/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

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

Depuis six semaines, le Charles-de-Gaulle est à quai, à cause d'un problème mécanique, comme nous le révélions alors. Une commission d'enquête, réunissant la Marine nationale et l'industrie DCNS, a été mise en place pour comprendre ce qui s'est passé et proposer des mesures correctives pour prévenir l'avenir. Les experts n'ont pas de réponse et le ministre de la Défense a reconnu que les travaux pourraient durer "de quatre à six mois". A l'heure actuelle, les deux navires de la classe Mistral sont également toutes les deux en panne, avec ce qu'ils appellent un "arrêt de cause".

L'appareil propulsif d'un bateau de guerre est en effet un système très précis et complexe. Il faut donc faire beaucoup de recherche pour mieux comprendre pourquoi il est plus difficile de dépanner un bateau de guerre que sa voiture.

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

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 ose je vous aide ?)

Marie On www.whoson.com says:

I need price [J'ai besoin de prix]

Stephen says:

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 vont à <http://www.whoson.com/installableorder.aspx>)

7

THE NEW REDDIT JOURNAL OF SCIENCE

hot new rising controversial top filter by field -

4389 15 hours ago by drewpwope 3720 comments share

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. ▶ phys.org

Top 200 Comments show 500 sorted by: best (suggested) ▾

[+] XIIcubed 1997 points 13 hours ago\* ⓘ

[+] old-tote 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 parent

[+] XIIcubed 1997 points 13 hours ago\* ⓘ

[+] Switch to nuclear energy.

edit: thanks for the gold nuclear energy ftw

permalink parent

[+] Mr\_Industrial 939 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

[+] climbrina 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

8

sign in subscribe search

dating more International - theguardian

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

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

Warnings of climate emergency after surface temperatures 1.95°C warmer than average temperature for the month

Great Barrier Reef Severe coral bleaching worsens

1,716

Japan US sailor arrested in Okinawa on suspicion of rape

1,489

German elections Anti-refugee AfD party makes dramatic gains

463

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

303

Brazil More than a million protest over 'horror' government

Thailand Eight die in bank after chemical fire extinguisher leak

Ivory Coast Gunmen open fire on tourist resort, killing 16

Egypt Justice minister sacked for saying he would arrest prophet Muhammad

3,2 sur 5 étoiles

5 étoiles 17

4 étoiles 14

3 étoiles 7

2 étoiles 8

1 étoile 13

Commentaires client

Hidden for obvious reasons

Evaluer cet article Écrire un commentaire

highlights

100 Best Nonfiction Books of All Time

No. 7: *The Right*

9

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

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.

Meilleur commentaire critique

Voir les 28 commentaires critiques

★★★★★ Déçu

Par julien [REDACTED] le 6 décembre 2015

Pas de campagne, juste un multi joueur qui se rattrape par de super graphisme mais ça 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...

10

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

Stéphanie Pouyflau @spouyl... Suivre Trouver des amis

Tendances · Modifier

#Envolure Sponsorisé par Allianz France #ASSEPG #JacquelineSavage #Camping #TheVoice #Malavita Milan Bordeaux Ronaldo Florian Thauvin Benoît Violier

11 nouveaux résultats

Marion @Marion\_Lec... 2 min Trop cool ce film de #LucBesson #Malavita #TF1

Gabi @gabi\_01 2 min Et putain... #Malavita

Mouna Camara @mouna\_camara 2 min Très bon film #Malavita

Stephanie L1 @SLidouren 2 min Après #Malavita place à #LOLUSA

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

Ree @HirRee 2 min Film d'action américain en normandie 😊 #Malavita

11

## Articles scientifiques

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

Anna Stavrianou\*, Julien Velcin\*\*, Jean-Hugues Chauvat\*\*  
 ERIC Laboratoire - Université Lumière Lyon 2, Université de Lyon,  
 5 avenue Pierre Mendès France 69676 Villeurbanne Cedex, France  
 \*anna.stavrianou@univ-lyon2.fr  
 \*\*jean-hugues.chauvat@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 utilisateurs du débat. Ici, nous proposons un nouveau modèle basé sur un graphe de messages : chaque noeud du graphe correspond à un des messages échangés et chaque arête relie 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 et facilite l'identification des idées principales et des thèmes. 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 du Web2.0 a provoqué la création d'un grand nombre de blogs, de forums et de discussions en ligne. L'analyse de ce type de discussions est très intéressante, tant pour des organismes publics que pour l'industrie ou le secteur commercial. Les discussions en ligne renforcent les intérêts et les avis des internautes, des critiques de produits, la présentation d'idées politiques, etc. En conséquence analyser ces dommages devient une problématique stratosphérique.

