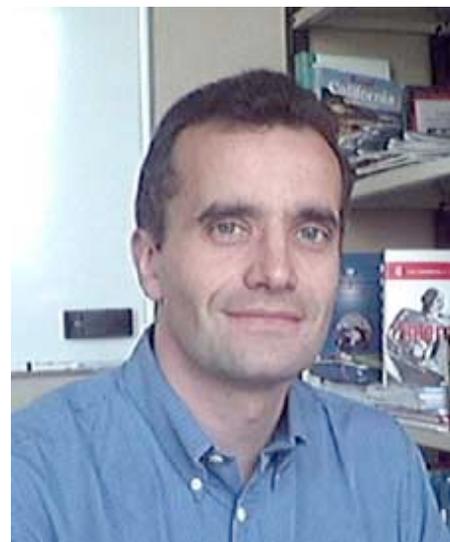


Stéphane Chaudiron (Ministère de la Recherche) : "les tendances de la recherche d'information vont vers le Web sémantique, l'indexation collective et les technologies linguistiques"

[Retour au sommaire de la lettre](#)

Stéphane Chaudiron est chargé de mission au sein du bureau de l'Information Scientifique et Technique du Ministère de la Recherche. Il revient avec nous, au cours de cet interview, sur les projets d'évaluation des outils de veille et de recherche d'information actuellement utilisés par les acteurs du tissu économique et le grand public. Cette évaluation sera, en France, le fondement de la mise en place d'une industrie de confiance pour la recherche de l'information stratégique, dans une perspective d'Intelligence Economique...

A partir de 2007, le Ministère de la Recherche mettra l'accent sur l'évaluation des outils de veille et de recherche d'information utilisés couramment par les acteurs du tissu économique et le grand public. Demain l'évaluation des outils de veille et de recherche d'information sera en France, le fondement de la mise en place d'une industrie de confiance pour la recherche de l'information stratégique, dans une perspective d'Intelligence Economique. Alors, l'évaluation assurera-t-elle à la veille de beaux lendemains ? Stéphane Chaudiron, chargé de mission au sein du bureau de l'Information Scientifique et Technique du Ministère de la Recherche, répond à nos questions sur ce point...



Bonjour Stéphane Chaudiron et merci de bien vouloir répondre à nos questions. Pouvez-vous vous présenter en quelques mots ?

Je suis professeur en Sciences de l'Information et de la Communication et chargé de mission au sein du bureau de l'Information Scientifique et Technique du Ministère de la Recherche. Ce bureau dépend de la DGRI (Direction Générale de la Recherche et de l'Innovation). Je suis responsable du secteur veille stratégique et technologies de la langue.

Quelle est la place de la recherche d'information dans ce contexte ?

Le bureau a deux missions dans le domaine de l'information spécialisée. D'une part, faire en sorte que les entreprises et organismes de recherche publics accèdent à une information de qualité. Cela signifie travailler sur l'offre d'information, notamment auprès des acteurs qui produisent de l'Information Scientifique et Technique. Cela veut dire aussi travailler sur l'accès à cette information et définir les outils les plus performants pour y parvenir. Il s'agit là de notre deuxième mission.

Au sein du nouveau dispositif qui inclut désormais l'ANR, quel est votre rôle exact ?

Je travaille à la mise en place d'un cadre global permettant par exemple de déposer des projets auprès de la Commission Européenne, de susciter l'émergence de nouveaux programmes au sein de l'ANR, et de participer à la mise en œuvre d'une industrie de confiance dans le domaine du traitement de l'information.

Vous vous situez donc en amont du processus décisionnel qui donne la tendance ?

On ne peut pas vraiment le formuler ainsi. L'ANR doit mettre en œuvre les décisions qui sont prises par le conseil d'administration au sein duquel le Ministère de la Recherche joue naturellement un rôle important. De notre côté nous nous situons en amont, dans la prospective et la définition des axes de recherche.

Pouvez-vous nous parler des outils actuels et des tendances en matière de veille et de recherche d'information ?

