

Bibliométrie : Apports et limites dans la compréhension de l'environnement scientifique et technologique, l'exemple des neurosciences computationnelles

Julia Soyez

► To cite this version:

Julia Soyez. Bibliométrie : Apports et limites dans la compréhension de l'environnement scientifique et technologique, l'exemple des neurosciences computationnelles. [Rapport Technique] RT-0356, INRIA. 2008, pp.32. inria-00327742

HAL Id: inria-00327742

<https://hal.inria.fr/inria-00327742>

Submitted on 31 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

***BIBLIOMETRIE : APPORTS ET LIMITES DANS LA COMPREHENSION DE L'ENVIRONNEMENT SCIENTIFIQUE
ET TECHNOLOGIQUE, L'EXEMPLE DES NEUROSCIENCES COMPUTATIONNELLES***

Julia Soyez
Service IST
Projet CORTEX

N° 0356

Octobre 2008

Thème COG

A large blue rectangular area containing the text 'Rapport technique' in a white, italicized serif font. To the left of the text is a large, light grey 'R' logo, and a horizontal grey brushstroke is positioned below the text.

R *apport
technique*

INRIA Rapport Technique

Julia Soyez¹, Service IST², Projet CORTEX³

Thème COG – Systèmes intelligents
Service IST & Projet CORTEX

Rapport technique n° 0356 – Octobre 2008 - 32 pages

Résumé:

Le but de cette démarche était de savoir si le service d'Information Scientifique et Technique (IST) devrait développer des activités bibliométriques visant à la compréhension de l'environnement scientifique et technologique en tant que service support de la recherche. Ce travail a été effectué en étroite collaboration avec les chercheurs de l'équipe Cortex et a permis la mise en place d'un processus itératif entre service IST et équipe de recherche. L'autre finalité de cette expérience était de savoir quels seraient les apports de cette méthode pour les scientifiques.

Dans ce document, seront présentés rapidement la méthode de collecte et de traitement des informations issues des bases de données bibliographiques. Les principaux résultats des études seront présentés.

Mots clés: veille scientifique et technologique, bibliométrie, neurosciences computationnelles

¹Julia Soyez, Service IST, INRIA Sophia Antipolis, Julia.Soyez@sophia.inria.fr ou juliasoyez@yahoo.fr

²Service d'Information Scientifique et Technique, INRIA Sophia Antipolis, doc@sophia.inria.fr

³Equipe CORTEX, LORIA, frederic.alexandre@loria.fr et Thierry.Vieville@sophia.inria.fr

INRIA Technical Report

Abstract:

This approach aims at knowing if a scientific and technical information department has to develop a bibliometric service for the researchers. This experience is due to determine if this method was adapted to understand the scientific and technical environment. This work has been done in cooperation with the research team, Cortex, and has allowed teamwork between it and the scientific and technical information department.

The other purpose was to know what contributions of this method could be getting by the researchers.

Firstly the collect and treatment's method will be exposed, and then the main results of the bibliometric studies will be described and explained.

Keywords: scientific and technical environmental scanning, bibliometrics, computational neuroscience

INTRODUCTION : PRESENTATION DE LA DEMARCHE	4
I. Méthode	6
1.1. Corpus des publications scientifiques	6
1.2. Corpus de brevets	8
II. Résultats	9
2.1. Les neurosciences computationnelles : une discipline en plein essor aux frontières floues	9
2.1.1. Les principaux journaux et éditeurs des neurosciences computationnelles	10
2.1.2. Les centres, universités et laboratoires de recherche qui publient le plus dans les neurosciences computationnelles	11
2.1.3. Les collaborations au sein des neurosciences computationnelles	12
2.2. Mesure et compréhension de la production scientifique à partir de données bibliométriques : quelques comparaisons.	15
2.2.1. Comparaison de corpus	15
2.2.2. Mise à l'écart du Journal Neurocomputing dans le corpus Dialog	18
2.3. Etude bibliométrique des brevets en neurosciences computationnelles	20
2.3.1. Les principaux déposants	21
2.3.2. Codes CIB : tendances de l'innovation en neurosciences computationnelles	23
2.3.3. Technologies et acteurs	24
CONCLUSIONS	25
BIBLIOGRAPHIE	27
ANNEXE : VEILLE DEFINITION ET TYPOLOGIE	28
1. La typologie	29
2. La veille, un processus	30

Introduction

INTRODUCTION : PRESENTATION DE LA DEMARCHE

Lors de l'implémentation d'une nouvelle fonction de veille au sein d'un service d'Information Scientifique et Technique (IST), une recherche de solutions a été effectuée pour mener à des tests. Cette recherche a été réalisée sur le Web à partir du moteur de recherche Google. Les mots-clés utilisés ont été : veille, solution, logiciel, collecte information, analyse information, intelligence économique, open source, plateforme collaborative.

Les recherches ont porté sur les logiciels de veille propriétaires, open source, les outils gratuits proposés sur le Net, les prestations de veille et les fonctionnalités que l'on peut ajouter sur le navigateur Firefox. Les sources d'information trouvées sont : les sites dédiés à la veille, les blogs de veilleurs, les sites d'éditeurs de solutions logicielles, les sites d'expert en veille comme l'INIST, les sites spécialisés en outils logiciels. A partir de ces recherches, un tri a été opéré, les solutions considérées intéressantes ont été retenues. Pour réaliser les tests des logiciels, une équipe de recherche, l'équipe Cortex, a été sélectionnée pour participer à la démarche de veille. Cette équipe a été choisie car lors d'un entretien, un des chercheurs a été enthousiasmé par la démarche. Etant donné que la phase de test nécessite une forte interaction veilleur et chercheurs de l'équipe, il s'agissait d'une condition importante de la réalisation du test.

Le domaine de recherche de l'équipe projet est les neurosciences computationnelles qui sont un domaine au carrefour de plusieurs disciplines scientifiques. Le souhait de l'équipe de recherche était de réaliser un état de l'art des neurosciences computationnelles.

Plusieurs méthodes ont été utilisées :

- ✓ Une veille automatique sur le Web avec :
 - la création d'alertes sur le web (sites et blogs) à partir de mots-clés,
 - l'abonnement à deux newsletters,
 - l'interrogation de trois moteurs de recherche (Google, Yahoo, MetaCrawler),
 - l'abonnement à une liste de diffusion,
 - le test de Copernic qui dans sa version bêta fait office de méta moteur.

De cette veille automatique, quinze informations ont été transmises dont trois ont été jugées très intéressantes.

- ✓ Deux études bibliométriques en neurosciences computationnelles ont été réalisées une sur les publications scientifiques et une sur les brevets (Cf. II. Résultats). Les études ont permis de tester les logiciels bibliométriques Mathéo Analyser et Mathéo Patent. Ces études ont par ailleurs mis en avant les limites des logiciels et de la méthode bibliométrique.

Les enjeux ayant motivé la réalisation des études bibliométriques sont :

- savoir si les études bibliométriques pouvaient intéresser les chercheurs et quels seraient les avantages de cette démarche pour eux.
- Savoir si la base de données ISI Web Of Knowledge (en la comparant avec DIA-LOG) possédait une couverture suffisante de nos domaines de recherche?
- Connaître les fonctionnalités et les couvertures offertes par les différentes bases de données en les comparant à partir d'une même requête.
- Savoir si le service IST serait prestataire d'études bibliométriques.

I. Méthode

1.1. Corpus des publications scientifiques

La **base de données interrogée est DIALOG** qui est une méta base de données qui donne accès à environ 1200 bases de données scientifiques et technologiques qui couvrent les brevets, les publications scientifiques, les rapports techniques, les conférences et les thèses. Par ailleurs, Dialog offre des informations de presse, des données boursières et des informations financières des entreprises.

La requête n'a pas été lancée sur Dialog dans son intégralité. Après discussion avec ma responsable de Master, il avait été convenu d'appeler via Dialog les bases SciSearch et Dissertation Abstract on Line, bases jugées révélatrices de la production scientifique en neurosciences computationnelles.

La base **SciSearch** est une base de données pluridisciplinaire et internationale. Cette base reprend toutes les publications de la base Science Citation Index et comprend des données en provenance des domaines suivants : ingénierie technologique, science informatique, physique, biologie, médecine clinique, pharmacologie, etc. Cette base est produite par Thomson Scientific et indexe plus de 6000 journaux scientifiques et techniques⁴.

La deuxième base appelée via Dialog est **Dissertation Abstracts Online**⁵ qui est la base américaine des thèses et travaux de recherche de master. Cette base comprend aussi les travaux universitaires de recherche de 50 universités anglaises et quelques uns en provenance de l'Europe.

La requête lancée dans ses bases est : **neuroscience ? (N) comput ?**

Le point d'interrogation signifie troncature donc dans ce cas neuroscience ? signifie recherche neuroscience au singulier ou au pluriel et pour comput ? cela signifie recherche computer / computation/ computational.

(N) signifie recherche ces deux termes ensemble sans qu'il y ait un mot entre eux, ce qui se traduit par rechercher neuroscience computational ou computational neuroscience.

L'échantillon de notices bibliographiques rapatriées est de 356 publications scientifiques. Ce corpus a été formalisé dans un document Word sur lequel des macros ont été réalisées afin d'harmoniser le corpus. Ensuite, le document a été enregistré en format .txt afin de pouvoir l'importer dans le logiciel bibliométrique Mathéo Analyser.

