

ACCÉLÉRER LA RECHERCHE, TRANSFORMER LES PRATIQUES, OUVRIR DE NOUVEAUX HORIZONS : LA FOUILLE DE DONNÉES TEXTUELLE AVEC GARGANTEXT

David Chavalarias

Directeur de Recherche CNRS à l'EHESS/CAMS

Directeur de l'Institut des Systèmes Complexes de Paris IdF

<http://chavalarias.com>

JNSO

8 Octobre 2019



A close-up photograph of a stone path in a forest. A dense line of small, dark ants is marching along the left edge of the path. The path is scattered with small, light-colored particles, likely food. The background shows lush green foliage and a stream of water on the right side.

STIGMERGIE
CATALYSEUR DE
L'INTELLIGENCE COLLECTIVE

Les formes générées par stigmergie dépendent de l'environnement

0 — 5m



Eciton hamatum



Eciton rapax



Eciton burchelli

TRACES STIGMERGIQUES & RÉVOLUTION DES DONNÉES SOCIALES

Velocity

TIME SCALES

Seconde

- > 6k pageview/s
- > 40k/s requêtes



Minutes

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



Hours

- > 4k/h blog post
- > 6,25k edits/h on Wikipedia



Days

- > 3,2k AFP news/d



Years

- 79k/y french papers
- (> 2M peer reviewed papers/an)



Audience

TRACES STIGMERGIQUES & RÉVOLUTION DES DONNÉES SOCIALES

Quelles sont les formes
générées par ce type de
stigmergie ?

Comment peut-on
exploiter leur
reconstruction dans le
cadre de la recherche ?

Velocity

TIME SCALES

Seconde

- > 6k pageview/s
- > 40k/s requêtes



Minutes

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



Hours

- > 4k/h blog post
- > 6,25k edits/h on Wikipedia



Days

- > 3,2k AFP news/d



Years

- 79k/y french papers
- (> 2M peer reviewed papers/an)



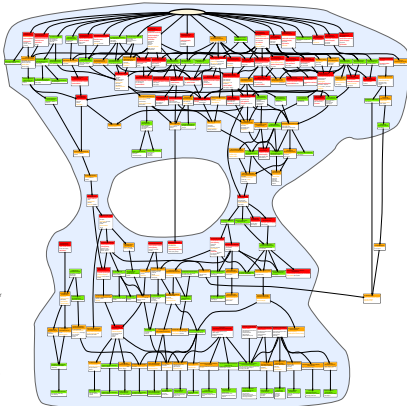
Audience

Science phylomemy

Chavalarias, D., 2016 (HDR) <https://hal.archives-ouvertes.fr/tel-01394843v1>

ORDINATEURS QUANTIQUES

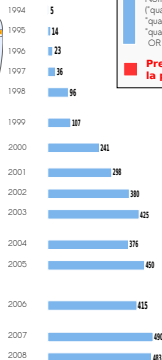
- 1994 - Peter Shor algorithm to factorize large integers.
- 1995 - First schemes for quantum error correction.
- First realization of a quantum logic gate
- 1996 - First quantum database search algorithm.
- First public call for research proposals in quantum information processing (US Gov. & Army).
- 1998 - First experimental demonstration of a quantum algorithm.
- First working 3-qubit NMR computer
- 2001 **Negative result**
Demonstration by Hootn Lidon 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.
- 2003 - 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://arxiv.org/abs/0706.2154>



