

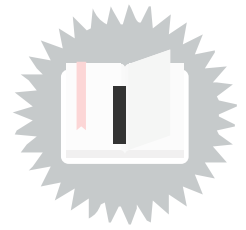
Une petite histoire d'Internet et du Web



Table des matières

I - Histoire d'Internet	3
II - Exercice	7
III - Naissance du Web (1989-1993)	8
IV - Exercice	10
V - Wayback Machine de l'Internet Archive	11
VI - Exercice	13
VII - Accéder au Web au XXe siècle : cybercafés, universités, connexions à domicile (1993-1999)	14
VIII - Exercice	17
IX - Des butineurs aux navigateurs web	18
X - Exercice	21
XI - Sites web personnels	22
XII - Exercice	24
XIII - Les moteurs de recherche et annuaires web	25
XIV - Exercice	27
XV - Économie du Web : bourse, publicité, commerce	28
XVI - Exercice	29
XVII - Informations et culture et sur le Web	30
Solutions des exercices	33
Abréviations	36
Bibliographie	37
Crédits des ressources	38

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.



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 télégraphique 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.37}) »

Updated by: [1349](#), [2474](#), [6864](#)
 RFC: 791

INTERNET STANDARD
 Errata Exist

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

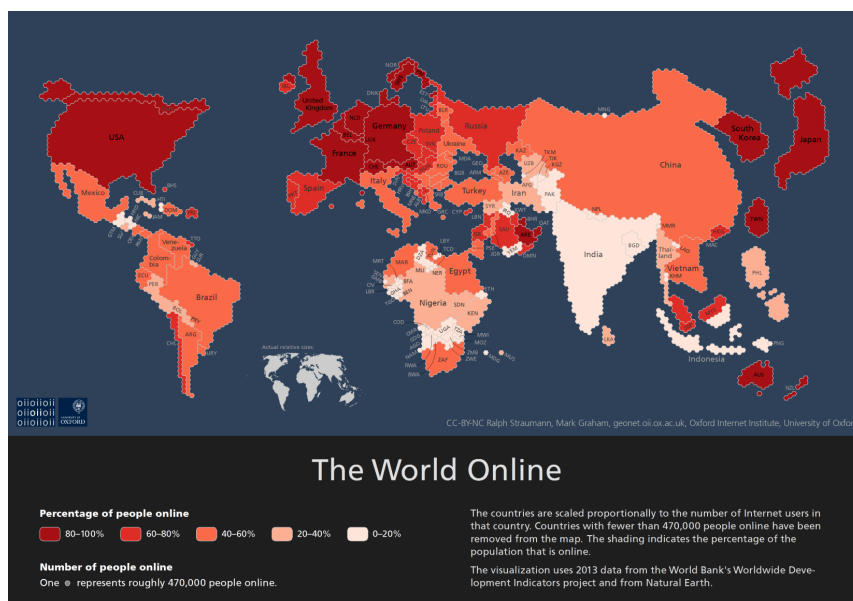
by

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 inégale : 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).

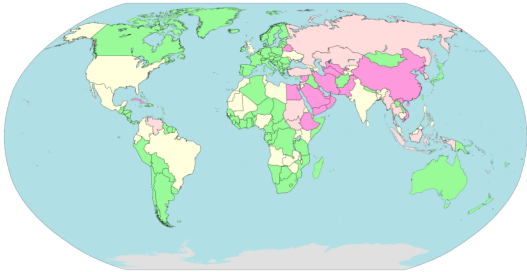


Répartition des internautes dans le monde en 2013

Internet : un enjeu politique et social fort



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



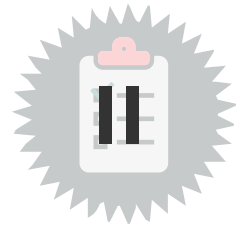
- Internet a aussi été un enjeu de pouvoir dans l'économie : la « **bulle Internet** » à la fin des années 1990 et le début 2000 a témoigné d'un engouement et d'une forte spéculation dans les entreprises des sites de commerces en lignes et les compagnies de services de télécommunications.
- La **neutralité 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 Xiommo (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. 33]



En 1990, le premier site web `info.cern.ch` voit le jour au CERN. Quelles technologies ont mises au point Tim Berners-Lee et Robert Cailliau pour y parvenir ?