Nous oeuvrons pour que, au sein de l'ANR il existe des programmes spécifiquement dédiés à la recherche d'information. Par ailleurs, la recherche d'information s'inscrit dans un cadre global qui

est celui de la veille. Il s'agit de la recherche de l'information stratégique et l'exploitation de celle-ci, aussi bien dans le domaine des grands groupes que celui des EPST, ou encore le tissu des PMI et PME. C'est donc dans ce cadre global de veille stratégique que s'inscrit notre action qui concerne la recherche d'information.

***Quelles sont les actions spécifiques dans ce domaine en direction des entreprises ?
Travaillez-vous en liaison avec le Haut Commissariat à l'Intelligence Economique ?***

Le maître mot est de faire en sorte qu'il existe en France une industrie du logiciel dite de confiance. Cela revient à permettre aux acteurs français souhaitant mettre en place un dispositif de veille, d'être assurés de la qualité de ce dispositif sur le plan technologique ainsi que de l'indépendance de l'entreprise qui les propose. Dans cette optique le rôle du Ministère consiste d'une part à favoriser l'émergence des innovations technologiques dans le domaine de la recherche d'informations et d'autre part en la mise en place de protocoles d'évaluation visant à caractériser les différentes technologies et les fonctionnalités des outils proposés. Le point fort du dispositif de confiance que nous mettons en place, c'est l'évaluation, conformément aux directives du Haut Commissariat à l'Intelligence Economique.

Quels sont les processus que vous mettez en place en matière d'évaluation ?

Un volet du programme Technolangue était consacré à l'évaluation entre 2002 et 2006. Ce programme a permis la mise en place de neuf campagnes d'évaluation de technologies au sein de la plateforme EVALDA. Les résultats de ce programme sont accessibles sur le site <http://www.technolangue.net/>. En 2007 nous souhaitons mettre en place une campagne d'évaluation dans le domaine des outils de veille. Si cette campagne a effectivement lieu, elle ne sera pas lancée par l'ANR mais par le Ministère directement. L'objectif de ce projet sera de permettre la réalisation d'un benchmark des outils de veille en fonction des tâches de contrôle qui auront été définies.

Sur quoi porteraient a priori ces tâches de contrôle ?

Il est encore un peu tôt pour les définir précisément. Cependant, il existe un certain nombre de points importants liés à l'analyse et au traitement de l'information. Nous nous intéressons beaucoup à la question de la cartographie et à la visualisation de l'information. Il existe très peu de benchmarks portant sur l'apport réel de la visualisation de l'information à la recherche d'information, ainsi qu'à son analyse. Il s'agit donc sans doute d'un bon candidat en tant que tâche de contrôle.

Quel est l'intérêt pressenti de la visualisation de l'information par rapport aux moteurs qui fournissent un information linéaire ?

Il y a deux aspects, l'aspect recherche d'une part et l'aspect analyse d'autre part. En ce qui concerne l'analyse, on se situe dans le cadre d'une cartographie décisionnelle. L'idée est de pouvoir extraire de corpus très volumineux une information essentiellement textuelle dans un premier temps, puis de la cartographier pour identifier les signaux faibles de façon plus ergonomique et plus rapide que ce que peuvent offrir des technologies classiques de recherche d'information.

Que pensez-vous du couplage de technologies statistiques et de technologies sémantiques dans ce domaine ?

Sur la phase d'analyse, il semble difficile de mettre en place une représentation graphique de gros volumes de données sans que soit également effectué un rapprochement sémantique entre les données. Les actions d'évaluations, pour peu qu'on puisse définir des protocoles les plus précis et les plus objectifs possibles, nous permettront peut-être d'en savoir un peu plus sur les effets du couplage des technologies.

Les tendances qui se dégagent en matière de recherche d'information stratégique nous orientent-elles donc vers des applications de visualisation ?

Non. Je dirais plutôt que la visualisation de l'information prend dans l'offre commerciale actuelle une place de plus en plus importante. Ce qu'il faut maintenant, c'est caractériser l'apport réel de la visualisation en termes d'usages. D'où la nécessité de mettre en place des actions d'évaluation

dans ce domaine. Nous voulons déterminer si ces technologies ne sont qu'un élément ergonomique superficiel ou si elles peuvent modifier profondément l'approche cognitive de l'information.