Turley, P. et M. Litman (2003). Measuring praise and criticism: inference of semantic orientation from association. *ACM TOIS* 2 (14), 315–346.  
 Zhang, J., M. Ackerman, et L. Adamic (2007). Expertise networks in online communities: Structure and algorithms. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pp. 321–322. ACM.  
 Zhou, D., J. Councill, H. Zhu, et C.-L. Giles (2007). Discovering temporal communities from social network documents. In *ICDM: International Conference on Data Mining*, pp. 745–750. IEEE Computer Society.

### 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 discussion analysis. It is based on a graph of messages: each node corresponds to a message and each edge (directed or undirected) connects two messages whose specific role is defined. The edges can be weighted by keywords that characterize the connected messages. This model allows a better representation of the content of the discussions and facilitates the identification of discussion chains. We compare the two representations and validate the proposed model on real debates. Our experiments show the additional information that can be extracted from the message-based graph.

# Applications (in particular to IR)

## Some applications

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

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

13

14

## Applications

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

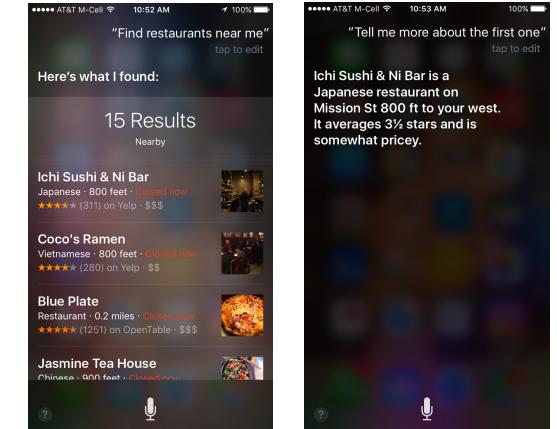


Watson gagne le Jeopardy! en 2011

A screenshot of a Google search results page for the query "text mining". The results include links to Wikipedia articles, academic papers, and news articles. A snippet from a Wikipedia page about text mining is visible, along with a snippet from a paper by Christian Faloutsos.