Une fois le fichier importé dans Mathéo Analyser, les différents champs importés (journal, éditeurs, auteurs, titre, affiliations) ont été nettoyés. Par exemple, si dans le champ auteurs, il y avait par exemple : « Deco G » et « Deco Gustavo », ils ont été fusionnés sous l'appellation : « Deco G ». Si cette opération n'est pas effectuée alors le logiciel interprétera Deco G et Deco Gustavo comme deux personnes différentes

⁴ Pour plus d'informations, se reporter à la « feuille bleue » Dialog concernant la base SciSearch, Cf. <http://library.dialog.com/bluesheets/html/bl0034.html>

⁵ Pour plus d'informations, se reporter à la « feuille bleue » Dialog, Cf. <http://library.dialog.com/bluesheets/html/bl0035.html>

alors qu'il s'agit de la même personne. Ce même traitement a été effectué sur les titres (pour voir les doublons), sur les équipes (regroupées les noms d'UMR, de laboratoire,...), sur les journaux (suppression des volumes, pages,... puis fusion par titre de journal) et éditeurs (exemple : Elsevier, Elsevier Science, ...) regroupés sous une seule appellation.

1.2. Corpus de brevets

La création du corpus de brevets en neurosciences computationnelles a été effectuée à partir du logiciel Mathéo Patent. Le logiciel Mathéo Patent permet d'appeler directement via l'outil les bases de données brevet et de rapatrier les documents dans l'outil bibliométrique. Les bases appelées sont Espacenet, le WIPO (via Espacenet) et UPS-TO.

☑ Espacenet⁶

Cette base est la base européenne de données des brevets, elle est en accès libre et offre un accès aux brevets européens (couverture de la zone européenne), aux brevets de couverture nationale de trente pays européens (ils ne sont pas encore tous indexés dans la base) et à une partie des brevets japonais. De plus, elle offre aussi un accès aux brevets PCT (Patent Cooperation Treaty) qui sont les brevets à couverture mondiale (plus de 30 000 brevets indexés dans la base).

☑ WIPO (World Intellectual Property Office)

Se dit en français OMPI, Office Mondial de la Propriété Intellectuelle.

La base de données créée par le WIPO en accès libre s'appelle « Patentscope⁷ ». Cette base de données offre un accès à tous les brevets PCT (mondiaux).

☑ UPSTO (United States Patent and Trade Office)

L'UPSTO est l'organisme gouvernemental américain de la propriété industrielle. Il possède une base de données brevets qui concerne tous les brevets déposés sur le sol américain ainsi que certains brevets PCT (mondiaux) qui sont en fait les extensions de brevets américains. Cette base est accessible librement en ligne⁸.

La requête utilisée pour créer le corpus de brevets est la suivante : « computational neuroscience » qui signifie recherche l'expression exacte : computational neuroscience.

Le corpus représente 71 brevets qui se répartissent ainsi :

- 7 EP (brevets européens)
- 8 PCT (brevets mondiaux)
- 15 US (brevets américains)
- 37 JP (brevets japonais : couverture nationale)
- 10 DE (brevets allemands : couverture nationale)
- 2 RU (brevets russes : couverture nationale)
- 2 KR (brevets coréens : couverture nationale)
- 1 SK (brevets slovaques : couverture nationale)

⁶Cf. le site Web d'Espacenet

<http://ep.espacenet.com/>

⁷ Cf. Site Web du WIPO : <http://www.wipo.int/pctdb/en/>

⁸ Cf. Site Web du UPSTO: <http://www.upsto.gov/patft/index.html>

II. Résultats

2.1. Les neurosciences computationnelles : une discipline en plein essor aux frontières floues

A partir de la requête envoyée sur la base de données Dialog, on peut constater (Cf. schéma ci-dessous) que le concept neurosciences computationnelles est pluridisciplinaire.

Il fait appelle aux disciplines des sciences dures (mathématiques, physique, biologie) à partir desquelles sont déduits des modèles (modèles numériques, modèles biologiques). Les neurosciences ont pour cœur de recherche l'étude du cerveau et du système nerveux (par exemple, les corrélations entre fonctions et structures).

Les neurosciences computationnelles ont pour fondement disciplinaire les STIC, Sciences et Technologies de l'Information et de la Communication et les Neurosciences. Les domaines de recherche en neurosciences computationnelles s'appuient sur l'observation, la compréhension, la simulation et la modélisation du réseau neuronal (dont les mots-clés sont neurone, axone, dendrite, synapse, impulsion électrique, codage-encodage-décodage de l'information des circuits neuronaux).

Les neurosciences computationnelles en tant que domaine des neurosciences cherche à comprendre le système cognitif de l'échelle moléculaire à comportementale. Pour y parvenir, elles interagissent avec de nombreuses disciplines.

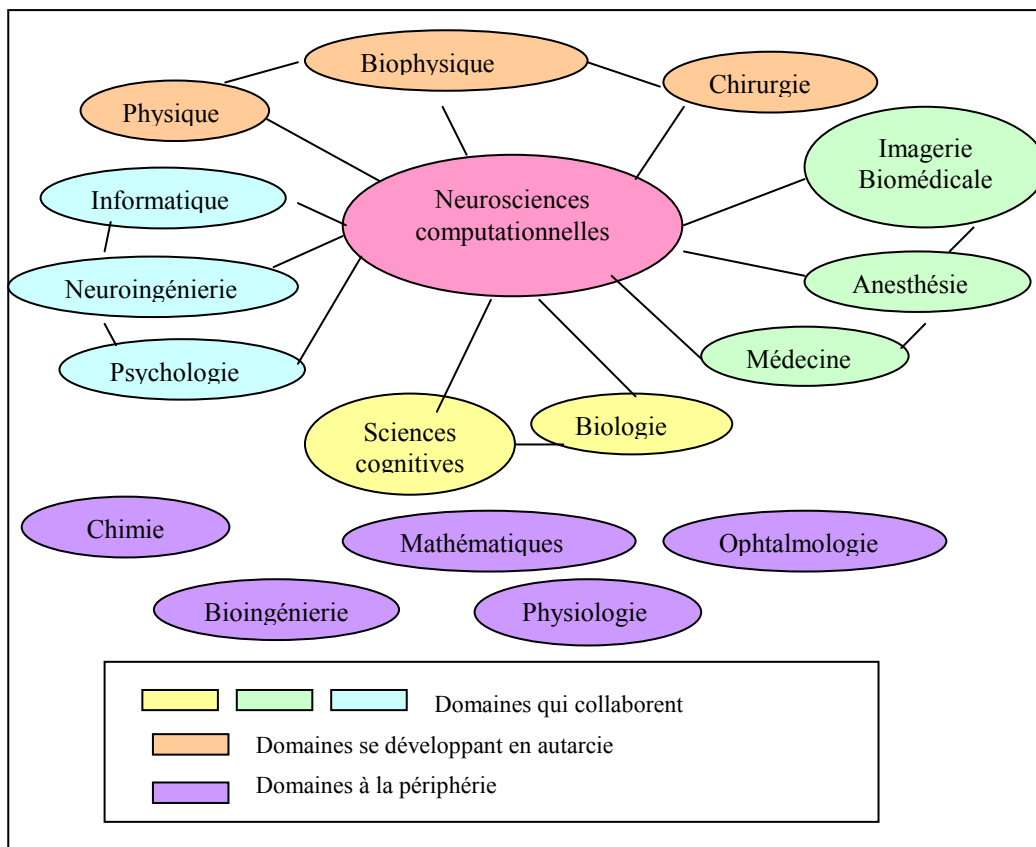


Figure 1- Les domaines scientifiques impliqués dans les neurosciences computationnelles

A partir des données obtenues via Dialog, nous pouvons estimer la naissance des neurosciences computationnelles aux alentours des années 1980. Nous pouvons proposer comme principaux cycles de la recherche scientifique en neurosciences computationnelles :

1982-1990 : Un démarrage lent

1990-2000 : Essor des neurosciences computationnelles

2001-2003 : Ralentissement de la production scientifique

2003-2007 : Une discipline en plein essor

2007 : Poursuite de l'accroissement de la production scientifique en neurosciences computationnelles (44 publications indexées dans les bases SciSearch et Dissertation Abstract Online). L'année 2007 a été la plus riche en termes de publications d'articles scientifiques et d'actes de conférences.

2.1.1. Les principaux journaux et éditeurs des neurosciences computationnelles

Le journal *Neurocomputing* qui représente 10% de la production scientifique du corpus crée en neurosciences computationnelles. La première position de *Neurocomputing* dans les journaux spécialisés en neurosciences computationnelles peut s'expliquer par sa politique éditoriale car ce journal publie beaucoup d'articles de conférences, dont C.N.S (Computational Neuroscience Meeting).

En seconde position avec plus de 2% de la production scientifique, on trouve les journaux :

Neural Computation, *Neural Networks*, *Proceedings of the National Academy of Science of the USA* et enfin le *Journal of Computational Neuroscience*.

Enfin les 80 journaux ont effectué moins de 0,5% de la production sur le sujet sur une période de plus de 10 ans.

Les principaux éditeurs dans le champ des neurosciences computationnelles sont Elsevier largement en tête avec 20% des publications, suivi de Springer avec 10% des publications. IEEE, le MIT Press, le KLUWER (éditeur académique aux Pays Bas, branche du groupe Wolters Kluwer Group), l'Académie Nationale des Sciences à Washington, l'Oxford University Press (Angleterre et Etats-Unis) se partagent les troisième et quatrième positions avec environ 2 à 3% des publications.

