



**HAL**  
open science

## La Lettre de l'URFIST de Toulouse, n°34 - 2005

Jo Link-Pezet, Élisabeth Lacombe-Carraud, Lise Herzhaft, Philippe Fabre,  
Olivier Ertzscheid, François Magnan, Alexandre Serres, Annie Léon

► **To cite this version:**

Jo Link-Pezet, Élisabeth Lacombe-Carraud, Lise Herzhaft, Philippe Fabre, Olivier Ertzscheid, et al..  
La Lettre de l'URFIST de Toulouse, n°34 - 2005. 2005. sic\_03250637

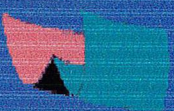
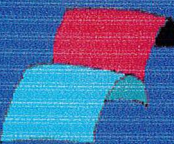
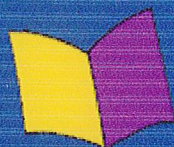
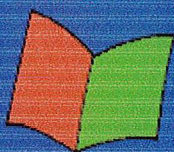
**HAL Id: sic\_03250637**

**[https://archivesic.ccsd.cnrs.fr/sic\\_03250637v1](https://archivesic.ccsd.cnrs.fr/sic_03250637v1)**

Submitted on 4 Jun 2021

**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.




# La Lettre de l'URFIST de Toulouse

## Sommaire

-  [La formation de formateurs en Rhône-Alpes, Auvergne, Bourgogne](#) Lise Herzaft, (Urfist de Lyon) et Philippe Fabre (Grand format)
-  [Faut-il avoir peur de Google ?](#) Olivier Ertzscheid (Urfist de Toulouse)
-  [Google Scholar](#) François Magnan (Urfist de Toulouse)
-  [Logiques marchandes](#) Olivier Ertzscheid (Urfist de Toulouse)
-  [Les 7 grandes tendances de la recherche d'information](#) Alexandre Serres (Urfist Rennes)
-  [Le texte intégral de Bases et Netsources sur Internet](#) (Weblog Réseau URFIST)
-  [Présentation d'Ecoline](#) Annie Léon (Urfist de Paris)
-  [Compte rendu du Congrès VSST](#) (Site du colloque VSST)
-  [La directive européenne sur le droit d'auteur](#) Elisabeth Lacombe (Urfist de Toulouse)
-  [L'environnement numérique de travail à UT1](#) Elisabeth Lacombe (Urfist de Toulouse)
-  [Les banques de données statistiques](#) Annie Léon (Urfist de Paris)



 [L'évaluation de l'information sur Internet](#) Elisabeth Lacombe (Urfist de Toulouse)

 [BibliOpéra, logiciel de gestion d'informations bibliographiques](#) François Magnan  
(Urfist de Toulouse)

L'ensemble de la lettre peut être [téléchargé](#) au format PDF.

Dans toutes les pages, le petit poucet situé en bas à gauche vous ramènera à ce sommaire.





# Formation de Formateurs

## Formation de formateurs La formation de formateurs en Rhône-Alpes, Auvergne, Bourgogne

Lise Herzaft, (Urfist de Lyon) et Philippe Fabre (Grand format)

### L'apparition des besoins

Depuis sa création en 1982, l'URFIST de Lyon propose des stages de recherche d'informations dans des bases de données précises ou sur un ensemble de bases d'un domaine des connaissances. Les stagiaires étaient souvent chargés de retransmettre dans leurs services ces informations et ressentait pour ce faire le besoin d'un minimum de notions de pédagogie. Aussi en 1997, à l'occasion d'une formation donnée à des doctorants en sciences de l'éducation, a été exprimé le besoin d'intervention d'un pédagogue connaissant le monde des bibliothèques.

### Les réponses

Deux intervenants se sont alors succédés, chaque collaboration n'ayant pu se poursuivre faute de disponibilité de l'animateur ou pour des raisons techniques.

Un enseignant de l'Ecole de management de Lyon à cette époque a travaillé avec l'Urfist sur un document (DVD) susceptible d'accroître la motivation des étudiants pour leurs études.

Si ce travail n'a pas abouti, l'Urfist a chargé cette personne d'organiser également des stages de pédagogie appliquée au monde de la documentation.

Le tout premier stage de deux jours sur les connaissances pédagogiques nécessaires pour être un 'bon formateur', associait une partie théorique sur la pédagogie, des travaux dirigés sur la création d'un stage, des réflexions de groupe. Les stagiaires ont immédiatement demandé une suite à ce travail et la formation a été organisée en deux sessions de deux jours.

### L'évolution des besoins

En parallèle, liés à la mise en place du LMD dans les universités, à l'insertion de la formation à la méthodologie documentaire dans les cursus, les services communs de documentation mettaient en place des responsables de la formation à

#### Sommaire de l'Article

- [L'apparition des besoins](#)
- [Les réponses](#)
- [L'évolution des besoins](#)
- [Les nouvelles réponses](#)
- [Quelques constantes remarquables](#)
- [Quelles perspectives ?](#)

[Imprimer la lettre](#)



l'utilisation des ressources de leur bibliothèque pour les étudiants; mais avec peu de connaissances pédagogiques, ces personnes se sentaient désemparées devant la tâche qui leur incombait.

Aussi, lorsqu'à l'occasion d'un audit de FORMIST nous avons rencontré un responsable d'une société de formation de formateurs, un partenariat s'est établi avec un pédagogue connaissant bien le milieu des bibliothèques.

Les demandes ont surtout porté sur la transmission de connaissances pédagogiques de base, la prise de conscience des diverses façons d'apprendre, mais aussi sur la nécessité de réaliser des supports de cours adaptés, attrayants, enfin sur l'intérêt d'évaluer le montant de connaissances transmis à la fin de la formation.

Mais d'autres services étaient intéressés par des formations plus approfondies, ou plus précisément centrées sur leurs questions dans une bibliothèque donnée. Des formations spécifiques ont donc été organisées pour eux.

### **Les nouvelles réponses**

Depuis 2002 des formations de deux jours suivies d'un second stage de deux jours d'approfondissement des connaissances et compétences ont été animées par le même formateur Philippe FABRE, de la société Grand Format, basée à Saint-Étienne, tandis qu'en parallèle, des réunions plus informelles étaient programmées

### **A - Le déroulement et les modalités de formation.**

La formule actuelle, qui paraît donner satisfaction, se déroule donc en 2 fois 2 journées.

La première partie, appelée "formation de base", permet aux participants de se "repérer" dans le paysage de la formation, de mettre des mots précis sur des notions ou des démarches qui sont le plus souvent perçues empiriquement. Elle leur permet de comprendre les logiques de différentes approches pédagogiques, de formaliser les progressions qu'ils proposent aux étudiants, de préciser leur positionnement institutionnel et de prendre mieux conscience de leur potentiel pédagogique. Pendant toute la durée du module, les méthodes actives sont privilégiées, les échanges d'expérience facilités. Ce premier module se termine par la construction et/ou l'optimisation, en sous groupes, de formations qui seront proposées aux étudiants avant la seconde session de formation.

Le second module se déroule plusieurs semaines plus tard, avec comme point de départ l'analyse et la recherche d'amélioration des formations qui ont été dispensées. Chaque personne, chaque groupe est alors amené à présenter à l'ensemble des participants les formations telles qu'elles se sont effectivement déroulées avec, bien sûr, les réussites mais également les doutes et les difficultés.... Les acquis relationnels et conceptuels du premier module assurent une grande qualité d'écoute et permettent des échanges constructifs grâce à un vocabulaire partagé. Ces "comptes rendus d'expérience" fournissent un matériau à partir duquel le programme de travail des 2 journées est établi. Le plus souvent ce sont les questions relatives à la dimension personnelle et relationnelle en formation qui émergent : prise de parole en public, gestion du temps, dynamique de groupe, conflits, évaluation des acquis et motivation des "apprenants". Chaque point fait l'objet d'un temps de travail particulier sous la forme, par exemple, d'un bref apport théorique, d'un échange entre les participants ou d'un exercice d'entraînement.