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 phyloméme**

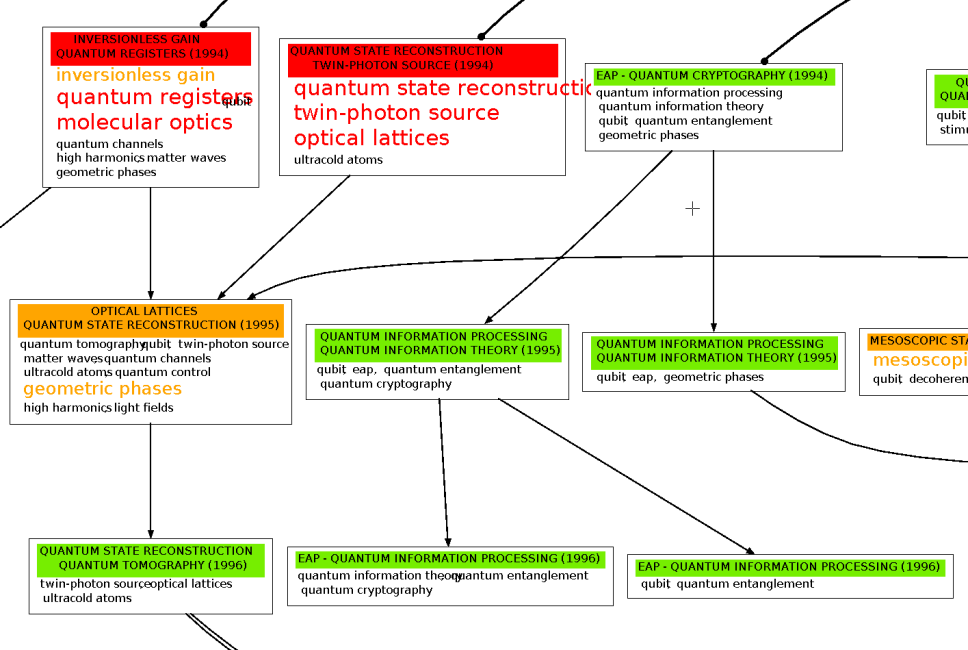


1999 Membrane computing, natural

2001 EUV, Quantum matter

2002 Quantum memory channel, topological quantum computing, magnetodielectric

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



Science phylomemy

(Chavalarias, D., 2016 (HDR) <https://hal.archives-ouvertes.fr/tel-01394843v1>)

ORDINATEURS QUANTIQUES

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 phylométrie

1994 5

1995 14

1996 23

1997 36

1998 96

1999 107

2000 241

2001 299

2002 380

2003 425

2004 376

2005 459

2006 415

2007 490

2008 483

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

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

1995 - 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
Quantum computation roadmap.

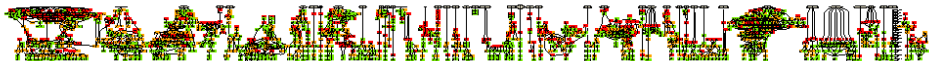
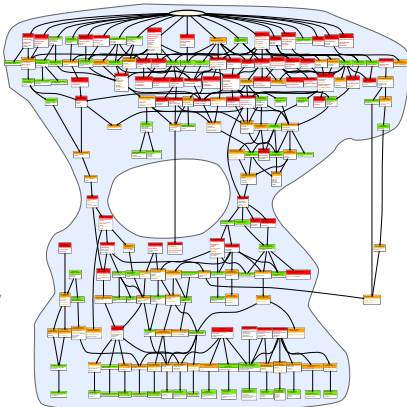
2003 - Quantum controlled-not gates using only linear optical elements.
- DARPA Quantum Network operational

2004 - First working pure state NMR quantum computer

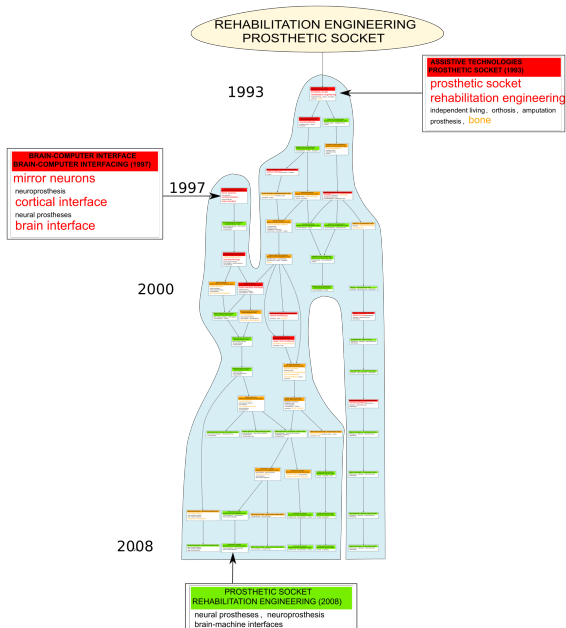
2005 - First quantum byte, or qubyte
- First transfer of quantum information between "quantum memories"

