HAL
open science

Evaluation des sites Internet. Construction d'indicateurs à partir des réseaux de "sitations

Hervé Rostaing, Eric Boutin

To cite this version:

Hervé Rostaing, Eric Boutin. Evaluation des sites Internet. Construction d'indicateurs à partir des réseaux de "sitations. Les systèmes d'informations élaborées, Société Française de la Bibliométrie Appliquée, Oct 1999, Ile Rousse, France. hal-01579939

HAL Id: hal-01579939

https://hal.science/hal-01579939v1

Submitted on 31 Aug 2017

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.

Proposition de communication pour le colloque Ile Rousse 99

Evaluation des sites Internet.

Construction d'indicateurs à partir des réseaux de " sitations "

Hervé ROSTAING
CRRM

Université Aix-Marseille
Centre Saint Jérôme
13397 Marseille Cedex 20

Tél : 04 91 28 86 77
Fax : 04 91 28 87 12
Mél : herve@crrm.univ-mrs.fr

Eric BOUTIN
LEPONT

Université de Toulon et du Var
IUT TC, BP 132
83957 La Garde

Tél : 04.94.14.22.16
Fax : 04.94.14.22.75
Mél : boutin@univ-tln.fr

Mots-clés :
Analyse Information Internet, Cartographie Internet, Cybermétrie, Evaluation des Ressources Internet, Analyse des Sitations, Bibliométrie, Analyse Réseaux.

Résumé :

L'Internet est de nos jours une nouvelle source d'information incontournable. Par contre, qui n'a pas été déjà confronté au phénomène de sur-information lors de la recherche sur un moteur ou de désorientation lors de la navigation sur Internet. Ces problèmes sont persistants sur le réseau Internet, du fait même de sa construction, et sont à l'origine de pertes de temps considérables.
Comment dans ce foisonnement de données juger l'importance des pages Internet et des sites les contenant ? Comment rationaliser la gestion de ces ressources et améliorer la performance de leur exploitation ?

La communication abordera dans un premier temps l'ensemble des différentes approches actuellement développées pour aider l'utilisateur à mieux gérer l'information présentes sur Internet. Ces techniques ont principalement pour principe l'évaluation de pages ou des sites Internet. Nous évoquerons les réflexions qui actuellement ont cours pour l'évaluation des ressources Internet [1]. Nous verrons que les principaux critères proposés font bien souvent appel à une évaluation purement intellectuelle.

Malgré tout, il parait indispensable de concevoir des outils de traitement automatique pour l'aide à l'évaluation de ces nouvelles ressources. Nous passerons rapidement en revue les différentes techniques d'évaluation actuellement disponible en essayant de les classer par catégorie :

- analyse du contenu lexical des pages Web : classement par pertinence (moteurs de recherche ou robots tels que DigOut4U, Search'97) ; simple classement (Search'97, Nomino) ; agrégation statistique suivie d'une cartographie soit parcellaire (Sampler) soit intégrale (WebSom, Neuronav+, SemioMap, Umap, TKS, Tétralogie)
- analyse du contenu multimédia des pages Web [2]
- analyse de la dimension hypertextuelle des pages Web : analyse de l'espace des relations entre pages [3]; analyse du phénomène de référencement entre sites, la " sitation " [4]

Nous positionnerons alors notre approche, l'analyse des réseaux de " sitations ", par rapport à ces techniques. Une description de l'ensemble des indicateurs et cartographie réseaux mise au point sera exposée.

Pour conclure, nous présenterons une étude de cas mettant en application cette méthode d'analyse réseaux des " sitations ". Cette expérimentation a été menée sur un corpus de données extrait de l'Internet provenant des sites traitant de la bibliométrie-scientométrie-infométrie. Ces données ont été collectées et mises en forme grâce au robot Auresys [5], puis ensuite traités par le logiciel Dataview [6] pour finalement être exploitées par le logiciel Matrisme [7]. La démarche complète de cette étude ainsi que les principaux résultats seront exposés. Cette méthodologie d'analyse de l'espace des cybernautes a déjà fait l'objet d'expérimentation sur de précédents jeux de données (voir figure 1) [8]