Au terme de ce travail, chaque groupe de formateurs, chaque personne en tire les enseignements pour l'animation et la conduite des formations proposées aux étudiants.

### **Quelques constantes remarquables**

Les formations réalisées ont concerné près d'une centaine de personnes. Sept formations de deux jours ont été organisées en 2002, six en 2003, de nouveau sept en 2004 (une à Dijon pour 12 personnes en formation de base, quatre à Grenoble pour 28 personnes en formation de base et pour 15 personnes en approfondissement et 2 à Lyon). Certaines caractéristiques communes



peuvent être soulignées :

Pour la grande majorité des formatrices et formateurs, le niveau d'implication dans l'activité de formation est élevé. Il y a un fort investissement qui est le plus souvent indépendant du temps effectivement passé à s'y consacrer.

Au-delà des aspects "techniques" de la formation, ces sessions offrent un espace pour s'informer mutuellement, confronter les expériences, échanger sur les pratiques, élaborer ensemble. L'activité de formation devient ainsi visible, partagée et reconnue. Ces temps constituent, semble-t-il, des moments forts de professionnalisation pendant lesquels se construisent des manières de faire et de penser la formation comme un élément constitutif du référentiel professionnel.

On peut également mentionner le passage progressif de projets de formation individuels à des projets d'équipe. Sous l'impulsion des équipes de direction, la question de la formation des étudiants est de plus en plus posée en terme de dispositif global propre à l'établissement. Des équipes de formateurs sont constituées et, dans cette optique la formation de formateurs qui leur est proposée évolue progressivement vers de la formation-action. Il s'agit alors d'outiller les formateurs afin de leur donner de quoi conduire et étayer leurs actions en prenant en compte de manière explicite les conditions (matérielles, institutionnelles, managériales) locales. En conséquence, le contenu des formations s'élargit à des thématiques complémentaires telles que les questions d'ingénierie et d'organisation de la formation et de développement du partenariat avec les autres acteurs de la formation

### **B - Les réunions informelles**

En parallèle, deux fois par an, les formateurs qui le souhaitent se réunissent à l'Urfist. L'objectif de départ était d'échanger sur les pratiques pédagogiques, méthodes et trucs, mais aussi sur le contenu des bibliographies distribuées, des supports en cours de réalisation, de tester les réactions des collègues-experts. Un second objectif était de recenser les besoins de formation à l'utilisation des ressources électroniques dans les différentes disciplines.

Puis ces formateurs ont été très impliqués dans les réflexions des universités sur la mise en place du LMD ; dans la majorité des universités des régions Auvergne, Bourgogne et Rhône-Alpes, des modules de formation à la recherche d'informations ont été mis en place. Aussi, une nouvelle préoccupation est apparue dans ces réunions : connaître de façon précise ce qui se fait dans les différentes universités des trois régions.

Une grille descriptive des formations à la méthodologie documentaire fut créée et remplie par certains formateurs. Ces grilles sont en ligne à l'adresse <http://urfist.univ-lyon1.fr/grillesformation>.

Ces réunions ont également servi d'appui et de test pour la réalisation d'ERUDIST référentiel des compétences documentaires pour l'étudiant de Licence (L du LMD).

Il est disponible à l'adresse provisoire <http://askonce.grenet.fr/refdoc3/>

### **Quelles perspectives ?**

Pour les formations pédagogiques, dans la continuité de leur évolution, il est probable que des thématiques nouvelles vont rapidement émerger et notamment les questions d'évaluation et de mutualisation.

En ce qui concerne l'évaluation, il s'agit non seulement de se doter de moyens d'évaluer les effets des formations dispensées aux étudiants mais également, et cela est au moins aussi important voire stratégique, d'être capable d'évaluer l'efficacité globale des dispositifs mis en place au niveau de l'établissement.

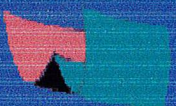
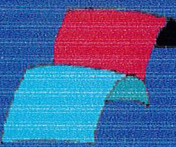
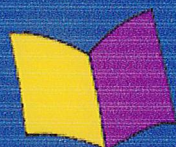
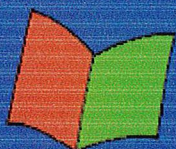
Dans le même temps, le développement des efforts de chaque SCD pour élaborer des cursus et constituer des équipes de formateurs ne peut pas indéfiniment se faire de manière isolée. Pour des raisons qui tiennent à la fois à l'économie de moyens et à la qualité des formations offertes aux étudiants, les équipes vont être amenées à mutualiser leurs processus pédagogiques et leurs stratégies.

[Sommaire](#)









## Faut-il avoir peur de Google ?

Olivier Ertzscheid (Urfist de Toulouse)

Nota Bene : une version courte de cet article est parue dans l'édition du 5 Janvier 2005 du journal "Le Monde". A ce titre merci de ne pas reproduire ce texte sans l'autorisation de l'auteur ([olivier.ertzscheid@gmail.com](mailto:olivier.ertzscheid@gmail.com))

Google est l'auteur de plusieurs "révolutions" en terme de recherche et d'accès à l'information. A l'inverse de ce que faisaient ces prédécesseurs, Google a fait l'hypothèse que la pertinence d'une page n'était pas nécessairement liée au nombre de fois où un mot clé y était présent, mais plutôt au nombre de liens en provenance d'autres pages, qui pointaient vers elle. Meilleur est cet "indice de popularité" (PageRank), meilleur sera le classement de la page dans le moteur. Il a aussi été le premier à offrir l'accès à des contenus du "web invisible", à savoir tous ces fichiers word, excel, powerpoint ou Acrobat que l'on a désormais l'habitude de trouver pêle-mêle parmi les résultats d'une recherche. Chose qui était impensable il y a de cela quelques années.

Google a également su se donner les moyens de son ambition : en terme d'infrastructure technique et même si sur ce point le moteur communique peu il est celui qui possède le plus de centres serveurs dans le monde, qui lui permettent de traiter et de mettre à disposition cette somme énorme d'information. C'est cette puissance technique qui, la nuit du 11 novembre 2004, lui a permis de proposer de chercher parmi non plus 4, mais 8 milliards de pages, simplement pour ne pas être dépassé par son concurrent Microsoft dont le nouveau moteur annonçait 5 milliards de pages indexées ! Remarquons cependant que derrière ce doublement se cachent nombre de doublons, de publicités (spam) ou de sites à caractère pornographique.

A côté, et à l'appui de cette puissance technique, Google est un champion du marketing et de la communication. De l'interface très sobre de sa page d'accueil jusqu'au nom et aux couleurs de la marque en passant par la devise de ses fondateurs, "*Don't be evil*", tout est soigneusement pesé. Mais il ne s'agit pas que d'une facade. Des projets comme "Ocean" ou "Google Print", doivent lui permettre de devenir le principal acteur de la numérisation des fonds documentaires de plusieurs universités et bibliothèques prestigieuses (Oxford, Stanford ...). Près de 15 millions d'ouvrages sont concernés et il ne s'agit là que d'une étape, l'objectif affiché de ses fondateurs étant d'"*Organiser et [de] donner accès à l'ensemble des connaissances humaines disponibles*", comme un lointain écho à la bibliothèque borgésienne.

Google est ainsi devenu, à l'échelle de la planète connectée, un véritable réflexe sociétal. Il se décline sur tous les terrains, dans presque toutes les langues et sur tous les continents. Trouver un restaurant près de chez soi, récupérer les articles d'un chercheur, imprimer une image de la Joconde, Google est le seul point commun de toutes ces recherches. Google à l'ambition et probablement les moyens de devenir un guichet d'accès unique à l'information, à la connaissance. Et même si une telle volonté centralisatrice dans un univers aussi décentralisé que l'est Internet, peut sembler à la fois une gageure et un non-sens, il en prend le chemin. Et ce indépendamment de la noblesse avérée ou supposée de cette ambition.