- Le protocole HTTP
- Un navigateur web
- Un serveur web
- Le langage HTML

Naissance du Web (1989-1993)



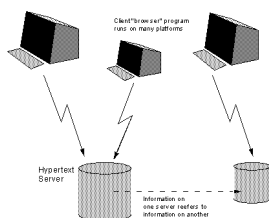
Le Web est né en 1990



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.37}) :

- 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.



« 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.37}



« Je suis sur Nantes Je reste 4 jours en rideau pas moyen de se connecter, je reconfigure les paramètres TCP IP en vain, n'étant plus connecté au réseau j'envoie un fax, sans réponse je téléphone à 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'accès de Nantes. Worldnet tu nous fatigue vraiment.

Post sur le newsgroup soc.culture.french en 1996, relaté par Valérie Schafer (2022)^{Schafer, 2022 p.37}



Au début, l'indifférence



- 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.37}



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.37}

Ricmac, 2021^{Ricmac, 2021 p.37}

Histoires d'avant le Web

 **Complément**

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

Histoire du Web

Exercice

[solution n°2 p. 33]



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

- L'intégralité des sites web non protégés par login est archivée par Internet Archive.
- Il est possible de demander à l'Internet Archive d'indexer ou de ne pas indexer son site web.
- La copie pour archive est toujours autorisée car l'Internet Archive est un organisme à but non lucratif.

Wayback Machine de l'Internet Archive



Internet Archive



Définition

Internet Archive (archive.org²) 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/about³ (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/>

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



BIENVENUE sur la toile du MONDE .Globe Online.

En raison de la non-distribution du journal

L'accès au journal du 19 juin, daté 20, est gratuit

En direct du salon du Bourget du 14 au 22 juin.

Les résultats des [Grandes Ecoles](#)



Les élections [Législatives 1997](#)



La fête de la [Musique](#)

Le Monde

Site en Europe / Site aux Etats-Unis





La [Une](#) du Monde

Le [multimedia](#)

Le [journal](#) complet

Solutions Internet IBM

Les [services](#)

Les [événements](#)

Page www.lemonde.fr du 22 juin 1997

L'UTC

Université de Technologie de Compiègne

Établissement d'Enseignement supérieur à caractère technologique, UTC est située à Compiègne (France).

Président : François Peccoud

BP 609 - 02206 Compiègne Cedex
Tél : 03 44 23 44 22 - Fax : 03 44 23 43 00

- [Enseignement](#), la Recherche, les Relations Industrielles,
- la Recherche Scientifique, l'UTC en Action,
- le [réseaux des entreprises](#),
- la [formation](#),
- [services](#) dédiés, en particulier [CASAUTO](#)
- [services](#) de service informatique.

Voir les [services](#) [adresses](#) et la [ville](#) de Compiègne.

[Gestionnaire du serveur \(Webmaster\): 20 juin 1996](#)

UTC -- Université de Technologie de Compiègne

<http://www.utc.fr/>

Gestionnaire du serveur (Webmaster)

Responsable de la publication : Françoise Cornet

Responsable de service communication de l'UTC:

Tel: 33 (0)3 44 23 43 42

E-mail: Francoise.Cornet@utc.fr

Contact technique : Michel Vaysse


Service Informatique

Tel: 33 (0)3 44 23 43 44

Fax: 33 (0)3 44 23 46 77

E-mail: im@utc.fr

Université de Technologie de Compiègne
Compiègne, France



Les premiers sites web à l'UTC

Exercice

[solution n°3 p. 33]



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 ?

- de l'ordre d'une seconde
- de l'ordre de dix secondes
- de l'ordre d'une minute
- de l'ordre d'une heure

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



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)



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

Les cybercafés en France en 1995



« 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.37



« [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.37



Modem



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.



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)



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

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, Comuserve** 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).