2.1.2. Les centres, universités et laboratoires de recherche qui publient le plus dans les neurosciences computationnelles

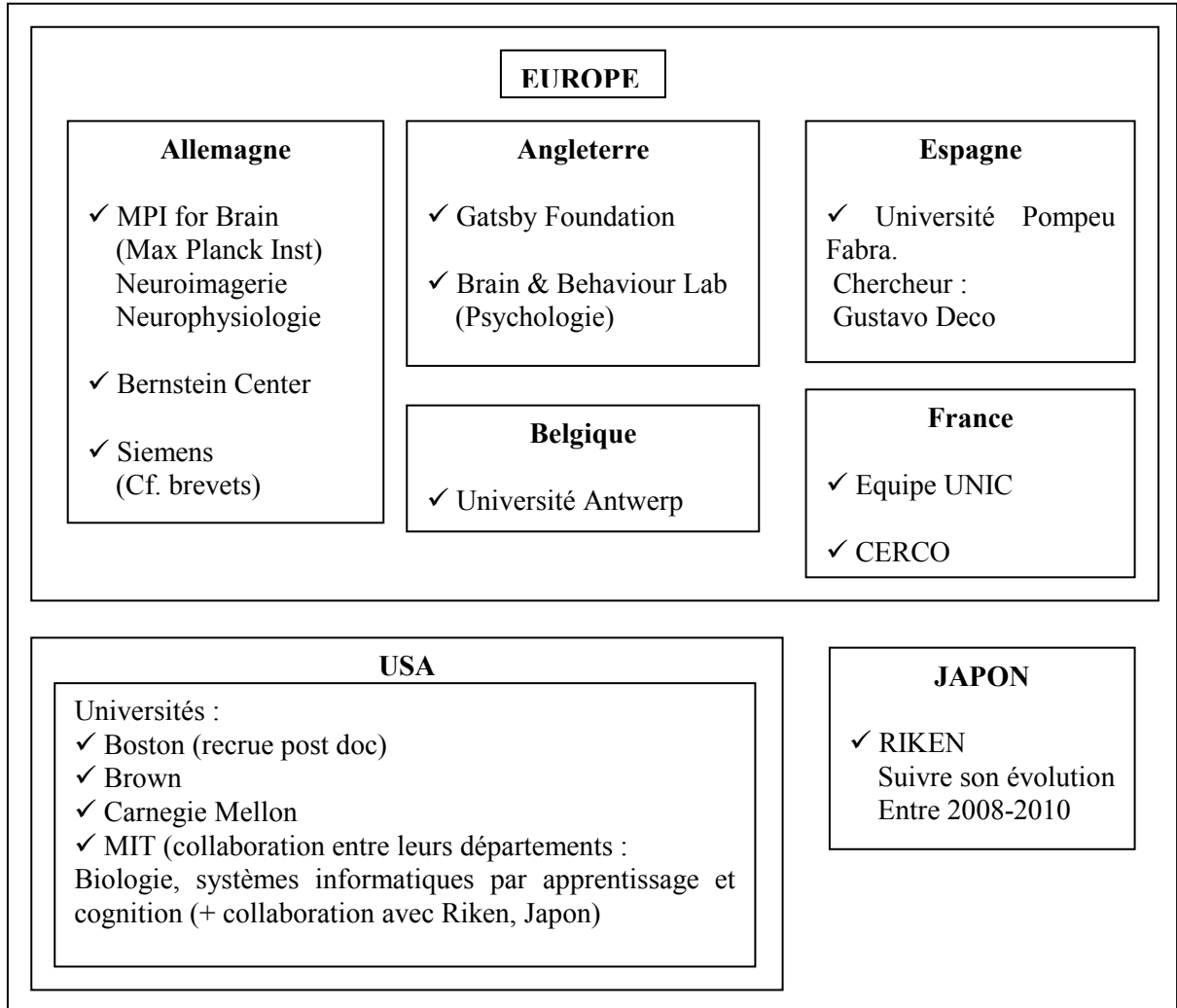


Figure 2- Institutions et laboratoires publiant le plus dans le champ des neurosciences computationnelles

2.1.3. Les collaborations au sein des neurosciences computationnelles

Ce corpus présente une caractéristique intéressante, on remarque que les pays collaborent majoritairement sur leur territoire national, en particulier sur les domaines scientifiques différents et/ou complémentaires alors que les collaborations avec l'étranger s'effectuent en général sur les domaines d'excellence des équipes (les mêmes domaines scientifiques).

Exemples de collaborations entre domaines scientifiques des neurosciences computationnelles :

- * Informatics, System and Computation & Neuroimaging (Canada)
- * Computer Science & Mental Health Studies (3 occurrences, Australia, Brazil, USA)
- * Computational Neuroscience & Psychology, Behavior and Brain (England, Germany, Spain, 6 occurrences)
- * Neuroimaging & Psychology (Scotland)
- * Anaesthesia Neuroscience & Biomedical Imaging & Learning and Memory & Electric Engine (USA, Harvard)
- * Neural Network & Brain Imaging (Canada)
- * Computational Neuroscience & Biology (Germany, 2 occurrences)
- * Computational Neuroscience & Medicine (Israel)
- * Neuroscience & Cognition Science & Brain & Biology (USA, MIT & Cambridge)
- * Neuroscience & Neural Surgery (USA, Portland University).
- * Neuroscience & Physics (USA, Brown University)

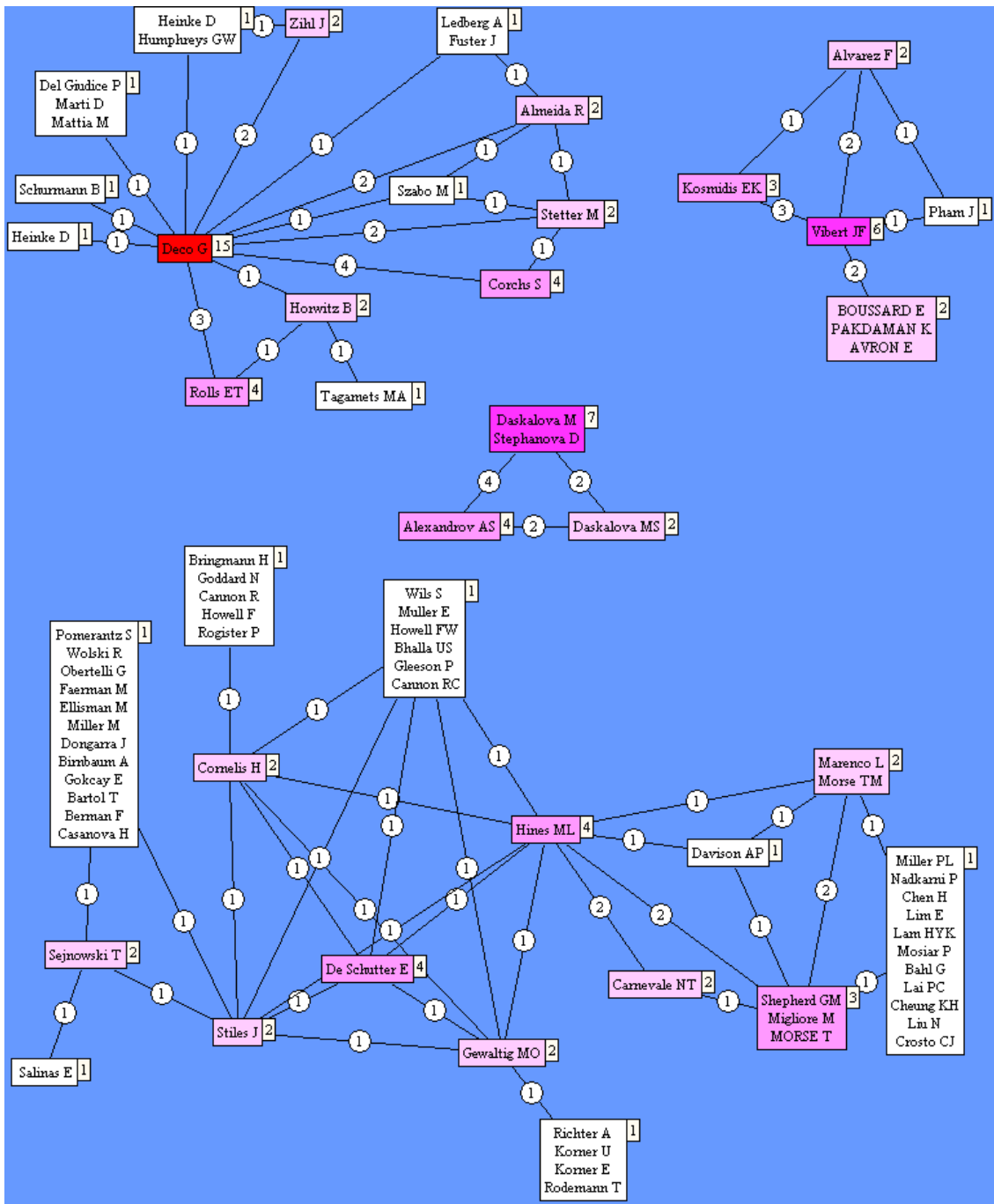


Figure 3- Collaborations entre auteurs (supérieures à 2)

Informations complémentaires pour la lecture du graphique:

- ✓ Une étiquette avec plusieurs noms d'auteurs signifie que ces auteurs ont coécrit leurs publications et qu'ils n'ont pas de publication en nom propre indexée dans la base Dialog).
- ✓ La couleur des étiquettes : blanc (1 publication), rose pâle (2 publications), rose (3 à 4 publications), fuchsia (7 publications), rouge (11 publications).
- ✓ Les bulles blanches chiffrées qui relient les auteurs représentent le nombre de copublications.

