Une petite histoire d'Internet et du Web

Table des matières

I - Histoire d'Internet 3

II - Exercice 8

III - Naissance du Web (1989-1993) 9

IV - Exercice 12

V - Wayback Machine de l'Internet Archive 13

VI - Exercice 15

VII - Accéder au Web au XXe siècle : cybercafés, universités, connexions à domicile (1993-1999) 16

VIII - Exercice 19

IX - Des butineurs aux navigateurs web 20

X - Exercice 23

XI - Sites web personnels 24

XII - Exercice 27

XIII - Les moteurs de recherche et annuaires web 28

XIV - Exercice 31

XV - Économie du Web : bourse, publicité, commerce 32

XVI - Exercice 34

XVII - Informations et culture et sur le Web 35

Solutions des exercices 39

Glossaire 43

Abréviations 44

Bibliographie 45

Crédits des ressources 46

Contenus annexes 48

I Histoire d'Internet

Objectifs

- Découvrir l'histoire d'Internet et du Web ;
- Découvrir les enjeux actuels d'Internet.

Mise en situation

Internet est apparu dans la seconde moitié du XX^{e} siècle afin de créer un réseau de grande envergure sans point central, notamment dans la perspective de résister à des pannes. C'est un réseau **décentralisé**.

Ce réseau a connu plusieurs applications, donc les plus utilisées aujourd'hui sont le **mail** et le **Web**.

Ces applications fonctionnent grâce à des **protocoles**, c'est à dire des langages standards que tous les serveurs parlent.

Le protocole du Web est **HTTP**. Son évolution récente HTTPS ajoute une couche cryptographique afin que les échanges sur le Web puissent être confidentiels.

Rappel

Avant l'arrivée d'Internet les premiers réseaux étaient des réseaux locaux, non connectés entre eux, organisés autour d'ordinateurs centraux.

IBM 9020 (London Air Traffic Control Centre, 1974)

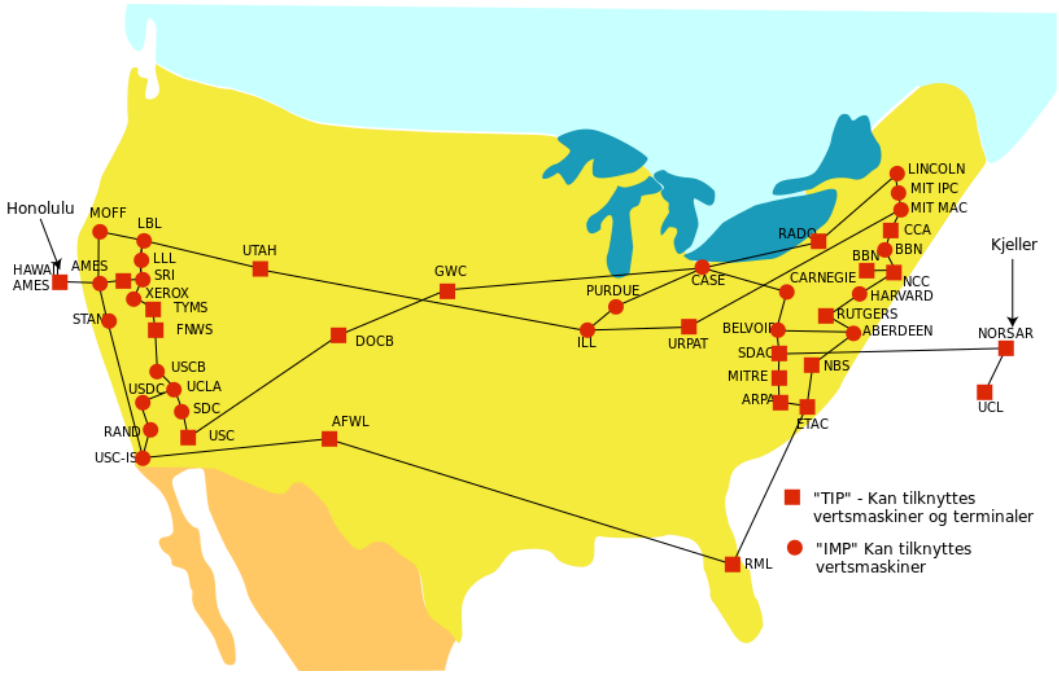
« Rien de plus télégénique qu'un ordinateur *mainframe*. Aujourd'hui, le terme est un peu désuet. On parle plutôt de serveurs ou de super-ordinateurs. Dans les années 1960, on les appelait « ordinateurs centraux », en bon français. De tels ordinateurs mêlaient haute performance et capacité à traiter de l'information en provenance ou à destination de plusieurs dispositifs : des mémoires sur bandes magnétiques situées dans des armoires équipées de lecteurs, cartes perforées (support d'instructions) placées dans des trieuses, terminal central ou autres terminaux distants, imprimantes... (*Masutti, 2020*[Masutti, 2020 p.45]) »

ARPANET : un réseau de communication américain (1969)

Les prémisses de la construction d'un réseau étendu de machines voient le jour à la fin des années 1960 aux États-Unis.

Porté par la recherche et financé par l'armée, l'enjeu est de concevoir un réseau :

- sans pouvoir central,
- capable de transmettre des messages sur de grandes distances.



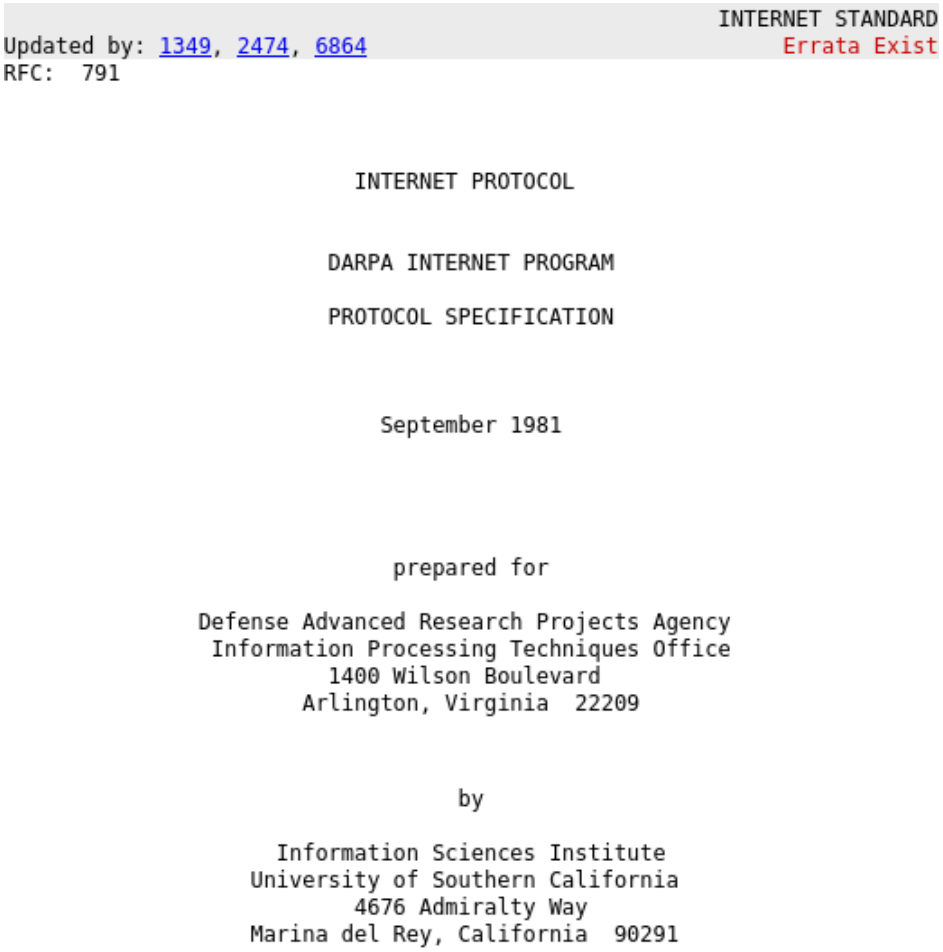
Sites des États-Unis reliés à ARPANET en 1974.

> 💬 Remarque
>
> L'absence de gestion centralisée permet de mieux résister aux pannes ou aux attaques.

> ⊕ Complément
>
> ARPANET fonctionne avec une partie isolée, réservée à l'armée, et une partie civile. Au début des années 80, ARPANET est progressivement abandonné au profit des technologies `TCP/IP`, qui forment les fondations d'Internet.
>
> Cet événement est majeur pour la **standardisation** des communications.

> **TCP/IP : l'unification du réseau (1983)** — 💡 Fondamental
>
> TCP/IP est le couple de *protocoles* p.43 qui est au cœur du fonctionnement d'Internet, c'est lui qui permet à des ordinateurs différents d'échanger des messages au sein de réseaux différents.

« En septembre 1981, le protocole de communication TCP/IP version 4 est décrit dans les Request for comments (RFC) 791, 792 et 793. Il est installé en 1983 sur ARPANET. (wikipedia[1]) »



INTERNET STANDARD
Updated by: 1349, 2474, 6864 Errata Exist
RFC: 791

INTERNET PROTOCOL

DARPA INTERNET PROGRAM

PROTOCOL SPECIFICATION

September 1981

prepared for

Defense Advanced Research Projects Agency
Information Processing Techniques Office
1400 Wilson Boulevard
Arlington, Virginia 22209

by

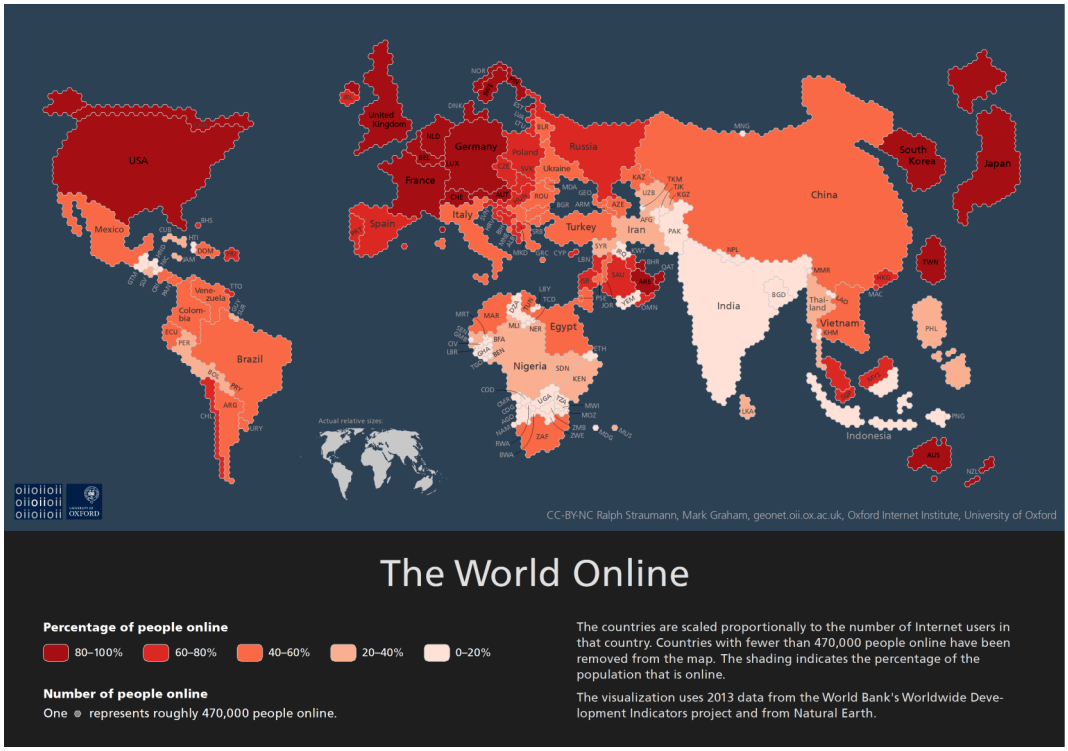
Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, California 90291

Internet : un réseau quasiment mondial

Internet et le Web sont omniprésents dans la plupart des sociétés. De nombreuses entreprises créent des services disponibles exclusivement sur le Web, si bien que les plus grandes capitalisations boursières sont des entreprises ne proposant que des services web. Plus de la moitié de l'humanité utilise le Web pour communiquer.

