W3Perl

Statistiques de consultation de sites Web

Laurent Domisse
(INRP)

Rôle

- **Transparence / Service**
- **Surveillance**
  - Bande passante
  - Impact événementiel
- **Amélioration du site**
  - Accroître la visibilité
  - Connaître son public (langues,références...)
  - Résoudre les problèmes

TOP SECRET

Différentes méthodes

- **Compteur**
  - Insertion de balises...
- **Fichier de log**
  - Serveur web...
- **Audit**
  - Application propriétaire...

Fichier de log

- **Serveur**
  - Unité de mesure : Requêtes/Pages
- **Gestion des logs**
  - Rotation
  - Compression
- **Format des logs**

Format des logs

- **De nombreuses variantes**
  - CLF : standard
  - ECLF,W3C, IIS, Lotus, Squid…
- **ECLF**

198.45.0.12 - - [24/Mar/2002:15:13:27 -0500] "GET /fr/ HTTP/1.1" 200 45 "http://www.mydomain.com/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)"

host login passwd date hour methode URL protocole status taille referer agent

Extraction

- **Stats de base**
  - Page, Site, Répertoires, Fichier, Téléchargement, Trafic, Pays, Scripts
- **Stats temporelles**
  - Heures, Jours, Semaines, Mois
- **Stats optionnelles**
  - Références, navigateur, OS, virus, session,erreur, URL mapping

W3Perl

- **Perl / ActivePerl**
  - Script par stats, librairie
- **Serveurs Unix-Linux / NT**
- **Fichier de configuration**
  - Définition des règles
- **Interface Web d 'administration**
- **Mode incrémental**

Installation

- **Perl / Fly**
- **Windows (9x/NT/XP/2000)**
  - IIS / Apache ou sans serveur
- **Unix**
  - Install.pl
    - Scripts --> /cgi-bin/w3perl
    - Ressources --> /htdocs/w3perl
  - RPM en cours

Fichier de config.

- **Un fichier par serveur / stats**
- **Gestion par interface web**
- **Etapes**
  - Serveur/Fichier de log
  - Affichage
  - Filtres
  - Préférences

Interface d'administration

- Fichier de configuration
  - Créer / Modifier / Effacer / Dupliquer
- Lancement
  - Contrôle à distance
- Mise à jour

Configuration INRP

Netscape Proxy (NT)
ECLF

Squid (Linux)
CLF + refer + agent

Transfert par FTP

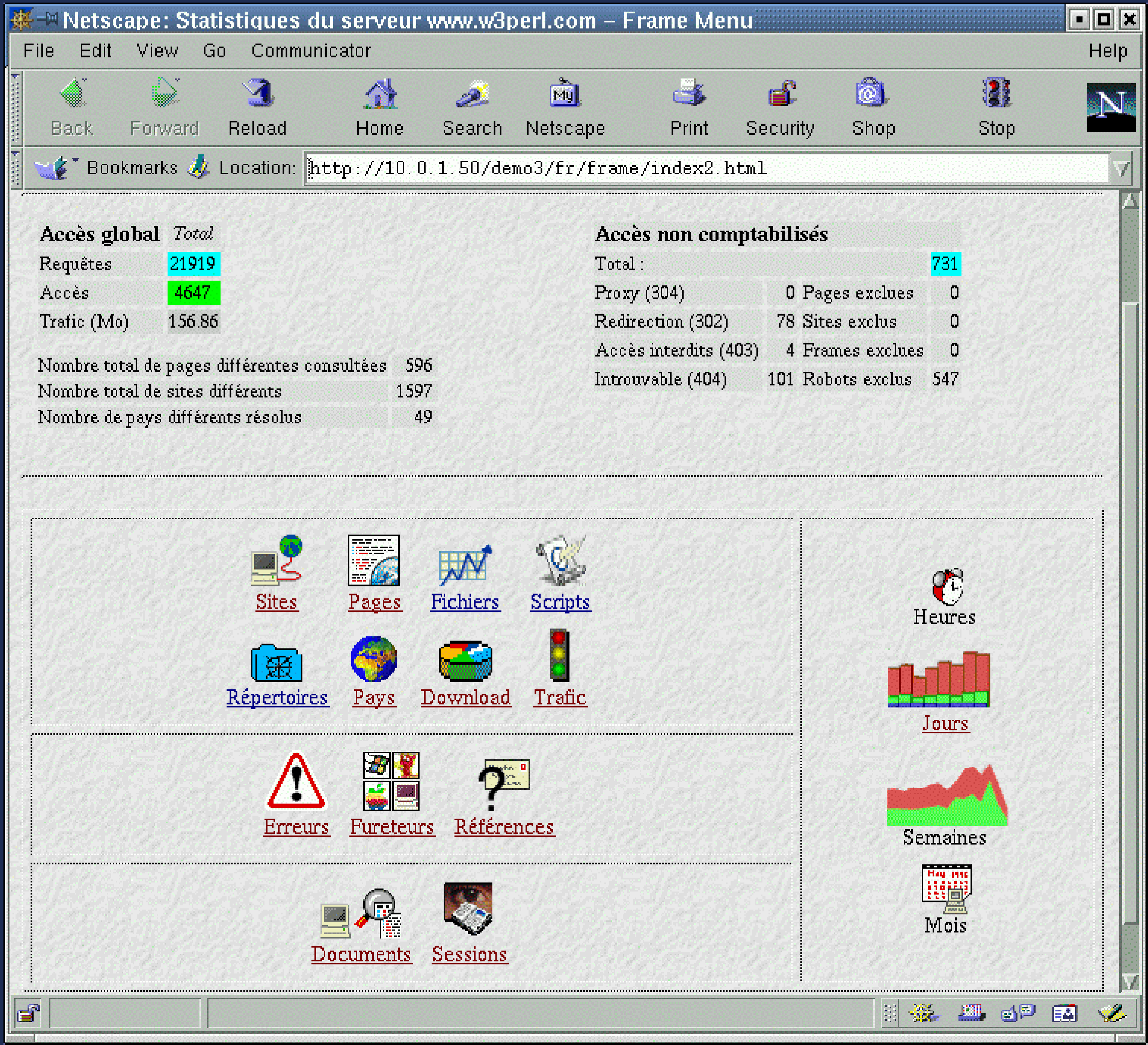
Titan (Linux)

Grep sur la date courante
Filtrage des images
Concaténation + Compression

Lancement

- **Initialisation / Incrémental**
- **Script de lancement**
  – Gère les autres scripts
  – Date de lancement
  – Option de lancement
  – Crontab
- **Interface web d'administration**