-
- Historique FAI. <https://www.abonnement-adsl.biz/historique-fai/>
 - Histoire d'AOL et de quelques fournisseurs d'accès Français. <http://leteneur.com/histoirefai.htm>

Exercice

[solution n°4 p. 34]



Classer dans leur ordre de création les navigateurs suivants :

1. Safari
2. NSCA Mosaic
3. Opera
4. WorldWideWeb
5. Chrome
6. Internet Explorer
7. Firefox
8. Netscape Navigator

Réponse : _____

Des butineurs aux navigateurs web



Navigateur web (web browser)



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://framsoft.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



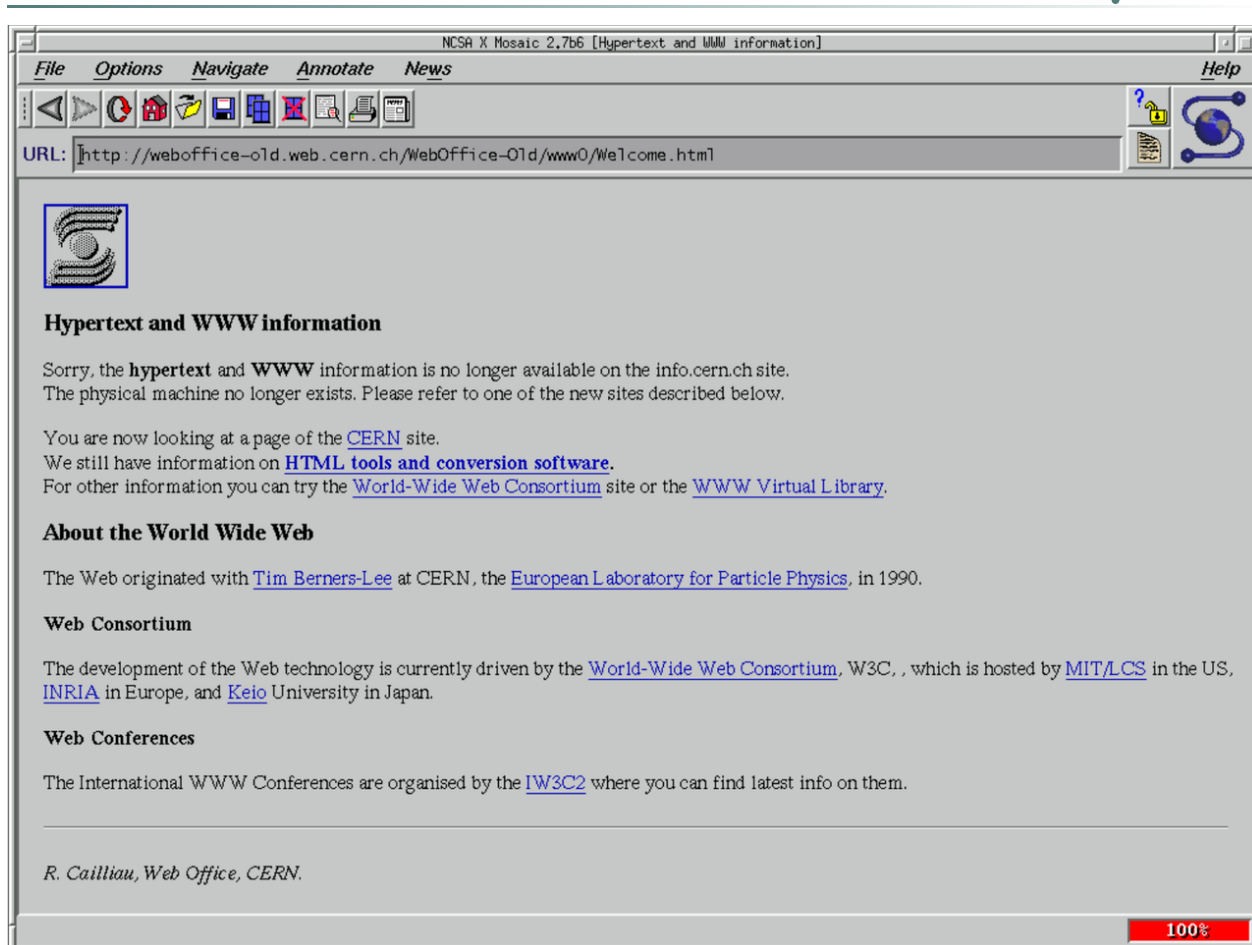
Le terme de butineur est utilisé en France jusque dans les années 2000.

