

Glossaire

Ce glossaire contient les termes techniques et de spécialité les plus employés dans cette thèse. Il emprunte, pour certaines d'entre elles, les définitions proposées par www.themanagerpage.org (© themanagerpage.org) soumises à la licence GNU FDL.

ADRESSE IP. Adresse de 32 bits utilisée par le protocole IP pour identifier de manière unique les hôtes (machines) sur réseau IP.

L'adresse IP est généralement présentée sous forme décimale « pointée » du type '127.0.0.1'. Elle comporte une partie identifiant le réseau (network ID), et une partie identifiant le numéro d'équipement dans le réseau (host ID).

Voir aussi : TCP/IP.

ANNUAIRE WEB. Site Web proposant une sélection de sites classés par thème. Les annuaires Web généralistes (ex : Yahoo, Voila, Nomade) ont vocation à couvrir l'ensemble des contenus disponibles sur le Web, et à proposer des liens pertinents pour chaque thème représenté.

APPLET. Application Java invoquée par une page Web. Les *applets*, contrairement aux vrais programmes Java ne peuvent pas accéder aux ressources système locales, telles que des fichiers ou périphériques. Les *applets* ne peuvent également pas communiquer avec d'autres serveurs que ceux dont elles sont issues. L'intérêt d'une *applet* est d'avoir des pages Web dynamiques, animées et interactives, en utilisant toute la puissance du Java. Les *applets* sont portables et offrent un bon niveau de fiabilité et de sécurité.

ASPIRATEUR WEB. Logiciel permettant de faire une copie locale d'une page ou d'un site Web.

CENTRE-SERVEUR. Les méthodes d'analyse des usages d'Internet opposent les approches centrées-serveur, qui reposent sur l'analyse de données collectées sur les serveurs (serveur Web notamment) à l'approche centrée-utilisateur, où les données sont recueillies au niveau de l'utilisateur.

Voir aussi : *DONNEES DE TRAFIC, LOGS*.

CENTRE-UTILISATEUR. Voir *CENTRE-SERVEUR*.

CLIENT/SERVEUR. Modèle fonctionnel logiciel dans lequel plusieurs programmes autonomes communiquent entre eux par échange de messages.

Le modèle est à l'origine dissymétrique, c'est toujours le client qui fait appel aux services du serveur. En aucun cas celui-ci ne peut fournir des services de sa propre initiative.

Voir aussi : *SERVEUR, REQUETE*.

COHORTE. Ensemble d'individus étudiés sur une période de temps donnée. Une cohorte permet de suivre de manière longitudinale les comportements de la population observée ainsi que sa réaction à un ou plusieurs événements donnés.

DNS. *Domain Name System/Server*

Système d'annuaire distribué sur l'Internet qui contient principalement les noms et les adresses IP des stations. Il sert à faire la conversion nom de machine-adresse IP.

Exemple : l'adresse symbolique www.themanagerpage.org est convertie en l'adresse IP

numérique 205.206.106.50.

DOMAINE. Le domaine identifie un groupe d'ordinateurs hôtes ou de réseaux locaux qui, sous une même entité administrative, sont branchés sur le réseau Internet. Le nom des domaines se compose de sections séparées par des points, qui définissent une arborescence (de droite à gauche) : au sein du domaine *.fr*, on trouve les sous-domaines *.lemonde.fr*, *.asso.fr*, etc. ; au sein du domaine *.asso.fr*, chaque association dispose d'un sous-domaine, etc.

L'achat d'un nom de domaine revient à disposer d'une entité au niveau -1 ou -2 des domaines de premier niveau. L'acheteur gère ensuite l'ensemble des sous-domaines du nom de domaine qui lui appartient (ex : *sport.tf1.fr*, *info.tf1.fr* font partie du domaine *tf1.fr*, géré par TF1).

Lors d'une communication entre deux ordinateurs du réseau Internet, les noms des ordinateurs et des domaines sont traduits en adresses numériques par un serveur de noms de domaine (ou Domain Name Server ou DNS).

DONNEES DE TRAFIC. Données techniques basées sur l'enregistrement de la communication sur un réseau.

Dans le cas du trafic Internet, selon l'endroit du réseau où est fait cet enregistrement, les données peuvent être centrées-utilisateur, centrées-serveur, ou recueillies à un endroit intermédiaire (routeurs, *proxy*, DSLAM, etc.).

Voir aussi : *CENTRE-SERVEUR*, *CENTRE-UTILISATEUR*, *LOGS*.

ECHANTILLON. Sous-ensemble caractéristique d'une population, dont l'analyse permet de décrire l'ensemble de la population.

FAI. Fournisseur d'Accès à Internet (ex : Wanadoo, Free, AOL, etc.).

FORUM. Ou *newsgroup*, *groupe de discussion*, *Usenet*.