Determine-t-on plusieurs types de visualisation en fonction de l'approche cognitive qu'elles privilégient ?

Différentes technologies sont effectivement proposées. Pour l'instant la plupart des logiciels de cartographie intègrent des technologies de type 2D. Mais il existe dans certains laboratoires des approches qui vont bien au-delà de cette approche traditionnelle. La question est de savoir s'il est utile d'évaluer dès maintenant des technologies très innovantes et qui n'ont pas encore rencontré l'offre commerciale, ou si on va se contenter dans un premier temps de s'intéresser uniquement à l'offre du marché.

Que pensez-vous des technologies de Mind Mapping ?

Je crois que, tout comme pour les moteurs graphiques, il faudrait mettre en place des tests d'utilisabilité de ces technologies afin de spécifier leur apport réel sur le plan cognitif.

Quand vous parlez de moteurs graphiques et de moteurs cartographiques pouvez-vous nous citer quelques éditeurs auxquels vous pensez ?

On peut citer Kartoo qui, aussi bien sur internet que dans sa version portail est à la pointe des technologies de visualisation. Kartoo fait partie des moteurs graphiques qu'il serait intéressant d'étudier de près en termes d'acceptabilité et de plus-value par rapport aux moteurs traditionnels. D'autres outils de recherche sont également couplés avec des technologies de visualisation telles que Mapstan ou Semiosys. Lors du prochain i-expo 2007 nous préparons avec le GFII une session qui sera d'ailleurs consacrée à la cartographie et à la visualisation de l'information, afin de faire un tour d'horizon des technologies et des offres existantes.

Quelles sont les technologies qui s'affirmeront dans les cinq années à venir ?

On observe une robustesse de plus en plus affirmée des technologies linguistiques. Une tendance forte est également dans le couplage entre technologies linguistiques et technologies statistiques. Des moteurs qui étaient il y a peu, uniquement statistiques, commencent aujourd'hui à intégrer des fonctionnalités linguistiques. Une autre tendance est au couplage des meta-données sur le web avec des technologies sémantiques.

Est-ce lié à l'émergence du Web 2.0 ?

La notion de web 2.0 est postérieure en réalité au début de l'articulation des différentes technologies. En termes de recherche d'information sur le web se dégagent trois grandes tendances. La première est dans l'affirmation du web sémantique. On trouve en effet de plus en plus de tags dans les documents structurés, qui facilitent à la fois la recherche humaine et la recherche automatique.

Une seconde tendance est à l'indexation collective qui passe notamment par des pratiques de folksonomie. L'indexation collective est une tendance forte dont on ne mesure pas encore vraiment les effets dans le monde professionnel. Ce qui est intéressant c'est qu'il s'agit d'une tendance qui vient du grand public et se diffusera sans doute dans le monde professionnel, alors qu'on a souvent observé le mouvement inverse. L'indexation collective n'est pas une notion nouvelle puisqu'elle nous vient du monde des bibliothèques et centres de documentation où elle a pris naissance notamment au travers de thesaurus partagés. Mais la folksonomie fait appel à une autre philosophie, beaucoup plus souple que l'utilisation de langages contrôlés.

Une troisième tendance est dans la recherche d'information qui fait appel à des technologies linguistiques. Dans le monde professionnel par ailleurs, existe une quatrième approche. On constate en effet dans les entreprises et administrations, un certain maintien du couplage indexation/recherche sur la base de langages documentaires. Cette dernière approche, à la fois ancienne et classique, se maintiendra certainement dans la mesure où plus il y a d'informations en ligne, plus le besoin de disposer d'une information qualifiée et pertinente est fort. Et le recours à des langages contrôlés, que ce soit sur la base de thesaurus, lexiques, dictionnaires, terminologies ou ontologies, fait partie des ressources métier dont on a besoin au sein de l'entreprise.