2006 - 2007

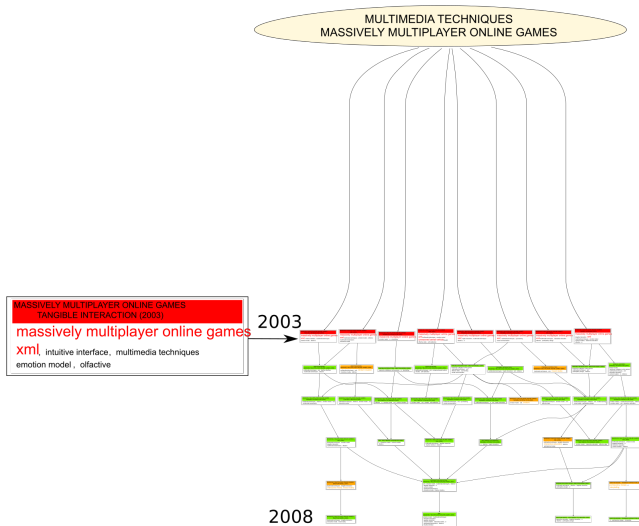
acceleration of discoveries.
Q: https://en.wikipedia.org/wiki/Timeline_of_quantum_computing



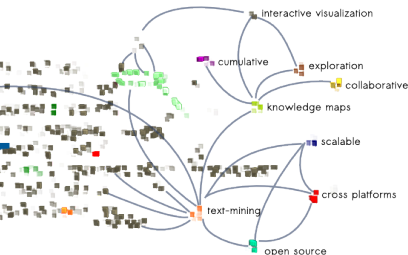
Hybridization de champs scientifiques



Emergence de champs scientifiques



Gargantext is a web service that offers 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 ISC-PIF gathering several institutions and projects for the interactive exploration



Alexandre Delanoë

project manager

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



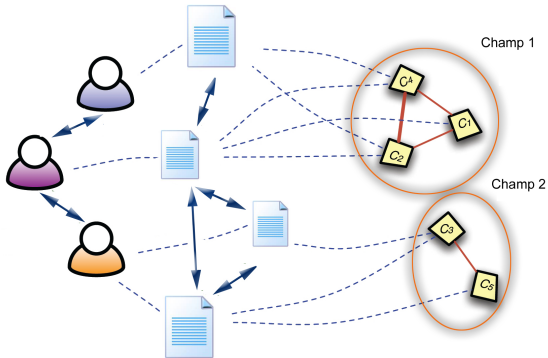
David Chavalarias

principal investigator

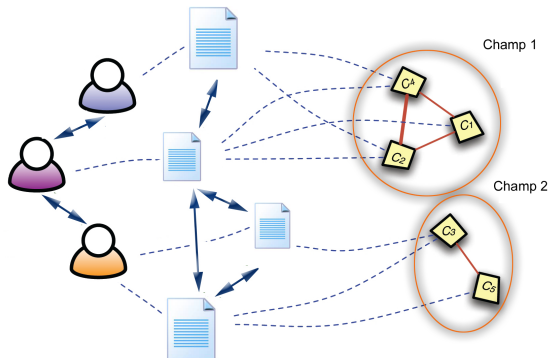
[Mail](#) [Website](#)

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

Objets d'étude

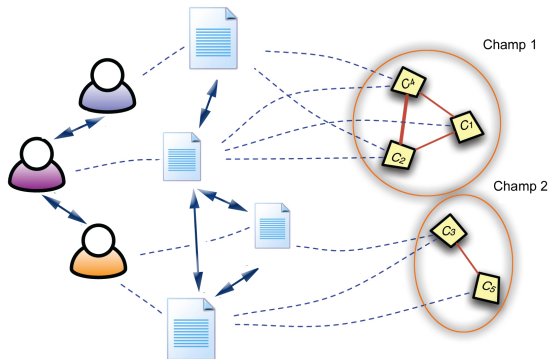


Objets d'étude



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

Objets d'étude



1. CSV files
2. Zotero (RIS Format)
3. Web Of Science (ISI Format)
4. Scopus (RIS Format)
5. Europress (HTML Format)
6. Jstor (RIS Format)
7. ...

Science phylomemy

(Chavalarias, D., 2016 (HDR) <https://hal.archives-ouvertes.fr/tel-01394843v1>)

ORDINATEURS QUANTIQUES

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 phylométrie

1994 5

1995 14

1996 23

1997 36

1998 96

1999 107

2000 241

2001 299

2002 380

2003 425

2004 376

2005 459

2006 415

2007 490

2008 483

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

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

1995 - 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
 Quantum computation roadmap.