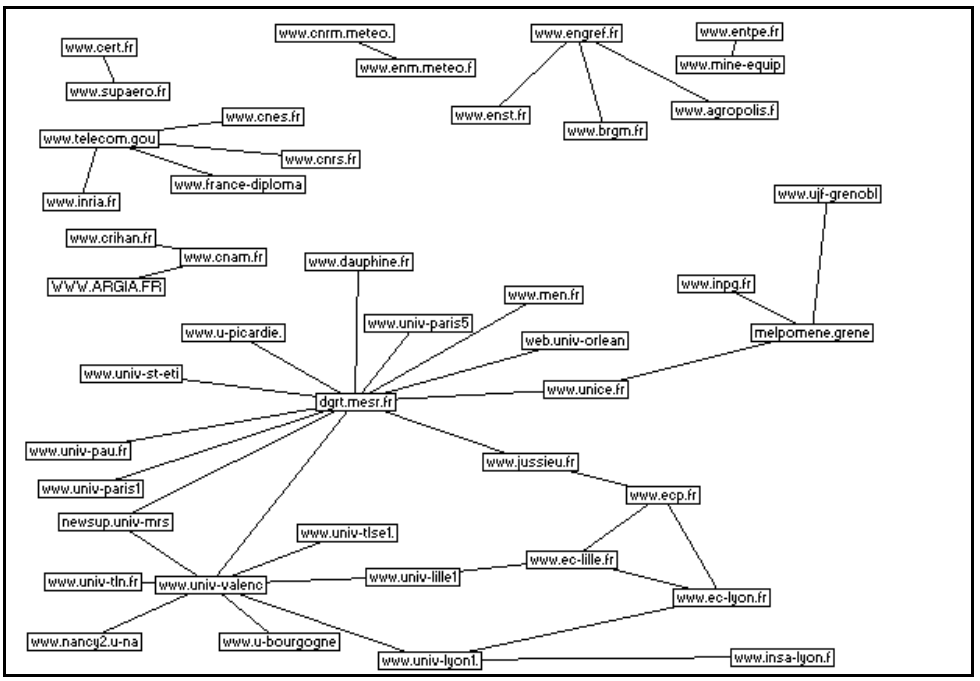

figure 1 : réseau des sitations réciproques entre les sites de l'administration française en 1996

Bibliographie :

[1] Monique Joly et Jean-Michel Mermet. " Evaluation de l'information présente sur Internet " [On-line]. Doc'INSA - INSA de Lyon 1998. Available from internet : URL:http://csidoc.insa-lyon.fr/sapristi/fristi36.html

" Critères d'évaluation des sources d'information d'Internet " [On-line]. Collège de Bois-de-Boulogne. Available form internet : URL:http://www.virtuel.collegebdeb.qc.ca/acritere.html

Smith, Alastair. " Evaluation of information sources " [On-line]. Mar. 1998. Available from internet : URL:http://www.vuw.ac.nz/~agsmith/evaln/evaln.htm

[2] Woodruff A, Aoki P, Brewer E, Hauthier P, Rowe L, " An investigation of documents from the World Wide Web. ", Actes du colloque : *Fifth International World Wide Web Conference, Paris, France, 1996* Available from internet : URL:.http://www5conf.inria.fr/fich_html/papers/P7/Overview.html

[3] Mukherjea Sougata, Foley James, "Visualizing the World-Wide Web with the Navigational View Builder", *Computer Networks and ISDN System*, Special Issue on the Third International Conference on the World-Wide Web '95, Darmstadt, Germany, 1995. [On-line] Available from internet URL:http://www.cc.gatech.edu/gvu/people/Phd/sougata/www/mukh.html

[4] Almind Tomas, Ingwersen Peter " Informetric analysis on the world wide web : methodological approaches to webometrics ", *Journal of documentation*, Vol 53, N°4, pp. 404-426

Larson R, " Bibliometrics of the world wide web : an exploratory analysis of the intellectual structure of cyberspace [On-Line]. Available from internet : URL: http://sherlock.berkeley.edu/asis96/asis96.html

McKiernan G, " CitedSites(sm): Citation Indexing of Web resources ". [On-line] Available froom internet : URL:http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm

[5] Mannina B, Dou H, Florence B, " Internet and the regional innovation : Auresys, the regional robot ", Acte du colloque : *IDT'97, Paris, 1997*

[6] Rostaing H, Dou H, Paoli C, " DATAVIEW: Bibliometric software for analysis of downloaded data ", Acte du colloque : *Fourth congress on bibliometrics, scientometrics, informetrics, Berlin, 1993*

[7] Boutin Eric, *Le traitement d'une information massive par l'analyse réseau : méthode, outils et applications,* Thèse de Doctorat soutenue en 1999 à l'Université d'Aix-Marseille

[8] Boutin Eric, Mannina Bruno, Rostaing Hervé, Quonaim Luc, “ Construction automatique de réseaux : un outil pour mieux appréhender l'information provenant de l'Internet ”, Acte de colloque : *JADT’98, Nice, 1998*