

# Chronique

de la recherche sur le web

par Marc Duval

26 février 2001, vol.1, n° 3

## L'étendue d'Internet

**D**epuis plusieurs années, des auteurs ont publié des études sur l'étendue du web.

Le 8 juillet 1999, MM. Steve Lawrence et C. Lee Giles annoncent que le web public a 800 millions de pages dans la revue Nature.

Le 18 janvier 2000, Inktomi et NEC Research Institute publient une étude selon laquelle le web compterait 1 milliard de pages web.

Le 11 juillet 2000, la société Cyveillance évalue à plus de 2 milliards de pages web.

Durant le même mois, la société BrightPlanet révèle que le web contiendrait environ 500 milliards, pas de pages web mais de documents. En mal de sensationnalisme, les échetiers transforment les 500 milliards de documents en 500 milliards de pages web, mettant en doute la capacité des moteurs de recherche à indexer le web.

Le 16 octobre 2000, une autre étude est publiée par la Online Computer Library Center ( OCLC ). Cette étude se base sur le nombre de sites. Selon celle-ci, la progression du nombre de sites est : 1,570,000 en 1997, 2,851,000 en

1998, 4,882,000 en 1999 et 7,399,000 en 2000. Elle divise le web en trois partis : le web public, le web privé et le web provisoire. Le web public est libre d'accès. Le web privé demande aux visiteurs un mot de passe. Enfin le web provisoire comprend des sites « en construction », au contenu non défini, vide de sens ou superficiel.

Selon les statistiques, il y aurait 7,1 millions de sites uniques. En terme de pourcentage, le web public représente 41%, le web privé, 21%, le web provisoire 37% et le web «adulte» 1%. Or, selon l'étude de 1999, les principaux moteurs de recherche couvrent 60% du web public.

Malgré les différences méthodologiques, les moteurs de recherche, tous confondus, indexent en grande partie le web public.

---

Marc Duval  
Service de recherche documentaire DSI  
1280-4, bd de Montarville.  
Boucherville. Québec. Canada  
J4B 8B4  
marc.duval@videotron.ca