15

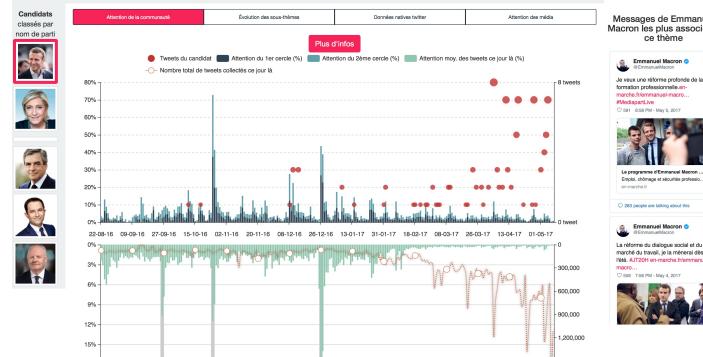
## Personal assistants



16

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

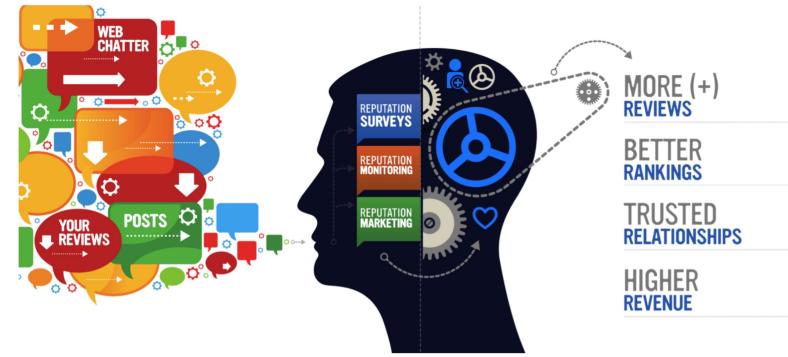
## Opinion mining



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

17

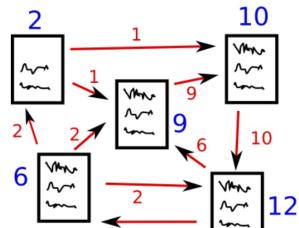
## E-reputation management



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

18

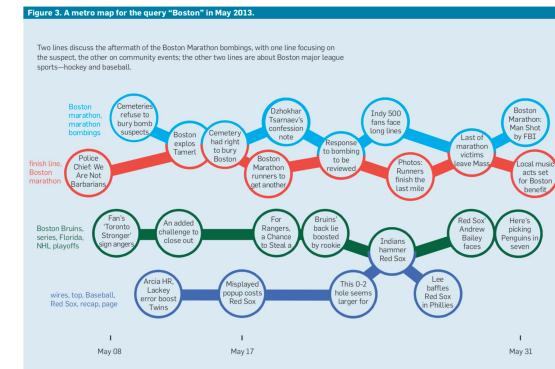
## Multi-document summarization



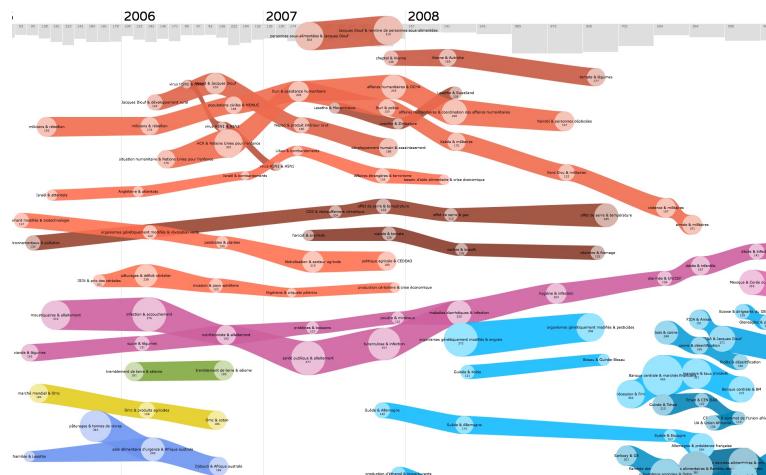
Source: <http://www.scottbot.net/HIAL/>

19

## Metromaps (Shahaf et al., 2015)



20

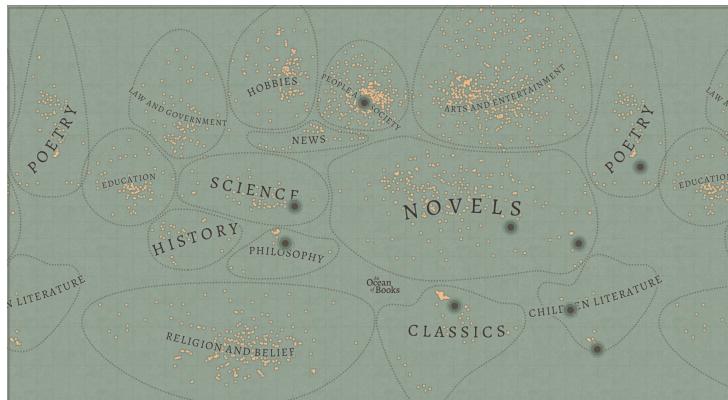


Project Pulseweb : <http://pulseweb.context>

21

# Projet LIFRANUM

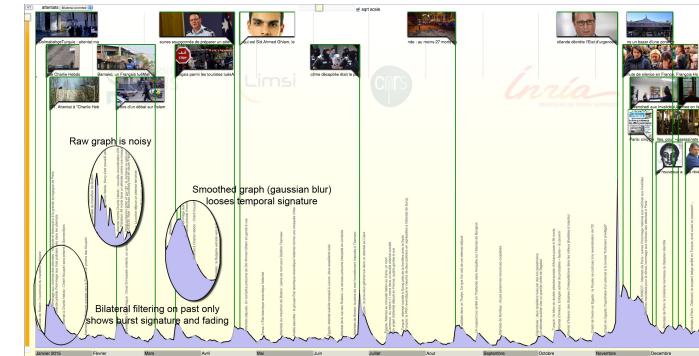
- Projet ANR  
LIFRANUM  
MARGE,  
ERIC,  
BnF



