

# GARGANTEXT

Plateforme collaborative pour l'analyse de corpus

projet EasISTEX

<http://gargantext.org>

David Chavalarias & Alexandre Delanoë  
Complex Systems Institute of Paris Île-de-France



# UBIQUITÉ DES DONNÉES TEXTUELLES & OPPORTUNITÉS POUR LE TDM

Vélocité

## ÉCHELLES DE TEMPS

### Secondes

- > 6k pageview/s
- > 40k/s requêtes



### Minutes

- > 347k/s WhatsApp messages
- > 280k tweets/min
- > 375k SMS/min



### Heures

- > 4k/h billets de blog
- > 6,25k edits/h sur Wikipedia



### Jours

- > 3,2k AFP/j



### Années

- 79k/an articles français
- (~ 1,6M articles/an total WoS)
- + littérature "grise"



Audience

## Comment représenter la connaissance ?

[...] rassembler dans le plus petit espace possible, & à placer, pour ainsi dire, le Philosophe **au-dessus de ce vaste labyrinthe dans un point de vûe fort élevé** d'où il puisse apercevoir à la fois les Sciences & les Arts principaux ; voir d'un coup d'œil les objets de ses spéculations, & les opérations qu'il peut faire sur ces objets ; distinguer les branches générales des connoissances humaines, les points qui les séparent ou qui les unissent ; & entrevoir même quelquefois les routes secrètes qui les rapprochent. C'est **une espèce de Mappemonde** qui doit montrer les principaux pays, leur position & leur dépendance mutuelle, le chemin en ligne droite qu'il y a de l'un à l'autre ; chemin souvent coupé par mille obstacles, qui ne peuvent être connus dans chaque pays que des habitans ou des voyageurs, & qui ne sauroient être montrés que dans des cartes particulières fort détaillées. Ces cartes particulières seront les différens articles de notre Encyclopédie, & l'arbre ou système figuré en sera la mappemonde.

## Comment représenter la connaissance ?

*Mais comme dans les cartes générales du globe que nous habitons, les objets sont plus ou moins rapprochés, & présentent un coup d'œil différent selon le point de vûe où l'œil est placé par le Géographe qui construit la carte, de même la forme de l'arbre encyclopédique dépendra du point de vûe où l'on se mettra pour envisager l'univers littéraire. On peut donc imaginer autant de systèmes différens de la connoissance humaine, que de Mappemondes de différentes projections ; & chacun de ces systèmes pourra même avoir, à l'exclusion des autres, quelque avantage particulier.*

## Comment représenter la connaissance ?

*Mais comme dans les cartes générales du globe que nous habitons, les objets sont plus ou moins rapprochés, & présentent un coup d'œil différent selon le point de vûe où l'œil est placé par le Géographe qui construit la carte, de même la forme de l'arbre encyclopédique dépendra du point de vûe où l'on se mettra pour envisager l'univers littéraire. On peut donc imaginer autant de systèmes différens de la connoissance humaine, que de Mappemondes de différentes projections ; & chacun de ces systèmes pourra même avoir, à l'exclusion des autres, quelque avantage particulier.*

*D'Alembert (1751) L'Encyclopédie/1re édition/Discours préliminaire*

# Comment interagir avec les représentations de la connaissance ?

- ▶ *“distinguer les branches générales des connaissances humaines”*  
et donner accès à *“des cartes particulières fort détaillées”*  
→ **représentations haut-niveau & navigation multi-échelle,**
- ▶ *“chemin souvent coupé par mille obstacles, qui ne peuvent être connus dans chaque pays que des habitants ou des voyageurs*  
→ favoriser le collaboratif & la diversité des expertises,
- ▶ *“autant de systèmes différents de la connaissance humaine, que de Mappemondes de différentes projections*  
→ naviger entre différents systèmes de représentation