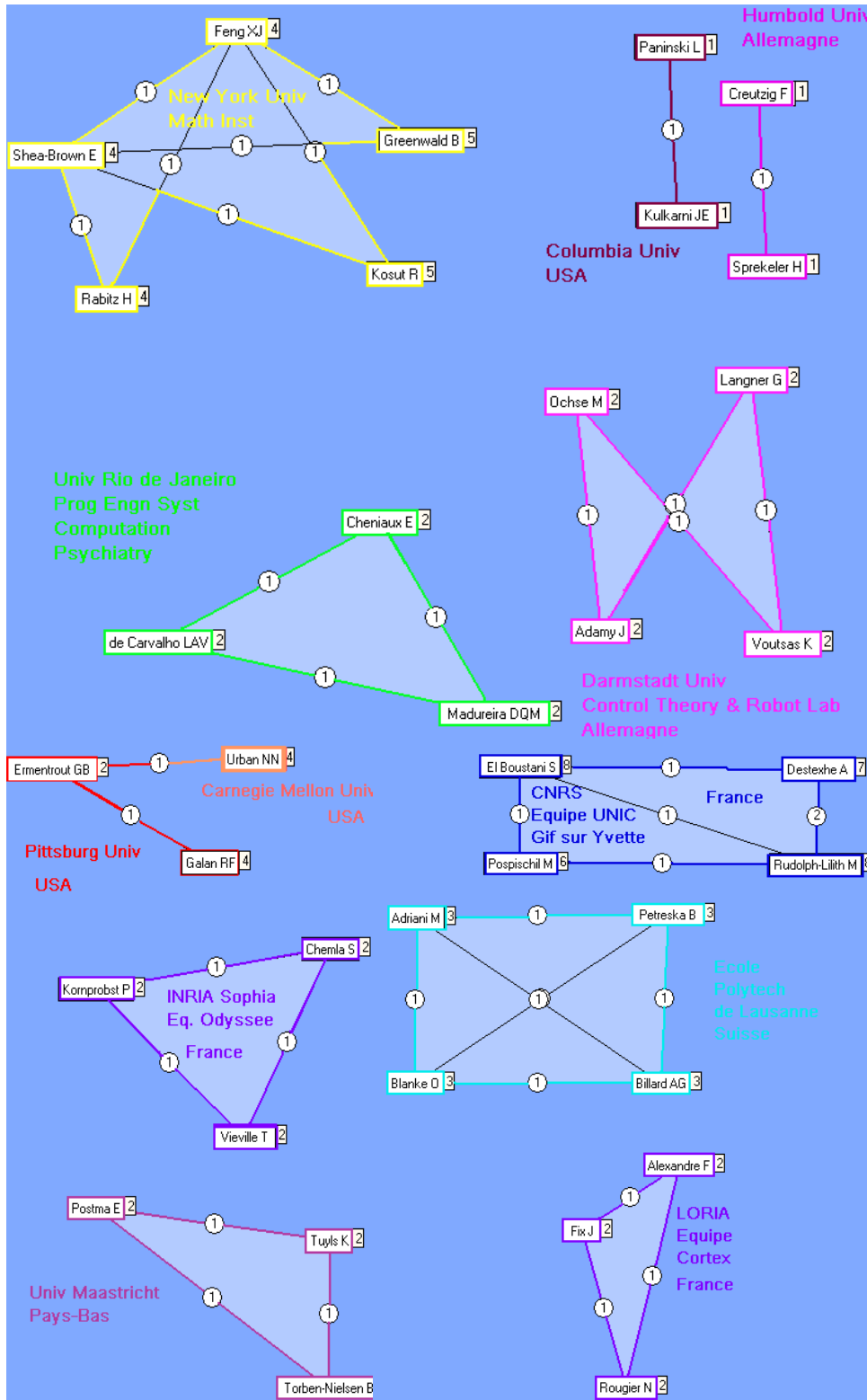


Figure 4- Collaborations entre auteurs (1 seule collaboration)

2.2. Mesure et compréhension de la production scientifique à partir de données bibliométriques : quelques comparaisons.

2.2.1. Comparaison de corpus

Etudier l'évolution d'une discipline à partir d'une étude bibliométrique est un bon point de départ pour faire un état de l'art de l'existant. Il faut tout de même noter qu'appeler une base de données spécifique pour réaliser une étude bibliométrique peut rendre compte d'une réalité partielle. La création d'un autre corpus peut s'avérer utile pour vérifier les informations obtenues ou pour comparer l'étendue et la couverture des bases de données.

Un deuxième corpus a donc été créé sous ISI Web Of Knowledge qui est un produit Thomson offrant un accès à plusieurs bases de données (c'est donc une méta base de données).

Dans notre étude bibliométrique sur les neurosciences computationnelles, les bases appelées via Web Of Knowledge sont : Web Of Science, Medline, ISI Proceedings. Chacune de ces bases possède ses caractéristiques : Web Of Science est une base pluridisciplinaire qui couvre les domaines scientifiques des sciences pures (mathématiques, physique, biologie, informatique, médecine,...) ainsi que les sciences humaines (psychologie, communication, sociologie,...) et les sciences de l'ingénieur (ingénierie mécanique, électrique, électronique, informatique,...). Medline est la base de la librairie nationale américaine de médecine, elle touche au domaine médical, biomédical et biologique. La base ISI Proceedings regroupe les publications des conférences dans tous les domaines scientifiques.

La requête lancée dans ISI Web of Knowledge est la suivante : « computational neuroscience » car la requête « neuroscience » & « comput » avait généré trop de bruit (2361 notices). Le corpus créé sous ISI Web of Knowledge représente 340 notices.

Nous verrons ici seulement l'information supplémentaire obtenue grâce à la création d'un deuxième corpus neurosciences computationnelles.

Au niveau des principaux journaux, nous avons vu apparaître en deuxième position le « Journal of Computational Neuroscience » avec 28 publications indexées dans la base ISI (contre seulement 7 dans Dialog). Globalement, nous avons pu à partir de cette étude sous ISI Web of Knowledge que la couverture de la base était suffisante étant donné que les informations trouvées dans les deux corpus sont très proches.

Cependant, si on opère un focus sur les auteurs et le nombre d'articles indexés dans les bases on remarque que le nombre d'articles par auteurs peut être fluctuant d'une base de données à une autre (Cf. tableau page suivante). Ayant constaté ceci, nous avons créé rapidement un corpus sous la base pluridisciplinaire Scopus d'Elsevier pour voir si des différences étaient observables sur le nombre de publications par auteurs.

Table 1- Nombre de publications par auteurs dans les bases de données Dialog, ISI Web of Knowledge et Scopus à partir de la requête : computational neuroscience

	Corpus Dialog	Corpus ISI	Corpus Scopus
Deco G	15	16	14
Stephanova D	7	7	8
Daskalova M	7	5	5
Vibert JF	6	6	6
Kawato M	2	5	5
Hines ML	4	5	4
Rolls ET	4	5	5
De Schutter E	4	4	7
Corchs S	4	4	4
Bower J	4	4	4
Morse T	3	4	4
Gerstner W	3	4	3
Choi CTM	3	4	6
Migliore M	3	3	2
Kosmidis EK	3	3	3
Carnevale NT	2	3	4
Ermentrout GB	2	4	2
Destexhe A	2	4	4
Barbieri R	1	3	2

Le tableau ci-dessous a pour vocation de montrer qu'à partir d'une requête comportant les mêmes mots-clés, le nombre de publications par auteurs peut être différent d'une base de données à une autre. Les explications de ce phénomène concernent la couverture des bases de données. Dans chacune de ces bases, le poids des journaux n'est pas identique, nous l'avons vu dans l'exemple du journal of Computational Neuroscience qui possédait 7 publications indexées dans la base de données Dialog contre 28 dans la base ISI Web Of Knowledge.

Cependant, il n'est pas certain qu'encourager une multiplication des comparaisons soit générateur d'informations à valeur ajoutée car une comparaison rapide avec la base Scopus donne des résultats similaires (342 notices de publications scientifiques indexées), les principaux journaux, éditeurs et auteurs sont les mêmes.

Nous pouvons conclure qu'une étude bibliométrique est représentative d'une production scientifique à un moment T. Par représentation, on comprend qu'il ne s'agit pas de la globalité de la production scientifique mais d'une partie que l'on considère comme la représentant.

Pour pouvoir faire une étude bibliométrique la plus complète d'une discipline, il faudrait alors interroger plusieurs bases de données (publications scientifiques, thèses, masters, rapports), nettoyer et harmoniser les différents champs (auteurs, affiliations, titre,...) des notices fournies par les différentes bases de données, ôter les doublons et à partir de ce travail, créer un « méta corpus » pour réaliser une étude bibliométrique ayant pour finalité un état de l'art.

2.2.2. Mise à l'écart du Journal Neurocomputing dans le corpus Dialog

Suite à la demande des chercheurs, il a été demandé de mettre à l'écart le journal Neurocomputing considéré comme un journal de faible qualité du fait de la publication de nombreuses conférences et d'articles non « reviewés ». Le corpus de référence pour cette comparaison est le corpus Dialog.

Voici les différences constatées entre le corpus d'origine et celui sans le journal Neurocomputing :

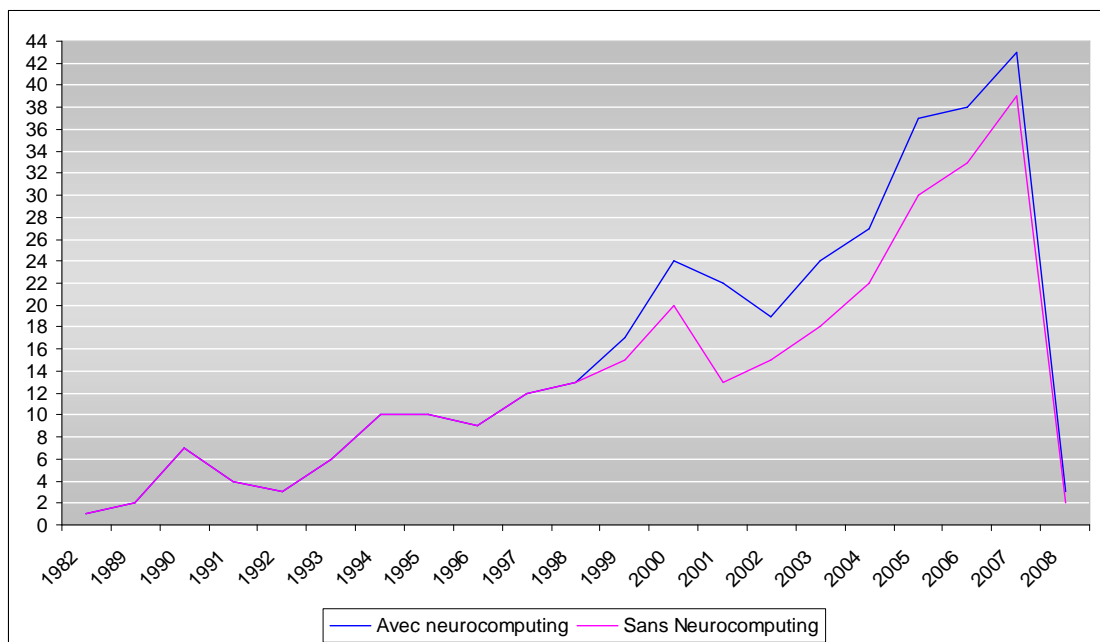


Figure 5- Tendances de la production scientifique avec et sans Neurocomputing

La production scientifique si l'on ôte le journal Neurocomputing (courbe rose) est minorée cependant les tendances sont quasi identiques (les mêmes pics et creux).