Dans le même temps, comme tout "numéro 1", Google est particulièrement exposé



aux attaques. Son fameux PageRank l'expose à des pratiques qui peuvent très vite cesser de devenir amusantes ou anecdotiques. Ainsi le "Google Bombing" permet assez facilement de classer première n'importe quelle page sur un mot clé pas très flatteur : des personnalités comme Jacques Chirac ("magouilleur"), Georges W. Bush ("miserable failure"), et Jean Dionis ("député liberticide") alors rapporteur de la Loi sur l'Economie Numérique (LEN) en ont fait les frais. Trop souvent utilisé comme une source unique et supposée fiable, il peut devenir une redoutable machine à désinformer. Ainsi pendant le début de la guerre en Irak certains sites ouvertement anti-Bush ont tout simplement "disparu" des résultats du moteur ... Ce même PageRank est également critiqué car il favorise les sites disposant déjà d'une bonne notoriété au détriment des autres. Il aurait ainsi pu y avoir un web à 2 vitesses. Mais Internet fonctionne comme un écosystème qui se réorganise constamment, qui s'auto-organise. Ainsi l'émergence du phénomène des weblogs permet désormais de "corriger" ce risque, ces derniers étant très rapidement référencés et posant par ailleurs d'autres problèmes aux algorithmes des moteurs.

Pour anticiper et contrer les attaques (nombreuses) de ses concurrents, Google se diversifie et propose une offre portail avec un webmail (Gmail), des groupes de discussion (GoogleGroups), une recherche sur vos documents (Google Desktop), etc. Ces services ne sont pas nécessairement meilleurs mais le phénomène de marque joue à plein et le public suit.

Pourtant Google n'est pas toujours le chevalier blanc qu'il prétend être. La position de Google fait qu'il s'autorise à passer outre un certain nombre de règles. Il conserve ainsi de manière tout à fait illégale une copie (cache) de certaines pages qui n'existent plus physiquement sur le réseau.

Alors ne nous y trompons pas : Les 8 milliards de pages indexées par Google sont loin de représenter la totalité de l'information accessible sur Internet et d'autres moteurs, bien qu'indexant moins de pages offrent un vrai gain qualitatif au chercheur d'information en présentant les résultats sous forme de cartes (Kartoo.com) ou en triant automatiquement ceux-ci dans des dossiers thématiques (Exalead.fr).

L'information et la connaissance ne sont pas dans Google. Elles sont sur la toile. Google est un télescope trop souvent utilisé comme une simple loupe. Donnez-lui un mot clé et il vous conduira vers des millions de pages dont nous ne verrons que les 3 ou 4 premières en ayant l'impression d'avoir fait le tour de la question. C'est alors bien du côté de la formation et de la pédagogie des usages qu'il faut faire porter l'effort si l'on ne veut pas que la formidable lentille de ce télescope informationnel ne devienne un prisme déformant nos connaissances.

[Sommaire](#)





# Société de l'Information

## Google Scholar

François Magnan (Urfist de Toulouse)

L'information a fait grand bruit chez nombre d'universitaires et de bibliothécaires : Google a inauguré le 18 novembre 2004 un service spécialisé dans la recherche de documents académiques et scientifiques, à destination des chercheurs comme des étudiants. Si Google Scholar (G.S.) utilise la même technologie que Google, la recherche se fait uniquement parmi un corpus de documents spécifiques. Autre innovation : G.S. cherche à aider le lecteur désireux d'obtenir au plus vite un document (article, livre) en proposant des liens vers le texte intégral des articles cités dans une réponse, ou en permettant de localiser une bibliothèque disposant d'un ouvrage mentionné : cela permet d'intégrer les différentes étapes d'une recherche d'information scientifique, depuis la formulation d'une requête jusqu'à l'obtention du document final.

Quelles sont les ressources interrogées par G.S. ? Des documents scientifiques accessibles sur Internet, très majoritairement de langue anglaise, issus en grand nombre des sciences dures et du domaine technique, qu'il est désormais possible d'interroger via une interface unique : des sites de revues scientifiques à comité de lecture, comme Nature, Science, The Lancet, des sites de Sociétés Savantes, d'Universités, des articles gratuits de sites participant au mouvement de "l'Open Access" comme BioMed Central (1) ou des bases d'agrégateurs de contenu scientifique payant comme Ingenta (2). Cette interface unique apporte un gain de temps appréciable pour le chercheur. A-t-elle des inconvénients ?

Professeur au département des Sciences de l'Information de l'Université d'Hawaï, Peter Jacso a développé un outil qui permet de tester et comparer les performances de G.S. "versus" l'interrogation directe de quelques sites de fournisseurs de contenu disposant de leurs propres outils (3) de recherche : l'utilisation de G.S. serait moins efficace que l'interrogation directe des sites originaux et connaîtrait (4) de gigantesques failles dans sa couverture des documents potentiellement disponibles.

Quelles ressources sont interrogées, lesquelles ont été écartées, avec quels critères de sélection ? Il est impossible de le savoir de façon précise, car Google - malgré un important battage médiatique - ne communique que très peu à ce sujet, comme le fait remarquer Gary Price (5), Bibliothécaire et Consultant en Recherche d'Information, responsable du site [resourceshelf.com](http://resourceshelf.com). Dans ses F.A.Q (6) (Frequently Asked Questions), G.S se contente de signaler aux éditeurs scientifiques qu'ils peuvent proposer d'ajouter leurs documents, sans mention claire de critères d'acceptation.

G.S. permet de trouver des thèses, des articles scientifiques, des extraits de livres, des rapports de recherche, des preprints, etc. Lorsque la référence d'un livre vient en réponse, en sélectionnant l'option "Library Search", il sera possible de localiser l'ouvrage dans une bibliothèque aux Etats-Unis et dans le monde entier, à condition que cette bibliothèque participe au programme "Open Worldcat" (7) de l' "Online Computer Library Center" (8) (OCLC). En sélectionnant l'option "Websearch", les références du livre (nom de l'auteur et titre) seront recherchées sur le moteur Google. En ce qui concerne les articles (que la fonction Websearch permet aussi d'obtenir sur



la totalité du Web) plusieurs cas de figure sont possibles : s'il s'agit de preprints (en accès gratuit) différentes versions peuvent être proposées : il ne faudra pas oublier qu'elles ne sont pas définitives. L'article sera présent en version intégrale s'il s'agit d'une revue en accès libre sur le Web. Dans le cas d'articles de publications scientifiques payantes, il faudra avoir au préalable souscrit un abonnement, mais il sera possible d'obtenir, gratuitement, un résumé.

G.S. trouve aussi des citations : analysées et extraites des articles, elles sont présentées comme des résultats distincts. Ainsi, pour chaque document trouvé, il sera possible de savoir combien de fois il a été cité, par qui, et dans quel contexte : à côté d'une réponse, on pourra trouver la mention : "cited by 10" (présent dans 10 citations) puis, après avoir cliqué sur cette mention, la liste complète des documents citant cette première référence apparaîtra à son tour. Cette possibilité d'analyse et d'extraction des citations, fonctionnalité courante des bases de données scientifiques, est d'ailleurs à l'origine de la fonction Pagerank - et donc du succès - du moteur de recherche Google.

Comment Google Scholar organise-t'il le classement des réponses ? Selon les F.A.Q. de G.S. - peu explicites à ce sujet - le classement des réponses prend en compte les paramètres suivants : texte de chaque article, auteur de l'article, publication dans laquelle apparaît l'article et nombre de fois où cet article est cité dans la littérature savante.

En plus des opérateurs booléens habituels, G.S. a sa propre syntaxe d'interrogation. Il est possible de faire une recherche :

- par auteur (author:dupont). Dans ce cas, Google suggère d'utiliser le dernier nom de l'auteur et d'ajouter le titre - ou une partie du titre - de l'article. Ainsi : author:witten "string theory dynamics" permettra de trouver l'article "string theory dynamics in various dimensions" écrit par Edward Witten
- par nom du site (site : harvard.edu). Par exemple, site : www.paris4.sorbonne.fr prospective permettra de trouver les pages du site de l'Université de Paris Sorbonne qui contiennent le mot prospective.
- par titre du document (intitle : astronomy). Ainsi, intitle : "histoire des mathématiques" permettra de trouver des dizaines de documents ayant comme titre "l'Histoire des Mathématiques".
- par date et titre de la publication en se servant du formulaire de recherche avancée
- par type de fichier indexé : par exemple, l'expression filetype:pdf permettra de trouver des documents au format Adobe PDF (Portable Document Format), très utilisé pour la publication d'articles. L'expression filetype:pdf airbus a380 permettra de trouver des dizaines d'articles au format PDF (ou leurs notices s'ils ne sont pas en accès gratuit) parlant de l'Airbus A380.