Page principale



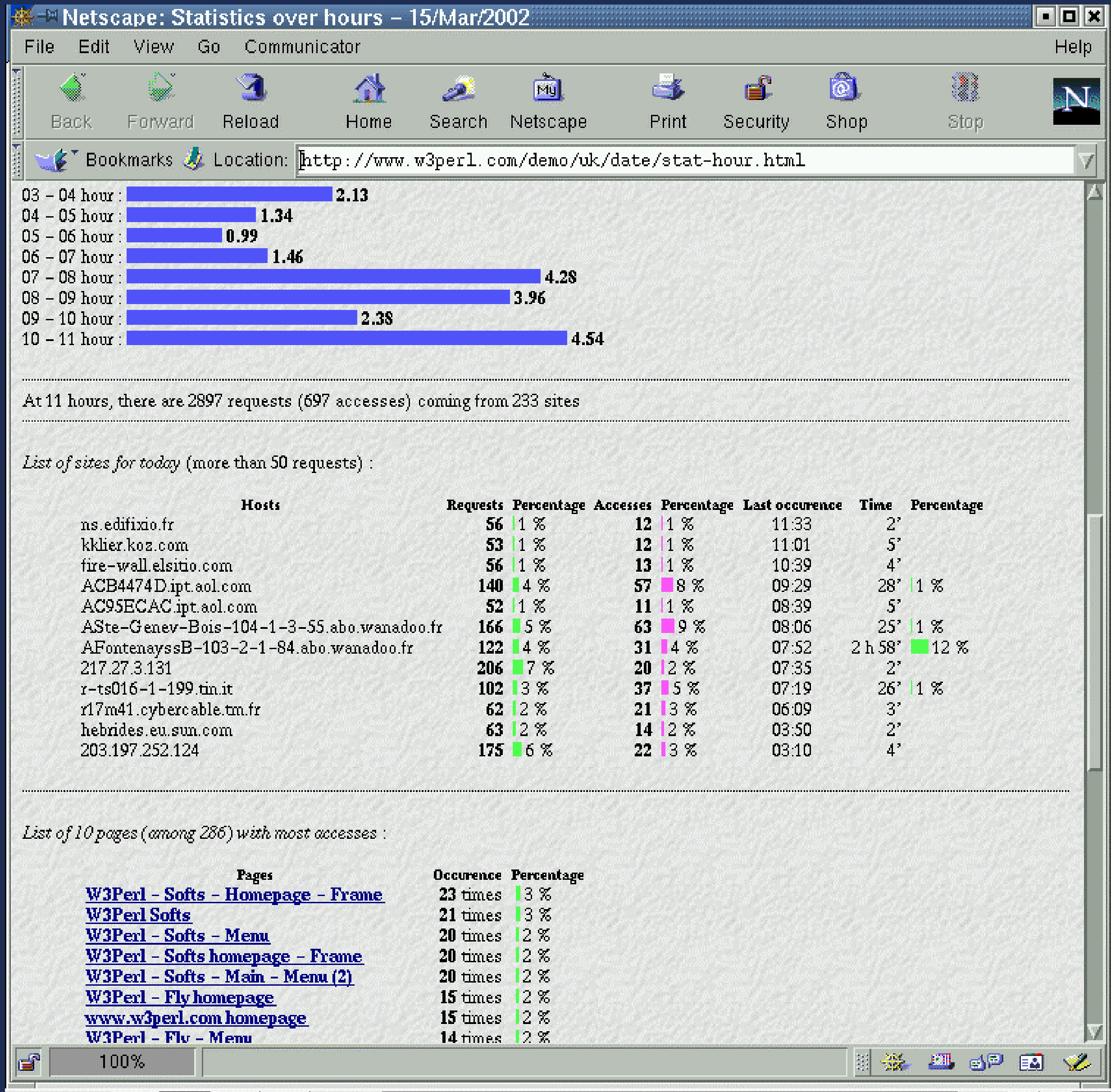
Bourse aux Outils

28-03-2002

Proxy

- **Cache web**
  - réduire la bande passante
  - sécurité
- **Transaction**
  - 200 : OK
  - 304 : Not modified

Stats par heures



Netscape: Statistics over hours – 15/Mar/2002

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://www.w3perl.com/demo/uk/date/stat-hour.html

03 – 04 hour : 2.13
04 – 05 hour : 1.34
05 – 06 hour : 0.99
06 – 07 hour : 1.46
07 – 08 hour : 4.28
08 – 09 hour : 3.96
09 – 10 hour : 2.38
10 – 11 hour : 4.54

At 11 hours, there are 2897 requests (697 accesses) coming from 233 sites

List of sites for today (more than 50 requests) :

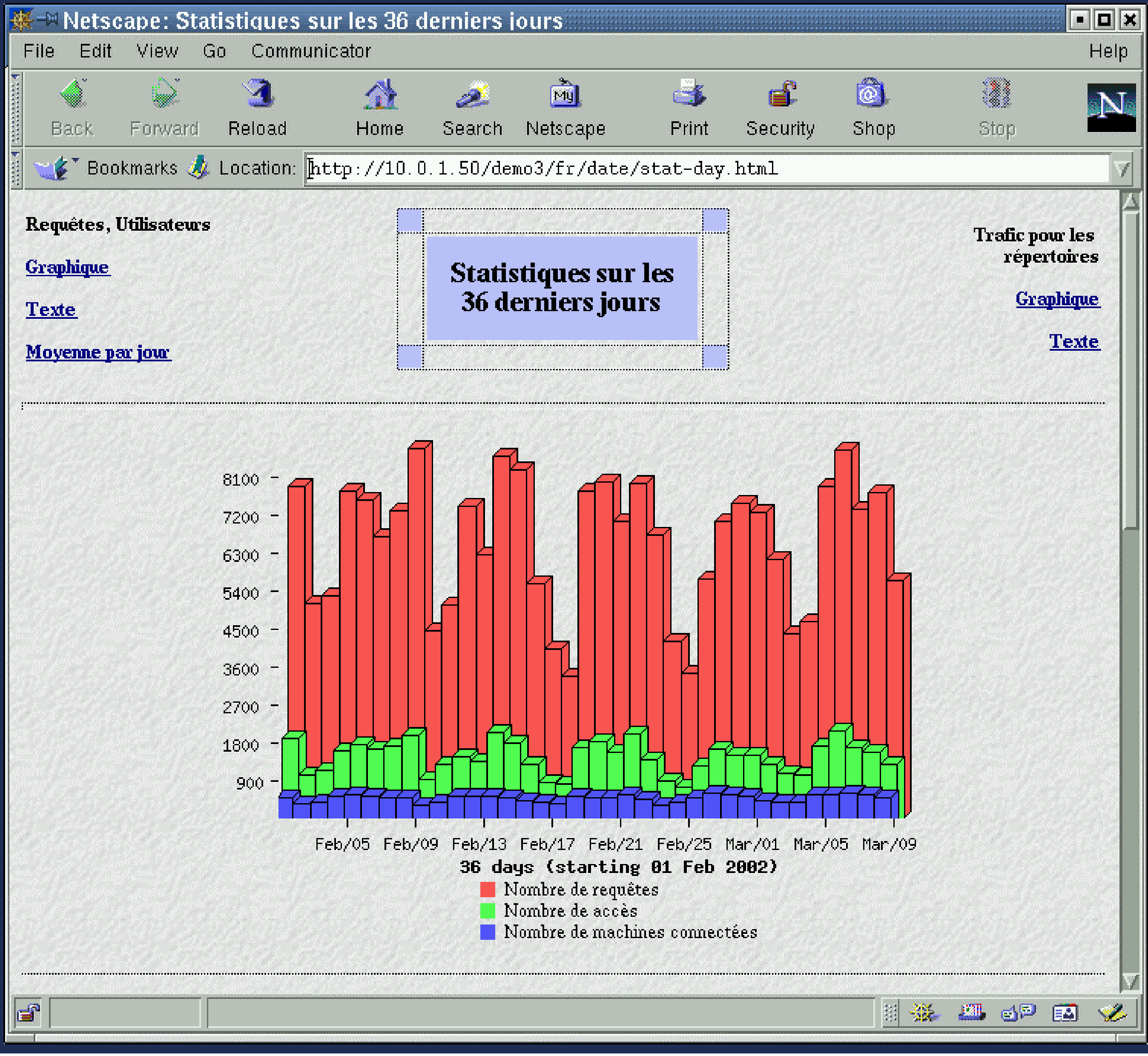
Hosts	Requests	Percentage	Accesses	Percentage	Last occurence	Time	Percentage
ns.edifixio.fr	56	1 %	12	1 %	11:33	2'	
kklier.koz.com	53	1 %	12	1 %	11:01	5'	
fire-wall.elsitio.com	56	1 %	13	1 %	10:39	4'	
ACB4474D.ipt.aol.com	140	4 %	57	8 %	09:29	28'	1 %
AC95ECAC.ipt.aol.com	52	1 %	11	1 %	08:39	5'	
ASte-Genev-Bois-104-1-3-55.abo.wanadoo.fr	166	5 %	63	9 %	08:06	25'	1 %
AFontenayssB-103-2-1-84.abo.wanadoo.fr	122	4 %	31	4 %	07:52	2 h 58'	12 %
217.27.3.131	206	7 %	20	2 %	07:35	2'	
r-ts016-1-199.tin.it	102	3 %	37	5 %	07:19	26'	1 %
r17m41.cybercable.tm.fr	62	2 %	21	3 %	06:09	3'	
hebrides.eu.sun.com	63	2 %	14	2 %	03:50	2'	
203.197.252.124	175	6 %	22	3 %	03:10	4'	