- Au niveau des éditeurs, la classification reste la même étant donné que Elsevier est le seul éditeur de la revue Neurocomputing qui passe à 45 publications indexées au lieu de 75 publications, il reste cependant le premier éditeur en neurosciences computationnelles.
- Au niveau des réseaux d'auteurs, le réseau Vibert disparaît car ce réseau n'a publié que dans cette revue et de nouveaux auteurs apparaissent (Cf. graphique page suivante)

Ces comparaisons nous apportent peu d'informations sur le corpus dont nous n'avions pas connaissance auparavant. Par contre, il est vrai que ces comparaisons ont permis une meilleure visibilité des réseaux d'auteurs.

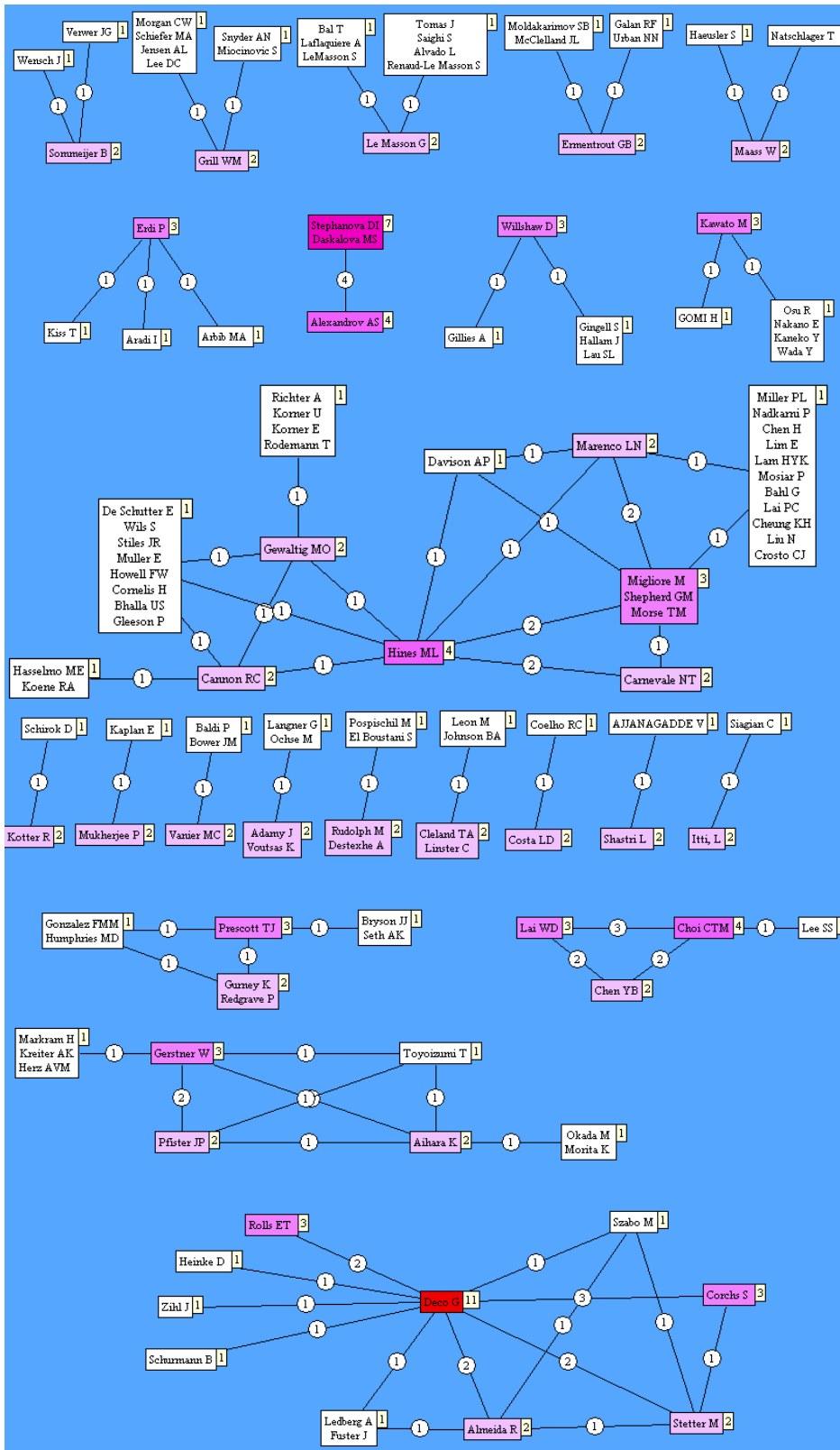
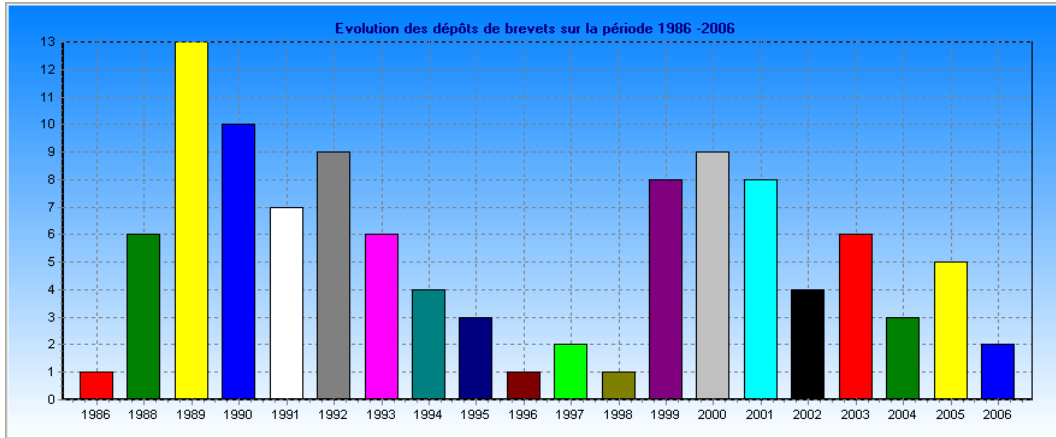


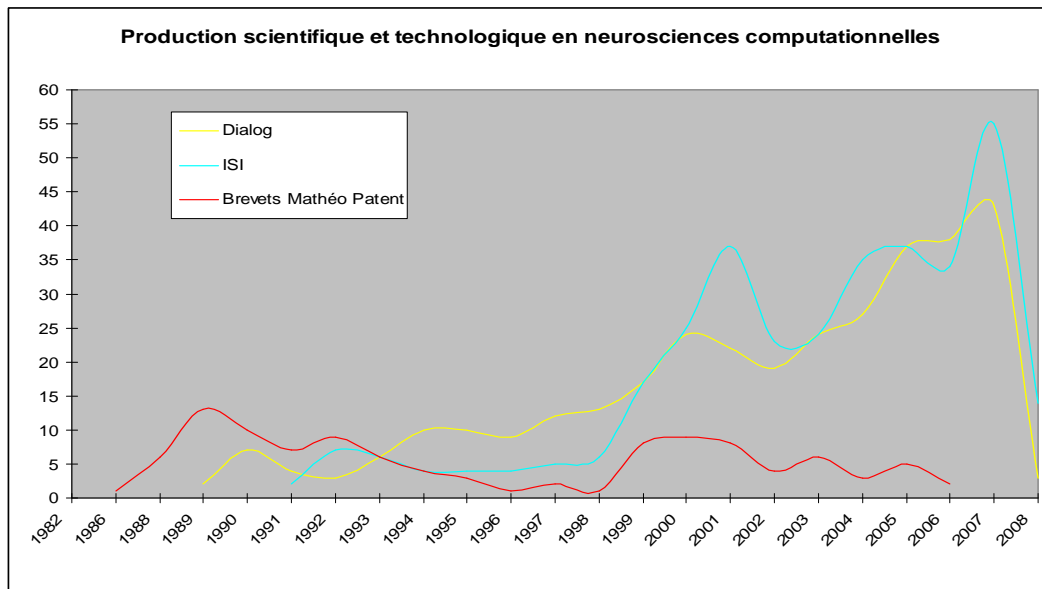
Figure 6- Réseau d’auteurs dont les collaborations sont supérieures à 1

2.3. Etude bibliométrique des brevets en neurosciences computationnelles



Le premier dépôt de brevet a eu lieu en 1986 soit 4 ans après la première publication scientifique en neurosciences computationnelles. On remarque un fort engouement au début des neurosciences computationnelles avec 13 brevets déposés en 1989 alors que le nombre de publications en neurosciences computationnelles s'élevaient seulement à 3 en 1989.

Entre 1989-1998, un fort déclin de la production technologique est observable (la production passe de 13 brevets / an à 1 brevet / an). En 1999, on assiste à un léger regain de l'activité R&D (Recherche et Développement) dans le champ des neurosciences computationnelles. Depuis 2001, on observe une tendance à la baisse avec un rythme saccadé.