# Comment interagir avec les représentations de la connaissance ?

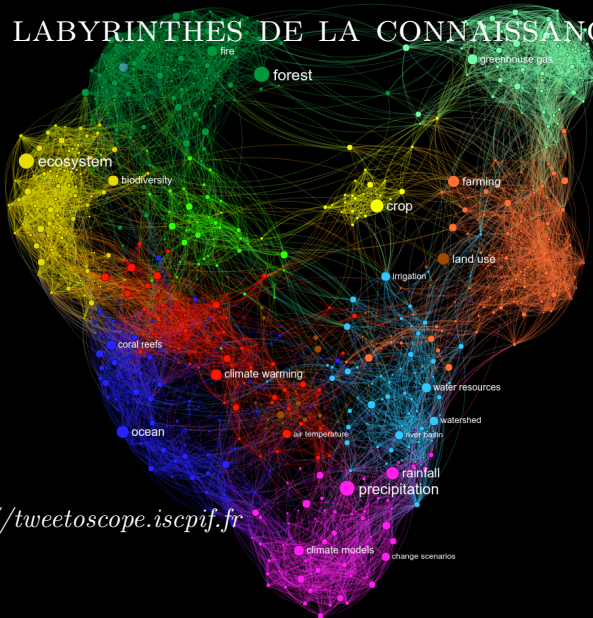
- ▶ *“distinguer les branches générales des connaissances humaines”*  
et donner accès à *“des cartes particulières fort détaillées”*  
→ **représentations haut-niveau & navigation multi-échelle,**
- ▶ *“chemin souvent coupé par mille obstacles, qui ne peuvent être connus dans chaque pays que des habitans ou des voyageurs*  
→ **favoriser le collaboratif & la diversité des expertises,**
- ▶ *“autant de systèmes différens de la connoissance humaine, que de Mappemondes de différentes projections*  
→ **naviger entre différents systèmes de représentation**

# Comment interagir avec les représentations de la connaissance ?

- ▶ *“distinguer les branches générales des connaissances humaines”*  
et donner accès à *“des cartes particulières fort détaillées”*  
→ **représentations haut-niveau & navigation multi-échelle,**
- ▶ *“chemin souvent coupé par mille obstacles, qui ne peuvent être connus dans chaque pays que des habitants ou des voyageurs*  
→ **favoriser le collaboratif & la diversité des expertises,**
- ▶ *“autant de systèmes différents de la connaissance humaine, que de Mappemondes de différentes projections*  
→ **naviger entre différents systèmes de représentation**

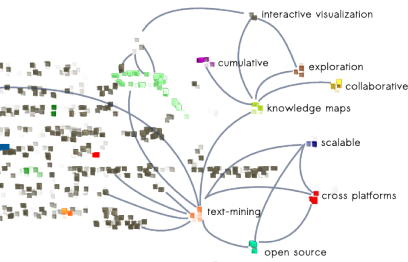


# LES LABYRINTHES DE LA CONNAISSANCE



<http://tweetoscope.iscpif.fr>

Gargantext is a web service that enables advanced text-mining, network analysis, and interactive visualisation to provide new interactions with your corporate Data Knowledge maps, science maps, write a state in few minutes. Gargantext a project of the ISC-PIF gathering several institutions and projects for the interactive exploration



Alexandre Delanoë

project manager

[Mail](#) [Website](#)



David Chavalarias

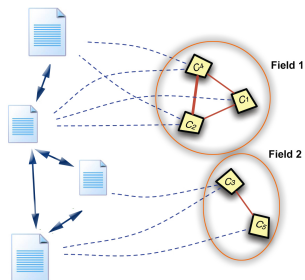
principal investigator

[Mail](#) [Website](#)

With contributions of Mathieu Rodic, Samuel Castillo, Maziyar Panahi, Elias Showk, Simon Murail, Nicolas Pouillard, Romain Loth, Constance Quatrebarbes.