List of 10 pages (among 286) with most accesses :

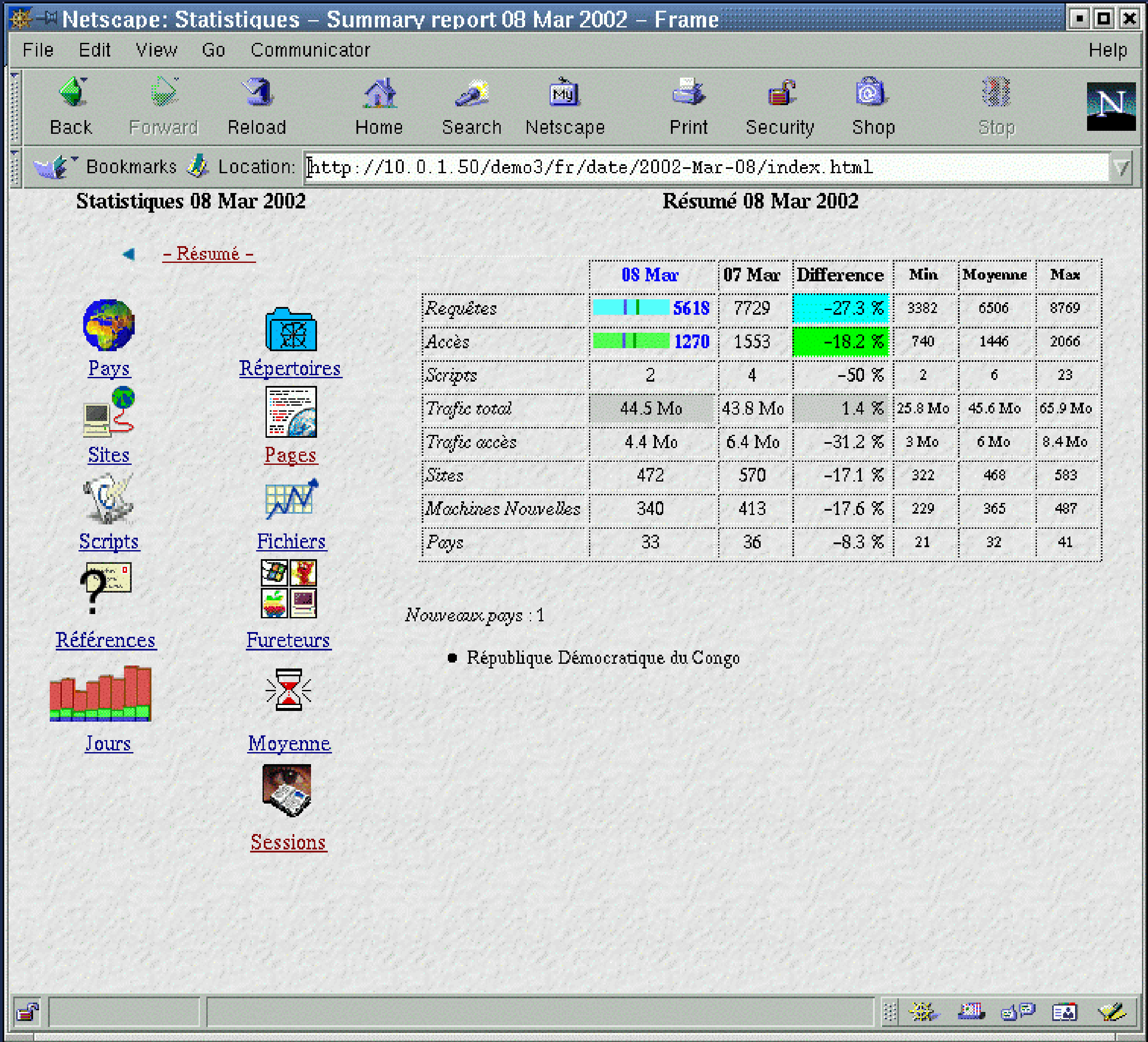
Pages	Occurence	Percentage
W3Perl - Softs - Homepage - Frame	23 times	3 %
W3Perl Softs	21 times	3 %
W3Perl - Softs - Menu	20 times	2 %
W3Perl - Softs homepage - Frame	20 times	2 %
W3Perl - Softs - Main - Menu (2)	20 times	2 %
W3Perl - Fly homepage	15 times	2 %
www.w3perl.com homepage	15 times	2 %
W3Perl - Fly - Menu	14 times	2 %

100%

Stats sur les jours



Stats quotidiennes



Statistiques 08 Mar 2002

- Résumé -

Pays
Répertoires
Sites
Pages
Scripts
Fichiers
Références
Fureteurs
Jours
Moyenne
Sessions