Au moment où ces lignes sont rédigées, il n'existe pas encore d'opérateur qui permettrait de ne rechercher que des ouvrages ou des citations, ou de distinguer les ressources gratuites des ressources payantes. Toutefois, cette syntaxe connaîtra certainement des évolutions : en janvier 2005, la version de G.S. n'est pas définitive.

Si l'arrivée de Google Scholar a été saluée par tous (9), et plus particulièrement par la communauté de l'Open Access, des éléments critiques peuvent être relevés chez certains commentateurs : Peter Jacso note (4) que certains résultats proposés par G.S. sont étonnants pour des documents censés être scientifiques et que sa communication est surtout orientée vers une stratégie de marketing, il relève que les F.A.Q. de G.S. ne reflètent pas les réelles questions les plus posées par les utilisateurs, mais constituent plutôt un argumentaire à caractère commercial destiné à mettre en valeur cet outil.

Gary Price regrette (5) un manque de transparence de la part de G.S. sur les critères de sélection des documents et note que G.S. n'a pas de définition claire de ce qu'est un document scientifique, particulièrement pour les sites Web inclus parmi ses résultats. Google est aussi peu disert sur la liste des ressources indexées, ainsi que sur la méthode de classement des réponses. On pourrait ajouter à ces critiques que G.S. agrège le travail de partenaires, comme Ingenta (10) ou l'OpenWorldcat de l'OCLC, sans les mentionner : mais Google, société commerciale, cherche essentiellement à promouvoir sa marque.

L'arrivée de G.S. parmi les moteurs de recherche représente un progrès certain car il démocratise l'accès à la connaissance : nombreux sont ceux qui, habitués à la simplicité du moteur de recherche Google, l'utilisent comme premier outil pour



trouver des ressources scientifiques, à ceux-là, cet outil ouvrira de nouveaux horizons, tandis que d'autres remarqueront que toutes les ressources disponibles sur G.S. l'étaient déjà avant, à condition toutefois de connaître les adresses et de maîtriser de nombreuses interfaces d'interrogation. Son propriétaire, Google, est d'ailleurs destiné à devenir un acteur incontournable de la recherche d'information scientifique car à côté de G.S. d'autres projets sont en cours ou déjà opérationnels :

- "Google Ocean", un projet de numérisation des fonds des plus prestigieuses universités : Harvard, Stanford, celle du Michigan, ainsi que la "New York Public Library".
- "Google Prints" : après avoir saisi un mot clé sur Google, un lien vers un livre numérisé est proposé si son éditeur a signé un contrat avec le moteur de recherche. Des extraits seront accessibles mais il faudra l'acheter en ligne pour obtenir la version complète. Cet outil est déjà opérationnel sur la version américaine de Google.

Ces différents projets peuvent provoquer de l'enthousiasme mais aussi la crainte d'un prochain monopole de Google sur l'accès à l'information scientifique.

Google Scholar : <http://scholar.google.com>

Sur le projet Google Ocean, on pourra lire cette page du Blog des URFIST : [http://urfistinfo.blogs.com/urfist\\_info/2004/12/nom\\_de\\_code\\_oce.html](http://urfistinfo.blogs.com/urfist_info/2004/12/nom_de_code_oce.html)

Les F.A.Q. de Google sur Google Print : <http://print.google.com>

Des liens pour en savoir plus :

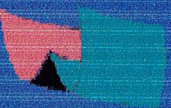
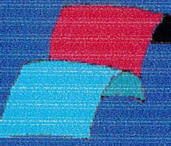
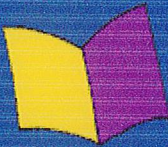
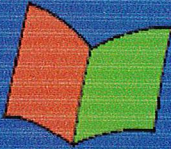
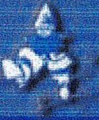
1. Sur Google Scholar  
Danny Sullivan, "Google Scholar Offers Access To Academic Information", <http://searchenginewatch.com/searchday/article.php/3437471>, 18 novembre 2004.  
Barbara Quint, "Google Scholar Focuses on Research-Quality Content", <http://www.infotoday.com/newsbreaks/nb041122-1.shtml>, 22 novembre 2004  
<http://schoogle.blogspot.com> : Weblog officieux
2. Sur Google :  
<http://www.google.com/googleblog/> : Weblog officiel de Google  
<http://www.prweaver.fr/blog> : Weblog plutôt technique et francophone sur Google  
<http://google-observatoire.blogspot.com> et <http://www.zorgloob.com> : Weblogs officieux, généralistes et francophones sur Google

- (1) <http://www.biomedcentral.com>.
- (2) <http://www.ingenta.com>
- (3) Peter Jacso, "Side - by - Side, Native Search Engines vs Google Scholar", <http://www2.hawaii.edu/~jacso/scholarly/side-by-side2.htm> - page consultée le 14/01/2005
- (4) Peter Jacso, "Peter's Digital Reference Shelf", (archives de décembre 2004) [http://www.galegroup.com/free\\_resources/reference/peter](http://www.galegroup.com/free_resources/reference/peter) - page consultée le 14/01/2005
- (5) Gary Price, "Google Scholar Documentation and Large PDF Files", message posté sur le blog de searchenginewatch.com le 01/12/2004, <http://blog.searchenginewatch.com/blog/041201-105511> - page consultée le 14/01/2005
- (6) Google Scholar, About Google Scholar - Frequently Asked Questions, <http://scholar.google.com/scholar/about.html>
- (7) <http://www.oclc.org/worldcat/open/default.htm>. Le programme de l'Open WorldCat veut rendre accessible la base Worldcat (catalogue coopératif de nombreuses bibliothèques du monde entier) de l'OCLC lors de simples recherches sur Internet
- (8) <http://www.oclc.org/> L'OCLC (Online Computer Library Center) est un organisme à but non lucratif qui regroupe des milliers de bibliothèques de tous les pays. WorldCat est le catalogue collectif en ligne de l'OCLC.
- (9) A l'exception de L'"American Chemical Society", en procès avec Google au sujet de Google Scholar, lui reprochant l'emploi du terme "Scholar" qu'il considère comme une violation du droit des marques. Il existe en effet un logiciel édité par l'ACS, nommé "SciFinder Scholar", bien connu des chimistes.
- (10) Un employé d'Ingenta, Leigh Dodds, raconte dans une page de son blog datée du 23/11/2004 les détails de sa collaboration avec les concepteurs de G.S. afin que ce dernier puisse correctement indexer le contenu proposé par Ingenta : or aucune mention de ce travail en commun ne se trouve parmi les pages



[Sommaire](#)





## Logiques marchandes

Olivier Ertzscheid (Urfist de Toulouse)

### **Objectif mon moteur ?? Logiques marchandes et des usages des moteurs : un bilan inquiétant.**

A l'heure où Jean Noël Jeanneney, président de la BNF, alerte, via les colonnes du journal "Le Monde", les politiques et les usagers sur les possibles dangers d'hégémonie du moteur Google il nous est apparu utile de rappeler quelques pratiques communes à tous les outils de recherche et nous paraissant, à court comme à long terme, beaucoup plus problématiques que les projets de numérisation de Google. Il s'agit des logiques marchandes auxquelles se soumettent ces outils.