« 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)

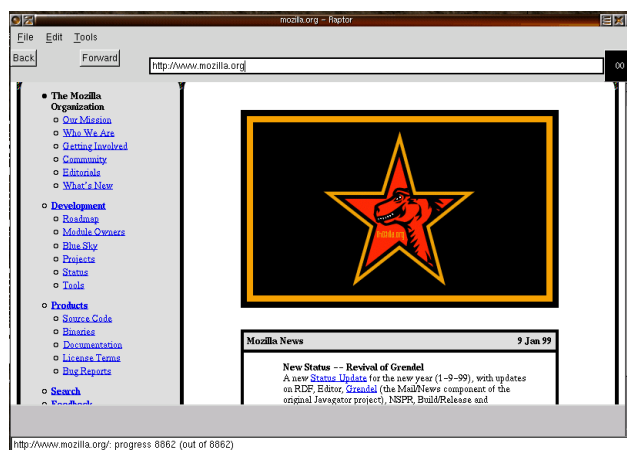


Les ancêtres

- **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.

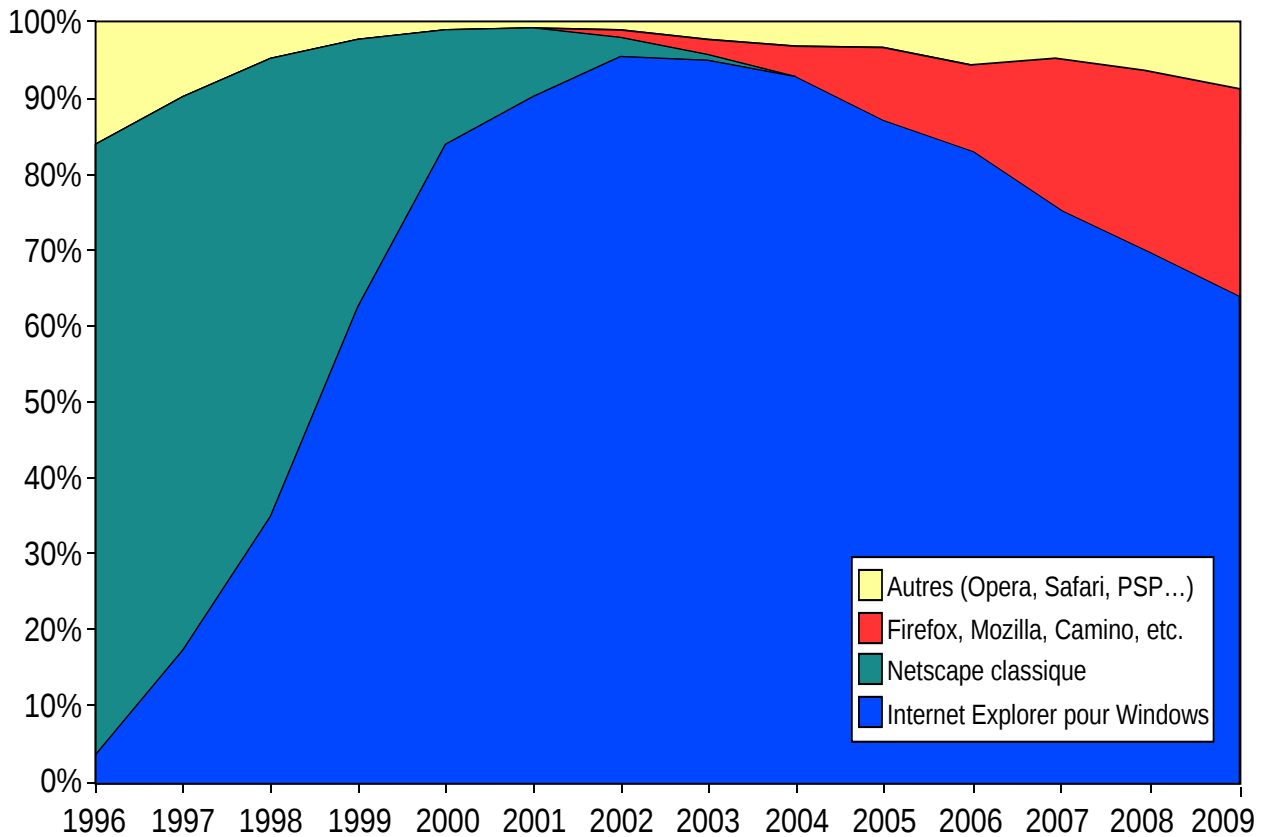


NCSA Mosaic (page de Robert Cailliau au Cern)



Mozilla (1999)

La guerre des navigateurs



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. 34]



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

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

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.37



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

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://framasoftware.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.37}



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 organisés 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 membres 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.36}.