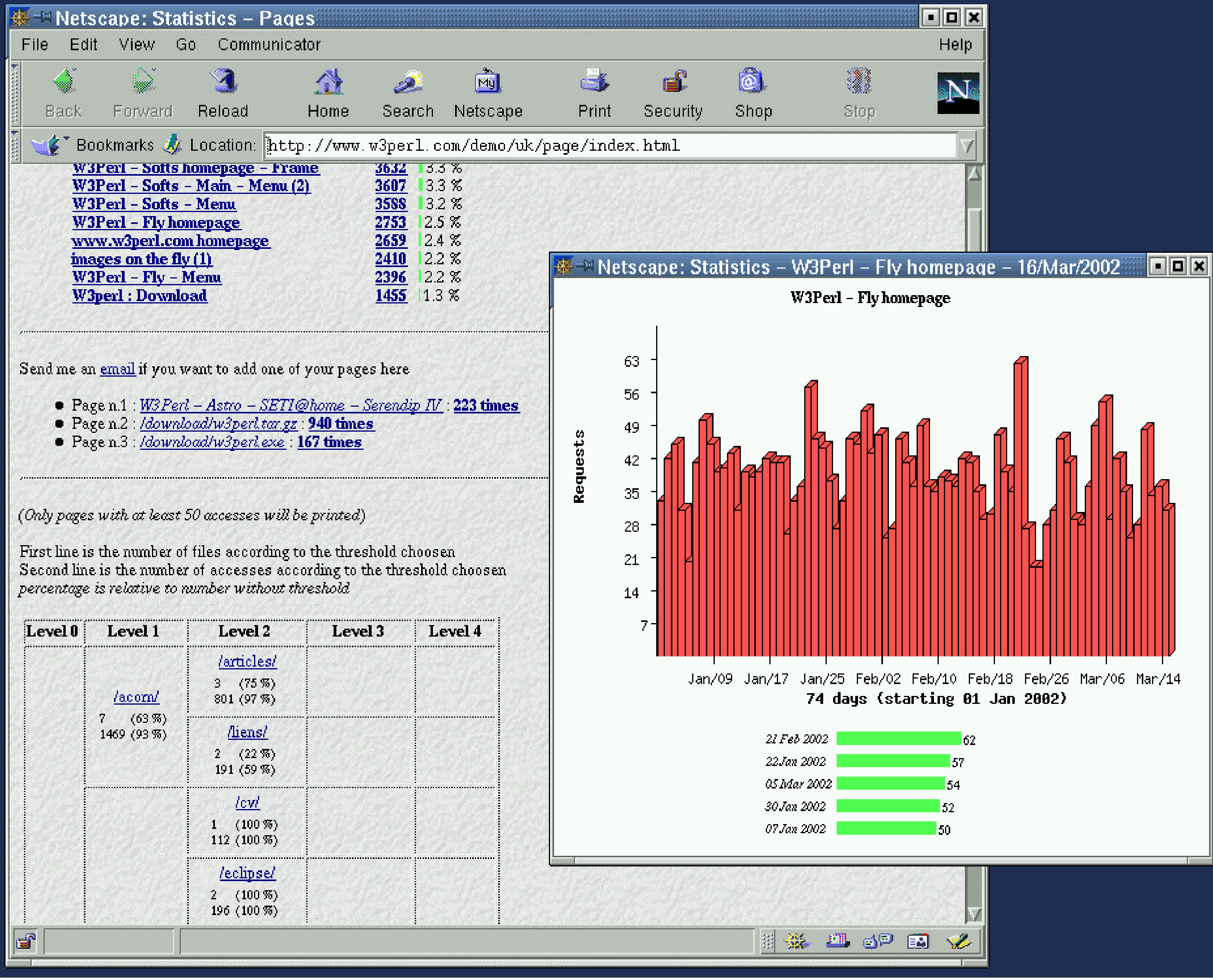
Résumé 08 Mar 2002

	08 Mar	07 Mar	Difference	Min	Moyenne	Max
Requêtes	5618	7729	-27.3 %	3382	6506	8769
Accès	1270	1553	-18.2 %	740	1446	2066
Scripts	2	4	-50 %	2	6	23
Trafic total	44.5 Mo	43.8 Mo	1.4 %	25.8 Mo	45.6 Mo	65.9 Mo
Trafic accès	4.4 Mo	6.4 Mo	-31.2 %	3 Mo	6 Mo	8.4 Mo
Sites	472	570	-17.1 %	322	468	583
Machines Nouvelles	340	413	-17.6 %	229	365	487
Pays	33	36	-8.3 %	21	32	41