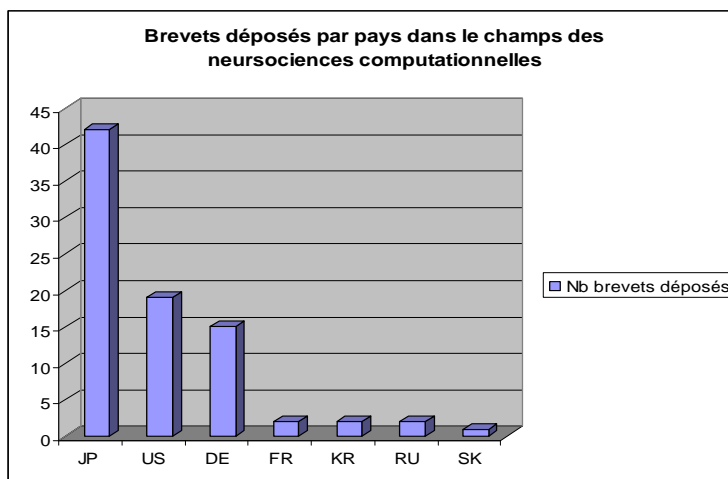
Si l'on compare la production technologique (en rouge) à la production scientifique, on remarque que les brevets sont en forte baisse contrairement à la production scientifique qui elle est en plein essor.

En 1989, il y avait 13 brevets déposés en neurosciences computationnelles alors que le nombre de publications a été atteint seulement en 1997 (soit 8 ans plus tard).

La baisse du nombre de brevet est difficile à interpréter ne sachant pas s'il s'agit d'un désintérêt des industriels ou s'ils n'ont pas pu obtenir les résultats escomptés. La question reste en suspens. Nous ne pouvons pas non plus estimer à partir des données existantes l'évolution de la production technologique en neurosciences computationnelles. De plus, il faut prendre en compte que toute demande de brevet suppose une latence de 18 mois entre la date de dépôt du brevet et sa publication. Nous avons donc qu'une partie des brevets 2006 et pas encore ceux de 2007 (sauf les demandes internationales (WO), qui sont en général des extensions de brevets déjà déposés).

2.3.1. Les principaux déposants du corpus sont :

Principaux pays :



Le **Japon** arrive en tête avec un total de 42 brevets déposés entre 1990 et 2001. Les brevets japonais sont majoritairement déposés sur le territoire japonais.

Le nombre de dépôts de brevets à l'extérieur du Japon est de 8 (4 aux Etats-Unis et 4 brevets européens). Par contre, ils n'ont fait aucune demande de PCT (couverture mondiale).

On peut supposer que le domaine des neurosciences computationnelles est actuellement en déclin au Japon car il n'y a pas eu de dépôt de brevets japonais depuis 2001.

En deuxième position se trouve les **Etats-Unis** qui ont déposé 18 brevets entre 1994 et 2007 répartis ainsi :

Les entreprises ont déposé 9 brevets en tout (7 sur le sol américain et 2 en PCT)

Les universités ont déposé 6 brevets (5 sur le sol américain et 1 en PCT)

Les particuliers ont déposé 6 brevets (5 aux USA et 1 en PCT).

Enfin l'**Allemagne** arrive en 3^{ème} position avec 14 dépôts de brevets, dont 10 ont été déposés sur le territoire allemand et 4 en PCT.

Sur les 10 brevets nationaux, 6 ont été déposés par des inventeurs en leur nom propre et les brevets restants ont été déposés par l'entreprise Siemens AG (4 en Allemagne et 4 en PCT) sur la période 1998-2007. Ainsi l'entreprise Siemens AG semble bien investie dans ce domaine de recherche.

Les autres pays ayant déposé un brevet dans les neurosciences computationnelles sont :

La France (2 brevets en PCT datant de 1998 et 2007), la Corée (2 brevets nationaux en 2000 et 2005), la Russie (2 brevets nationaux en 1995 et 2005), et la Slovaquie (1 seul brevet national en 1997).

Les principales institutions (entreprise et universités) déposantes sont :

Fujitsu Ltd (JP), 18 brevets (1990-1998)
Hitachi Ltd (JP), 9 brevets (1990-2001)
Siemens AG (DE), 9 brevets (1993-2007)
California Institute of Technology (US), 2 brevets (2001 et 2004)
Mitsubishi Corp (JP), 2 brevets (1990 et 1994)
Nippon Telegraph & Telephone (JP), 2 brevets (1990 et 1991)
Mendensha Electric MFG CO Ltd (JP), 2 brevets (1991 et 1992)

Les particuliers (« unassigned ») ont déposé 18 brevets (entre 1996 et 2008).

Tous les autres déposants (entreprises et universités) ont un seul brevet à leur actif.

Si l'on compare ces données avec les informations issues de l'étude bibliométrique des corpus scientifiques en neurosciences computationnelles, on remarque que certains organismes sont présents dans les 2 corpus, il s'agit de :

- Siemens AG (avec Deco Gustavo)
- California Institute of Technology (appelé Caltech dans le corpus scientifique)
- Brown University

Les seules collaborations entre institutions observables sont des collaborations internes aux entreprises, par exemple : les laboratoires R&D français et néerlandais de Philips Electronics qui ont co-déposé un brevet. Le peu de collaborations constatées peut s'expliquer du fait des intérêts économiques engendrés par le dépôt d'un brevet.

Le brevet confère au déposant une fois validé et publié, un monopole sur l'exploitation commerciale de l'invention pour une durée maximale de 20 ans à condition qu'il paie ses annuités. Le brevet est concédé pour un territoire désigné par le déposant, ce peut-être : un pays, une zone par exemple Europe ou Etats-Unis ou une multitude de pays dans le cas des brevets déposés en PCT.

2.3.2. Codes CIB : tendances de l'innovation en neurosciences computationnelles

Table 2- Codes CIB, innovations et tendances

Code CIB	Catégorie de la CIB	Tendance (évolution du nombre de brevets)
G09 G09B	Matériel éducatif ✓ Matériel de communication destiné aux aveugles, sourds et muets. ✓ Graphiques, diagrammes, globes	2002-2006 : Naissance (5 brevets) 2006 : En Stand-by
G06	Calcul, comptage	
G06T	Traitement ou génération de données d'image (modification d'image, modélisation 2D et 3D (2 dimensions et 3 dimensions))	1991-2001 : en augmentation 2001-2008 : en déclin
G06N G06N7 G06N5 G06N3	Systèmes de calculateurs basés sur les modèles mathématique, biologique et sur la connaissance. Modèles mathématiques Modèles basés sur la connaissance Modèles biologiques (réseaux neuronaux, méthodes d'apprentissage, simulation)	Pas de dépôt de brevet depuis 1994 1 brevet en 2006 Forte production entre 1990 et 1996, déclin de 1998-2002, à partir de 2003 ré-augmentation de la R&D (Recherche et Développement).
G06K G06K9	Reconnaissance, présentation des données et supports d'enregistrement. Reconnaissance des caractères imprimés, des formes (captation, extraction, reconnaissance et correction de l'image par des moyens électroniques et/ou optiques)	En stand-by en R&D dans les neurosciences computationnelles depuis 2001.
G06G7	Calculs effectués à partir de variations électriques ou magnétiques.	Forte production R&D jusqu'en 1996, puis fort ralentissement. (5 brevets entre 2000 et 2005)
G06F G06F19 G06F17 G06F15 G06F15 /18	Traitement électrique de données numériques Equipements et/ou calculs numériques destinés à des fins administratives, financières, de prévision et/ou de surveillance. Méthodes de calcul et traitement de données (équations, fonctions, codes, statistiques, ...) Calculateurs, équipement de traitement de données. Apprentissage (programmes et machines auto apprenantes)	En augmentation depuis 2001. 2 brevets (2005 et 2006) Forte production R&D de 1990 à 1996, en déclin aujourd'hui.
G05 G05B	Commande, régulation Dispositifs de contrôle, d'essai pour systèmes de commande et de régulation	3 brevets en 2005
A61N1	Electrothérapie (usage sur le corps ou in vivo, exemple : les électrodes).	2 brevets (2006 et 2003).
A61B5 /04	Mesures de signaux bioélectriques du corps	Forte production R&D depuis 2005

2.3.3. Technologies et acteurs

Nous pouvons proposer le tableau suivant pour représenter les domaines, technologies et acteurs des neurosciences computationnelles à partir du corpus créé avec l’outil Mathéo Patent.

Table 4- L’innovation par acteurs et par domaine

DOMAINE	TECHNOLOGIE	ACTEURS R&D	PAYS
Réseau neuronal	Statistiques, pronostic, réseau multicouche	Siemens AG	Allemagne
	Classification d'image, création d'arborescence	France Telecom	France
	Prédiction d'événements	Ramot at Tel Aviv University	Israel
	Contrôle des machines	Unassigned	Allemagne
Activité neuronale	Analyse des signaux et des stimuli (encéphalographie)	Siemens AG Brown University California Institute of Technology	Allemagne US
	Simulation, modélisation	George Masson Int. Property Unassigned Unassigned	US Allemagne
	Stimulation de neurones et signaux électriques	Science Application Int Corp	US
	Appareil de contrôle de l'activité neuronale	University of Pennsylvania	
Diagnostic / Médecine	Diagnostics des troubles neurologiques à partir des ERP	Thuris Corp	US
	Surveillance neurophysiologique	Unassigned	
	Diagnostic de maladies chez les animaux	Unassigned	Allemagne
	Tests cognitifs et mesure des temps de réaction	CNS Vital LLC Corp	US
	Prothèse musculaire contrôlée par le cerveau	California Institute of Technology	
DOMAINES A LA PERIPHERIE			
Reconnaissance	Vocale	Siemens AG Unassigned	Allemagne
	Formes et signaux	Canon	Japon
Connaissance	Gestion des connaissances	IP Century	Allemagne
Expérimentation	Obtention d'informations multi-niveaux par contrôle du voltage	Matsushita	Japon

CONCLUSIONS SUR LES ÉTUDES BIBLIOMÉTRIQUES

Les études bibliométriques nous ont permis de tirer les conclusions suivantes :

- Une base de données appropriée est représentative de la production scientifique d'un domaine à un moment T
- Interroger plusieurs bases de données permet de pouvoir comparer les résultats. Quelques différences existent entre les corpus créés, notamment sur le nombre de publications par auteurs. Cependant dans le cadre d'une étude bibliométrique scientifique qui a pour but d'identifier les tendances scientifiques, cet écart ne justifie pas la création d'un même corpus sous différentes bases de données à moins qu'il s'agisse d'identifier la base de données la plus adéquate pour un domaine de recherche.
- De plus, une étude bibliométrique ayant pour but d'établir l'état de l'art d'un domaine scientifique peut justifier la création d'un « méta corpus », c'est-à-dire d'un corpus créé à partir de plusieurs bases de données dont les notices auraient été au préalable « dé doublonnées » avant le traitement du corpus via un outil bibliométrique.
- Les résultats obtenus avec les corpus Dialog et ISI Web Of Knowledge nous permis de comparer les deux bases (Cf. tableau ci-dessous).