Dans l'article "Des machines pour chercher au hasard : moteurs de recherche et recherche d'information". Gabriel Gallezot et moi-même faisons remarquer que (...) "Devant l'ampleur de la toile mondiale et la difficulté du recensement de l'ensemble des informations disponibles, de plus en plus d'acteurs majeurs de la recherche d'information sur le web fusionnent et se regroupent, ce qui donne lieu à un échange ou à une vente de tout ou partie de leurs bases d'index et de leurs bases de données. Ainsi, une recherche dans la partie annuaire de Google donnera exactement les mêmes résultats qu'une recherche similaire sur l'Open Directory, ce dernier fournissant à Google sa base d'annuaire. Ces pratiques ne constitueraient pas en elles-mêmes un danger pour l'objectivation de la recherche d'information sur le web si les usagers (novices) en avaient connaissance ou pouvaient le soupçonner, ce qui est loin d'être le cas. D'autant que ces logiques marchandes, qui s'expliquent par des nécessités autant techniques que pragmatiques sont à leur tour biaisées par des aspects de politique marketing également déroutantes pour le chercheur. Mentionnons également au premier rang de ces logiques marchandes, l'arrivée de l'indexation payante qui s'affirme comme le seul modèle économiquement viable pour les différents outils de recherche. Il va sans dire que cette logique et les pratiques qui lui sont associées (achat de mots-clés auprès de certains moteurs, garanties de "positionnement" dans la liste affichée de leurs résultats, etc.) constitue là encore un biais évident pour l'objectivation du déroulement d'une procédure de recherche d'information, même si - le monde francophone ayant exprimé sa défiance vis à vis de ce système - les résultats relevant de "liens sponsorisés" ou de "mots-clés achetés" apparaissent de manière différente et plus ou moins repérable sur les sites francophones de ces outils, et qu'il est donc en théorie toujours possible de faire fi de ces résultats pour aller consulter ceux ne relevant pas de cette approche."

La morale semblait alors sauve. "Semblait" car la plupart des utilisateurs de ces moteurs ignorent tout de ces pratiques. Du coup la manière dont ils prennent pour argent comptant ce que leur renvoie leur moteur pose clairement de sérieux problèmes d'usage de l'information. Un rapport vient de faire le point sur cette question. Edité par ConsumerWebWatch (une sorte de "50 millions de consommateurs" américain) et baptisé : In Search of Disclosure: How Search Engines Alert Consumers to the Presence of Advertising in Search Results . A partir de plusieurs études basées sur 1500 utilisateurs adultes il ressort que 60 % des personnes ignorent l'existence de ces liens sponsorisés.

Une autre étude du PewInternet sur les usages de nos chers outils de recherche



abonde dans le même sens. En voici quelques chiffres qui, à tout le moins, font question et posent problème (ou l'inverse) :

- 92 % des gens interrogés font confiance aux moteurs (52% leur font même "très confiance")
- 68 % des gens interrogés pensent que les moteurs de recherche sont des "sources d'information légitimes et non-biaisées"
- 1 personne sur 6 (seulement ...) est consciente de l'existence de liens sponsorisés et d'informations commerciales dans les résultats des moteurs ...
- 44 % des gens interrogés n'utilisent ... qu'un seul moteur et seulement 7% en utilisent plus de trois ...
- 57 % des gens interrogés ignorent que les moteurs conservent des données relatives à la (leur) navigation (tracking) et que ces données peuvent être réutilisées dans un cadre commercial.

Et bien entendu, la majeure partie de ces "personnes interrogées", sont plutôt jeunes (20-30 ans), plutôt d'un bon niveau d'éducation et passent beaucoup de temps sur le Net.

Vous avez dit "il y a urgence pour une pédagogie des usages ?" Ajoutez maintenant "citoyenne" à "pédagogie".

Voir aussi :

- [Webwatch](#)
- [Critical Thinking](#)
- [Difference in People's Online Skills](#)

[Sommaire](#)





## Les sept grandes tendances de la recherche d'information sur Internet

Alexandre Serres (Urfist Rennes)

Le propos d'Alexandre Serres est d'essayer dans cet article, de discerner quelques-unes des grandes tendances de la recherche d'information sur Internet, en cours et à venir, au vu de l'explosion des outils, des incessants progrès des techniques de recherche, des nouvelles applications (comme les weblogs, le RSS ...), des travaux du Web sémantique... qui montrent un univers de la recherche d'information en constante ébullition, en transformation permanente depuis 10 ans.

L'article intégral est paru dans Savoirs CDI sous le titre : "Recherche d'informations sur Internet : Où en sommes-nous ? Où allons-nous ?"

Le regard rétrospectif, le souci de l'histoire, toujours nécessaire, sont de plus en plus oubliés ou minorés aujourd'hui, notamment à cause de l'emballement technologique, où une innovation chasse l'autre tous les six mois, en effaçant les traces des techniques passées. Sans faire ici un historique des outils et des techniques, nous évoquerons quelques unes des grandes tendances qui peuvent résumer les principaux bouleversements de la recherche d'information depuis les débuts de l'informatisation documentaire, il y a plus d'une quarantaine d'années.

En prenant en compte les différents "composants" de la recherche d'information, nous en avons relevé sept, sans aucune prétention d'exhaustivité, que l'on peut résumer ainsi :

- de la dépendance à l'autonomie des usagers
- de la maîtrise des stocks à la surabondance des flux
- de la validation a priori à la validation a posteriori
- de la rareté et de la distinction à l'explosion et à l'hybridation des outils et des modes de recherche
- du "retrouvage" booléen à la "sérendipité"
- du modèle de l'accès à celui du traitement de l'information
- de la gratuité à la commercialisation de la recherche

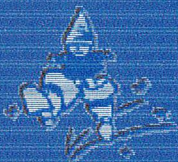
### Du côté des usagers : de la dépendance totale à l'autonomie relative

Il s'agit là sans aucun doute de l'évolution majeure, profonde, de la recherche d'information, qui résume toutes les autres : depuis les premières recherches "en différé" des années 60, où l'utilisateur posait sa question au bibliothécaire ou documentaliste qui la transmettait à l'informaticien, jusqu'à l'utilisation actuelle des moteurs de recherche, en passant par l'interrogation des

#### Sommaire de l'Article

- [Du côté des usagers : de la dépendance totale à l'autonomie relative](#)
- [Du côté de l'offre informationnelle : de la maîtrise des stocks à la surabondance des flux](#)
- [Du côté de la "chaîne de production" de l'information : de la validation a priori à la validation a posteriori](#)
- [Du côté des outils : de la rareté à l'explosion, de la distinction à l'hybridation des outils et des modes de recherche](#)
- [Du côté des processus de recherche : du "retrouvage" booléen à la "sérendipité"](#)
- [Du côté des modèles de la recherche d'information : de l'accès au traitement de l'information](#)
- [Du côté de l'économie de l'information : de la](#)





banques de données par le minitel, les usagers sont passés d'une situation de dépendance totale vis-à-vis des professionnels à une interaction directe avec les outils. Cette autonomisation croissante des utilisateurs est la conséquence directe d'une tendance lourde de l'évolution des outils : la simplification d'usage. Simplification des accès, des interfaces, des procédures... : toute la complexité et l'intelligence technique sont de plus en plus "enfouies" en amont, dans la technologie même des outils, et ceux-ci deviennent des "boîtes noires", auto-simplifiantes, utilisables par le grand public (cf le succès de Google). Nous sommes loin d'avoir tiré toutes les leçons de ce phénomène irréversible, qui marque par ailleurs un extraordinaire phénomène de démocratisation dans l'accès à l'information et de popularisation de notions et de pratiques jusqu'alors réservées aux professionnels. Les problèmes de la recherche d'information concernent aujourd'hui tout le monde et sont inséparables des questions et enjeux politiques, culturels, sociaux..., liés à l'utilisation des technologies de l'information, à la "fracture numérique". L'usage de masse des outils de recherche pose également de nombreux (et relativement nouveaux) problèmes, sur les modifications des pratiques informationnelles (comme le "zapping"), sur les nouveaux modes de connaissance induits par les logiques des outils de recherche, sur les dangers de "l'info-pollution", etc.