En 2020, environ 58% de la population mondiale a accès à Internet. La répartition est inégalitaire : l'Europe et l'Amérique du Nord sont plus connectés que l'Afrique et l'Asie. Déployer Internet dans ces zones est l'objet d'enjeux qui sont à la fois technologiques (avoir les infrastructures pour se connecter au réseau) et politique (volonté du gouvernement d'ouvrir le pays à ce réseau).

1. https://fr.wikipedia.org/wiki/Internet

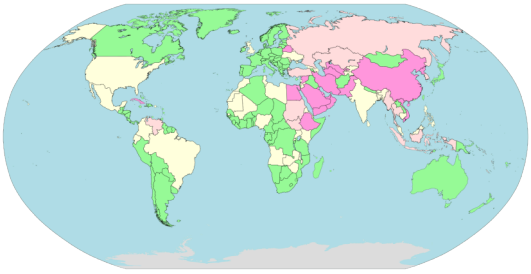


Répartition des internautes dans le monde en 2013

Internet : un enjeu politique et social fort

⊕ Complément

Internet, grâce à l'absence de pouvoir central et l'équipement massif de la population, permet en théorie le **libre accès à l'information** et l'exercice de la **liberté d'expression**. Si la gestion d'Internet ne dépend directement d'aucun état, certains pays peuvent appliquer leur politique locale sur le réseau, en restreignant ou bloquant l'accès à Internet à leurs citoyens, ou en surveillant les communications sur Internet.



Internet censorship and surveillance by country (2018)

Les pays en rose surveillent et contrôlent le plus Internet ; les pays en blanc et en vert exercent un contrôle et une surveillance sélective.

Des enjeux qui dépassent la question numérique

⊕ Complément

- Internet a aussi été un enjeu de pouvoir dans l'économie : la « **bulle Internet** » à la fin des années 1990 et le début 2000 a témoigné d'un engouement et d'une forte spéculation dans les entreprises des sites de commerces en lignes et les compagnies de services de télécommunications.
- La **neutralité du net** désigne l'idée qu'aucune communication n'est privilégiée plutôt qu'une autre sur Internet, et est un des fondements du réseau Internet. En l'absence de neutralité du net, on peut faire une analogie entre des autoroutes à péages d'un côté, et

des routes de campagne de l'autre. La neutralité du net n'est plus garantie aujourd'hui : certains opérateurs imposent des coûts supplémentaires pour l'accès à certains services.

- Si Internet a été conçu pour n'avoir aucun pouvoir central, une grande partie des services les plus utilisés d'Internet sont détenus par les géants américains Google, Amazon, Facebook, Apple et Microsoft (GAFAM) et les géants Chinois Baidu, Alibaba, Tencent et Xiaomo (BATX) : on parle de **centralisation d'Internet**, ce qui soulève de nouvelles problématiques politiques et sociales.
- Enfin, Internet est enclin à des enjeux géo-politiques, en particulier depuis les déclarations du lanceur d'alertes Edward Snowden, qui révèle en 2013 au monde le système de surveillance globalisé mis au point par les services de renseignements des États-Unis et les grandes entreprises du numérique nord-américaines.

À retenir

- Internet est issu du projet militaire américain ARPANET. Il a été conçu et utilisé par les universitaires et l'armée, puis s'est standardisé et étendu au monde pour devenir Internet.
- Internet pose des enjeux politiques, économiques, sociaux et géo-politiques importants.

Exercice

[solution n°1 p. 39]

Le premier site web `info.cern.ch` a vu le jour au CERN. Quelles technologies ont mises au point Tim Berners-Lee et Robert Cailliau pour y parvenir ?

- **A** Le protocole HTTP
- **B** Un navigateur web
- **C** Un serveur web
- **D** Le langage HTML

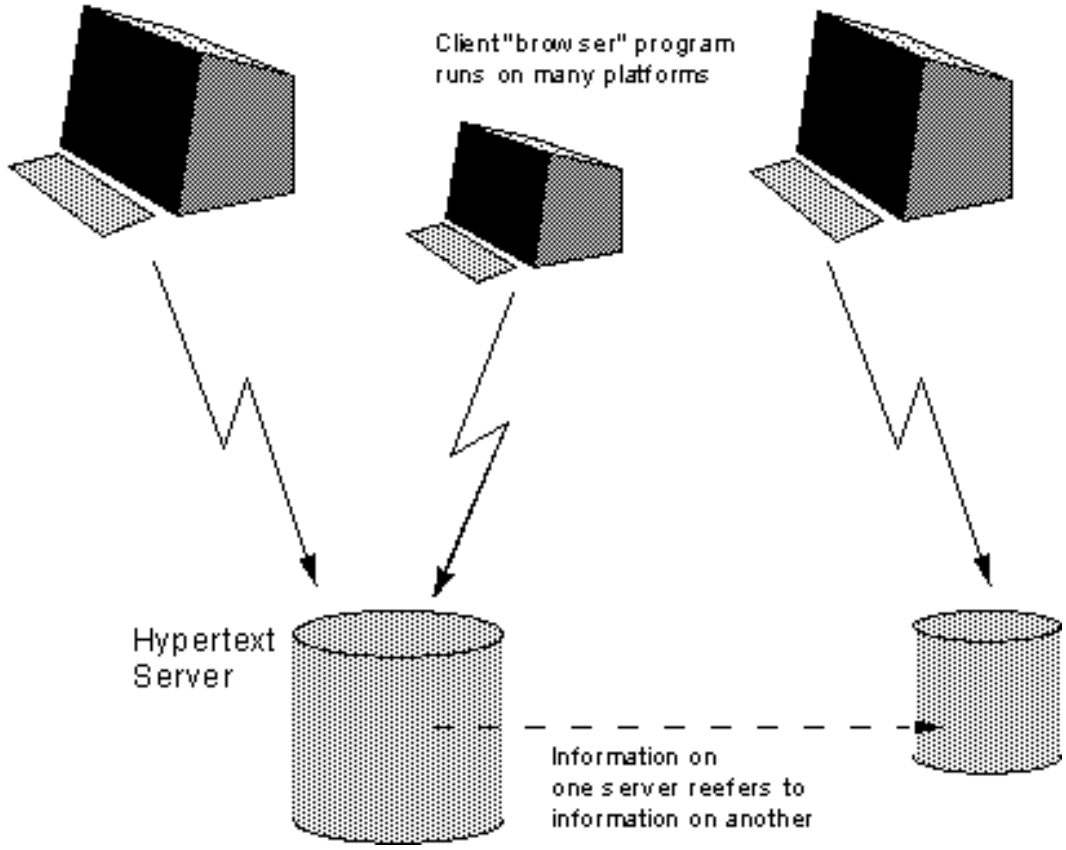
III Naissance du Web (1989-1993)

Le Web est né en 1990 Fondamental

En 1989 et 1990, les travaux dirigés par Tim Berners-Lee et Robert Cailliau font émerger le Web (*Berners-Lee, 1989*[Berners-Lee, 1989 p.45]) :

- Le protocole HTTP
- Le premier client qui implémente HTTP (ce premier navigateur est nommé WorldWideWeb)
- Le premier serveur qui implémente HTTP
- Le langage HTML

Le premier site web `info.cern.ch` est en ligne en décembre 1990.



Remarque

« Comprendre Internet et le Web des années 1990 implique de se défaire de notre vision faite d'hyperconnectivité, de forfaits illimités, de navigation fluide, de moteurs de recherche puissants, qui créent un environnement numérique bien différent de celui que connaissent les premiers internautes.

Schafer, 2022[Schafer, 2022 p.45] »

Exemple

« Je suis sur Nantes Je reste 4 jours en rideau pas moyen de se connecter, je reconfigure les parametres TCP IP en vain, n'etant plus connecté au reseau j'envoie un fax, sans réponse je télephone a la hot-line 20 minutes d'attente avec de la musique pour rester calme, à 2.23F la minute ; Bénéficiaire de l'appel worldnet (36 68 52 60) cout de l'appel environ 60F, et la le technicien m'annonce gentillement que ca vient du point d'acces de Nantes. Worldnet tu nous fatigue vraiment.

Post sur le newsgroup soc.culture.french en 1996, relaté par Valérie Schafer (2022)[Schafer, 2022 p.45] »

Au début, l'indifférence

Complément

- Le projet de Tim Berners-Lee and Robert Cailliau n'est pas supporté par le CERN.
- Le premier article proposé à la conférence ACM Hypertext 91 (Texas, USA) est refusé.

Les premiers sites

- **1991**, les premières **universités** ouvrent leurs sites web : Toronto (Canada), Stanford (Californie) puis en France en 1992 (à Lyon).
- **1992**, il existe **entre 10 et 20 sites** web dans le monde.

Exemple

« On avait un NeXT à Lyon, on a dit qu'on allait l'installer, on a fait le premier site Web en France qui s'appelait info.in2p3.fr [...] il était installé en septembre 1992 au centre de calcul de l'IN2P3 à Lyon.

(Schafer, 2022)[Schafer, 2022 p.45] »

Libération du Web

Fondamental

1993, le CERN dépose les spécifications et codes sources du Web dans le domaine public afin de maximiser sa dissémination.

Complément

À la fin de l'année 1993, Matthew Gray (chercheur au MIT) estime qu'il existe 623 pages web.

Sources

Complément

Hoffmann, 2022[*Hoffmann, 2022* p.45]

Ricmac, 2021[*Ricmac, 2021* p.45]

Histoires d'avant le Web

Complément

Histoire d'Internet *(cf. p.3)*

Histoire du Web *(cf. p.48)*

Exercice

[solution n°2 p. 39]

La Wayback Machine permet de consulter des contenus web copiés par l'organisme Internet Archive (pages web, images, vidéos...).

A L'intégralité des sites web non protégés par login est archivée par Internet Archive.

B ☐ Il est possible de demander à l'Internet Archive d'indexer ou de ne pas indexer son site web.

C ☐ La copie pour archive est toujours autorisée car l'Internet Archive est un organisme à but non lucratif.

V Wayback Machine de l'Internet Archive

Internet Archive Az Définition

Internet Archive (archive.org[2]) a été créé en 1996 pour conserver des copies de pages et de médias présents sur le Web (images, vidéos...). Les copies sont mises à disposition du public 6 à 12 mois après leur capture. Internet Archive est un organisme à but non lucratif.

Quelles pages web sont collectées ? Méthode

L'archivage n'est pas exhaustif, ce serait matériellement impossible.

- Une partie de la collecte est automatisée via des *crawlers*, c'est-à-dire des programmes autonomes qui parcourent les pages web les plus populaires et naviguent de liens en liens.
- Les humains peuvent également signaler manuellement des contenus à conserver.
- Seules les pages publiques accessibles sans contrôle d'accès (*login*) peuvent être copiées.
- Toutes les versions ne sont pas archivées (des intervalles de plusieurs jours ou semaines séparent deux copies).

Exemple

«

- 735 billion web pages
- 41 million books and texts
- 14.7 million audio recordings (including 240,000 live concerts)
- 8.4 million videos (including 2.4 million Television News programs)
- 4.4 million images
- 890,000 software programs

archive.org/about[3] (2023)

»

Complément

- En France, la BNF et l'Ina sont également en charge de l'archivage du Web français (au sens des contenus produits par les Français).
- Software Heritage s'occupe d'archiver les logiciels (par exemple ceux présents sur les dépôts Git).

2. https://archive.org/
3. https://archive.org/about/

Wayback Machine

Az Définition

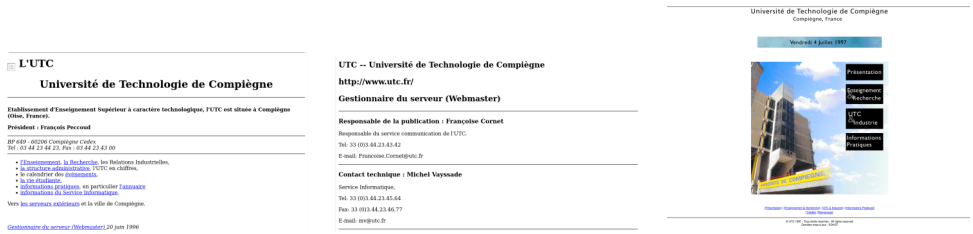
La Wayback Machine permet aux utilisateurs de consulter les contenus archivés par Internet Archive.

INTERNET ARCHIVE
WayBackMachine

Exemple

Page www.lemonde.fr du 22 juin 1997

Exemple



Les premiers sites web à l'UTC

Exercice

[solution n°3 p. 40]

Quelle est le temps approximatif de chargement d'une page web de 80 ko (le poids moyen d'une page à la fin des années 1990) avec un modem 56k ?