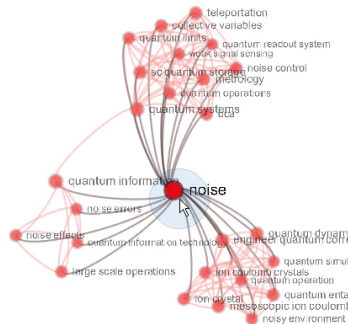
# Cartographie interactive de la connaissance

- ▶ **Manipulation de corpus non structurés**
- ▶ **Interactivité** Vous vous faites votre propre idée, l'utilisateur reste dans la boucle de production de représentations,
- ▶ **Collaboratif** et connaissance cumulative,
- ▶ **Interopérabilité** avec d'autres logiciels et bases de données plusieurs formats d'importation et d'exportation, comparaison de corpus et système d'APIs.



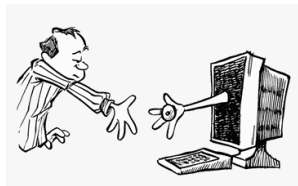
# Cartographie interactive de la connaissance

- ▶ **Manipulation de corpus non structurés**
- ▶ **Interactivité** Vous vous faites votre propre idée, l'utilisateur reste dans la boucle de production de représentations,
- ▶ Collaboratif et connaissance cumulative,
- ▶ **Interopérabilité avec d'autres logiciels et bases de données** plusieurs formats d'importation et d'exportation, comparaison de corpus et système d'APIs.



# Cartographie interactive de la connaissance

- ▶ **Manipulation de corpus non structurés**
- ▶ **Interactivité** Vous vous faites votre propre idée, l'utilisateur reste dans la boucle de production de représentations,
- ▶ **Collaboratif** et connaissance cumulative,
- ▶ **Interopérabilité avec d'autres logiciels et bases de données** plusieurs formats d'importation et d'exportation, comparaison de corpus et système d'APIs.

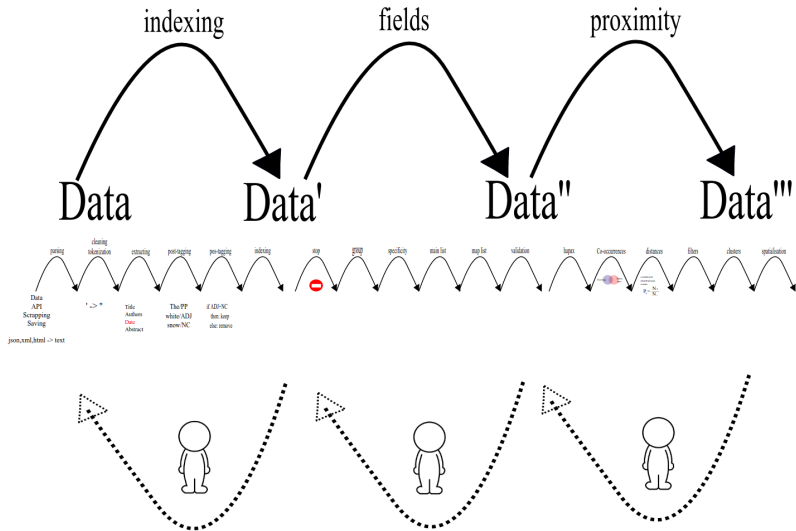


# Cartographie interactive de la connaissance

- ▶ **Manipulation de corpus non structurés**
- ▶ **Interactivité** Vous vous faites votre propre idée, l'utilisateur reste dans la boucle de production de représentations,
- ▶ **Collaboratif** et connaissance cumulative,
- ▶ **Interopérabilité avec d'autres logiciels et bases de données** plusieurs formats d'importation et d'exportation, comparaison de corpus et système d'APIs.



# Workflow



# Gargantext est agnostique

## **APIs :**

1. ISTEEX: littérature académique interdisciplinaire
2. PubMed : principale archive de la recherche bio-médicale
3. SCOAP : archives pour la physique des particules
4. REPEC : archives pour les sciences économiques
5. HAL : archives des EPST français,
6. ISIDORE : accès unifié aux données de la recherche dans les sciences humaines et sociales.