Cela exclut-il l'approche sémantique ou peut-on envisager un couplage de technologies ?

Je crois qu'on peut aller encore plus loin et envisager un couplage des quatre approches. Certains établissements publics et organismes de recherche ont mis en place des systèmes de recherche d'information fondés à la fois sur un thesaurus, ainsi qu'une approche collaborative s'inspirant du web 2.0, et une approche complémentaire fondée sur les technologies linguistiques. Et l'ensemble donne un système d'accès à l'information à la fois souple, cohérent et pertinent.

Quelle sera d'ici quelques années la place réservée au multimedia au sein de ces approches ? On observe un couplage de plus en fort entre le texte et l'image en particulier dans le domaine de la production de données scientifiques. Auparavant existait une dichotomie entre la publication scientifique traditionnelle qui pouvait éventuellement intégrer des images de façon uniquement illustrative et l'exploitation de données brutes. On cherche maintenant à rendre accessibles à une communauté scientifique élargie, les données de la recherche et notamment les images sur lesquelles ont travaillé les chercheurs. Il peut s'agir, de façon non exhaustive de données satellitaires, d'images d'archéologie, de modélisation de molécules chimiques ou encore d'imagerie médicale. Cette accessibilité des données scientifiques, c'est-à-dire l'accès à la donnée brute qui a permis d'élaborer les hypothèses à la source d'une publication scientifique, sera une priorité du bureau de l'Information Scientifique et Technique à partir de 2007 et dans les années à venir.

Nous pouvons en ce sens citer l'action en faveur des archives ouvertes qui s'est notamment concrétisée par un protocole qui a été signé en juillet 2006 par huit organismes publics de recherche, la conférence des présidents d'universités et la conférence des grandes écoles. Dans les années à venir, le Ministère de la Recherche va promouvoir ce protocole qui vise à coordonner l'archivage ouvert de la production scientifique au niveau national. Parmi les acteurs fortement impliqués dans la question des archives ouvertes, on peut citer l'INIST ainsi que quelques universités françaises.

Quelles sont les tendances sur le plan Européen ?

Les idées que nous venons d'exprimer sont très en phase avec la préparation du 7ème PCRD. On a observé en Allemagne et en Angleterre notamment, qu'une grande attention était apportée à la valorisation de l'Information Scientifique et Technique. Les archives ouvertes sont un point transnational et la France n'est pas particulièrement en avance dans ce domaine. Par contre nous accordons depuis fort longtemps en France une place importante aux technologies linguistiques et cela depuis plus longtemps que beaucoup d'autres pays. On peut d'ailleurs regretter que la Commission Européenne, au sein du 7ème PCRD, n'ait pas accordé aux technologies linguistiques la place qu'elles méritent.

La France est-elle moteur dans le domaine des technologies linguistiques ?

Sur le plan politique, oui. Elle est à tout le moins l'un des moteurs. Nous avons été à l'origine d'un projet qui visait à regrouper les initiatives nationales dans le domaine des technologies de la langue pour chaque pays de l'Union et qui mettait en œuvre le principe de subsidiarité. Malheureusement ce projet n'a pour l'instant pas retenu l'attention de la Commission.

Croyez-vous que les campagnes classiques d'évaluation de technologies de recherche d'information ont vécu ?

Non. Car les technologies évoluent. Ce qui a peut-être vécu, même s'il s'agit d'un paradigme fort, ce sont les méthodologies d'évaluation de la recherche d'information telles que nous les connaissons aujourd'hui, c'est-à-dire dans une approche exclusivement technocentrée. En particulier, il est difficile de ne pas tenir compte désormais des nouvelles pratiques de recherche. Nous avons cité plus haut les approches collaboratives. Une évolution des campagnes actuelles semble nécessaire. Cela dit le protocole TREC est tellement ancré dans les pratiques évaluatives qu'il a certainement encore de l'avenir. Aujourd'hui nous sommes face à un défi et de nouvelles pratiques sont à mettre en place dans le domaine de l'évaluation des usages de la recherche d'information. Je pense notamment à la cartographie, ainsi qu'à l'approche collaborative et à la folksonomie. Ces pratiques sont laissées de côté par les protocoles d'évaluation classiques. Or il est absolument nécessaire de les prendre en compte dans les protocoles à venir.