A de l'ordre d'une seconde

B de l'ordre de dix secondes

C de l'ordre d'une minute

D de l'ordre d'une heure

VII Accéder au Web au XXe siècle : cybercafés, universités, connexions à domicile (1993-1999)

Az Définition

Un cybercafé est un lieu qui propose des ordinateurs connectés à Internet que l'on peut utiliser contre paiement.

Aux débuts d'Internet, avant la diffusion massive des ordinateurs personnels, et a fortiori des ordiphones, c'était un moyen privilégié pour accéder au Web ou envoyer des mails. Aujourd'hui encore dans certains pays, cela reste un outil pour les populations les plus modestes.

Les premiers cybercafés (1994)

Exemple

En 1994, The Binary Cafe ouvre à Toronto, puis Cyberia ouvre à Londres.

Les cybercafés en France en 1995

Exemple

> « À l'initiative prise par le Centre Pompidou d'ouvrir un cybercafé équipé de quinze terminaux et de matériel permettant les visioconférences répondent en effet des espaces plus modestes avec parfois un ou deux ordinateurs connectés dans un espace de convivialité plus ancien, le café.
>
> *Schafer, 2022*[Schafer, 2022 p.45] »

> « [Ces ordinateurs] permettent notamment de consulter des sites aux États-Unis (trouver des paroles de chansons des Rolling Stones, écouter les miaulements du chat de Bill Clinton), de visiter un musée, de prendre connaissance du CAC 40, de « se confesser ou se pencher sur la plastique de Pamela Anderson », ou encore pour Isabelle de communiquer avec des amis canadiens via une adresse mail, pour 60 francs de l'heure.
>
> *Schafer, 2022*[Schafer, 2022 p.45] »

Modem

Az Définition

Un modem est un appareil qui permet de connecter un ordinateur en réseau via le téléphone. Il transforme pour cela les données numériques en signal analogique (et inversement) : modem signifie modulateur-démodulateur.

Fondamental

Les premiers accès depuis son domicile au Web se font grâce à des connexions modem, qui atteignent 56 kbit/s à la fin des années 1990.

- La page "modem" sur Wikipédia fait 277 ko (donc ~2200 kbits) sans les images. En 2022, il aurait fallu environ 1 minute pour la charger.
- Une connexion ADSL "bas débit" d'1 Mbit/s est environ 20 fois plus rapide.

Connexion intermittente

Attention

Le modem monopolise la ligne téléphonique : si quelqu'un est connecté au Web, plus personne ne peut émettre ou recevoir d'appel sur la même ligne.

On ne peut donc pas rester connecté en permanence sans se couper de la communication téléphonique par ailleurs (à moins de disposer de deux lignes, ce qui est coûteux).

FAI (Fournisseur d'accès à Internet)

Az Définition

Un FAI est une organisation (entreprise, association...) qui s'occupe de raccorder un utilisateur final au réseau Internet.

- Il est en charge de la boucle locale, c'est-à-dire le lien entre l'utilisateur et ses propres équipements (via le réseau téléphonique, la fibre, des ondes électromagnétiques...).
- Il est en charge de raccorder ses équipements au réseau Internet (par peering ou transit).
- Il offre parfois des services associés (hébergement web, mails...).

Complément

Les interactions entre acteurs du réseau : peering et transit (cf. p.50)

FAI des entreprises (1991)

Exemple

- **1991**, **Oléane** (détenu à 20% par le FAI états-unien WorldCom) propose de l'accès Internet aux entreprises.
- **1994**, **Transpac** (filiale de France Télécom) propose de l'accès.

Les deux sociétés ont depuis été rachetées et intégrées à Orange.

FAI des universités

1993, le **Renater** (Réseau National de Télécommunications pour la technologie, l'Enseignement et la Recherche) relie les universités et centres de recherche français.

Renater apporte également l'accès à Internet et au Web ; les enseignants, chercheurs et étudiants disposent par exemple d'une adresse mail ou de la possibilité de mettre en ligne leurs sites web.

Le réseau Renater offre dès ses débuts une connexion très qualitative (débits, disponibilité de service...).

FAI des particuliers (1993)

- **1993**, **Compuserve** propose **un premier forfait de 5h par mois** via le FAI Transpac ; la connexion téléphonique reste payante et dépendante de la distance d'appel (les centraux ne sont disponibles que dans certaines villes).
- **1994**, **FranceNet** propose **des services en plus de l'accès (hébergement de site web, courrier...)** ; FranceNet a été racheté depuis par British Télécom.
- **1994**, **WorldNet** propose des kits de connexion de 3h offerts dans les magazines spécialisés (InfoPC, Le Monde Informatique...) puis un **forfait bas débit illimité hors coût des communications** à 99 F par mois.
- **1995**, **Club Internet** (filiale du groupe Lagardère) propose également un **forfait à tarif réduit de 77 F** hors coût des communications.
- **1996**, **Wanadoo** (filiale de Pages Jaunes Groupe) propose un **accès 2,4 kbit/s** ; Wanadoo est devenu Orange.
- **1998**, **AOL** (filiale du groupe états-unien Time Warner et de SFR) propose pour la première fois un **accès bas débit illimité tout compris** pour 99 F ; AOL compte 520 000 abonnés en 2000 et propose de nombreux services associés via son portail.
- **1999**, **Free** propose un **accès bas débit gratuit hors coût des communications téléphoniques** ; le service inclut un service de webmail et d'hébergement web gratuit.

FFDN

1992, l'association loi 1901 **French Data Network** propose des accès et des services sur un modèle non marchand.

2011, La **FFDN** (fédération des fournisseurs d'accès à internet associatifs) est créée pour rassembler d'autres associations similaires.

À l'UTC :

- **2011**, l'association **Rhizome** est un FAI membre de la FFDN ;
- **2016**, l'association **Picasoft** héberge une partie de ses serveurs chez le FAI et hébergeur Tetaneutral (à Toulouse).

⊕ Complément

- Historique FAI. https://www.abonnement-adsl.biz/historique-fai/
- Histoire d'AOL et de quelques fournisseurs d'accès Francais. http://letenneur.com/historefai.htm

Exercice

[solution n°4 p. 40]

Classer dans leur ordre de création les navigateurs suivants :

A Chrome | **B** Internet Explorer | **C** NSCA Mosaic | **D** WorldWideWeb | **E** Netscape Navigator | **F** Safari | **G** Firefox | **H** Opera

Réponse : ___ ___ ___ ___ ___ ___ ___ ___

IX Des butineurs aux navigateurs web

Navigateur web (web browser) Az Définition

Un navigateur web est un logiciel qui exerce notamment les fonctions de :

- client HTTP : il sait envoyer des requêtes HTTP (comme "envoie-moi la page située à l'adresse *https://framasoft.org/index.html*") ;
- interpréteur HTML : il sait afficher de façon lisible par les humains un code HTML ;
- interpréteur CSS et JavaScript : il sait exécuter les codes CSS et JavaScript contenus dans les pages HTML.

Butineur web 💬 Remarque

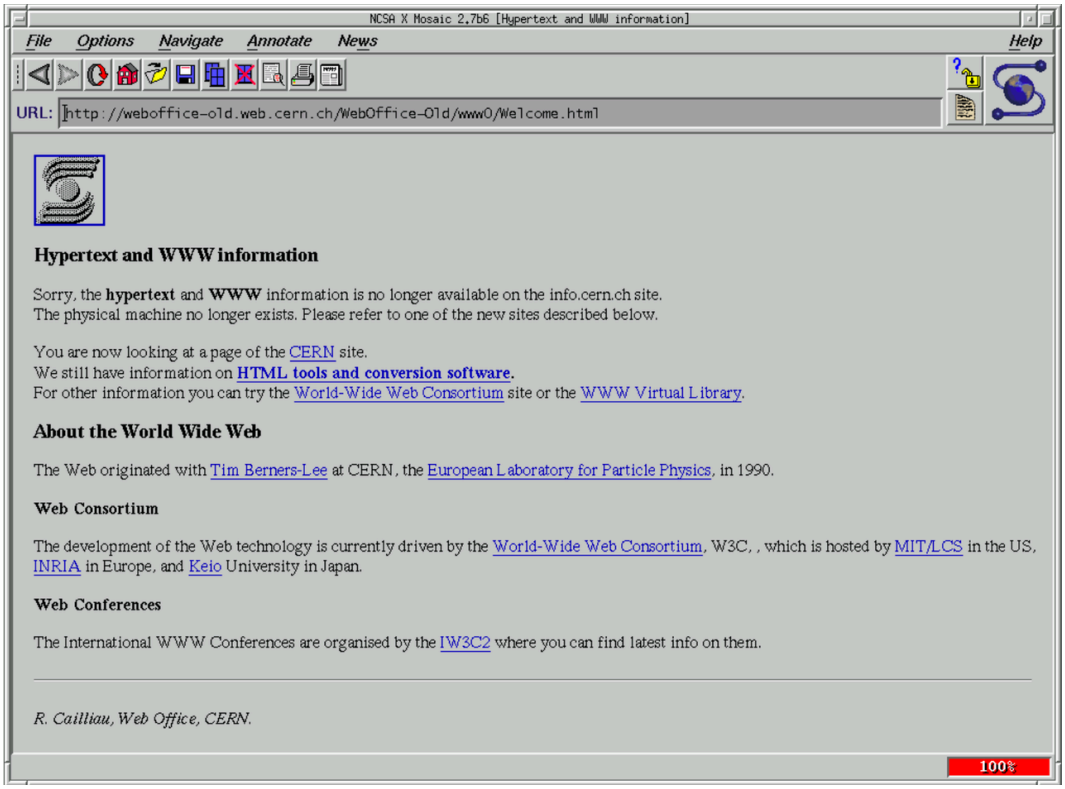
Le terme de butineur est utilisé en France jusque dans les années 2000.

« Je cherche un butineur web qui soit beaucoup plus léger que tous les Mozilla ...
https://linuxfr.org/users/drac/journaux/un-butineur-web-leger (2003) »

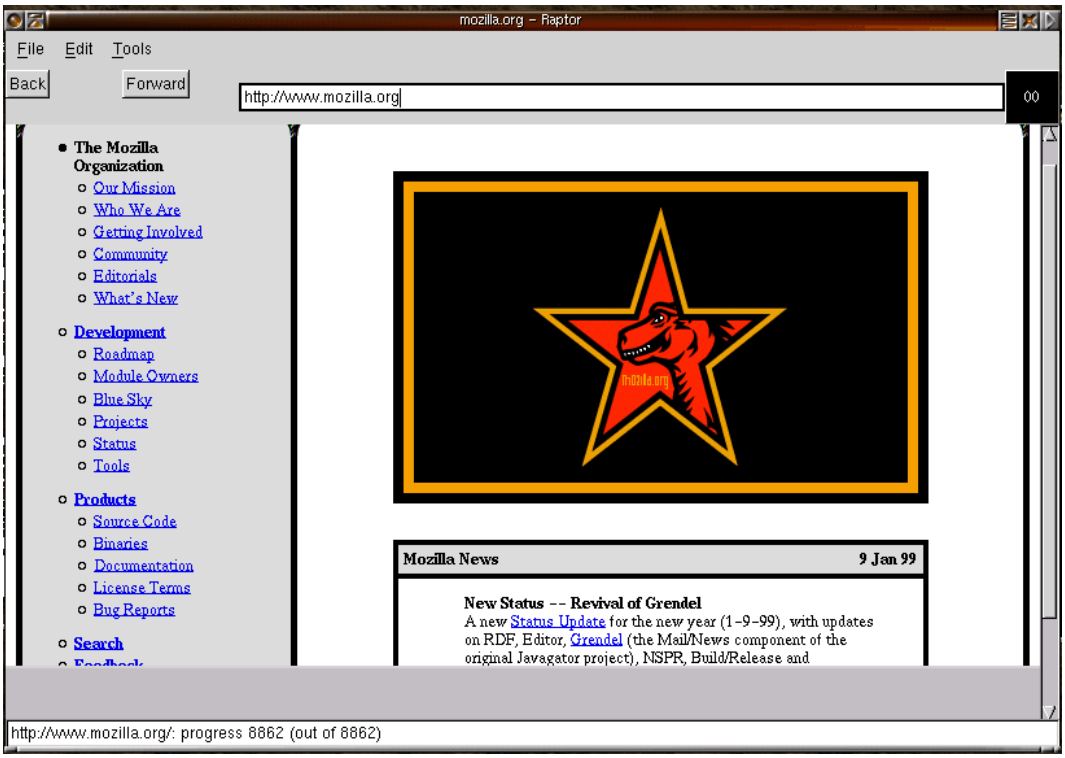
Les ancêtres

- **1990**, **WorldWideWeb** est le premier navigateur développé au Cern ; il est disponible sur les systèmes **NeXT** (qui deviendra Apple).
- **1992**, **ViolaWWW** est développé à l'université de Berkeley (Californie) et *Erwise* est développé à l'université de technologie d'Helsinki (Finlande) ; les deux sont disponibles sur les systèmes **Unix** dotés de l'environnement graphique X Window.
- **1993**, **NSCA Mosaic** est développé par l'université de l'Illinois, toujours pour Unix et X Window, les versions Windows et Apple arriveront plus tard ; il est le premier navigateur à avoir affiché les images directement à l'intérieur des pages ; plusieurs développeurs de Mosaic rejoindront progressivement Netscape.
- **1994**, **Netscape Navigator** (qui deviendra ensuite Netscape Communicator) est développé par la société Netscape Communications (AOL Time Warner) ; il est disponible sur Windows, Unix et Mac ; c'est un navigateur gratuit mais non libre ; fin 1997 Netscape possède 62% de parts de marché ; le code est libéré en 1998 (mais il ne sera pas utilisé).
- **1995**, **Internet Explorer** est développé par Microsoft pour Windows 95 ; il possède 95 % de parts de marché en 2004 grâce à sa livraison pré-installée avec le système d'exploitation Windows ; il est remplacé progressivement par Microsoft Edge (base Chromium) à partir de 2015 et abandonné officiellement en 2022.