	DIALOG	ISI Web Of Knowledge
Interrogation des bases	Plus précis grâce à l'opérateur (N) ⁹ de Dialog	
Traitement du corpus avant importation dans l'outil Mathéo.	Création de macros (7 au total pour harmoniser les notices)	Aucun
Le corpus	Meilleure couverture temporelle (à partir de 1982)	Meilleure couverture à partir de l'année 2000
Export du corpus	Notices complètes	Perte des champs affiliations, mots-clés auteurs et domaines scientifiques
Richesse des corpus	Meilleure couverture temporelle	Mise en lumière d'un journal et d'un réseau d'auteurs.

Grâce aux comparaisons entre ces deux bases de données, nous savons aujourd'hui que la couverture de la base ISI Web of Knowledge est suffisante bien que la base Dialog du fait de la base Dissertation Abstract (thèses américaines et une partie des thèses européennes) offre une plus grande couverture temporelle.

⁹ L'opérateur (N) signifie : recherche ces deux mots ensemble sans aucun mot entre eux.

Exemple : computational (N) neuroscience signifie rechercher « computational neuroscience » ou « neuroscience computational »

Aux questions posées avant la réalisation des études bibliométriques, nous pouvons répondre que les chercheurs de l'équipe Cortex ont montré beaucoup d'intérêt pour les études bibliométriques, ils ont qualifié ces études de « photo instantanée » de la production scientifique et technologique. Les conclusions tirées par l'équipe Cortex sont que le champ de recherche des neurosciences computationnelles est un domaine de recherche se définissant « de manière intrinsèque », qu'il nécessiterait un travail collectif des chercheurs du domaine pour définir les grandes questions scientifiques relatives aux neurosciences computationnelles. Par ailleurs, l'étude bibliométrique a permis de mettre en lumière l'évolution de la production scientifique des neurosciences, de voir les principaux journaux éditeurs dans le domaine bien qu'un bémol ait été souligné par les scientifiques concernant l'approche quantitative pour qualifier les journaux. En effet, le premier journal (Neurocomputing qui représente 10% du corpus et publie essentiellement des actes de conférences) occupe une place prépondérante dans la diffusion de la production scientifique alors que ce journal n'est pas considéré par les experts du domaine comme le journal phare des neurosciences computationnelles. Ainsi les résultats quantitatifs obtenus sur les journaux sont à soumettre aux scientifiques afin que les experts puissent qualifier les informations des études bibliométriques. Les experts ont considéré que le corpus d'environ 356 notices est « représentatif » de la production scientifique des neurosciences computationnelles, cependant ils ont précisé que pour réaliser un état de l'art complet des neurosciences computationnelles il faudrait élargir la requête d'interrogation des bases de données. Ils présupposent la taille du corpus permettant de réaliser l'état de l'art 10 fois supérieure à celle traitée (soit environ 3500 notices).

Suite aux conclusions des chercheurs sur la taille du corpus, nous allons relancer une étude bibliométrique scientifique sur une autre base de données afin de pouvoir confirmer ou contester l'hypothèse d'un corpus dix fois supérieur. Cette hypothèse suppose que le choix des bases de données n'a pas été optimum ou que les mots-clés utilisés dans la requête ne soient pas suffisamment révélateur du domaine des neurosciences computationnelles.

En conclusion de cette expérience, le service IST pourrait donc proposer des études bibliométriques scientifiques car elles sont dans la continuité du métier de documentaliste habitué à interroger différentes bases de données, ayant une connaissance des thèmes et équipes de recherche, des journaux scientifiques et principaux éditeurs.

Le développement d'une telle méthode au sein d'un service IST suppose une participation des équipes de recherche au sein du processus bibliométrique pour l'analyse et la validation du corpus ainsi que pour la qualification des informations obtenues via les traitements statistiques et mathématiques. Cette activité sous-tend une forte collaboration entre fournisseurs et commanditaires des études bibliométriques. Une communication et valorisation des informations à double sens serait la clef de voute d'une bonne étude bibliométrique.

BIBLIOGRAPHIE

E. Giménez Toledo, A. Romàn Romàn, H. Rostaing, *Analyse du transfert de l'information scientifique entre le secteur public et privé. Etude des co-publications dans les revues espagnoles*, in : Information Science for Decision Making, n° 6, 2003.

Y. Gingras, *La fièvre de l'évaluation de la recherche. Du mauvais usage de faux indicateurs bibliométriques*, Note de recherche du CIRST, Mai 2008.

A-M. Kemarrec, E. Faou, J-P. Merlet, P. Robert, L. Segoufin, *Que mesurent les indicateurs bibliométriques?* , Document d'analyse de la commission d'évaluation de l'INRIA, Septembre 2007.

J.-D. Pierret, F. Dolfi, L. Quoniam, E. Boutin, *Découverte de connaissances dans les bases de données bibliographiques. Modèles expérimentaux autour de la première hypothèse de Swanson*, ISDM N°20, 2005.

Y. Okubo, *Indicateurs bibliométriques et analyse des systèmes de la recherche : Méthodes et exemples*, Document de travail sur la science, la technologie et l'industrie, Editions OCDE, 1997, pp3-31, 44, 47, 54.

P. Volland-Nail, P. Monget, J. Saumande, F. Laville, C. Paoli, *L'apoptose : apport des outils bibliométriques de la station ATLAS pour l'étude du sujet et l'orientation des recherches*, in : Journées d'Etudes de la Société Française de Bibliométrie appliquée sur Les systèmes d'Information Elaborée : Analyses et Produits Bibliométriques, Colloque Ile Rousse, France, 9-11 Juin 1993.

ANNEXE : VEILLE DEFINITION ET TYPOLOGIE

Pour appréhender les évolutions rapides des marchés, des technologies, des sciences, les entreprises ont mis en place dans les années 1980 un processus de management de l'information. L'information est une notion à multiple facette possédant plusieurs définitions. L'information peut-être définie comme l'action d'informer, comme une donnée utile, comme « *une donnée acquise d'une source* »¹⁰, elle peut être aussi perçue comme vecteur de pouvoir dans les organisations.

La définition la plus adéquate dans le contexte de la veille semble être une ou plusieurs données traitées et analysées dans le but de répondre à une question. Manager l'information signifie donc analyser et utiliser les informations dans un but d'action.

La veille peut être définie comme un processus permettant la compréhension et la surveillance de l'environnement afin de répondre à la question : qui fait quoi, comment et où dans le but de prendre une décision ou de planifier une action.

En veille, il existe trois niveaux d'accès à l'information :

- ✓ Les informations blanches sont les informations qui sont accessibles à tous de manière simple et licite, par exemple : les informations diffusées par les médias, les rapports annuels des grandes sociétés cotées en Bourse, les textes juridiques publiés sur les sites gouvernementaux.
- ✓ Les informations grises sont elles aussi licitement accessibles mais le nombre de personnes les connaissant est plus restreint, il s'agit des bulletins des Ambassades Françaises, les thèses des universités et centres de recherche, les brevets, (...).
- ✓ Les informations noires sont à diffusion restreinte, par exemple les informations internes d'une entreprise, les secrets de fabrication d'un produit, (...). Elles sont protégées et leur collecte de manière frauduleuse est illicite.

La veille s'effectue au niveau des informations blanches et grises. L'accès aux informations noires n'est pas du ressort du veilleur mais de l'espionnage ou du contre espionnage. Ces deux métiers sont différents ainsi que leurs méthodes.

On associe à la veille, deux types de sources :

- ✓ Les sources formelles qui sont des sources validées (souvent payantes) et classées par domaines (technique, juridique, scientifique,...), ou par types de documents (colloques, ouvrages,...), ou par type de support (papier, électronique,...), (...).
- ✓ Et les sources informelles qui par nature ne sont pas fixées sur un support (papier ou électronique) mais qui proviennent des personnes, des salons, des colloques. Les informations issues des sources informelles peuvent être validées en étant recoupées avec des informations formelles. Elles peuvent être formalisées sur un support comme par exemple un rapport d'étonnement (dans le cas de visites en pays étranger) ou de comptes rendus.

¹⁰ Cf. Norme XP X50-053, Prestation de veille.

1. La typologie

Parler de veille implique la connaissance de la typologie. Il existe plusieurs types de veille : la veille scientifique, technologique, concurrentielle, commerciale, juridique, sociale, sociétale et environnementale. Nous aborderons dans la cas présent les veilles scientifique et technologique.

➤ La veille scientifique

Elle a pour but d'identifier les nouveaux sujets de recherche, de connaître les organismes et équipes concurrents, d'identifier les nouvelles publications et ouvrages relatifs à une thématique de recherche et de savoir quelles sont les colloques et conférences où il faut être présent. Elle permet à une équipe de pouvoir s'assurer du caractère innovant de ses recherches, de bénéficier des travaux de recherche effectués par d'autres pouvant lui apporter des connaissances, elle peut aussi dans le cas où d'autres équipes travaillent sur les mêmes thématiques permettre l'élaboration d'un partenariat ou un changement d'axe de recherche. Elle est un outil stratégique offrant une meilleure vision de l'environnement scientifique sur une thématique donnée. Elle peut aussi dans le cas d'un domaine de recherche récent permettre un état de l'art des recherches.