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

23

## 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.

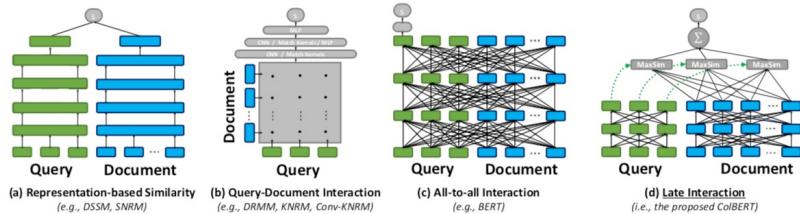
22

## Importance of indexing

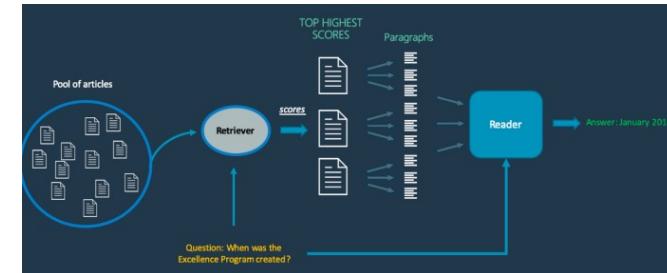
- What we're looking for?
    - words (and n-grammes)
    - authors
    - sources
    - categories (predefined? discovered? both?)
  - An heterogenous perspective : building/using (latent) spaces that link various objects (eg., documents and authors, topics and words, etc.)