Exemple

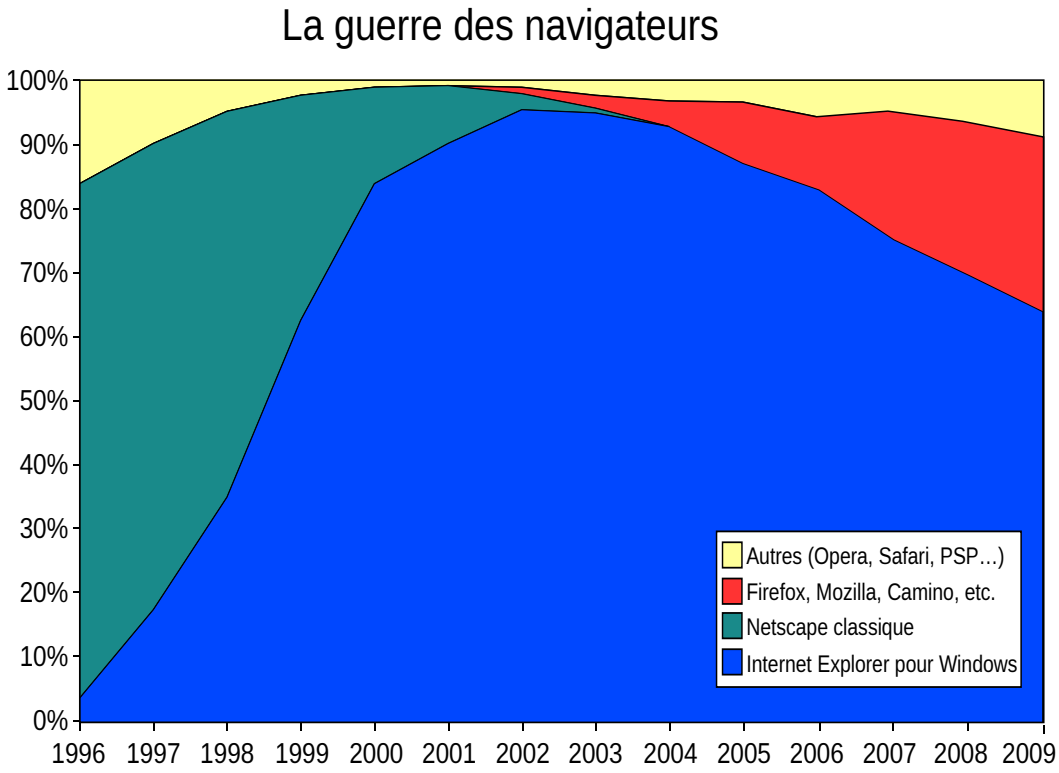


NCSA Mosaic (page de Robert Cailliau au Cern)



Mozilla (1999)

Complément



Évolution de la part respective des navigateurs entre 1996 et 2009

Les modernes

- **1995**, **Opera** est un navigateur multi-plateforme gratuit mais non libre ; Opera a toujours été minoritaire (~2% de parts de marché en 2020) et depuis 2013 Opera est basé sur Chromium.
- **2002**, **Firefox** est créé par la fondation Mozilla (elle-même créée par l'entreprise Netscape) ; c'est un logiciel libre et multi-plateforme ; en 2010 Firefox est le navigateur le plus utilisé en Europe.
- **2003**, **Safari** est créé par Apple pour les systèmes Mac OS X et Windows XP (le support pour Windows a été abandonné depuis) ; Safari est le second navigateur le plus utilisé en 2022.
- **2008**, **Chrome** est créé par Google ; c'est un logiciel propriétaire basé sur Chromium, un navigateur libre développé par l'organisation Chromium Project également créée par Google en 2008 ; Chrome est le navigateur le plus utilisé en 2022.

Exercice

[solution n°5 p. 40]

De quoi a-t-on nécessairement besoin pour mettre en ligne un site web personnel en France ?

A Qu'une entreprise spécialisée nous mette à disposition un espace sur un de ses serveurs

B D'un accès à un ordinateur doté d'une adresse IP fixe

C D'une autorisation de publication délivrée par l'état

D D'un logiciel de serveur web

XI Sites web personnels

> Alors qu'on assiste à une certaine nostalgie du « Web 1.0 », celle-ci a des soubassements politiques : pouvoir de l'utilisateur-producteur des années 1990, éclectisme et foisonnement d'expressions individuelles et collectives subversives, avant-gardisme et indépendance, pouvoir du code et de l'informatique, face à un Web aujourd'hui perçu comme davantage standardisé et contrôlé. Derrière le kitsch des gifs et animations flash affleure la valorisation de formes vernaculaires et de l'empowerment permis.
>
> *Schafer, 2022*[Schafer, 2022 p.45]

Fondamental

Dès lors qu'on dispose d'une machine connectée à Internet, on peut installer un logiciel de serveur web et mettre à disposition son site personnel sur le Web.

Attention

En fait, il faut également :

- disposer d'une adresse IP fixe (ce qui n'est pas systématique pour les contrats proposés par les FAI aux individus) ;
- de préférence disposer d'un nom de domaine (l'accès à un site par IP est possible, mais peu usuel et peu pratique à mémoriser).

Exemple

- On peut utiliser une machine chez soi (si l'on a une IP fixe), par exemple un Raspberry Pi ;
- On peut louer une machine chez un hébergeur externe, comme Gandi.net par exemple.

Complément

Adresses IP et noms de domaine *(cf. p.52)*

Serveur web Définition

Un serveur web est un logiciel qui exerce notamment la fonction de serveur HTTP :

- il sait envoyer un fichier correspondant à une requête comme "envoie-moi la page située à l'adresse *https://framasoft.org/index.html*") ;
- ou une erreur sinon (comme *404 not found*).

Serveurs web

1990, **CERN httpd** est le premier serveur web développé par Tim Berners Lee et ses collègues.

1993, **NCSA HTTPd** est le premier serveur développé hors du Cern ; il est gratuit et sous licence libre ; il servira de base au serveur Apache.

1995, **Apache** est développé par la Apache Software Foundation ; c'est un logiciel multi-plateforme (pour Linux et Windows notamment) sous licence libre ; Apache est resté l'un des deux serveurs les plus utilisés, initialement avec IIS, aujourd'hui avec Nginx.

1995, **IIS** est développé par Microsoft (licence propriétaire) et distribué avec Windows NT et XP), il détient environ 10% de parts de marché en 2022.

2004, **Nginx** est serveur multi-plateforme sous licence libre ; il détient environ un tiers des parts de marché et se partage la première place avec Apache.

« En 1997, Reuter estime le nombre de sites français à 21 367 contre 825 385 aux États-Unis. »

Schafer, 2022[Schafer, 2022 p.45]

Fondamental

Si l'on ne dispose pas de son propre serveur, on peut utiliser des fournisseurs d'hébergement de sites web sur des serveurs externes.

Exemple

1994, **Geocities** est un service d'hébergement web gratuit. Les sites sont organisé en "cités" thématiques (par exemple Paris référence les pages liées à la littérature, la culture, Vienne la musique classique, Nashville la musique country et SiliconValley l'informatique) ; Geocities a été racheté en 1999 par Yahoo! puis fermé en 2009.

Exemple

- Les abonnés du FAI Free disposent d'un espace personnel pour publier leurs pages web ; par exemple : http://martinetl.free.fr/suzannebriet/questcequeladocumentation/
- Les membre d'une université ont souvent accès à un espace personnel de publication, par exemple : https://www.utc.fr/~crozatst/

WordPress

Complément

À partir du début des années 2000, la publication personnelle migre en grande partie sur des systèmes qui prennent en charge à la fois l'hébergement et l'édition des pages HTML, les *WCMS* [p.44].

Le plus populaire d'entre eux, WordPress, est créé en 2001.

On estime qu'aujourd'hui encore ce système alimente plus d'un tiers de tous les sites web du monde.

Complément

CMS *(cf. p.57)*

Exercice

[solution n°6 p. 41]

Un moteur de recherche intègre nécessairement les fonctions suivantes :

- [ ] **A** Un programme qui parcours le Web pour référencer les pages existantes (*crawler*)
- [ ] **B** Un programme qui décrit les pages référencées dans une base de données (indexation)
- [ ] **C** Une interface qui permet à l'utilisateur d'exprimer une recherche (formulaire de recherche)
- [ ] **D** Une présentation ordonnée des résultats de recherche (pages des résultats)
- [ ] **E** Une personnalisation des résultats en fonction d'un modèle de la personne qui fait la requête

XIII Les moteurs de recherche et annuaires web

Avant le Web

1988, **WAIS** (Wide Area Information Servers) permet de se connecter (via Internet à partie de 1991) à un serveur annuaire pour :

- accéder à une base de données de serveurs de documents ;
- faire des recherches et consulter ces documents.

À la naissance du Web

1991, **WWW Virtual Library** est une page sur le site du Cern contenant une liste de liens, que Tim Berners-Lee tient à jour. On pouvait lui envoyer un email pour demander à être ajouté.

Yahoo! : premier annuaire élargi avec moteur de recherche

Exemple

1994, **Yahoo!** propose un annuaire initialement nommé "Jerry and David's Guide to the World Wide Web". À l'origine, l'annuaire est constitué par du travail humain (consulter, trier et filtrer), tandis que des programmes informatiques exécutés sur le serveur permettent de gérer les recherches.

Yahoo! en 1996 (depuis Internet Archive)

AltaVista : intégration de l'indexation automatique

Exemple

1995, **Lycos** et **AltaVista** intègrent l'indexation automatique des pages : des programmes (robots) explorent le Web en sautant de lien en lien et gèrent automatiquement les filtres et classements. AltaVista est racheté par Yahoo! en 2003.

Google : PageRank

Exemple

1998, **Google** se différencie grâce à PageRank, un algorithme de classement qui ordonne la présentation des résultats de recherche. La fonction est devenue déterminante avec la croissance du nombre de sites Web.

La recherche et le biais publicitaire, par les fondateurs de Google en 1998

Complément

« Currently, the predominant business model for commercial search engines is advertising. The goals of the advertising business model do not always correspond to providing quality search to users. For example, in our prototype search engine one of the top results for cellular phone is "The Effect of Cellular Phone Use Upon Driver Attention", a study which explains in great detail the distractions and risk associated with conversing on a cell phone while driving. This search result came up first because of its high importance as judged by the PageRank algorithm, an approximation of citation importance on the web [Page, 98]. It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers. For this type of reason and historical experience with other media [Bagdikian 83], we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.