Nouveaux pays : 1

- République Démocratique du Congo

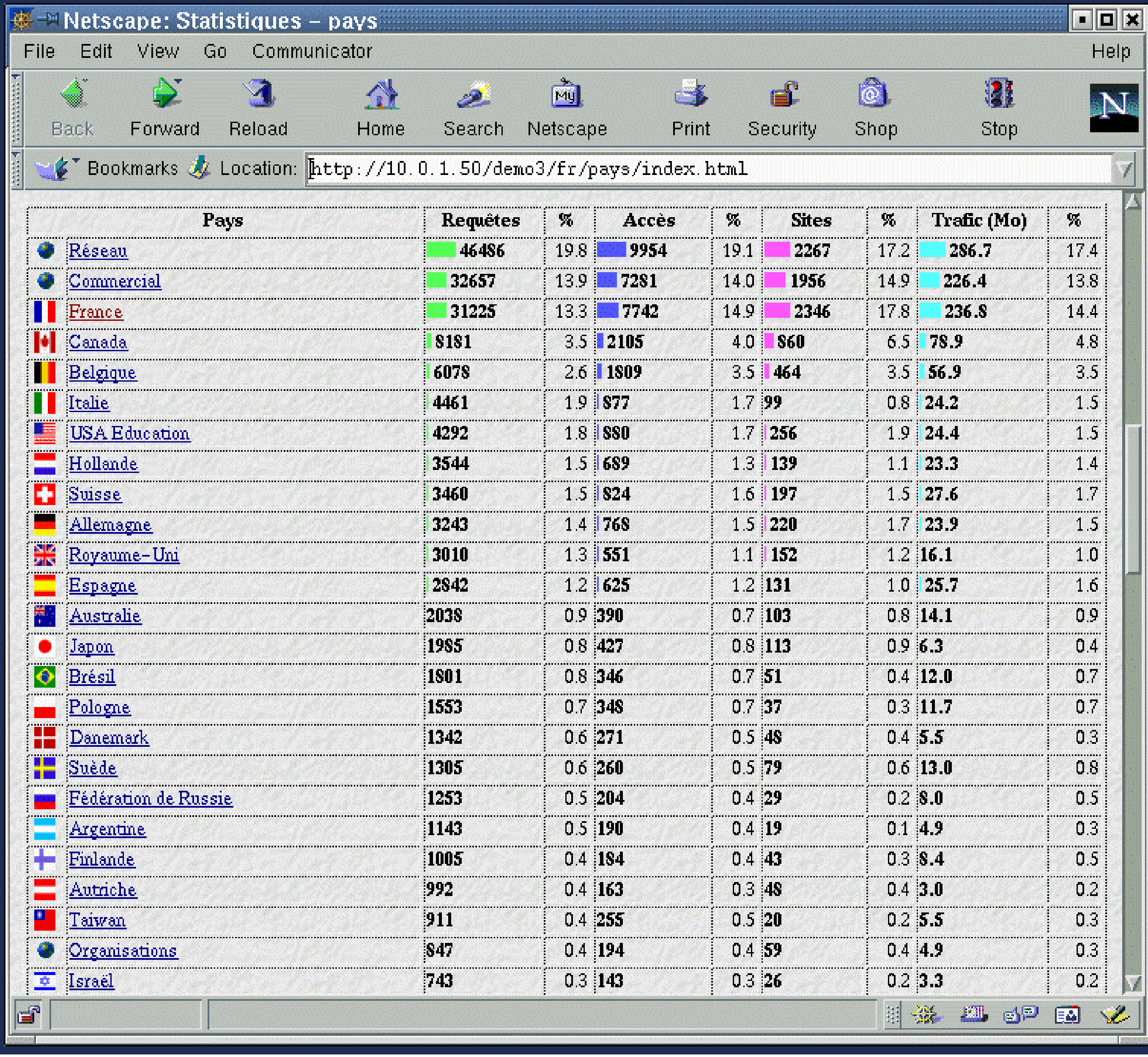
Stats sur les pages



Bourse aux Outils

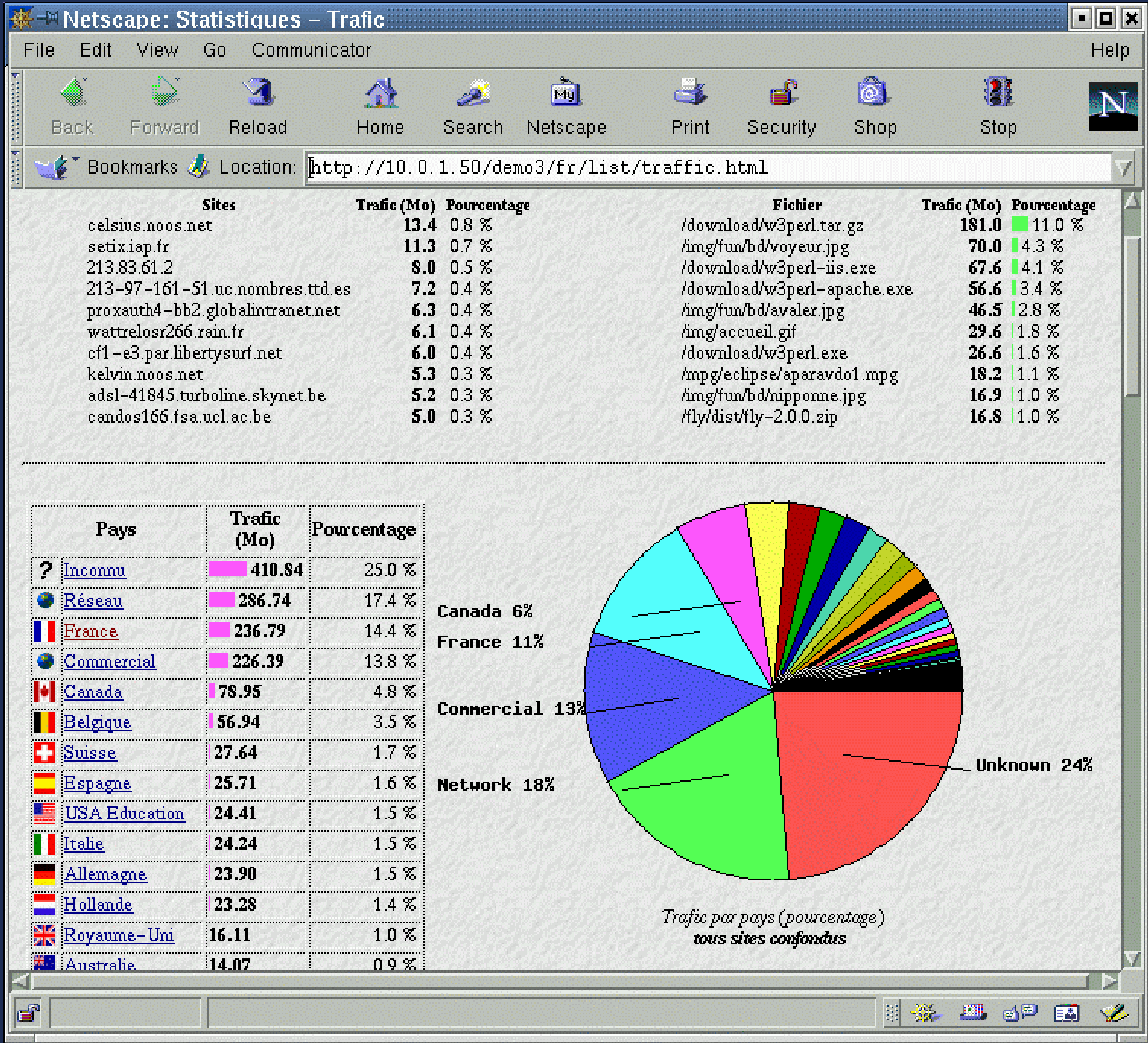
28-03-2002

Stats sur les pays



Pays	Requêtes	%	Accès	%	Sites	%	Trafic (Mo)	%
Réseau	46486	19.8	9954	19.1	2267	17.2	286.7	17.4
Commercial	32657	13.9	7281	14.0	1956	14.9	226.4	13.8
France	31225	13.3	7742	14.9	2346	17.8	236.8	14.4
Canada	8181	3.5	2105	4.0	860	6.5	78.9	4.8
Belgique	6078	2.6	1809	3.5	464	3.5	56.9	3.5
Italie	4461	1.9	877	1.7	99	0.8	24.2	1.5
USA Education	4292	1.8	880	1.7	256	1.9	24.4	1.5
Hollande	3544	1.5	689	1.3	139	1.1	23.3	1.4
Suisse	3460	1.5	824	1.6	197	1.5	27.6	1.7
Allemagne	3243	1.4	768	1.5	220	1.7	23.9	1.5
Royaume-Uni	3010	1.3	551	1.1	152	1.2	16.1	1.0
Espagne	2842	1.2	625	1.2	131	1.0	25.7	1.6
Australie	2038	0.9	390	0.7	103	0.8	14.1	0.9
Japon	1985	0.8	427	0.8	113	0.9	6.3	0.4
Brésil	1801	0.8	346	0.7	51	0.4	12.0	0.7
Pologne	1553	0.7	348	0.7	37	0.3	11.7	0.7
Danemark	1342	0.6	271	0.5	48	0.4	5.5	0.3
Suède	1305	0.6	260	0.5	79	0.6	13.0	0.8
Fédération de Russie	1253	0.5	204	0.4	29	0.2	8.0	0.5
Argentine	1143	0.5	190	0.4	19	0.1	4.9	0.3
Finlande	1005	0.4	184	0.4	43	0.3	8.4	0.5
Autriche	992	0.4	163	0.3	48	0.4	3.0	0.2
Taiwan	911	0.4	255	0.5	20	0.2	5.5	0.3
Organisations	847	0.4	194	0.4	59	0.4	4.9	0.3
Israël	743	0.3	143	0.3	26	0.2	3.3	0.2