# Gargantext est agnostique

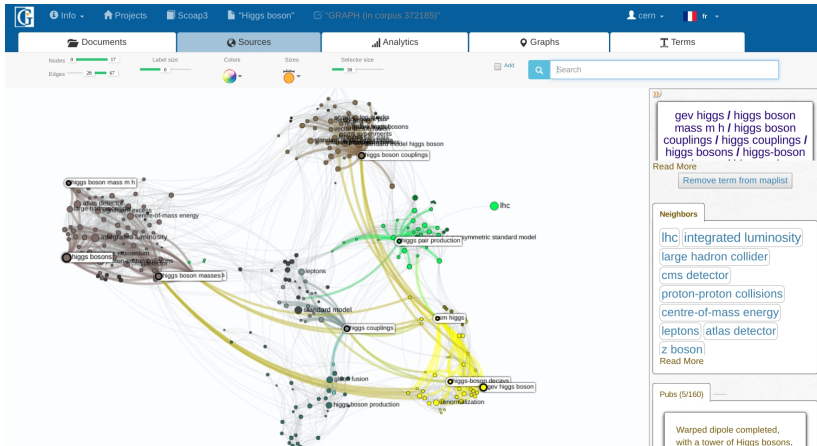
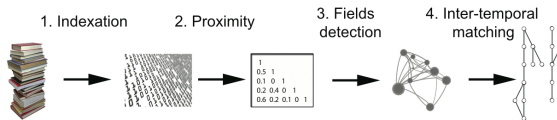
## **APIs :**

1. ISTEEX: littérature académique interdisciplinaire
2. PubMed : principale archive de la recherche bio-médicale
3. SCOAP : archives pour la physique des particules
4. REPEC : archives pour les sciences économiques
5. HAL : archives des EPST français,
6. ISIDORE : accès unifié aux données de la recherche dans les sciences humaines et sociales.

## **En import uniquement**

1. Web Of Science (ISI Format)
2. Scopus (RIS Format)
3. Europress (HTML Format)
4. Zotero (RIS Format)
5. Jstor (RIS Format)
6. CSV files

# Mapping projects with Gargantext



# Perspective dans EasIstex

- ▶ Développer l'aspect collaboratif,
- ▶ Développer l'aspect cumulatif et l'intelligence collective (communicabilité des éléments qui constituent une carte, mutualisation des annotation, etc.)
- ▶ Passage à l'échelle de la plateforme Gargantext.
- ▶ Analyse de phylométries des sciences

# Phylomemie des sciences (Chavalarias 2016 - HDR)

## ORDINATEURS QUANTIQUES

1994 - Peter Shor algorithm to factorize large integers  
 - First schemes for quantum error correction.  
 - First realization of a quantum logic gate

1996 - First quantum database search algorithm.  
 - First public call for research proposals in quantum information processing (US Gov. & Army).

1998 - First experimental demonstration of a quantum algorithm.  
 - First working 3-qubit NMR computer

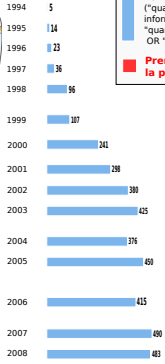
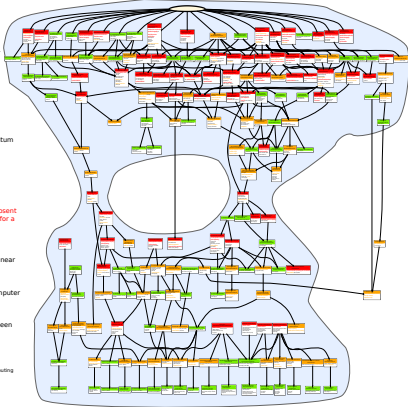
**2001 Negative result**  
 Demonstration by Noah Linden and Sandu Popescu that entanglement (so far absent from experiments) is a necessary condition for a large class of quantum protocols.  
 - First execution of Shor's algorithm