Mais si les usagers, et donc les étudiants, deviennent de plus en plus autonomes, faut-il pour autant les laisser "seuls face à Internet" ? Quelle est la nature de cette autonomie ? Cette question de l'autonomisation (surtout procédurale) des usagers pose la question de leur formation intellectuelle.

### Du côté de l'offre informationnelle : de la maîtrise des stocks à la surabondance des flux

Du côté de l'offre informationnelle, nous sommes passés de la métaphore de "l'explosion documentaire" des années 60, qui concernait surtout l'information scientifique et technique (essor des banques de données, accroissement des revues scientifiques, etc.) à celle du "déluge informationnel" d'Internet. Mais le changement est loin d'être métaphorique, car il s'agit tout à la fois :

- **d'un changement d'échelle**, dans la production documentaire, mesurée désormais en milliards et non plus en millions (sur le web "visible", i.e. indexé par les moteurs de recherche, et impossible à évaluer précisément, le nombre de pages web serait entre 10 et 15 milliards ; quant au web "invisible", il serait estimé à 550 milliards de documents !) ;
- **d'un changement de support**, avec la numérisation généralisée des textes, des sons, des images et de tous types de traces, l'internet devenant un gigantesque espace "multimédia" ;
- **d'un changement de système éditorial**, le web étant avant toute chose un vaste système d'auto-publication, permettant à chacun de publier ce qu'il veut, pour le meilleur et pour le pire.

Tous ces aspects sont connus et il est inutile d'y revenir. Mais ce "déluge informationnel" a entraîné une conséquence, tout à fait essentielle pour les professionnels de l'information, qu'il ne faut jamais perdre de vue : le renversement de la problématique de la validation documentaire.

### Du côté de la "chaîne de production" de l'information : de la validation a priori à la validation a posteriori

On sait que contrairement aux bibliothèques et aux CDI, ces espaces documentaires protégés, surveillés, balisés et aux allées (généralement!) bien droites, le web est une jungle, un océan, un fouillis ou une poubelle, selon l'appréciation. Ce qui est patent, et qui constitue d'ailleurs l'un des enjeux éducatifs les plus forts, c'est bien ce retournement de la validation de l'information : jusqu'alors effectuée "en amont" de la chaîne de production de l'information, d'abord par les chercheurs et les auteurs, qui n'écrivent pas (théoriquement) n'importe quoi, puis par les éditeurs, qui ne publient pas tout ce qui s'écrit, ensuite par les libraires, qui ne vendent pas tout ce qui se publie et enfin par les bibliothécaires-documentalistes, qui n'achètent pas tout ce qui se vend, la validation de l'information (terme générique sous lequel on mettra l'évaluation, la sélection, le filtrage...) s'opérait à différents niveaux, par différents acteurs, selon différentes modalités et pour différentes finalités. Ce schéma, toujours valable dans le monde "traditionnel" de l'édition ou de la production scientifique, n'est plus celui du web : la validation de l'information est ici généralement reportée sur l'utilisateur, "en aval", avec tous les problèmes, les risques et les dégâts possibles. Le web ou la crise des médiateurs...

### Du côté des outils : de la rareté à l'explosion, de la distinction à l'hybridation des outils et des modes de recherche

Une première observation très simple frappe l'observateur, devant le paysage actuel des outils et des méthodes de recherche : nous sommes passés, en deux décennies à peine, d'une relative pauvreté, ou rareté des outils, à une abondance, **une explosion, une prolifération d'outils de recherche** ; dans les années 80, par exemple, la recherche d'information se limitait aux logiciels documentaires (environ une vingtaine), aux logiciels des serveurs de banques de données, et à

[gratuité à la vente des mots-clés](#)

[Imprimer la lettre](#)



quelques outils documentaires spécifiques. Aujourd'hui, les outils de recherche, uniquement sur Internet, se comptent par milliers, il en meurt et il en naît chaque année plusieurs dizaines et la diversité, dans la spécialisation, est infinie.

Deuxième observation : **l'imbrication, l'hybridation des modes de recherche et des outils.** On peut distinguer, schématiquement, **quatre modalités de recherche d'information** : la **navigation arborescente** (dans les annuaires thématiques, les classifications), la **navigation hypertextuelle** (dans les sites web, les CD-ROM, les encyclopédies), la **recherche par requête sur des mots-clés dans des champs délimités** (l'interrogation des banques de données, des catalogues...) et la **recherche par requête sur le contenu** (recherche en texte intégral, moteurs de recherche). A chacune de ces modalités correspondaient des pratiques, des usages de recherche, mais aussi des outils, jusqu'alors bien distincts. Ainsi, aux débuts du web, la distinction entre annuaires et moteurs était-elle parfaitement claire, du seul point de vue de la modalité de recherche. Or l'une des évolutions profondes de la "RI" a consisté à entremêler ces diverses modalités ainsi que les outils sur lesquels elles s'appuient. Depuis quelques années, la mixité entre annuaires et moteurs (combinant recherche arborescente et sur le contenu), ou le développement des portails, proposant tous les types de recherche, témoignent de cette imbrication d'outils, de techniques et de modalités de recherche différents, ajoutant parfois de la confusion au paysage et rendant les distinctions de plus en plus difficiles. Pour autant, les typologies d'outils restent fondamentales à maîtriser et à expliquer aux étudiants, pour sortir des "apparences de l'écran".

## **Du côté des processus de recherche : du "retrouage" booléen à la "sérendipité"**

Toutes ces évolutions, touchant aussi bien à l'offre, aux outils et aux modalités de recherche sur Internet, ont induit une autre transformation profonde, tenant à la fois aux procédures, aux processus mêmes et aux usages de la recherche d'information.

Dans l'univers documentaire -familier aux professionnels de l'information-, des bases de données, des catalogues de centres documentaires ou de bibliothèques, c.a.d. dans le monde de ce qu'on appelait la "RDI" (Recherche Documentaire Informatisée), les recherches se font avant tout selon la logique booléenne (par l'utilisation des opérateurs booléens, de troncature, éventuellement de proximité) et selon des règles de syntaxe plus ou moins formelles et complexes (cf les équations de recherche). Mais la principale caractéristique de la "RDI" tient au fait qu'il s'agit toujours de retrouver des références de documents préalablement saisies : la recherche porte toujours sur un fonds ou une base fermée (quelle que soit la taille de cette base), dont on peut connaître à l'avance le contenu exact ou la composition, et elle fait peu de place au hasard et à l'intuition. On sait ce qu'on (re)cherche. Et les notions de bruit et de silence s'appliquent parfaitement à cet univers documentaire.

La recherche sur le web offre à cet égard un tout autre aspect : le contenu est, par définition, infini, impossible à caractériser, à cerner et les modes de recherche sont variés, comme on l'a vu (logique booléenne, mais aussi navigation hypertexte, recherche sur le texte intégral...).

On peut certes (et on doit !) maîtriser toute la gamme des opérateurs, utiliser pleinement les fonctionnalités et les astuces de recherche des outils. Mais quiconque a fait l'expérience d'une recherche sur le web sait que les plus belles découvertes se font souvent par hasard, au gré des navigations de site en site, ou dans la liste des résultats d'un moteur.

Un terme, très employé aujourd'hui sur le web, désigne ce nouveau type de recherche, la **sérendipité**, traduction (non encore officielle, semble-t-il) du terme d'origine, *serendipity*, définie comme *"la découverte par chance ou sagacité de résultats que l'on ne cherchait pas"* (sur la sérendipité, voir les articles de Ertzscheid-Gallézet et S. Catellin).

Fondée sur l'intuition, l'association d'idées, mais aussi l'abduction, la sérendipité, qui caractérise également le mécanisme de l'invention et de la découverte scientifique, est devenue l'un des modèles de la recherche d'information sur internet, remettant au premier plan les questions de la navigation hypermédia, de l'apprentissage par exploration, de la conversion des informations en connaissances... Il ne faut pas y voir pour autant la fin de la nécessaire maîtrise des principes et des procédures de recherche ou l'inutilité de tout savoir technique sur les outils de recherche.