Stats téléchargements



Sites	Trafic (Mo)	Pourcentage
celsius.noos.net	13.4	0.8 %
setix.iap.fr	11.3	0.7 %
213.83.61.2	8.0	0.5 %
213-97-161-51.uc.nombres.ttd.es	7.2	0.4 %
proxauth4-bb2.globalintranet.net	6.3	0.4 %
wattrelosr266.rain.fr	6.1	0.4 %
cf1-e3.par.libertysurf.net	6.0	0.4 %
kelvin.noos.net	5.3	0.3 %
adsl-41845.turboline.skynet.be	5.2	0.3 %
candos166.fsa.ucl.ac.be	5.0	0.3 %

Fichier	Trafic (Mo)	Pourcentage
/download/w3perl.tar.gz	181.0	11.0 %
/img/fun/bd/voyeur.jpg	70.0	4.3 %
/download/w3perl-iis.exe	67.6	4.1 %
/download/w3perl-apache.exe	56.6	3.4 %
/img/fun/bd/avaler.jpg	46.5	2.8 %
/img/accueil.gif	29.6	1.8 %
/download/w3perl.exe	26.6	1.6 %
/mpg/eclipse/aparavdo1.mpg	18.2	1.1 %
/img/fun/bd/nipponne.jpg	16.9	1.0 %
/fly/dist/fly-2.0.0.zip	16.8	1.0 %

Pays	Trafic (Mo)	Pourcentage
Inconnu	410.84	25.0 %
Réseau	286.74	17.4 %
France	236.79	14.4 %
Commercial	226.39	13.8 %
Canada	78.95	4.8 %
Belgique	56.94	3.5 %
Suisse	27.64	1.7 %
Espagne	25.71	1.6 %
USA Education	24.41	1.5 %
Italie	24.24	1.5 %
Allemagne	23.90	1.5 %
Hollande	23.28	1.4 %
Royaume-Uni	16.11	1.0 %
Australie	14.07	0.9 %

Trafic par pays (pourcentage)
tous sites confondus

Bourse aux Outils

28-03-2002

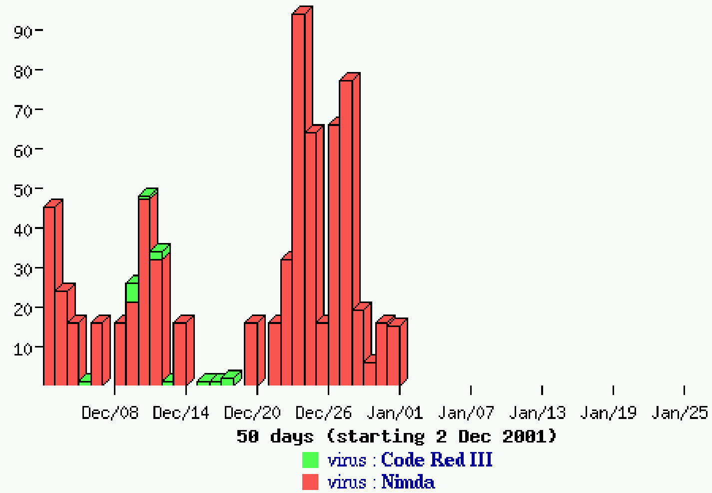
Stats sur les virus

Netscape: Statistiques – Virus

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

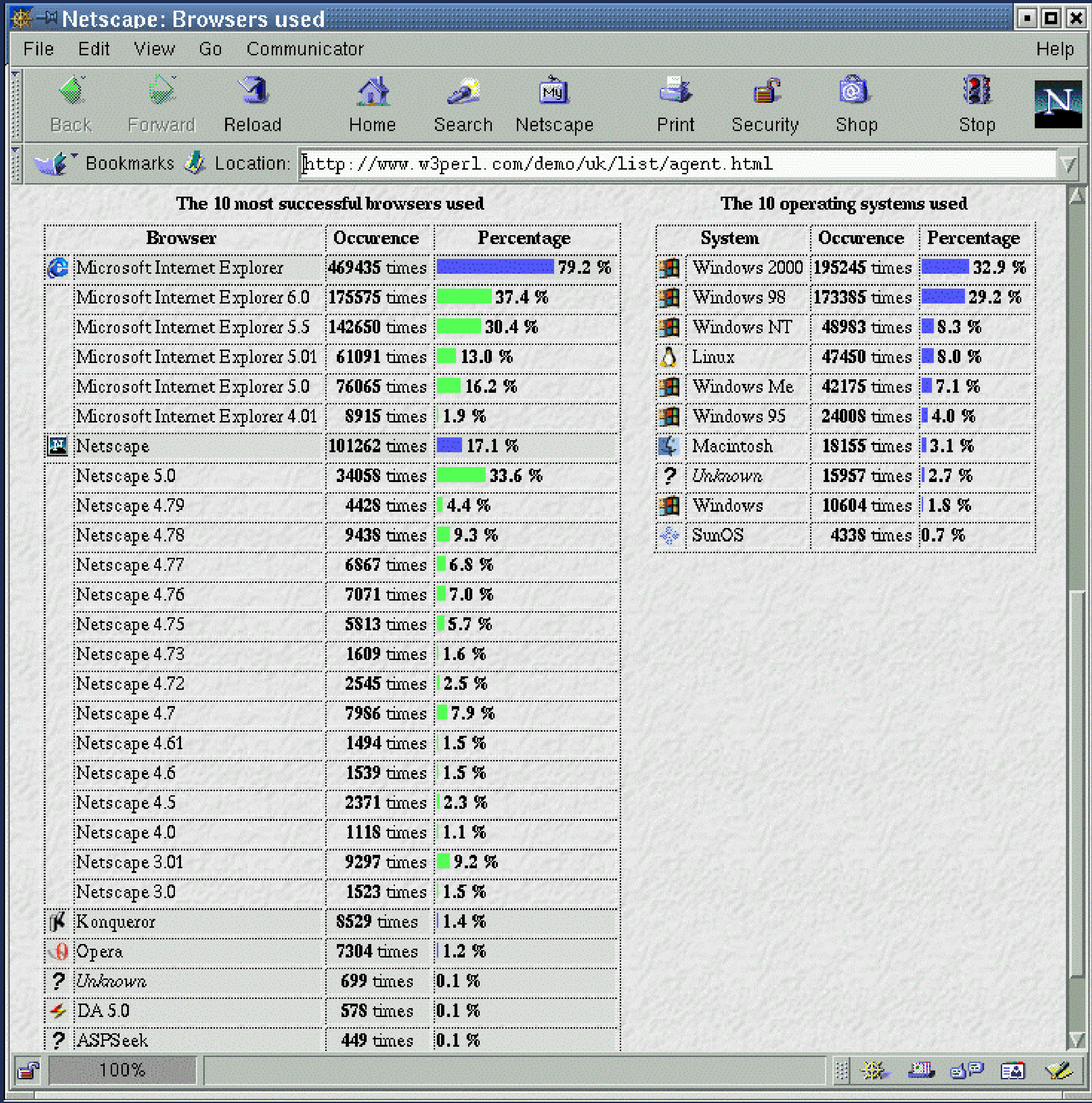
Bookmarks Location: http://titan.inrp.fr/stats/2001/fr/list/virus.html