Since it is very difficult even for experts to evaluate search engines, search engine bias is particularly insidious. A good example was OpenText, which was reported to be selling companies the right to be listed at the top of the search results for particular queries [Marchiori 97]. This type of bias is much more insidious than advertising, because it is not clear who "deserves" to be there, and who is willing to pay money to be listed. This business model resulted in an uproar, and OpenText has ceased to be a viable search engine. But less blatant bias are likely to be tolerated by the market. For example, a search engine could add a small factor to search results from "friendly" companies, and subtract a factor from results from competitors.

Brin and Page, 1998[*Brin and Page, 1998* p.45] »

Exercice

[solution n°7 p. 41]

Parmi les organisations suivantes, lesquelles ont été introduites en bourse une ou deux années après leur création seulement, alors qu'elles étaient déficitaires :

A Amazon

B Wikimedia Foundation (qui gère Wikipédia)

C Netscape Communications

D Iliad (maison mère de Free)

XV Économie du Web : bourse, publicité, commerce

« *Quarante-deux.org* est un des premiers sites Web personnels français et Ellen et Dominique se souviennent que l'étroitesse de l'offre rend populaire leur réalisation : « On était très recherché, on voulait nous mettre de la pub, nous payer ».

Schafer, 2022[Schafer, 2022 p.45] »

Introduction en bourse (logique des start-up)

Exemple

Les créateurs de Mosaic sont des étudiants et universitaires. Une partie d'entre eux, dont Marc Andreessen partent pour la Silicon Valley et créent Netscape Communication Corporation en 1994.

Netscape est introduite en bourse en 1995 :

- le cours de l'action double le premier jour
- la valeur de l'entreprise est de 2 milliards de dollars
- soit plus de 20 fois son chiffre d'affaires de 85 millions
- malgré une perte de 6,6 millions

Publicité

Exemple

Hotwired est un magazine en ligne adossé à Wired traitant des nouvelles technologies.

En 1994, il intègre une bannière publicitaire pour AT&T.

Wired.com invents the web banner ad (Oct. 27, 1994)

Vie privée

Complément

En 1995, la *Data Protection Directive* de l'Union Européenne restreint la collecte de données personnelles non nécessaire et impose aux sites d'expliciter quelles données sont collectées. C'est l'ancêtre du RGPD.

Commerce en ligne

Exemple

- **1994**, **Amazon** lance la vente de livres à distance ; elle est introduite en bourse en 1997 ; la filiale française est ouverte en 2000.
- **1995**, AuctionWeb, qui deviendra **eBay** en 1997, est un site de commerce en ligne qui permet aux particuliers de vendre des biens de seconde main, notamment aux enchères.

- **1995**, **Hoteldiscounts.com**, qui deviendra *Hotels.com*, propose la réservation d'hôtels en ligne.

Exercice

[solution n°8 p. 42]

Exercice

Combien d'articles existaient sur Wikipédia en anglais à la fin de l'année de sa création en 2001 ?

- **A** Environ 2.000
- **B** Environ 20.000
- **C** Environ 200.000
- **D** Environ 2.000.000

Exercice

Combien d'articles existaient sur Wikipédia en anglais à la fin de l'année 2020, 20 ans après sa création ?

- **A** Environ 60.000
- **B** Environ 600.000
- **C** Environ 6.000.000
- **D** Environ 60.000.000

Exercice

En quelle année a été créée la version de Wikipédia en Français ?

XVII Informations et culture et sur le Web

Informations sur le Minitel (1981)

> « La page d'accueil du Parisien Libéré se présentait sous la forme d'un tableau austère qui faisait correspondre des rubriques à des chiffres. En tapant 1, la météo. En tapant 2, la Bourse. En tapant 3, les dépêches.
>
> *Bellanger, 2012*[Bellanger, 2012 p.45] »

Musique

1993, **IUMA** (Internet Underground Music Archive) permet à des artistes sans maison d'édition de faire connaître leur musique. IUMA commence avant l'émergence du Web avec les technologies Gopher et FTP.

Site de l'IUMA en 1996 (depuis Internet Archive)

Presse

1996, **Slate** est un magazine créé par Microsoft qui n'existe que sur le Web (*pure player*).

Encyclopédie

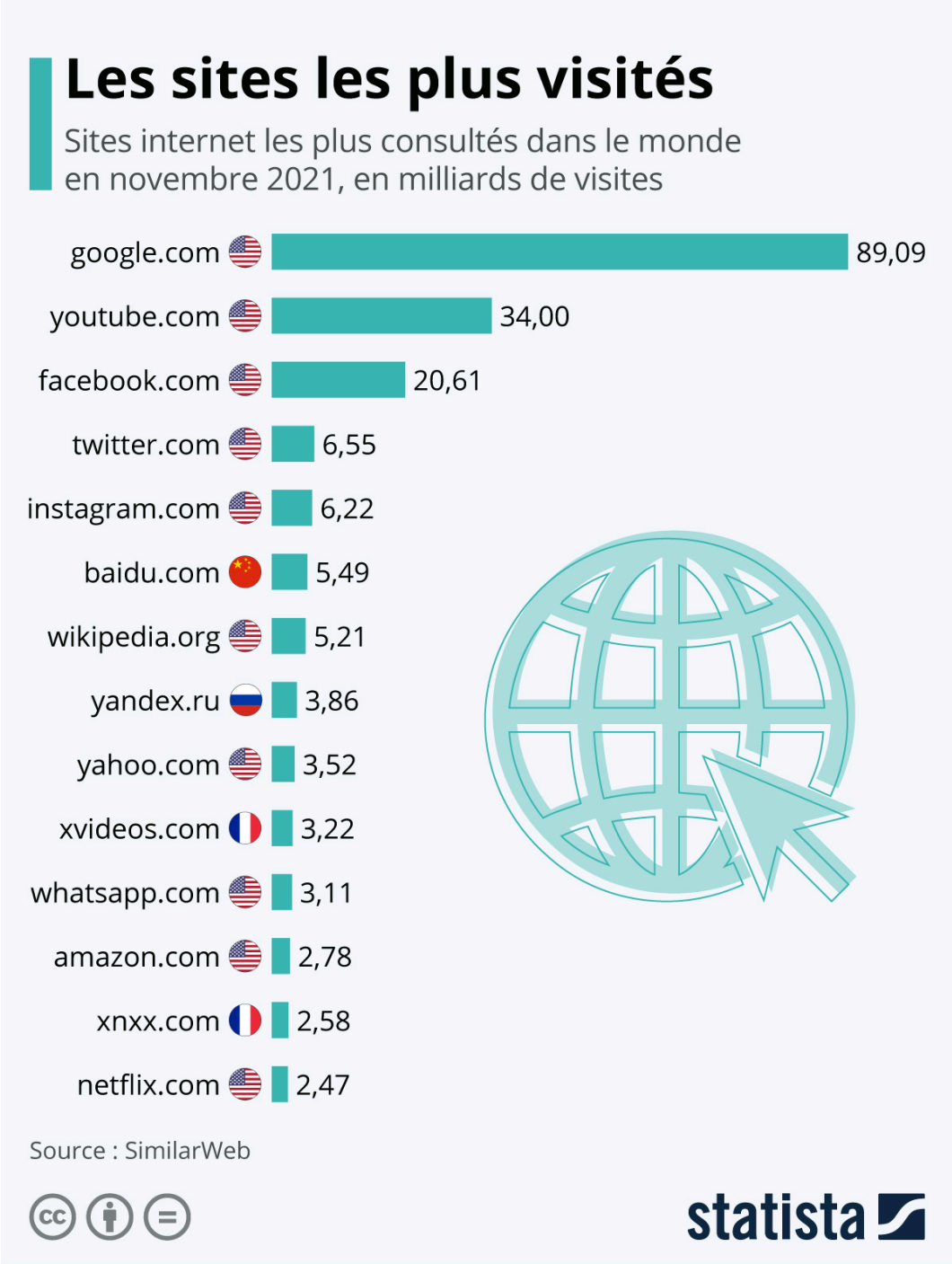
- **2000**, **Nupedia** est une encyclopédie commerciale, sous licence propriétaire, avec comité éditorial.
- **2001**, Nupedia devient **Wikipedia**, les contenus sont créés collaborativement via un Wiki et sont placés sous licence libre.

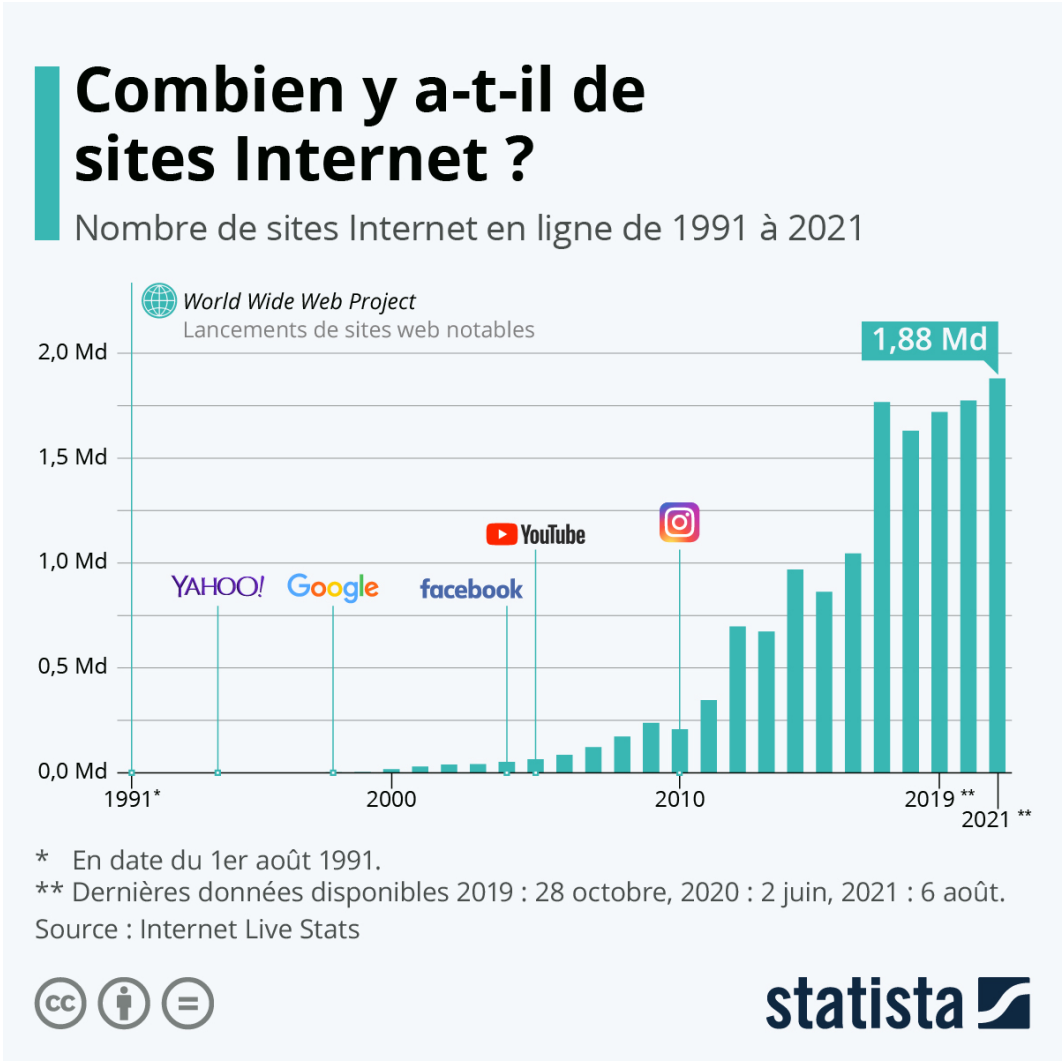
Wikipedia fait partie des 10 sites les plus visités au monde.

Photos et vidéos

- **2004**, **Flickr** permet de diffuser gratuitement des photographies.
- **2005**, **YouTube** permet de partager des vidéos sur le Web, alors que le coût de l'hébergement web est encore élevé.

> 💬 Remarque
>
> Flickr et YouTube comportent des fonctions de de type "aimer", "partager", "commenter" qui les classent également dans le domaine des médias sociaux.