Les sources formelles de la veille scientifique sont les bases de données payantes, (par exemple : ISI Web of Science, Dialog, Scopus), les grands organismes de recherche (CNRS, CEA, INSERM, INRA, INRIA,...), les pages Web des équipes de recherche, les pages Web ou blogs des chercheurs, les sociétés savantes, les communautés scientifiques, les articles de publications scientifiques, les ouvrages scientifiques, les conférences, les bulletins des Ambassades de France et éventuellement les brevets.

Les sources informelles sont les colloques et conférences, les discussions entre chercheurs, la présence de chercheurs dans les comités de lecture, les universités et grandes écoles.

Ce type de veille est réalisé par les scientifiques et en général le centre de documentation.

➤ La veille technologique

Elle a pour finalité la création ou l'amélioration des produits, la connaissance des projets concurrents et des entreprises dynamiques du domaine. Elle permet d'identifier les innovations passées et en cours afin de connaître les tendances technologiques et d'évaluer les tendances futures, de transférer une technologie d'un secteur d'activité à un autre ou d'effectuer un benchmarking (comparer les sociétés, leurs domaines d'activités et leurs innovations technologiques).

Les sources formelles sont les bases de données brevet¹¹, les publications scientifiques, les salons, les bulletins des Ambassades Françaises, les journaux spécialisés, les comptes-rendus des salons, les webzines, les newsletters spécialisées.

¹¹ Les trois bases de données brevet les plus importantes sont :

Espacenet, base de données européenne des brevets, Cf. : <http://ep.espacenet.com/>

USPTO, US Patent Trade Office, brevets américains, Cf. : <http://www.uspto.gov/patft/index.html>

WIPO, World Intellectual Property Office, brevets à couverture mondiale Cf. : <http://www.wipo.int/pctdb/en/>

Les sources informelles de la veille technologique sont les experts du domaine (chercheurs et ingénieurs R&D), les fournisseurs, les employés des firmes concurrentes, les salons.

La veille technologique peut être réalisée par le centre de documentation assisté par les ingénieurs du service R&D (Recherche et Développement) et de la propriété intellectuelle dans le cas d'étude de brevets et les scientifiques experts du domaine.

2. La veille, un processus

La veille est un processus symbolisé par la définition d'axes de surveillance, la collecte des informations recherchées, la validation, le traitement et l'analyse des données, suivie de la diffusion des résultats.

1. Définition des axes de veille

Plusieurs méthodes existent pour définir les axes de surveillance. Nous aborderons les travaux de Porter qui a proposé un modèle intitulé « les forces de la concurrence » permettant de modéliser l'environnement de l'entreprise en le divisant en cinq catégories :

- ✓ Les concurrents de l'entreprise (qui proposent un produit similaire)
- ✓ Les clients qui peuvent négocier les prix ou s'intégrer dans la filière en amont (c'est-à-dire devenir producteur)
- ✓ Les fournisseurs qui peuvent développer leurs activités en aval et devenir concurrents ou augmenter les prix des matériaux (influence les coûts d'achat).
- ✓ Les substituts qui sont les entreprises pouvant fournir le même produit à un prix plus compétitif.
- ✓ Les entrants potentiels sont les entreprises pouvant s'intégrer dans le même secteur d'activité ou développer les mêmes produits.

Bien que ce modèle soit approprié au cadre industriel, il semble cependant que dans le cadre d'un centre de recherche, il comporte des limites. Un centre de recherche a pour concurrent les autres centres de recherche et universités ainsi que les laboratoires R&D des grandes entreprises innovantes.

La première mission de l'INRIA est de favoriser l'évolution des connaissances scientifiques et d'assurer un transfert technologique auprès des entreprises. Son but est de produire des innovations scientifiques et technologiques.

2. Collecte des informations

La collecte d'informations s'effectue sur les informations formelles et informelles. Collecter des informations signifie comprendre le domaine que l'on étudie, connaître le vocabulaire, être capable de définir le sujet et les enjeux de la veille à réaliser. La collecte d'informations peut être effectuée à partir des bases de données payantes, des médias, des sites Internet, Les compétences requises sont la connaissance de l'interrogation des bases de données, la construction d'une requête, les outils permettant une collecte ponctuelle en réponse à un problème ou automatique pour le suivi

d'une tendance ou dans le but de prospective (flux RSS, alertes mail, newsletter et outils de collecte).

Les informations obtenues via les diverses sources doivent être qualifiées sur les critères de fiabilité, accessibilité, nouveauté et utilité.

3. Validation de l'information

Le processus de veille est un processus interactif et itératif, ce qui signifie que la richesse d'une veille est assurée par la participation de différents acteurs.

Un veilleur n'est pas forcément l'expert du domaine mais il est l'expert de la recherche d'informations. Dans le cas où le veilleur n'est pas l'expert du domaine, il doit être épaulé par les experts pour valider les informations recueillies et les qualifier (utiles, importantes, intéressantes, pouvant servir, inutiles ou hors champ). Ce travail de collaboration permet au veilleur de pouvoir cibler les informations utiles pour son destinataire et de pouvoir augmenter sa connaissance du sujet traité.

Parler de processus itératif signifie que ce processus est évolutif, qu'il est amené à s'améliorer. Il fonctionne un peu à la manière d'une macro avec des boucles conditionnelles :

✓ « if (...) end if » : si la condition (...) est remplie, alors (...)

✓ «Do while i =(...) loop » : tant que la condition n'est remplie pas alors effectue (...)

Telle une « méthodologie expérimentale » que l'on teste et que l'on améliore au fil des usages. Dans l'exemple, ci-dessus la condition serait tant que l'information n'est pas jugée pertinente par le destinataire, alors le veilleur doit continuer ses recherches en affinant ses requêtes et/ou en modifiant ses sources.

4. Traitement et analyse des données

Cette étape intervient après la validation des données ou informations collectées. Il s'agit de créer du sens à partir d'une multitude d'informations pouvant provenir d'une seule ou plusieurs sources.

Le traitement des données sera dépendant des sources. Dans le cas de sources hétérogènes en provenance du Web par exemple, le veilleur peut éditer une liste de liens (url), peut effectuer une synthèse des documents en citant ses sources, rédiger une newsletter ou encore envoyer un mail selon une certaine fréquence (quotidienne, hebdomadaire, mensuelle, trimestrielle). Dans le cas de source homogène en particulier les bases de données, le veilleur peut opérer un traitement bibliométrique des données car les informations contenues sont normalisées et organisées sous la nomenclature d'une notice. A l'intérieur d'une notice, on trouve le titre (de l'article, de la conférence, du livre), le ou les auteurs, l'affiliation, la date, le nombre de pages, l'ISBN, les mots-clés auteurs, le résumé, la langue... . Certaines bases de données ajoutent leurs mots-clés et les classent par domaines scientifiques.

L'avantage des bases de données réside dans l'organisation des données qui permet un traitement mathématiques et/ou statistiques des données. Effectuer une requête dans une base de données revient à créer un corpus (un ensemble de notices). Ce corpus peut représenter jusqu'à plus de 1000 notices et nécessite un traitement permettant au veilleur de pouvoir prendre en compte les informations utiles et de pouvoir écarter les

informations à la marge. Il existe des outils logiciels appelés outils bibliométriques ayant pour fonction d'assister le veilleur dans cette tâche.

L'analyse des données est effectuée par le veilleur sur les données dites unidimensionnelles (par exemple : l'évolution de la production, les équipes les plus productives, les journaux scientifiques dans lesquels paraissent les articles,...). Les données bidimensionnelles (qui révèlent les liens entre auteurs, laboratoires, thèmes de recherche,...) seront étudiées dans un premier temps par le veilleur et devront être complétées par les renseignements détenus par les experts. Dans la phase d'analyse, l'interaction entre veilleur et experts est essentielle, elle permet un feedback sur le travail effectué et de pouvoir réorienter au besoin la veille jusqu'à ce que l'étude soit validée par les commanditaires.

5. Diffusion des résultats

Une fois les résultats validés, il est important de les diffuser auprès des commanditaires. En termes de diffusion, plusieurs solutions sont envisageables : si la veille effectuée concerne un acteur ou un groupe d'acteurs définis de l'entreprise et qu'il n'est pas utile à d'autres acteurs alors la diffusion peut être restreinte comme par exemple par mail ou sous forme de rapport écrit. Par contre, si les résultats de cette veille peuvent être utilisables par d'autres, le mode de diffusion peut être envisagé différemment. Il peut s'agir d'une réunion entre membres d'un service, d'une fonction GED intégrée au SI de l'entreprise, d'une newsletter à laquelle les destinataires s'abonnent, d'une liste de diffusion, d'un outil de communication de type blog ou d'une plateforme collaborative de type wiki. Ce choix repose sur les manières de communiquer et sur les habitudes de travail des membres de l'organisation. Par exemple, une entreprise dont l'organisation est pyramidale, optera probablement pour des newsletters ou encore des listes de diffusion avec différents niveaux d'informations alors qu'une organisation fonctionnant en réseau optera davantage pour une solution de type collaboratif. Chaque organisation possède une structure organisationnelle propre, une culture d'entreprise, des normes et des valeurs qui font qu'elle est unique. Chaque type de diffusion doit convenir et être en corrélation avec les valeurs et attentes de l'organisation.

La diffusion des résultats pose le problème de l'accès aux ressources et donc du stockage des documents: quelle sera la durée de vie des documents de veille produits?

En résumé, un service ou une fonction de veille que l'on développe au sein d'une organisation doit répondre aux besoins en veille de l'organisation. Pour définir ces besoins, l'utilisation de la typologie de la veille pour identifier les domaines de surveillance est conseillée. La veille en tant que processus recouvre les fonctions de définition d'axes de surveillance, collecte des données, traitement et analyse de données et enfin diffusion des résultats.