	Pourcentage	Occurence	Dernière tentative
Nimda	**90.6 %**	**9911**	**31/Dec/2001**
195.151.159.231	0.2 %	15	*31/Dec/2001:12:27:19*
195.92.43.82	0.2 %	16	30/Dec/2001:21:19:25
abn167-38.ank-ports.kablonet.net.tr	0.1 %	6	29/Dec/2001:01:32:41
195.144.53.19	0.2 %	16	28/Dec/2001:03:17:23
195.117.2.131	0.2 %	19	28/Dec/2001:02:57:10
195.251.210.149	0.2 %	16	27/Dec/2001:09:08:18
195.65.131.59	0.2 %	16	27/Dec/2001:15:05:15
195.22.177.89	0.1 %	13	27/Dec/2001:17:18:07
195.239.125.141	0.2 %	16	27/Dec/2001:21:49:16
195.205.200.225	0.2 %	16	26/Dec/2001:19:43:00

100%

Stats sur les navigateurs



The 10 most successful browsers used

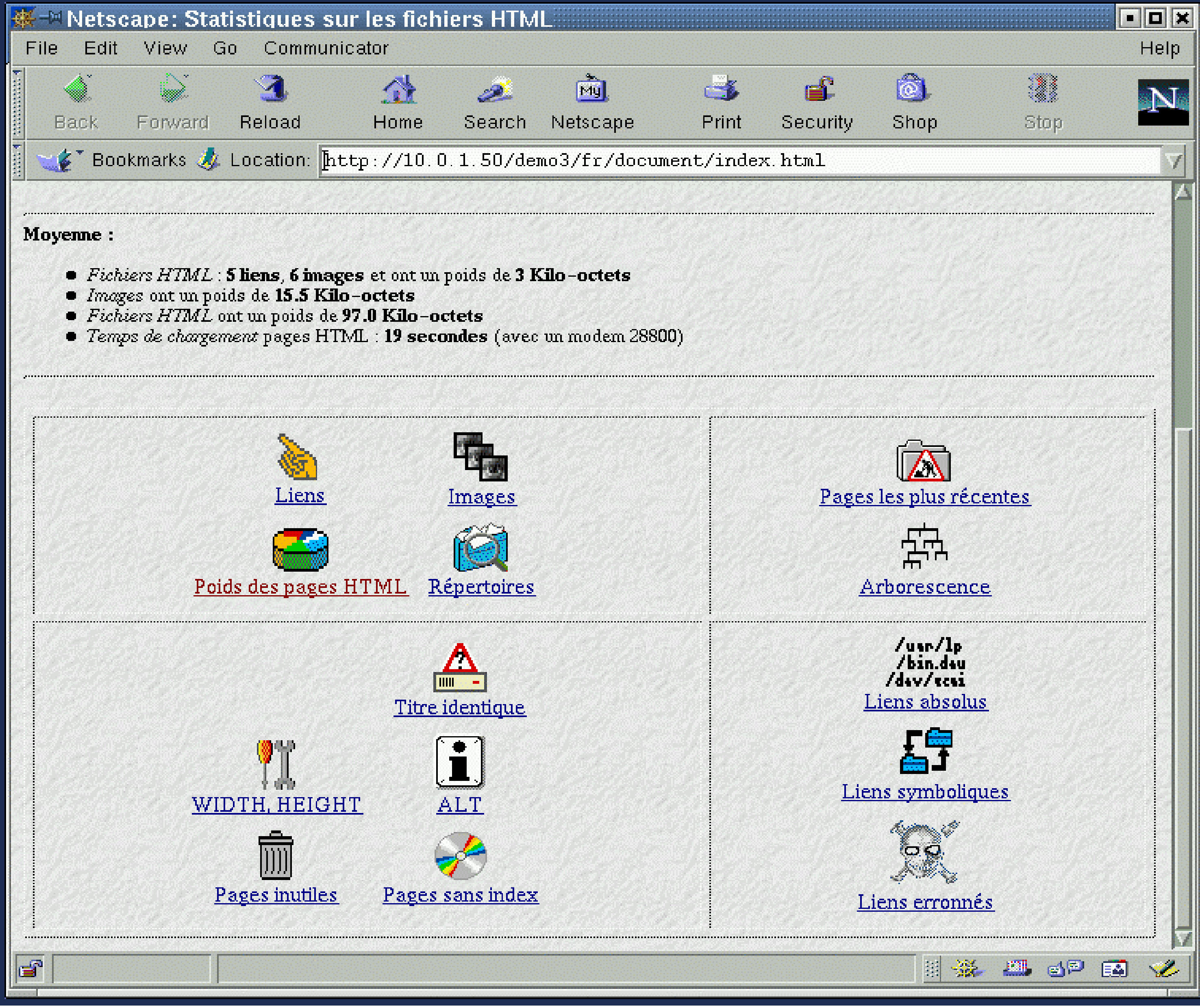
Browser	Occurence	Percentage
Microsoft Internet Explorer	469435 times	79.2 %
Microsoft Internet Explorer 6.0	175575 times	37.4 %
Microsoft Internet Explorer 5.5	142650 times	30.4 %
Microsoft Internet Explorer 5.01	61091 times	13.0 %
Microsoft Internet Explorer 5.0	76065 times	16.2 %
Microsoft Internet Explorer 4.01	8915 times	1.9 %
Netscape	101262 times	17.1 %
Netscape 5.0	34058 times	33.6 %
Netscape 4.79	4428 times	4.4 %
Netscape 4.78	9438 times	9.3 %
Netscape 4.77	6867 times	6.8 %
Netscape 4.76	7071 times	7.0 %
Netscape 4.75	5813 times	5.7 %
Netscape 4.73	1609 times	1.6 %
Netscape 4.72	2545 times	2.5 %
Netscape 4.7	7986 times	7.9 %
Netscape 4.61	1494 times	1.5 %
Netscape 4.6	1539 times	1.5 %
Netscape 4.5	2371 times	2.3 %
Netscape 4.0	1118 times	1.1 %
Netscape 3.01	9297 times	9.2 %
Netscape 3.0	1523 times	1.5 %
Konqueror	8529 times	1.4 %
Opera	7304 times	1.2 %
Unknown	699 times	0.1 %
DA 5.0	578 times	0.1 %
ASPSeek	449 times	0.1 %