Nombre de sites Internet entre 1991 et 2021

	Websites	Launched
2018	1 630 322 579	
2017	1 766 926 408	
2016	1 045 534 808	
2015	863 105 652	
2014	968 882 453	
2013	672 985 183	
2012	697 089 489	
2011	346 004 403	
2010	206 956 723	Pinterest Instagram
2009	238 027 855	
2008	172 338 726	Dropbox
2007	121 892 559	Tumblr
2006	85 507 314	Twitter
2005	64 780 617	YouTube Reddit
2004	51 611 646	Facebook Flickr
2003	40 912 332	WordPress LinkedIn
2002	38 760 373	
2001	29 254 370	Wikipedia
2000	17 087 182	Baidu
1999	3 177 453	PayPal
1998	2 410 067	Google
1997	1 117 255	Yandex Netflix
1996	257 601	
1995	23 500	Altavista Amazon
1994	2 738	Yahoo
1993	130	
1992	10	
1991	1	

Nombre de sites web entre 1991 et 2018

Solutions des exercices

Solution n°1

[exercice p. 8]

Le premier site web `info.cern.ch` a vu le jour au CERN. Quelles technologies ont mises au point Tim Berners-Lee et Robert Cailliau pour y parvenir ?

- **A** Le protocole HTTP
- **B** Un navigateur web
- **C** Un serveur web
- **D** Le langage HTML

Ces 4 éléments fondateurs ont été posés dans leurs premières versions dès 1990 par les inventeurs du Web.

Solution n°2

[exercice p. 12]

La Wayback Machine permet de consulter des contenus web copiés par l'organisme Internet Archive (pages web, images, vidéos...).

A

L'intégralité des sites web non protégés par login est archivée par Internet Archive.

Cela représente une quantité trop importante de données, a fortiori si l'on considère toutes les versions de ces sites.

B

☐ Il est possible de demander à l'Internet Archive d'indexer ou de ne pas indexer son site web.

On peut demander à l'Internet Archive d'indexer son site, ou au contraire demander la désindexation et la suppression des copies qui auraient été faites automatiquement.

C

La copie pour archive est toujours autorisée car l'Internet Archive est un organisme à but non lucratif.

Dans la plupart des pays du monde, dont la France, l'auteur d'un contenu original est seul détenteur des droits sur son œuvre et seul lui peut autoriser la copie de ce contenu.

Solution n°3

[exercice p. 15]

Quelle est le temps approximatif de chargement d'une page web de 80 ko (le poids moyen d'une page à la fin des années 1990) avec un modem 56k ?

- **A** de l'ordre d'une seconde
- **B** de l'ordre de dix secondes
- **C** de l'ordre d'une minute
- **D** de l'ordre d'une heure

80 ko ~ 500 kbits ~ 10 secondes à condition que le modem ne soit utilisé que pour cet accès et que le serveur soit en mesure de servir à cette vitesse également.

Solution n°4

[exercice p. 19]

Classer dans leur ordre de création les navigateurs suivants :

A WorldWideWeb | **B** NSCA Mosaic | **C** Netscape Navigator | **D** Opera | **E** Internet Explorer | **F** Firefox | **G** Safari | **H** Chrome

Des butineurs aux navigateurs web (1990-2008) *(cf. p.20)*

NB : première version de Opera et Internet Explorer selon Wikipédia

- Opera : avril 1995
- IE : août 1995

Solution n°5

[exercice p. 23]

De quoi a-t-on nécessairement besoin pour mettre en ligne un site web personnel en France ?

- **A** Qu'une entreprise spécialisée nous mette à disposition un espace sur un de ses serveurs
- **B** D'un accès à un ordinateur doté d'une adresse IP fixe
- **C** D'une autorisation de publication délivrée par l'état
- **D** D'un logiciel de serveur web

Solution n°6

[exercice p. 27]

Un moteur de recherche intègre nécessairement les fonctions suivantes :

- [x] **A** Un programme qui parcours le Web pour référencer les pages existantes (*crawler*)
- [x] **B** Un programme qui décrit les pages référencées dans une base de données (indexation)
- [x] **C** Une interface qui permet à l'utilisateur d'exprimer une recherche (formulaire de recherche)
- [x] **D** Une présentation ordonnée des résultats de recherche (pages des résultats)
- [ ] **E**

Une personnalisation des résultats en fonction d'un modèle de la personne qui fait la requête	Google Search utilise cette technique, mais ce n'est pas le cas de Qwant ou DuckDuckGo. Outre la question de la collecte des données personnelles, la recherche personnalisée engendre la création de bulles de filtres. Voir aussi : *Chercher des contenus de référence* (cf. p.58)

Solution n°7

[exercice p. 31]

Parmi les organisations suivantes, lesquelles ont été introduites en bourse une ou deux années après leur création seulement, alors qu'elles étaient déficitaires :

- [ ] **A**

Amazon	Amazon ouvre en 1995, entre en bourse en 1997 et réalise ses premiers bénéfices en 2001.

- [ ] **B**

Wikimedia Foundation (qui gère Wikipédia)	La Wikimedia Foundation est une organisation à but non lucratif.

- [ ] **C**

Netscape Communications	Netscape Communications est créée en 1994 et entre en bourse en 1995 (6,6 millions de dollars de perte pour un chiffre d'affaires de 85 millions)

- [ ] **D**

Iliad (maison mère de Free)	Free est créé en 1999, Illiad entre en bourse en 2004, l'entreprise est bénéficiaire depuis 1999 (6,7 millions d'euros de bénéfices pour 46,6 millions de chiffre d'affaires).

Sources : Wikipédia

Solution n°8

[exercice p. 34]

Exercice

Combien d'articles existaient sur Wikipédia en anglais à la fin de l'année de sa création en 2001 ?

- **A** Environ 2.000
- (**B**) Environ 20.000
- **C** Environ 200.000
- **D** Environ 2.000.000

Exercice

Combien d'articles existaient sur Wikipédia en anglais à la fin de l'année 2020, 20 ans après sa création ?

- **A** Environ 60.000
- **B** Environ 600.000
- (**C**) Environ 6.000.000
- **D** Environ 60.000.000

Exercice

En quelle année a été créée la version de Wikipédia en Français ?

2001

Glossaire

Protocole

Un **protocole informatique** est un ensemble de règles techniques permettant à des ordinateurs d'échanger des informations via un langage commun.

Abréviations

CMS : Content Management System
SGC : Système de Gestion de Contenu
WCMS : (ou WMS) Web Content Management System
WCMS : (ou WMS) Web Content Management System

Bibliographie

[(Levie, 2006)] Levie F. (2006). L'homme qui voulait classer le monde : Paul Otlet et le Mundaneum. Les impressions nouvelles.

[(Otlet, 1934)] Otlet P. (1934). Traité de documentation : le livre sur le livre, théorie et pratique. Editions Mundaneum, Bruxelles. http://search.ugent.be/meercat/x/view/rug01/000990276.

[Bellanger, 2012] Bellanger Aurélien. 2012. *La Théorie de l'information*. Gallimard. https://fr.wikipedia.org/w/index.php?title=La_Th%C3%A9orie_de_l%27information_(roman)&oldid=171794737.

[Berners-Lee, 1989] Berners-Lee Tim. 1989. *Information Management: A Proposal*. https://www.w3.org/History/1989/proposal.html.

[Brin and Page, 1998] Brin Sergey, Page Lawrence. 1998. *The Anatomy of a Search Engine*. Université de Stanford. https://archive.wikiwix.com/cache/index2.php?url=http%3A%2F%2Finfolab.stanford.edu%2Fpub%2Fpapers%2Fgoogle.pdf%2Findex.html#federation=archive.wikiwix.com&tab=url.

[Defosse et al., 2006] Defosse Marie-Françoise, Follet Marianne, Maniez Dominique. 2006. *Rechercher l'information*. C2IMES. http://www.c2imes.org/MODULES/B2.

[Hoffmann, 2022] Hoffmann Jay. 2022. *The Web's Timeline*. in The History of the Web. https://thehistoryoftheweb.com/timeline/.

[Masutti, 2020] Masutti Christophe. 2020. *Affaires privées : Aux sources du capitalisme de surveillance*. C&F Éditions.

[Ricmac, 2021] Ricmac. 2021. *Web Development History*. in Web Development History. https://webdevelopmenthistory.com/.

[Schafer, 2022] Schafer Valérie. 2022. *En construction : La fabrique française d'Internet et du Web dans les années 1990*. Ina. https://ebooks.appeldulivre.fr/1433593/En-construction.ebook.

Crédits des ressources

IBM 9020 (London Air Traffic Control Centre, 1974) p. 3
NATS (OGL v1.0[4]) — https://en.wikipedia.org/wiki/IBM_9020

Sites des États-Unis reliés à ARPANET en 1974. p. 4
Licence : Domaine Public - Yngvar

p. 5
Internet Protocol, RFC 791, 1981. rfc-editor.org[5]

Répartition des internautes dans le monde en 2013 p. 6
Attribution - Pas d'Utilisation Commerciale - Ralph Straumann, Mark Graham

Internet censorship and surveillance by country (2018) p. 6
Attribution - https://commons.wikimedia.org/wiki/File:Internet_Censorship_and_Surveillance_World_Map.svg

p. 9
Licence : Domaine Public - Tim Berners-Lee, 1989 · https://www.w3.org/History/1989/proposal.html

Tiroirs du Mundaneum p. 48
Par fdecomite — Flickr: Drawers, CC BY 2.0, commons.wikimedia.org/w/index.php?curid=15888697

Adresse de la page d'accueil de Wikipédia qui utilise le protocole HTTPS p. 49
Attribution - https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal

Page www.lemonde.fr du 22 juin 1997 p. 14
https://web.archive.org/web/19970622152113/http://www.lemonde.fr/

NCSA Mosaic (page de Robert Cailliau au Cern) p. 21
Universel - Transfert dans le Domaine Public - Selbst Erstellt · https://commons.wikimedia.org/w/index.php?curid=75617889

Mozilla (1999) p. 21
web.archive.org[6]

Évolution de la part respective des navigateurs entre 1996 et 2009 p. 22
Attribution - Partage dans les Mêmes Conditions

Répartition géographique des différents RIR p. 55
Attribution - Partage dans les Mêmes Conditions - Dork, Canuckguy et al., Sémhur

4. https://nationalarchives.gov.uk/doc/open-government-licence/version/1/
5. https://www.rfc-editor.org/rfc/rfc791
6. https://web.archive.org/web/19990508132717/http://www.mozillazine.org/screenshots/image/linux08.gif

Yahoo! en 1996 (depuis Internet Archive) p. 29
Yahoo! home page · web.archive.org[7]

Wired.com invents the web banner ad (Oct. 27, 1994) p. 32
wired.com[8]

Site de l'IUMA en 1996 (depuis Internet Archive) p. 35
Licence : Domaine Public - IUMA · https://web.archive.org

Nombre de sites Internet entre 1991 et 2021 p. 37
Statista / Live Stats · fr.statista.com[9]

Nombre de sites web entre 1991 et 2018 p. 38
InternetLiveStats[10]

7. https://web.archive.org/web/19961017235908/http://www2.yahoo.com/
8. https://www.wired.com/2010/10/1027hotwired-banner-ads/
9. https://fr.statista.com/infographie/19063/nombre-de-sites-web-internet-dans-le-monde/
10. https://www.internetlivestats.com/

Contenus annexes

1. Histoire du Web

Objectifs

- Découvrir l'histoire du Web

Mise en situation

Le *protocole* p.43 du Web est **HTTP**. Son évolution récente HTTPS ajoute une couche cryptographique afin que les échanges sur le Web puissent être confidentiels.

Classification documentaire, fin XIXe (à l'origine des moteurs de recherche) Rappel

- La classification décimale de Dewey (CDD) est un système visant à classer l'ensemble du fonds documentaire d'une bibliothèque, développé en 1876 par Melvil Dewey. (https://fr.wikipedia.org/wiki/Classification_d%C3%A9cimale_de_Dewey)
- Dans la continuité, Paul Otlet crée en 1905 avec Henri La Fontaine le système de Classification Décimale Universelle. Ils travaillent aussi à l'établissement d'un répertoire de l'ensemble des « faits » et des « écrits » concernant la société. En 1895, celle-ci comprend 400 000 notices (https://fr.wikipedia.org/wiki/Paul_Otlet, *(Otlet, 1934)*(Otlet, 1934) p.45).

Mundaneum, 1920 (à l'origine de Wikipédia) Rappel

Tiroirs du Mundaneum

Paul Otlet construit le Mundaneum, une bibliothèque universelle ayant vocation à archiver toute la connaissance du monde sous toutes ses formes documentaires *(Levie, 2006)*(Levie, 2006) p.45.