## **Du côté des modèles de la recherche d'information : de l'accès au traitement de l'information**

Selon une étude californienne, 16 % du temps de travail en entreprise serait consacré à rechercher l'information. *"Soit, si l'on transpose à un travailleur français soumis aux 35 heures, un peu plus de 5 heures 30 par semaine..."* Qui peut encore nier que la recherche d'information n'a pas un coût économique et ne représente pas un enjeu majeur ?

Et la question centrale aujourd'hui, face au "déluge informationnel", n'est plus tant la recherche elle-même que le traitement, l'exploitation des résultats. A quoi peuvent servir les 300 000 documents trouvés sur Google sur un sujet quelconque ? Comment filtrer, comment réduire le nombre de références, comment exploiter les listes de résultats de manière plus "intelligente", comment avoir un aperçu du contenu de telle page web, comment obtenir une analyse de tel corpus de données, etc., bref, comment mieux exploiter et gérer les informations : le véritable défi est là. Un certain nombre d'outils de recherche, parmi les plus innovants, travaillent depuis quelques années sur ces problématiques et ont apporté des innovations parfois spectaculaires (cf plus loin). Et il est intéressant d'observer une sorte de chassé-croisé entre les deux univers technologiques, que sont les logiciels documentaires et les outils du web : d'un côté les logiciels de



gestion documentaire "en local" (de type Superdoc, BCDI, JLB, etc.) évoluent de plus en plus vers les technologies du web, basculant parfois complètement sur une interface web et rendant transparentes la création et la diffusion de bases de données locales sur le web ; de l'autre côté, certains outils de recherche du web ont intégré des fonctionnalités de gestion documentaire, notamment d'analyse de corpus, ou des notions issues de la documentation (comme les thésaurus) et les appliquent à l'univers du web.

## **Du côté de l'économie de l'information : de la gratuité à la vente des mots-clés**

Enfin, une autre tendance lourde, parfois insuffisamment connue, se développe depuis plusieurs années et est en passe de transformer en profondeur le paysage de la recherche d'information : l'inscription, au cœur même des procédures de recherche, de la commercialisation. Il ne s'agit pas de la fin d'un prétendu âge d'or de la gratuité de l'information, qui n'a jamais vraiment existé. L'information a toujours eu un coût (le plus souvent supporté par le fournisseur) et souvent un prix pour l'utilisateur. Les banques de données professionnelles et scientifiques sont le plus souvent payantes (et souvent très chères), et ce depuis leur naissance.

La nouveauté réside ici dans cette nouvelle forme d'économie et de marché, apparue autour des outils de recherche privés du web et des enjeux financiers énormes, à la mesure du trafic généré par ces outils. Liens sponsorisés, liens commerciaux, "adwords"..., les techniques de ce qu'on appelle le "*positionnement payant*" ne cessent de se diversifier et de se développer, ajoutant un nouveau défi et un nouveau risque pour les usagers parfois distraits : savoir distinguer du premier coup d'œil un lien "sponsorisé" d'un résultat "normal". Rappelons que le positionnement payant consiste en un système compliqué de vente aux enchères de mots-clés, par des sociétés spécialisées (comme [Overture](#), [Espotting](#)) ou certains moteurs de recherche (comme [Google](#)). Cette vente de mots-clés permettra par exemple à un site commercial, spécialisé dans le voyage, d'apparaître en haut d'une page de résultats pour toute requête comprenant le mot "voyage". Avec le positionnement payant, c'est la notion même de pertinence qui est atteinte.

[Imprimer la lettre](#)

[Sommaire](#)





# Echo des Stages

## Utilisation des banques de données statistiques

Annie Léon (Urfist de Paris)

[L'Urfist de Paris](#) a réalisé le 11 Janvier 2005 un stage sur "l'utilisation des banques de données statistiques" à l'intention des personnels des bibliothèques responsables des acquisitions des ressources électroniques ou chargés de renseigner le public.

Objectifs du stage :

Un recensement des principaux produits a semblé utile, partant des constatations suivantes :

- les personnels des Bibliothèques universitaires ont rarement une connaissance du domaine des statistiques économiques et rencontrent donc des difficultés pour assister les lecteurs.
- le choix des banques de données statistiques n'est pas facile car les produits sont de niveaux variés et il faut trouver ce qui correspond aux besoins des publics.
- une évolution apparaît dans ce domaine avec l'ouverture en accès public et gratuit de banques de données importantes ([Eurostat](#) avec [Comext](#), [OCDE](#)).
- de nouveaux produits offrent des interfaces plus conviviales pour traiter différentes banques de données ([Ecowin](#)).
- la réforme LMD prévoit la formation des étudiants à la recherche documentaire.
- le développement des portails universitaires tend à offrir les ressources directement accessibles aux Centres de recherches ou au domicile des étudiants et des chercheurs : il faudra les aider à utiliser les ressources au mieux.

Le programme :

- Présentation du site **Eurostat** (Comext et New Cronos) par M. Noam Leandri d'INSEE Info Service (European Statistical Data Support France) avec des démonstrations et des travaux dirigés.
- **Chelem** par Mme Colette Herzog, responsable de la banque de données Chelem au CEPII : couverture du cédérom, modalités d'interrogation avec de très nombreux exemples et interprétation des résultats d'après les présentations graphiques des données extraites.
- Le site de l'**OCDE** par M. François Barnaud, responsable des ventes et de la promotion à l'OCDE : description du site et de son utilisation avec des exemples.
- **Ecowin** par Mme Oela Rasolofonjatovo d'Ecowin. Ecowin est une banque de données économiques et financières qui propose des données macroéconomiques internationales et de l'information financière quotidienne sur plus de 100 pays. Son logiciel permet d'interroger de manière uniforme toutes ces données.



Les 24 stagiaires se sont montrés satisfaits de l'information apportée, mais ont regretté le manque de temps pour des travaux pratiques. Une session supplémentaire entièrement consacrée à des travaux dirigés avec les producteurs va être organisée à leur intention. D'autres journées pourront aussi être consacrées à des banques de données statistiques qui n'ont pas été couvertes le 11 Janvier, si le besoin se manifeste.

[Imprimer la lettre](#)

[Sommaire](#)





# Sources et Méthodes

L'URFIST  
de  
Toulouse



## Présentation d'Ecoline (Janvier 2005)

Annie Léon (Urfist de Paris)

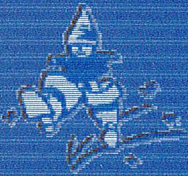
Le guide gratuit en Sciences économiques et de gestion, [Ecoline](#) a été réalisé à l'intention des étudiants des différents cycles universitaires. Il peut rendre service également dans le cadre de formations ou en auto-formation.

- Ses auteurs sont issus de grandes écoles (E.M. Lyon, ENS-LSH, HEC), d'universités (Bordeaux IV, IUT de Paris, Marne la Vallée, Paris 1, Paris 9 Dauphine, Paris 12 Créteil (et Nantes), Pôle universitaire Léonard de Vinci, (Reims), de la Bibliothèque publique d'information, d'URFIST (Bordeaux, Lyon, Paris) et la coordination est assurée par l'URFIST de Paris.
- Les matières traitées peuvent être abordées sous plusieurs angles :
  - Ecoline est un répertoire de ressources rigoureusement sélectionnées : des fiches détaillées décrivent le contenu et les usages de dictionnaires, d'encyclopédies et de glossaires ainsi que de ressources électroniques et de banques de données; une bibliographie structurée de manuels essentiels est proposée ; une importante liste de cours universitaires librement accessibles sur internet a été constituée ; des indications sur les grands réservoirs de thèses et de papiers de recherche sont fournies.
  - Ecoline est également un guide méthodologique d'accès à la documentation en Sciences économiques et de gestion : le chapitre sur la recherche d'information statistique propose une orientation au sein de sources multiples ; des "questions/réponses", s'appuyant sur des exemples concrets et commentés, développent une méthodologie de recherche sur les entreprises, les secteurs, les pays. On trouve aussi des "parcours", exercices commentés aidant à se familiariser avec des ressources électroniques incontournables.
- Pour aborder la recherche d'information dans Ecoline, plusieurs méthodes peuvent être utilisées :
  - par approches successives en suivant les menus
  - de manière systématique par les "thèmes"
  - par les mots-clef à l'aide du moteur de recherche.