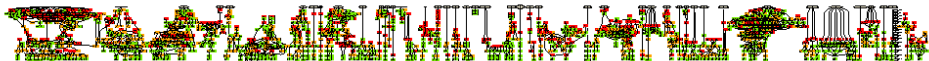
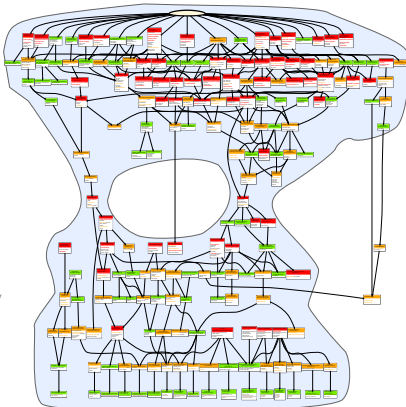
2003 - Quantum controlled-not gates using only linear optical elements.
 - DARPA Quantum Network operational

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



Le Tweetoscope Climatique

QUESTIONS :

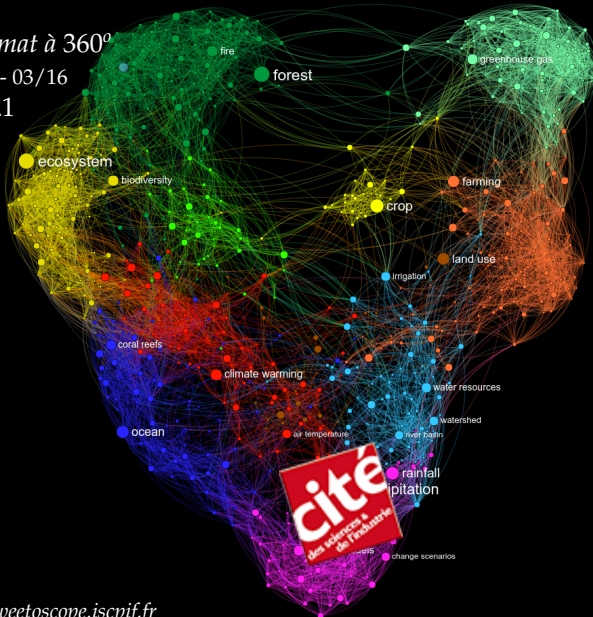
- ▶ DE QUOI PARLENT LES SCIENTIFIQUES LORSQU'ILS ÉTUDIENT LE CHANGEMENT CLIMATIQUE ?
- ▶ DANS QUELLE MESURE LES THÈMES ASSOCIÉS ET LEURS ENJEUX SONT-ILS RELAYÉS SUR LE WEB ?
- ▶ COMMENT RÉCONCILIER LES DIFFÉRENTES DYNAMIQUES ATTENTIONNELLES ENTRE LES SPHÈRES ACADÉMIQUE, CITOYENNE ET POLITIQUE ?

TWEETOSCOPE CLIMATIQUE

Le Climat à 360°

10/15 - 03/16

COP21



<http://tweetoscope.iscpif.fr>

ET SI VOUS POUVIEZ FAIRE VOTRE
PROPRE CARTE DES MOTS DU CLIMAT
EN QUELQUES MINUTES ?

Créer un référentiel commun

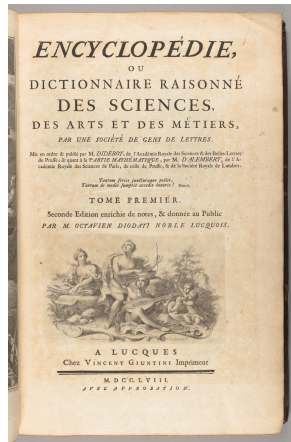
*[...] rassembler dans le plus petit espace possible, & à placer, pour ainsi dire, le Philosophe **au-dessus de ce vaste labyrinthe dans un point de vûe fort élevé** d'où il puisse appercevoir à la fois les Sciences & les Arts principaux ; voir d'un coup d'œil les objets de ses spéculations, & les opérations 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 secretes 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 particulieres fort détaillées. Ces cartes particulieres seront les différens articles de notre Encyclopédie, & l'arbre ou système figuré en sera la mappemonde.*

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

Créer un référentiel commun

“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/1ère édition/Discours
préliminaire



Les labyrinthes de la connaissance

<http://risk.iscpif.fr> & <http://gargantext.org>



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 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 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**

Quelles opportunités de la science ouverte ?

- ▶ Démultiplier l'intelligence collective (meilleure coopération, meilleure coordination),
- ▶ Changer notre rapport à la connaissance en donnant un point de vue réflexif et "temps réel" sur son évolution,
- ▶ Proposer un "GPS" de la connaissance et des innovations.

MERCI POUR VOTRE ATTENTION



David CHAVALARIAS



Rendez-nous visite
iscpif.fr