Le Web : la partie documentaire d'Internet

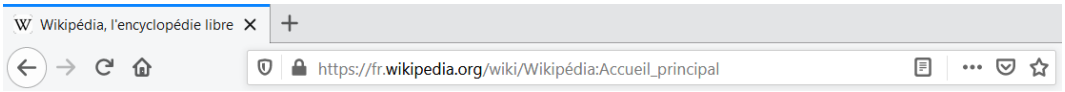
Le Web (ou WWW, *World Wide Web*) est une application d'Internet mise au point au CERN en 1989, par Tim Berners Lee et Robert Cailliau.

C'est une **toile de ressources** composée de documents - les pages web - et d'un système de navigations entre les documents - les **liens hypertexte**. Le Web est donc l'ensemble des sites internets, auxquels on accède en utilisant un navigateur (Google Chrome, Mozilla Firefox, etc.).

HTTP et HTTPS, les protocoles du Web

Le fonctionnement du Web se base sur le protocole HTTP, pour *HyperText Transfert Protocol*.

Le protocole HTTP a été étendu par le protocole HTTPS, pour *HyperText Transfer Protocol Secure*. De plus en plus utilisé, il sécurise les communications en empêchant un tiers indiscret d'accéder à leur contenu en espionnant le réseau.



Adresse de la page d'accueil de Wikipédia qui utilise le protocole HTTPS

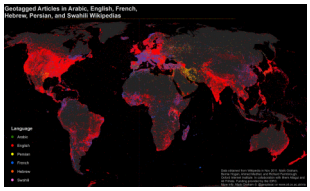
On voit l'utilisation d'HTTPS dans l'adresse de la page (`https://`) et par la présence du petit cadenas à gauche de l'adresse.

40 maps that explain the internet (2014)

Complément

« The internet increasingly pervades our lives, delivering information to us no matter where we are. It takes a complex system of cables, servers, towers, and other infrastructure, developed over decades, to allow us to stay in touch with our friends and family so effortlessly. Here are 40 maps that will help you better understand the internet — where it came from, how it works, and how it's used by people around the world.

https://www.vox.com/a/internet-maps (2014) »



Maps that explain the internet

Complément

- « *Information Management: A Proposal* », proposition initiale de Tim Berners-Lee pour la réalisation de son prototype : https://www.w3.org/History/1989/proposal.html
- Rubrique du CERN sur la naissance du Web : https://home.cern/science/computing/birth-web

À retenir

- Le Web est l'application la plus connue d'Internet. Elle repose sur le protocole HTTP.

2. Les interactions entre acteurs du réseau : peering et transit

Mise en situation

Ce qui fait d'Internet un réseau mondial est l'**interconnexion** entre ses acteurs. Les clients d'un FAI à Paris peuvent communiquer des données à un hébergeur aux États-Unis, à travers une chaîne d'intermédiaires (transitaires, autres FAI, etc.).

Cette interconnexion se manifeste par deux méthodes : le **peering** et le **transit**.

Peering

Az Définition

Le *peering* (de l'anglais *peer*, pair) interconnecte deux acteurs.

Chaque acteur donne ainsi accès à son réseau à l'autre.

L'accès est

- direct (*peering* **privé**)
- ou réalisé par l'intermédiaire d'un IXP (*peering* **public**).

Le *peering* interconnecte en général des acteurs de taille comparable et il n'y a pas d'échange d'argent si les deux acteurs pensent qu'il est équitable.

Transit

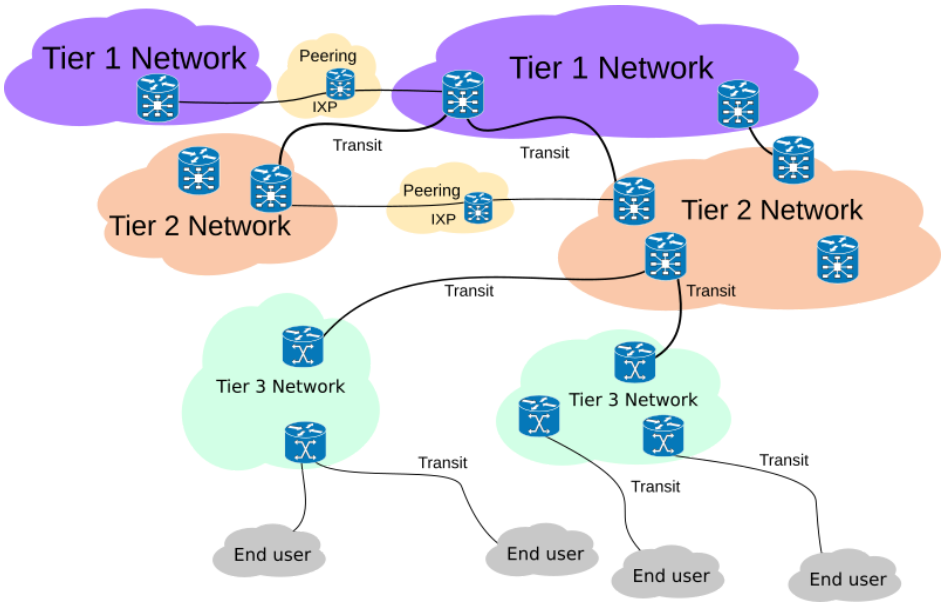
Az Définition

Le transit permet à un acteur plus modeste d'acheter du transit à un acteur plus important.

Cet achat de transit permet de connecter son réseau à l'ensemble d'Internet.

Classement des opérateurs d'Internet

Az Définition



Transit et peering entre opérateurs Internet de niveau 1 à 3

Les opérateurs d'Internet sont classés selon leur type d'interconnexion. Les premiers sont en général moins connus du grand public mais de taille très importante.

- *Tier* 1 : ils n'achètent pas de transit et accèdent à la totalité d'Internet via des peering avec d'autres grands opérateurs également de *tier* 1.
- *Tier* 2 : ils ont des accords de peering mais dépendent aussi d'une offre de transit.
- *Tier* 3 : n'ont aucun accord de peering et dépendent totalement d'une offre de transit.

Chemin emprunté par une ressource

Exemple

La requête pour consulter un article sur Wikipédia depuis un ordinateur relié à une box Internet pourra par exemple :

- Partir du réseau local pour rejoindre le réseau du FAI (en général de *tier* 3),
- Être reçue par le FAI qui détermine la meilleure route à emprunter, en fonction de ses accords de peering et/ou de transit,
- Être transférée à d'autres FAI, de proche en proche, en changeant éventuellement de *tier*,
- Être finalement transmise au FAI utilisé par Wikipédia,
- Être reçue par Wikipédia.

Peering et transit

Remarque

Le peering et le transit sont des méthodes qui se complètent.

En général :

- Les petits acteurs, lorsqu'ils ne sont pas trop éloignés, se connectent via du peering.
- Autrement, ils achètent du transit pour se connecter l'ensemble d'Internet.
- Les transitaires s'interconnectent entre eux avec du peering, et la boucle est bouclée.

IXP

Complément

Un point d'échange Internet (IXP, pour *Internet Exchange Point*) est une des infrastructures physiques permettant aux différents acteurs d'Internet de se connecter.

Un IXP est une alternative aux transitaires : les acteurs souhaitant s'échanger des données (FAI, entreprises, états, etc.) s'interconnectent à un IXP plutôt que d'acheter du trafic à un transitaire.

Les IXP permettent d'interconnecter plusieurs réseaux.

Pour aller plus loin ⊕ Complément

www.nextinpact.com/news/108821-on-vous-explique-simplement-comment-fonctionnent-tuyaux-dinternet.htm.

À retenir

- Le peering permet à deux acteurs de partager leurs réseaux respectifs, généralement de manière équitable et gratuite.
- Le transit permet à un acteur d'avoir accès à l'ensemble d'Internet via un accord avec un autre acteur de taille plus importante.

3. Adresses IP et noms de domaine

Adresses IP Az Définition

Une adresse IP est un **numéro d'identification d'une machine**.

Il est attribué de façon à chaque appareil relié à Internet.

Les adresses ont un format différent selon la version d'IP utilisée :

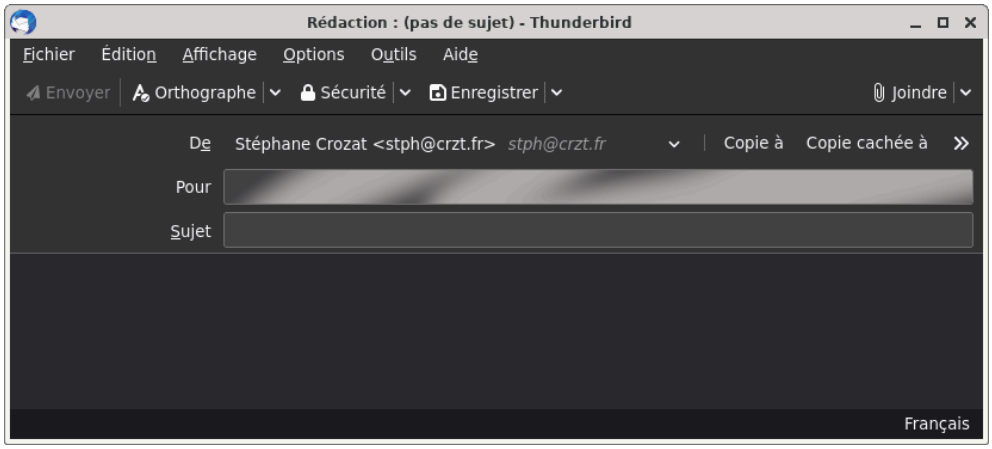
- pour IPv4, les adresses sont sur 4 octets (exemple : `192.168.3.200`),
- pour IPv6, les adresses sont sur 16 octets (exemple : `2001:0db8:0000:85a3:0000:0000:ac1f:8001`).

Nom de domaine Az Définition

Un nom de domaine est une **adresse symbolique** permettant d'éviter de devoir retenir des adresses IP.

Ainsi, à un nom de domaine tel que `wikipedia.org` on peut associer une adresse telle que `91.198.174.192`.

Exemple



DNS

Az Définition

Un **serveur DNS** (pour *Domain Name System*) est un serveur qui a connaissance de la **correspondance entre noms de domaines et adresses IP**. Il est capable de répondre à des requêtes qui lui demandent de **résoudre** un nom de domaine en une adresse IP.

Complément

Adresses IP (cf. p.53)

Noms de domaine (cf. p.55)

4. Adresses IP

IPv4 et IPv6

Rappel

Il existe deux versions du protocole IP : IPv4 et IPv6.

Adresses IP

Az Définition

Une adresse IP est un **numéro d'identification d'une machine**.

Il est attribué de façon à chaque appareil relié à Internet.

Les adresses ont un format différent selon la version d'IP utilisée :

- pour IPv4, les adresses sont sur 4 octets (exemple : `192.168.3.200`),
- pour IPv6, les adresses sont sur 16 octets (exemple : `2001:0db8:0000:85a3:0000:0000:ac1f:8001`).

Remarque

On dit souvent une IP pour parler d'une adresse IP

IP statique ou dynamique

Attention

L'adresse IP attribuée à un appareil peut être statique (fixe) ou dynamique (temporaire).

- Seuls les appareils avec des IP statiques peuvent correctement fonctionner en temps que serveur (il est difficile de savoir comment les contacter sinon).
- Les appareils avec IP dynamiques sont donc essentiellement des consommateurs de contenus (c'est typiquement le cas des smartphones).

Pour un Internet plus équilibré il serait intéressant que la majorité des utilisateurs disposent d'IP fixes pour pouvoir proposer du contenu, mais la pénurie d'IPv4 rend cet objectif difficile à atteindre.

IP locales

Az Définition

Les adresses IP sont des ressources (devenue rares) qui ont une valeur. Pour limiter les coûts, à l'intérieur d'un réseau local, les ordinateurs ont des adresses IP spéciales réservées aux réseaux locaux appelées adresses **IP locales**.

Ces adresses sont libres d'utilisation et ne nécessitent pas de demander d'autorisation à qui que ce soit.

Synonyme : IP privée

Translation d'adresse

Méthode

Comme les adresses privées ne sont pas utilisées de manière unique, une technique appelée **translation d'adresse** associe une adresse IP publique unique à un ensemble d'adresse IP privées d'un réseau local donné.

Ainsi, les appareils à l'intérieur du réseau local peuvent tout de même être contactés depuis Internet à travers l'IP publique.