Comment évaluer une pratique de ce type ?

Il faut intégrer des critères ou métriques d'évaluation de fonctionnalités et non plus seulement de technologies. Cela fait partie des objectifs que nous nous sommes fixés dans le domaine de l'évaluation des outils de veille.

Comment déterminer un référentiel dans ce domaine ?

C'est très difficile. Dans le domaine de la cartographie par exemple, on ne voit pas encore clairement comment comparer différents moteurs graphiques. Nous parlons tout-à-l'heure par exemple du Mind Mapping. Il serait intéressant de spécifier la façon d'évaluer différents outils présentant cette même fonctionnalité. On peut envisager de mettre en place une approche qui serait orientée vers les usages plus que vers les technologies et qui fonctionnerait sur la base de panels d'utilisateurs.

On peut arguer bien entendu que les panels d'utilisateurs ne sont pas des évaluations objectives. Mais existe-t-il des évaluations strictement objectives et scientifiques ? Il faut dans tous les cas maîtriser les effets de bord et savoir que toute évaluation véhicule sa part de subjectivité. Le but est en réalité de réaliser une photographie d'un système à un moment donné et qui permet d'émettre des hypothèses en vue de son amélioration.

A quoi sert finalement l'évaluation ? Qui en sont les bénéficiaires finaux ?

L'évaluation de technologies, telle qu'elle est pratiquée actuellement, permet de connaître la performance d'une technologie par rapport à une tâche donnée. Cela permet ainsi d'orienter la recherche, soit directement par les laboratoires, soit par les agences de financement. C'est d'ailleurs l'objectif du NIST et de la DARPA aux Etats-Unis au travers des campagnes TREC. Les campagnes d'évaluation servent notamment à fournir des indicateurs pour le financement de la recherche. C'est ce que nous avons fait en France dans le cadre des campagnes EVALDA et il s'agit d'un point important. Les campagnes d'évaluation c'est aussi pour les industriels et les laboratoires une bonne façon de confronter un prototype à l'usage. Les industriels pratiquent l'évaluation en interne et participer aux campagnes leur donne la possibilité de disposer de ressources pour réaliser leurs propres évaluations, notamment de non-régression. Les campagnes permettent aussi de faire évoluer les protocoles existants. S'interroger sur la pertinence des protocoles a également un effet, certes indirect, mais réel, sur la dynamique d'innovation. En ce sens, l'évaluation est aussi un objet de recherche pour la communauté scientifique, en particulier en sciences de l'information.

Avez-vous quelques mots en guise de conclusion ?

La mise en place d'une industrie de confiance dans le domaine des outils de veille à des fins d'Intelligence Economique est une absolue nécessité. Il est donc nécessaire de mettre l'accent sur l'évaluation des outils dans les années à venir. Jusqu'à présent les campagnes d'évaluation réunissaient essentiellement des experts d'un domaine technologique. Le prolongement naturel de ces campagnes passera par une prise en compte croissante des usages, notamment en recherche d'information. Les campagnes passées ont certes un lien indirect avec la création d'une industrie de confiance dans le domaine de l'Intelligence Economique et de la veille. Au GPIE elles ont d'ailleurs suscité le plus grand intérêt. Mais il s'agit d'un premier niveau et il faut renforcer aujourd'hui ce lien.

Le second niveau passera par la mise en place de protocoles d'évaluation des fonctionnalités de la veille, dans une perspective d'Intelligence Economique. Le troisième niveau passera par la compréhension des usages des différents outils de veille et l'identification des freins à l'usage de ces outils, par les acteurs du tissu économique. Nous en sommes au premier niveau. Les deux autres niveaux demeurent à construire et il est de notre responsabilité d'y parvenir.

Interview réalisée par Marianne Dabbadie

*Directrice Innovation i-KM
Laboratoire GERIICO – EA 1060*