Ecoline est évolutif : certaines rubriques seront renseignées prochainement, et une mise à jour régulière des ressources est prévue.

[Sommaire](#)









# Echo des Stages

## L'évaluation de l'information sur Internet

*Elisabeth Lacombe (Urfist de Toulouse)*

On peut consulter le support de cours à l'adresse :

[http://www.uhb.fr/urfist/Supports/StageEvalInfo/EvalInfo\\_cadre.htm](http://www.uhb.fr/urfist/Supports/StageEvalInfo/EvalInfo_cadre.htm)

Alexandre Serres (Maître de conférence à l'Urfist de Rennes) a animé à Toulouse le 6 octobre 2004 un stage sur le thème de l'évaluation de l'information sur Internet.

Le contexte général : numérisation qui a complètement transformé la chaîne des savoirs, diminution des valeurs de vérité au profit de celles de l'efficacité, survalorisation de la pertinence, virtualisation du texte..., met en avant la nécessité et la difficulté de l'évaluation de l'information.

Les risques de l'info-pollution sont soulignés : surabondance d'informations qui tue l'information, désinformation, contamination, invasion publicitaire. Les enjeux de l'information sont immenses pour les professionnels de l'information comme pour tous les usagers. Une étude américaine montre qu'Internet est la première source d'information pour 70% des utilisateurs d'Internet, alors que 50% de ceux-ci pensent qu'elle n'est pas fiable... Le modèle traditionnel a changé pour les professionnels de l'information, c'est l'utilisateur qui valide les informations ; cependant ceux-ci ont toujours un rôle à jouer par la sélection des sources et par leur rôle de formateurs.

Cette problématique de l'évaluation est prise en compte par des réseaux d'acteurs comme Centrale Santé, [Net Scoring](#) (évaluation de sites médicaux), des entreprises et sites spécialisés dans l'évaluation comme [Temesis](#), des institutions adoptant une démarche qualité : [Cadre de qualité de Bruxelles](#) (les sites culturels), des communautés d'utilisateurs comme les Web Ring.

L'évaluateur de la qualité de l'information sur Internet se trouve devant un certain nombre de difficultés d'ordre psycho-cognitif et technique (caractéristiques d'Internet, spécificités du document numérique) et d'ordre méthodologique (multiplicité des domaines, difficultés méthodologiques de l'évaluation, confusions sur la notion d'information, flou autour des critères de qualité et de pertinence). Voici quelques-unes de ces difficultés à prendre en compte :

- le document numérique mélange "les strates du document", les compétences et les genres
- Internet, c'est le déluge informationnel, l'hétérogénéité des sources et leur fiabilité incertaine, l'absence de normes de présentation...
- la complexité et l'ambiguïté inhérentes à la notion d'information : elle est sociale, elle n'existe pas en soi (définition de Jeanneret), elle est contextualisée, elle a un caractère matériel qui souligne l'importance de la mise en forme pour trouver l'information
- la notion de pertinence qui est "le degré de corrélation entre une question et la réponse



apportée", et qui s'applique à la fois au système, à l'utilisateur et au thème (selon S. Dalbin)

- la notion de qualité : une information doit être identifiée (source), fiable, novatrice (40% de redondance sur Internet), utile, précise...

En étant conscient de ces aspects caractéristiques d'Internet et de l'importance de l'information pertinente, on peut établir une méthodologie dans le processus d'évaluation :

- respecter les étapes : identifier, évaluer, valider
- se fixer des objectifs généraux et spécifiques : le recensement de ressources pertinentes (évaluation du contenu), l'amélioration de la visibilité et de la notoriété (mise en forme et diffusion)
- définir ce qui est évalué : information, document, site (structuration, graphisme, accessibilité...)
- établir, lister et hiérarchiser des critères d'évaluation selon ses objectifs
- identifier et utiliser des normes de référence : normes de la qualité documentaire (qualité de l'information, normes d'identification, de description...)
- utiliser des grilles d'évaluation dans la pratique, en les adaptant à l'objet et à l'objectif, sans s'enfermer dans des grilles figées.

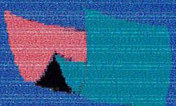
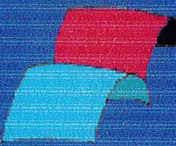
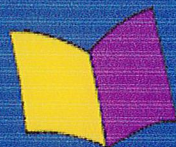
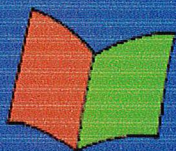
Des exemples de grilles d'évaluation sont proposées.

[Imprimer la lettre](#)

[Sommaire](#)



L'URFIST  
de  
Toulouse



# Echo des Stages

## BibliOpéra, logiciel de gestion d'informations bibliographiques

*François Magnan (Urfist de Toulouse)*

d'après la présentation du logiciel BibliOpera sur le site <http://bibliopera.mom.fr>

Une présentation du logiciel BibliOpera a eu lieu à l'Urfist de Paris le 1<sup>er</sup> février 2005 par Marjorie Burghart, conceptrice de ce projet. BibliOpera est un logiciel libre et gratuit, basé sur PHP 4, MySQL et XML/XSLT, qui permet la gestion des informations bibliographiques et leur diffusion sur Internet.

Conçu pour répondre aux besoins spécifiques des chercheurs et des laboratoires, BibliOpera veut se placer à la frontière des logiciels de gestion d'informations bibliographiques (information dématérialisée, importance de la présentation, etc.) et des logiciels de gestion de bibliothèque (respect des normes, environnement multi-utilisateurs). BibliOpéra gère la construction et la diffusion de bibliographies, mais il n'est pas conçu pour la gestion bibliothéconomique, même si certaines de ses fonctionnalités s'en approchent.

Par rapport au logiciel de gestion de références bibliographiques Endnote, qu'il ne prétend pas remplacer, BibliOpéra est :

- Multi-utilisateur : une équipe peut y travailler, avec un administrateur qui supervisera le travail collectif
  - Orienté Web : il permet une publication directe sur Internet ou sur un Intranet, accessible depuis n'importe quel navigateur, il sera ainsi possible de travailler sur Mac, Linux, PC/Windows, etc.
  - Un logiciel libre, dont le code source est accessible et modifiable, ce qui peut permettre de nombreux paramétrages.

Ses autres caractéristiques fonctionnelles sont :

- Une présentation de l'information paramétrable : une même bibliographie peut être diffusée sous différentes mises en forme, grâce à des feuilles de style XSL, que l'on pourra modifier pour correspondre aux habitudes des lecteurs : ainsi, à partir des mêmes données, on pourra diffuser une bibliographie abrégée sur un site Internet, et une version plus détaillée (avec la localisation des documents) sur un réseau Intranet.
- Une possibilité d'indexation hiérarchique ou par mots-clés : il est possible d'indexer les notices bibliographiques soit en utilisant un système classique de mots-clés, soit, de façon plus originale, en utilisant une liste hiérarchique.
- Une gestion étendue des "listes d'autorités" : les informations bibliographiques nécessitent une certaine uniformité au sein d'une base (noms d'auteurs, titres de revues et de collections, etc.) BibliOpera permet de contrôler ces "autorités" sur des listes, ce qui évite d'introduire trop de variantes dans l'orthographe de ces entrées.



- En plus des informations strictement bibliographiques, il sera possible de préciser l'origine de la notice, ce qui peut servir lorsque différentes personnes construisent en commun une bibliographie. Il sera aussi possible de préciser la localisation, ce qui permet d'indiquer où est disponible l'oeuvre concernée par la notice.
- La possibilité d'importer des références bibliographiques : BibliOpera permet de consulter des catalogues de bibliothèque, et d'importer les notices, lorsque ces catalogues respectent la norme Z3950.

[Imprimer la lettre](#)

[Sommaire](#)