24

## Deep learning and IR (1)



## Deep learning and IR (2)

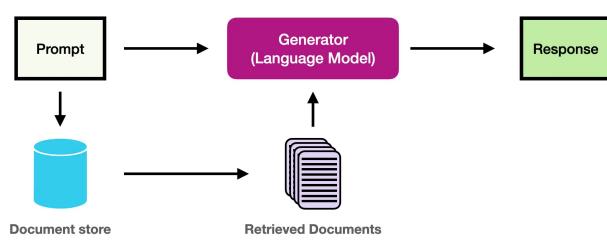


<https://medium.com/@varun030403/colbert-a-complete-guide-1552468335ae>

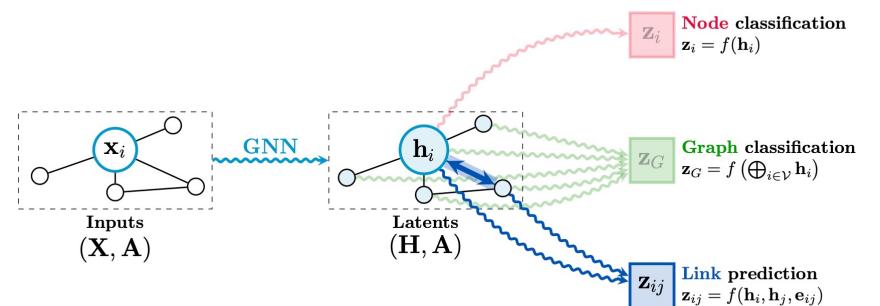
25

26

## Retrieval Augmented Generation (Lewis et al., 2021)



## Geometrical deep learning (Bronstein et al., 2021)



27

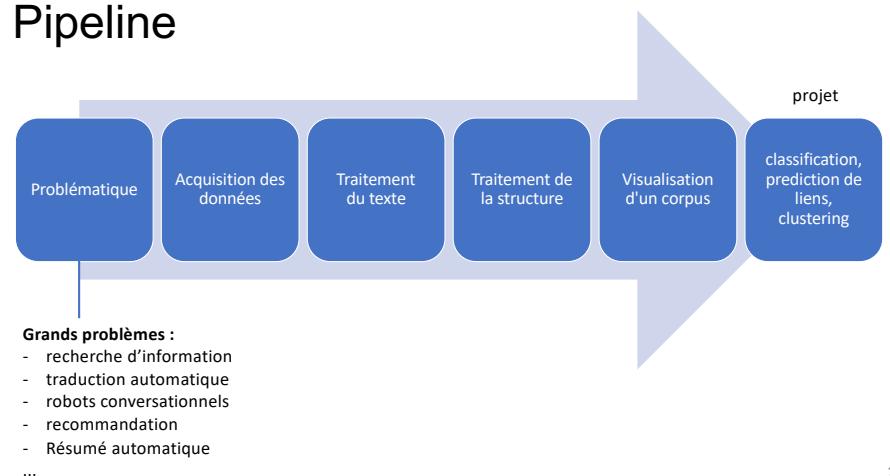
28

## Suggested pipeline

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

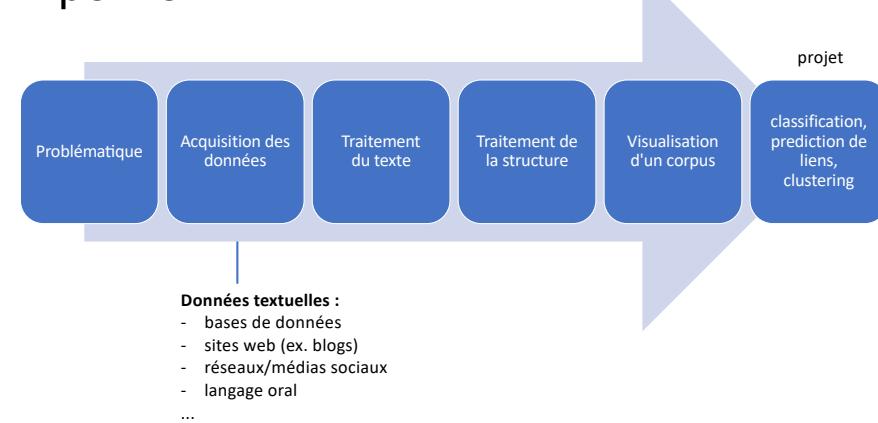
29

## Pipeline



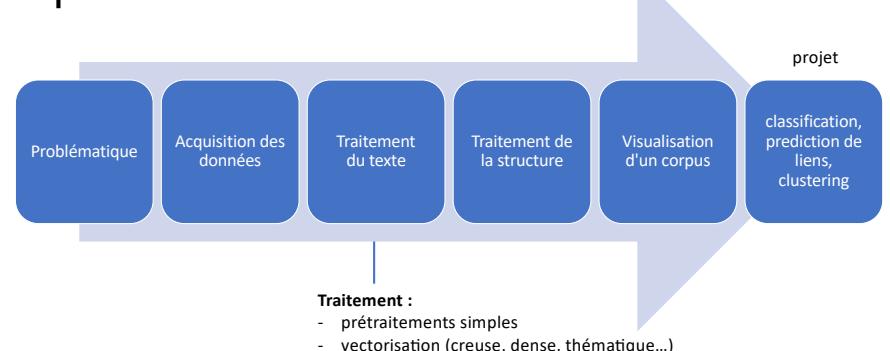
30

## Pipeline



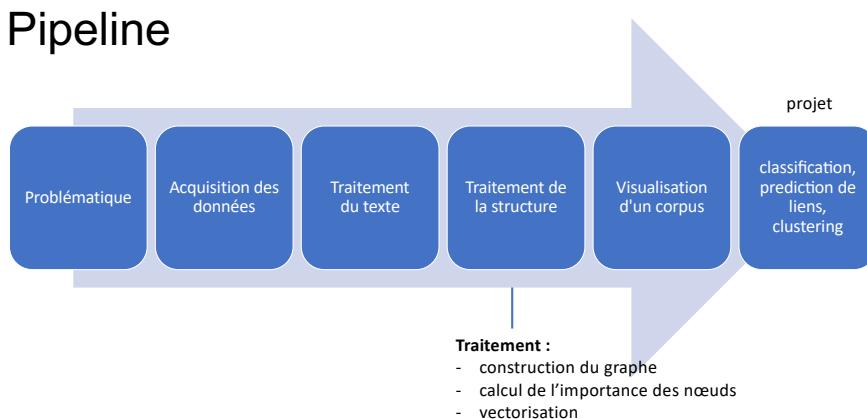
31

## Pipeline

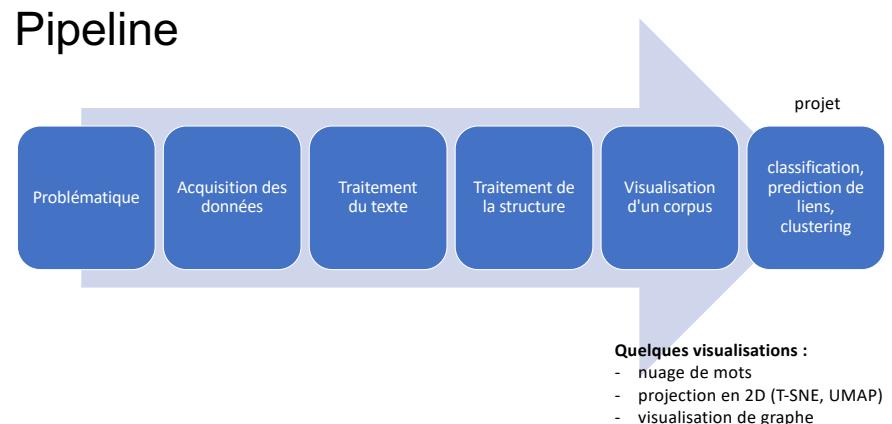


32

## Pipeline



## Pipeline



33

34

## What's next?

- Representation of documents
  - sparse representations
  - dense representations
  - topic models
- Network analysis
  - spectral clustering, modularity
  - representation learning for graphs
- Analyzing information networks
  - Graph Neural Networks

35