2002 Quantum computation roadmap.  
 - Quantum controlled-not gates using only linear optical elements  
 - DARPA Quantum Network operational

2004 - First working pure state NMR quantum computer

2005 - First quantum byte, or qubyte  
 - First transfer of quantum information between "quantum memories"

**2006 - 2007**  
 acceleration of discoveries.  
 Cf. [https://en.wikipedia.org/wiki/Timeline\\_of\\_quantum\\_computing](https://en.wikipedia.org/wiki/Timeline_of_quantum_computing)



**LEGENDE**

■ Nombre de documents du WoS pour la requête ("quantum communication" OR "quantum information" OR "quantum computing" OR "quantum computer" OR "quantum processing" OR "quantum algorithm")

■ Première apparition du terme dans la phylomemie

**1999 Membrane computing, natural**

**2001 EUV, Quantum matter**

**2002 Quantum memory channel, topological quantum computing, magnetodielectric**

**2003 One-way quantum computing, superconducting nanocircuits, pulse engineering, cluster states**

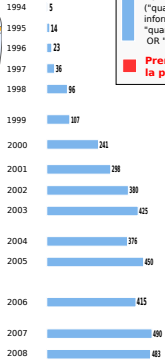
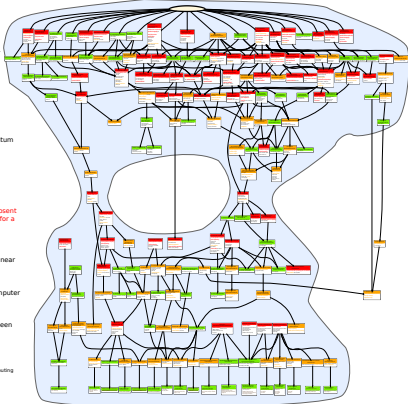
# Phylomemie des sciences (Chavalarias 2016 - HDR)

## ORDINATEURS QUANTIQUES

- 1994 - Peter Shor algorithm to factorize large integers
- 1995 - First schemes for quantum error correction.
- 1996 - First quantum database search algorithm.
- 1998 - First experimental demonstration of a quantum algorithm.
- 2001 - Negative result
- 2002 - Quantum computation roadmap.
- 2003 - Quantum controlled-not gates using only linear optical elements
- 2004 - First working pure state NMR quantum computer
- 2005 - First quantum byte, or qubyte
- 2006 - 2007 - acceleration of discoveries.

- First realization of a quantum logic gate
- First public call for research proposals in quantum information processing (US Gov. & Army).
- First working 3-qubit NMR computer

- Demonstration by Noah Linden and Sandu Popescu that entanglement (so far absent from experiments) is a necessary condition for a large class of quantum protocols.
- First execution of Shor's algorithm
- DARPA Quantum Network operational
- First transfer of quantum information between "quantum memories"
- CT: [https://en.wikipedia.org/wiki/Timeline\\_of\\_quantum\\_computing](https://en.wikipedia.org/wiki/Timeline_of_quantum_computing)

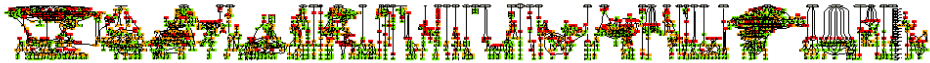


**LEGENDE**

■ Nombre de documents du WoS pour la requête ("quantum communication" OR "quantum information" OR "quantum computing" OR "quantum computer" OR "quantum processing" OR "quantum algorithm")

■ Première apparition du terme dans la phylomemie

- 1999 Membrane computing, natural
- 2001 EUV, Quantum matter
- 2002 Quantum memory channel, topological quantum computing, magnetodielectric
- 2003 One-way quantum computing, superconducting nanocircuits, pulse engineering, cluster states



# THANKS FOR YOUR ATTENTION



David CHAVALARIAS &  
Alexandre DELANOE

Web : <http://gargantext.org>



**Rendez-nous visite**  
**[iscpif.fr](http://iscpif.fr)**