The 10 operating systems used

System	Occurence	Percentage
Windows 2000	195245 times	32.9 %
Windows 98	173385 times	29.2 %
Windows NT	48983 times	8.3 %
Linux	47450 times	8.0 %
Windows Me	42175 times	7.1 %
Windows 95	24008 times	4.0 %
Macintosh	18155 times	3.1 %
Unknown	15957 times	2.7 %
Windows	10604 times	1.8 %
SunOS	4338 times	0.7 %

Bourse aux Outils

URL mapping



Bourse aux Outils

28-03-2002

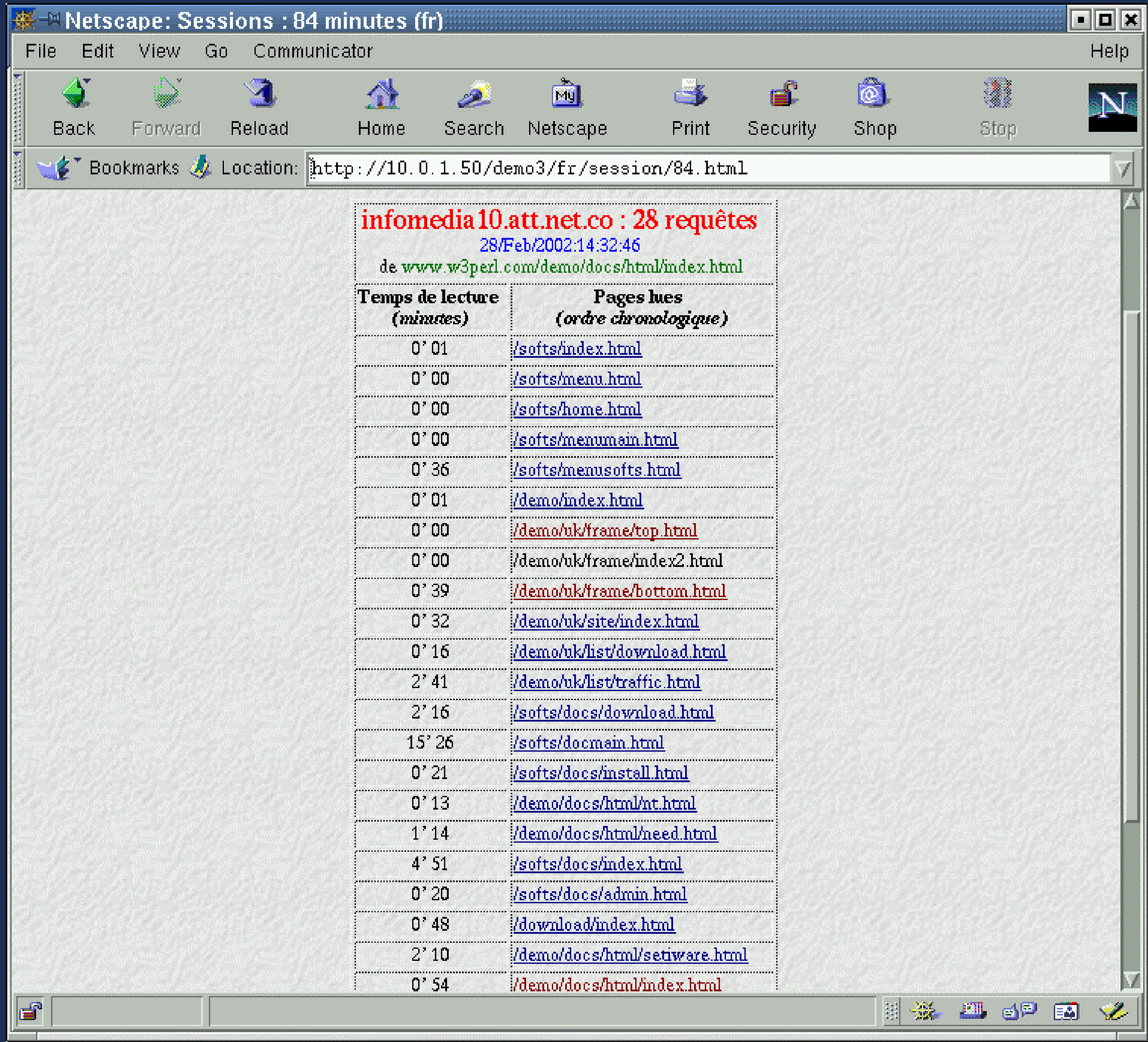
Session : problème

- **Pas de support de début / fin**
- **Visite = Machine?**
  - Même machine (IP)
    - Proxy, Provider, Foyer
  - Même personne
    - Provider, DHCP, multi-postes

Session : solution

- **Période d'inactivité / Timeout**
- **Filtrage**
  - Basé sur les types de fichiers
- **Résultats**
  - Exploration précise de chaque visite
  - Robots, Login, Fidélité
  - Durée de connexion, par pages

Exemple de visite



Netscape: Sessions : 84 minutes (fr)

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: http://10.0.1.50/demo3/fr/session/84.html

infomedia10.att.net.co : 28 requêtes

28/Feb/2002:14:32:46

de www.w3perl.com/demo/docs/html/index.html

Temps de lecture *(minutes)*	Pages lues *(ordre chronologique)*
0' 01	/softs/index.html
0' 00	/softs/menu.html
0' 00	/softs/home.html
0' 00	/softs/menumain.html
0' 36	/softs/menusofts.html
0' 01	/demo/index.html
0' 00	/demo/uk/frame/top.html
0' 00	/demo/uk/frame/index2.html
0' 39	/demo/uk/frame/bottom.html
0' 32	/demo/uk/site/index.html
0' 16	/demo/uk/list/download.html
2' 41	/demo/uk/list/traffic.html
2' 16	/softs/docs/download.html
15' 26	/softs/docmain.html
0' 21	/softs/docs/install.html
0' 13	/demo/docs/html/nt.html
1' 14	/demo/docs/html/need.html
4' 51	/softs/docs/index.html
0' 20	/softs/docs/admin.html
0' 48	/download/index.html
2' 10	/demo/docs/html/setiware.html
0' 54	/demo/docs/html/index.html

Bourse aux Outils

28-03-2002

Futur

- **Moyen terme**
  - RPM
  - Squid natif, FTP, IRC….
  - Détection des pirates
- **Long terme**
  - MySQL
  - PDF

Autres outils gratuits

	(+)	(-)
AWStats 3.2	Temps réel	CGI
Analog 5.21	Vitesse	Affichage
Webalizer 2.0	Robuste	Session
W3Perl 2.86	Complet	Vitesse

Documentation

Vous trouverez sur le site de l'application toute la documentation online (y compris cette présentation avec les commentaires associes)

http://www.w3perl.com/softs/