IPv4 et pénurie

Complément

Aujourd'hui, la quasi-totalité des 4 milliards d'adresses IPv4 sont attribuées. Cet état de fait pousse le monde à se tourner vers IPv6 qui dispose d'un espace d'adressage bien plus vaste.

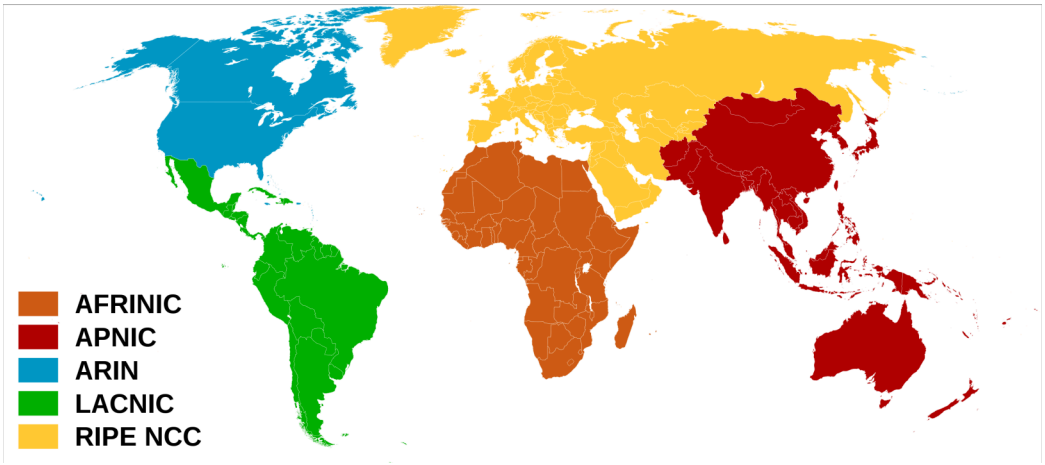
La translation d'adresses est un palliatif temporaire mais très utilisé pour le manque d'adresses IPv4.

Attribution des adresses IP

Complément

L'attribution des adresses IP est gérée par l'IANA (pour *Internet Assigned Numbers Authority*), un département de l'ICANN, qui administre les ressources numériques d'Internet.

L'IANA a segmenté les adresses IP disponibles en 256 blocs distribués aux RIR (Registres Internet Régionaux). Les RIR gèrent les blocs d'adresses au niveau local en définissant des politiques d'attributions d'adresses IPv4 à ceux qui le demandent.



Répartition géographique des différents RIR

5. Noms de domaine

Nom de domaine

Az Définition

Un nom de domaine est une **adresse symbolique** permettant d'éviter de devoir retenir des adresses IP.

Ainsi, à un nom de domaine tel que `wikipedia.org` on peut associer une adresse telle que `91.198.174.192`.

Les noms de domaines ont une structure bien particulière : ils forment une arborescence dont la racine est un point, et dont les niveaux sont séparés par des points. Cette arborescence permet de connaître l'endroit où chercher pour trouver l'adresse IP correspondante au nom de domaine concerné.

Utilité d'un nom de domaine

💬 Remarque

- Un nom de domaine est destiné à être utilisé par des utilisateurs humains. Il n'est fondamentalement pas nécessaire mais permet de donner du sens et est aussi plus facile à retenir.
- De plus un nom de domaine peut être changé là où une adresse IP ne peut généralement pas l'être.

Pour autant, cela ajoute un niveau de complexité : l'appareil qui veut communiquer avec le serveur de Wikipédia a besoin de transformer `www.wikipedia.org.` en une adresse IP.

Pour cela, il fait appel à un **serveur de noms de domaine**.

DNS

Az Définition

Un **serveur DNS** (pour *Domain Name System*) est un serveur qui a connaissance de la **correspondance entre noms de domaines et adresses IP**. Il est capable de répondre à des requêtes qui lui demandent de **résoudre** un nom de domaine en une adresse IP.

DNS, un système hiérarchique

Les serveurs DNS sont organisés hiérarchiquement et sont habilités à établir la correspondance entre IP et nom de domaines pour seulement certains noms de domaines et un certain niveau de l'arborescence. Quand un DNS est autorisé à établir cette correspondance, il est **fait autorité** pour un ensemble de noms de domaines donnés.

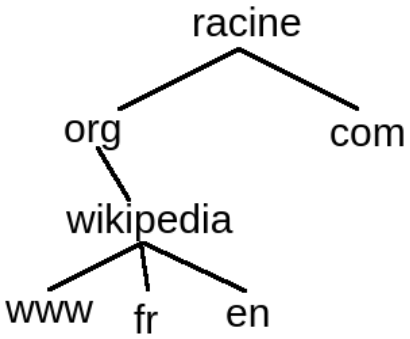
Si un serveur DNS n'a pas l'adresse IP correspondante pour un nom de domaine, il va faire appel à un autre serveur DNS.

Les serveurs DNS responsables du premier niveau d'arborescence (.fr, .com, .org, etc.) sont appelés TLD, pour *Top-Level Domain*.

Résolution d'une requête DNS

Exemple

Les navigateurs web ont besoin de résoudre des noms de domaine. Lorsque `www.wikipedia.org` est consulté, un **résolveur** DNS entre en jeu.



Chaque nœud de cet arbre correspond à un serveur DNS, responsable d'une partie de l'arborescence.

Pour résoudre `www.wikipedia.org`, le résolveur commence donc le haut de l'arborescence : le **serveur racine**. Ce dernier n'est pas responsable de l'adresse IP de Wikipédia mais connaît les serveurs DNS TLD pour `.org`.

Les serveurs DNS autoritaires pour `.org` ne sont pas non plus responsables de l'adresse IP de Wikipédia. À leur tour, ils connaissent les adresses IP des serveurs DNS faisant autorité pour Wikipédia.

Ces serveurs fourniront finalement l'adresse du sous-domaine `www` de `wikipedia.org`.

Résoudre une IP

Méthode

Il existe des résolveurs DNS en ligne de commande, comme `host`.

```
$ host www.wikipedia.org
www.wikipedia.org is an alias for dyna.wikimedia.org.
dyna.wikimedia.org has address 91.198.174.192
dyna.wikimedia.org has IPv6 address 2620:0:862:ed1a::1
```

`host` retourne les enregistrements DNS, ici on voit l'adresse IPv4 actuelle de Wikipédia, qui est `91.198.174.192`

Réponse ne faisant pas autorité

⊕ Complément

L'exemple ci-dessus mentionne que la réponse ne fait pas autorité, ce qui signifie qu'elle émane d'un serveur n'ayant pas autorité sur le nom de domaine.

Ceci s'explique par un système de **cache** : pour éviter que les serveurs DNS de Wikipédia soient surchargés, d'autres serveurs hébergent une copie des correspondances entre IP et nom de domaine. Ces serveurs ne sont en revanche pas autorisés à modifier ces correspondances.

Comment est-ce que les noms de domaines sont enregistrés ?

Fondamental

Un nom de domaine peut être acheté auprès d'organismes en charge de l'administration des nom de domaines, appelés **registrar**. Le registrar va s'occuper d'ajouter le nom de domaine à un de ses DNS faisant autorité puis va aller indiquer aux DNS TLD correspondants qu'un nouvel enregistrant a eu lieu.

Résolution locale de nom de domaine via le fichier hosts

⊕ Complément

Les systèmes d'exploitation maintiennent un fichier `hosts` contenant les informations de résolution de certains nom de domaines. Ce fichier a généralement cette forme :

```
127.0.0.1       localhost
127.0.0.1       monPC
208.80.154.224  www.wikipedia.org.
```

Ce fichier est interrogé en premier : il est ainsi possible de forcer la résolution d'un nom de domaine en n'importe quel adresse, au niveau de l'ordinateur.

À retenir

- Un nom de domaine est un alias pour une adresse IP.
- Les noms de domaines décrivent une arborescence.
- Le système permettant d'associer adresse IP et nom de domaine s'appelle DNS.
- Les serveurs DNS sont organisés de manière hiérarchique et s'occupent d'un ensemble donné de noms de domaines.

6. CMS

Si le terme d'ECM est sensé inclure les outils dédiés à la production de sites Web, dans les faits ces outils constituent plutôt une classe à part que l'on regroupe sous le terme de WCMS.

WCMS

Az Définition

Un *Web Content Management System* (*WCMS* p.44, souvent appelé simplement *CMS* p.44 ou *SGC* p.44) est un outil de production de pages HTML, généralement adossé à une base de données relationnelle, à des logiques de gabarits de pages (*templates*) et des feuilles de style CSS.

Cette classe d'outils née au tout début des années 2000 a permis de démocratiser la réalisation de sites Web et d'intranets d'entreprise, en délestant les auteurs de contenus des tâches techniques d'encodage HTML et de programmation client (JavaScript) ou serveur (PHP généralement).

Exemple de CMS Exemple

Parmi la pléthore de systèmes existants, la plupart open source, l'on pourra citer les système les plus couramment rencontrés tels que SPIP, Wordpress, Drupal ou Joomla!.

http://fr.wikipedia.org/wiki/Liste de systèmes de gestion de contenu[11]

7. Chercher des contenus de référence

Objectif

Savoir chercher des contenus de référence en utilisant des sites généralistes et des bibliographies.

- La recherche superficielle est à proscrire (moteur de recherche avec bulle de filtre, articles de presse...).
- Découvrir un domaine via Wikipédia ou des articles de blog permet de remonter aux sources.

Trouver, qualifier, documenter l'information Fondamental

« *Tout travail universitaire doit être documenté (Defosse et al., 2006*[Defosse et al., 2006 p.45]*)* » : donc clairement **sourcé**, on doit savoir qui parle et d'où.

L'important n'est pas d'avoir raison mais que le lecteur soit en mesure de situer et critiquer l'information que l'on fournit.

Attention

L'usage de référence journalistiques superficielles n'est en général pas une bonne idée (sauf s'il s'agit d'une enquête originale ou d'un article de fond).

En revanche on peut se servir d'un (bon) article d'information journalistique pour trouver des sources (remonter aux sources citées donc).

11. http://fr.wikipedia.org/wiki/Liste_de_syst%C3%A8mes_de_gestion_de_contenu

Découverte d'un domaine

Conseil

Wikipédia et une recherche web superficielle (premiers résultats) sont un bon moyen d'entrer dans un domaine. En revanche, il s'agit d'un point de départ, et non d'un point d'arrivée.

Les premiers mots-clés ne suffisent pas en général :

- suivre les traces, c'est à dire utiliser les premiers résultats pour mieux comprendre et mieux définir ce qui est cherché ;
- affiner en ajoutant des mots du domaine visé collectés au long de la recherche ;
- utiliser les opérateurs `"..."` (entre guillemets informatiques) pour chercher des expressions complètes et `-` (moins) pour exclure des termes.

Exemple

Recherche Wikipédia : identifier et suivre les références

Bulles de filtre

Attention

Les moteurs qui utilisent les données personnelles pour sélectionner les résultats de recherche (comme *Google Search*) créent des bulles de filtres, c'est à dire que les premiers résultats tendent à être choisis pour être déjà familiers de l'utilisateur, pour **lui faire plaisir**.

Dans le cas de l'exploration d'un nouveau domaine, on vise habituellement l'inverse, c'est à dire explorer de nouveaux territoires.

Quelques solutions :

- utiliser un moteur sans profilage comme **Duckduckgo** ou **Qwant** ;
- utiliser un VPN ou Tor pour masquer son profil au moteur de recherche (il reste le risque d'hériter du profil du VPN ou du nœud de sortie Tor, mais celui-ci sera tout de même dilué par la masse d'utilisateurs empruntant le même chemin).

Autres sources d'information (au-delà de la recherche superficielle)

Méthode

- Les bibliographies issues des premières recherches
- Les sites web de référence liés au domaine de recherche

- Les moteurs de recherche spécialisés permettant la recherche de publications scientifiques :
  - Google Scholar (https://scholar.google.com[12])
  - Semantic Scholar (https://www.semanticscholar.org)

Exemple

Recherche sur Google Scholar

Complément

- Les archives ouvertes : HAL, arXiv...
- Les bases de données partiellement ouvertes (recherche gratuite, accès ouvert à certains articles, payant à d'autres) : Cairn.info, OpenEdition, JSTOR...
- Bases de données fermées à péage : Scopus, Web of Science...

12. https://scholar.google.com/