Système de discussion écrite asynchrone, sorte de salon électronique d'accès libre où se tiennent les discussions relatives à un ou plusieurs thèmes définis lors de la création du groupe.

Généralement, le nom du forum renseigne sur son contenu. Par exemple, le nom *fr.comp.sys.windows* désigne un forum francophone (*fr*) traitant d'informatique (*comp* pour 'computer'), plus précisément des systèmes d'exploitation (*sys*) et plus spécifiquement du système Windows.

Il existe plusieurs types de forums, selon la façon dont les messages sont transmis : certains se trouvent sur le Web, d'autres sur l'Usenet et les serveur de news.

FRAME. Concept inventé par Netscape, consistant à diviser la fenêtre d'un navigateur Web en plusieurs sous-fenêtres, dans chacune desquelles on affiche une page Web différente (page HTML ou autre). Chaque *frame* possède sa propre URL, la position et le contenu des *frames* étant définis par une page HTML appelée *frameset*.

FTP. *File Transfer Protocol* (Protocole de Transfert de Fichiers)

Protocole orienté vers le transfert de fichiers, en envoi comme en réception, fonctionnant sur les protocoles TCP/IP.

HTML. *Hypertext Markup Language*

Langage utilisé pour écrire les pages Web d'Internet.

C'est une version simplifiée de la norme SGML (Standard Generalized Markup Language), langage de document structuré, avec liens hypertexte, utilisé en gestion documentaire. Il a été inventé dans les années 80 par Tim Berners-Lee qui cherchait à l'époque un moyen simple et efficace pour mettre à disposition sur réseau la document du CERN.

HTTP. *Hyper Text Transfer Protocol*

Protocole inventé par Tim Berners-Lee à la même époque que le HTML et spécialement conçu pour accéder aux documents HTML. Par extension, ce protocole peut être utilisé pour accéder à presque tout type de ressource Web *via* Internet.

INTERNET. Réseau de portée mondiale interconnectant des centaines de réseaux spécifiques et auquel sont reliés quelques centaines de millions d'utilisateurs individuels et professionnels.

Ce réseau est le support de multiples activités et protocoles : consultation de sites Web (HTTP), messagerie (POP/SMTP), *peer-to-peer*, *chat*, jeux, etc.

IRC. *Internet Relay Chat*

Protocole permettant de dialoguer en mode texte en direct avec plusieurs personnes.

LOGS. Enregistrement technique de traces d'activité. Dans le cas des serveurs HTTP, les fichiers de *logs* conservent la liste de tous les accès aux ressources du serveur.

NAVIGATEUR. Logiciel de navigation sur le Web. Formellement, il s'agit d'un client pour le protocole HTTP, qui a comme fonctionnalité centrale l'interprétation et la visualisation en local ou à distance du langage HTML.

C'est en grande partie la réalisation du premier navigateur Web graphique Mosaic par un étudiant américain de NCSA (National Center for Supercomputing Applications, Université d'Illinois) nommé Marc Andreessen et la commercialisation des logiciels qui s'en sont inspiré, qui explique l'explosion de l'Internet à au milieu des années 90. À partir de cet instant, le navigateur devient de plus en plus le client Internet universel. Les navigateurs possèdent en outre la faculté d'exécuter du code javascript et des *applets* Java. D'autres types d'applications peuvent être exécutées grâce à des *plugins* (composants logiciels additionnels).

Parmi les navigateurs les plus connus, on retrouve Internet Explorer, Netscape Navigator ou Mozilla/Firefox.

NTIC. Nouvelles Technologies de l'Information et de la Communication.

PAGE WEB. Unité ergonomique élémentaire d'un site Web, désignée par une URL unique.

Une page contient souvent plusieurs éléments (images, *frames*, etc.), qui occasionnent chacun une requête auprès d'un serveur Web, et qui sont assemblés par le navigateur. Une page peut être dite statique (le serveur Web renvoie le contenu d'un fichier) ou dynamique (le serveur Web renvoie le résultat d'un traitement particulier).

Voir aussi : *FRAME*, *SITE*, *SERVEUR*.

PANEL. Échantillon de personnes représentatif d'une population. Suivi dans le temps, un panel est corrigé régulièrement pour conserver sa représentativité.

POP. *Post Office Protocol*

Protocole permettant l'accès aux messages E.Mail se trouvant sur un serveur de messagerie ; il est utilisé conjointement avec SMTP (envoi de messages).

Voir aussi : *SMTP*.

PROTOCOLE. Description formelle de règles et de conventions à suivre dans un échange d'informations. Les protocoles peuvent définir ces échanges au niveau des couches Internet (IP), transport (TCP, UDP, etc.) ou applicatives (HTTP, FTP, Telnet, etc.).

Voir aussi : *HTTP*, *FTP*, *POP*, *SMTP*, *IRC*.

REQUETE. Dans un modèle client-serveur, envoi d'une instruction d'un client vers un serveur.

Dans le cas des serveurs Web, une requête HTTP demande à un serveur l'envoi du contenu d'un document (ex. : une page HTML) ou du résultat de l'exécution d'un traitement par le serveur (ex. : rechercher dans une base de données).

Voir aussi : *CLIENT/SERVEUR, SERVEUR*.

SERVEUR. Ressource informatique (machine ou programme) capable de délivrer une information où d'effectuer un traitement à la requête d'autres équipements.

Voir aussi : *CLIENT-SERVEUR, REQUETE*.

SESSION. Période d'activité cohérente d'un utilisateur sur Internet.

À partir des données de trafic centrées-utilisateur, on identifie les sessions sur la base d'une période d'inactivité (aucune trace d'activité enregistrée) de plus de trente minutes.

SITE WEB. Ensemble de pages et de services Web sous la responsabilité d'une même entité éditoriale (individu, société, organisme, etc.).

Il est identifié par un nom de domaine, par exemple *tf1.fr*. La plupart du temps, le site correspond à l'adresse du serveur Web sur le domaine spécifié (ex : *www.globz.net*), mais il peut être réparti sur plusieurs (*sport.tf1.fr, info.tf1.fr, etc.* pour le domaine *tf1.fr*). Dans le cas de sites personnels, le site, en tant qu'entité éditoriale, correspondre à un sous-domaine (comme chez Free, ex : *restaurefour.free.fr*) ou a un sous-répertoire sur un domaine dédié (comme chez Wanadoo, ex : *perso.wanadoo.fr/4pat/*).

SMTP. *Simple Mail Transfer Protocol*

Protocole de messagerie permettant d'envoyer des mails à partir d'un logiciel client vers un serveur de messagerie. Il est utilisé conjointement avec POP (réception des messages).

Voir aussi : *POP*.

TCP/IP. *Transmission Control Protocol / Internet Protocol*

Les deux protocoles de communication qui forment les fondements de l'Internet. TCP assure le transfert des données, IP l'adressage.

Voir aussi : *ADRESSE IP*.

TIC. Technologies de l'Information et de la Communication.

URL. *Uniform Resource Locator*

Identifiant de ressource à travers le réseau Internet permettant la détermination du protocole à utiliser pour transférer le document, l'adresse du serveur sur lequel se trouve le document, et le chemin d'accès du document sur le serveur.

La structure d'un URL est la suivante :

[protocole]://[Nom du serveur][:numéro de port]/[chemin d'accès]

W3C. *World Wide Web Consortium*

Consortium créé début 1995 dont le principal objectif est la mise au point de normes et de protocoles ouverts et libres pour le Web, dans un souci d'interopérabilité maximale. Il est géré conjointement par le MIT aux États-Unis, l'INRIA en France et l'université Keio au Japon. Son directeur est Tim Berners-Lee, père du Web.

WEB. Ou *World-Wide-Web* : littéralement « toile d'araignée mondiale ». Système d'information réparti, basé sur des documents en hypertexte (au format HTML). Service

d'informations sur le réseau Internet, créé au CERN (Centre Européen de la Recherche Nucléaire) à Genève en 1993 par Tim Berners-Lee, et mettant à la disposition des utilisateurs un ensemble distribué de documents multimédias composites reliés entre eux par des liens hypertextes. Chaque document est identifié par une adresse appelée URL. C'est ce service qui explique en grande partie le succès récent d'Internet, et son utilisation par le grand public. Les sites et les documents sont accessibles et visualisables grâce à des logiciels clients appelés navigateurs.

Voir aussi : *NAVIGATEUR, HTTP, URL*.

WEBCCHAT. Accès à des services de *chat* à l'aide d'un navigateur Web, au lieu d'un logiciel spécifique.

WEBMAIL. Accès à des services de messagerie à l'aide d'un navigateur Web, au lieu d'un logiciel spécifique. Certains services de WebMail permettent d'accéder à un compte de messagerie « classique » (c'est le cas pour la plupart des FAI), d'autres n'existent que sur le Web.

WEB MINING. Champ de recherche apparu dans les années 90 se centrant sur l'analyse et la fouille de données relatives au Web. Il se divise en trois catégories : analyse de la structure du Web (*Web Structure Mining*), fouille de son contenu (*Web Data Mining*) et étude de ses usages (*Web Usage Mining*).

WEB USAGE MINING. Champ de recherche apparu dans les années 90, dont l'objectif est l'analyse des usages du Web. Il se concentre principalement sur l'étude soit centrée-serveur, soit centrée-utilisateur.

Voir aussi : *WEB MINING*, *CENTRE-UTILISATEUR*, *CENTRE-SERVEUR*.

WWW. Voir *WEB*.

Placé au début d'une URL : convention désignant l'adresse d'un serveur HTTP sur un domaine donné.