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

Exercice

[solution n°6 p. 34]



Un moteur de recherche intègre nécessairement les fonctions suivantes :

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

Les moteurs de recherche et annuaires web



Avant le Web

1988, **WAIS** (Wide Area Information Servers) permet de se connecter (via Internet à partir 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



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.

NEW COOL RANDOM EXTRA HEAD LINES INFO ADD URL

Yahoo! Deutschland CLICK HERE TO VISIT THE STARS Yahoo! LOS ANGELES Weekly Picks

Search Options

Yellow Pages - People Search - City Maps -- News Headlines - Stock Quotes - Sports Scores

- **Arts** -- [Humanities](#), [Photography](#), [Architecture](#), ...
- **Business and Economy [Xtra!]** -- [Directory](#), [Investments](#), [Classifieds](#), ...
- **Computers and Internet [Xtra!]** -- [Internet](#), [WWW](#), [Software](#), [Multimedia](#), ...
- **Education** -- [Universities](#), [K-12](#), [Courses](#), ...
- **Entertainment [Xtra!]** -- [TV](#), [Movies](#), [Music](#), [Magazines](#), ...
- **Government** -- [Politics \[Xtra!\]](#), [Agencies](#), [Law](#), [Military](#), ...
- **Health [Xtra!]** -- [Medicine](#), [Drugs](#), [Diseases](#), [Fitness](#), ...
- **News [Xtra!]** -- [World \[Xtra!\]](#), [Daily](#), [Current Events](#), ...
- **Recreation and Sports [Xtra!]** -- [Sports](#), [Games](#), [Travel](#), [Autos](#), [Outdoors](#), ...
- **Reference** -- [Libraries](#), [Dictionaries](#), [Phone Numbers](#), ...
- **Regional** -- [Countries](#), [Regions](#), [U.S. States](#), ...
- **Science** -- [CS](#), [Biology](#), [Astronomy](#), [Engineering](#), ...
- **Social Science** -- [Anthropology](#), [Sociology](#), [Economics](#), ...
- **Society and Culture** -- [People](#), [Environment](#), [Religion](#), ...

Yahoo! New York - Yahoo! Shop - Yahoooligans!
Yahoo! Japan - Yahoo! Internet Life - Yahoo! San Francisco

Yahoo! en 1996 (depuis Internet Archive)

AltaVista : intégration de l'indexation automatique



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



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



« 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.37}



Exercice

[solution n°7 p. 35]



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

- Amazon
- Wikimedia Foundation (qui gère Wikipédia)
- Netscape Communications
- Iliad (maison mère de Free)

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.37*



Introduction en bourse (logique des start-up)



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 :

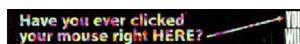
- le cours de l'action double le premier jour
- la valeur de l'entreprise est de 2 milliards de dollars
- soit plus de 20 fois son chiffre d'affaires de 85 millions
- malgré une perte de 6,6 millions

Publicité



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



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

Commerce en ligne



- **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. 35]



Exercice

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

- Environ 2.000
- Environ 20.000
- Environ 200.000
- 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 ?

- Environ 60.000
- Environ 600.000
- Environ 6.000.000
- Environ 60.000.000

Exercice

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

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.37}



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.



Check our catalog of merchandise on the [Ordering Info](#) page.
Reduce your bandwidth with [IUMA-lite](#) or if it's your first time try the [Guided Tour](#).

News Flash: Another IUMA band gets signed.

[EuPHoria](#) signs with **Six Degrees**, a new **Island Records** label. The album will be out early next year.

Enjoy our [artists](#) in real-time with 



© 1995 Internet Underground Music Archive - All rights reserved - IUMA 2.0

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

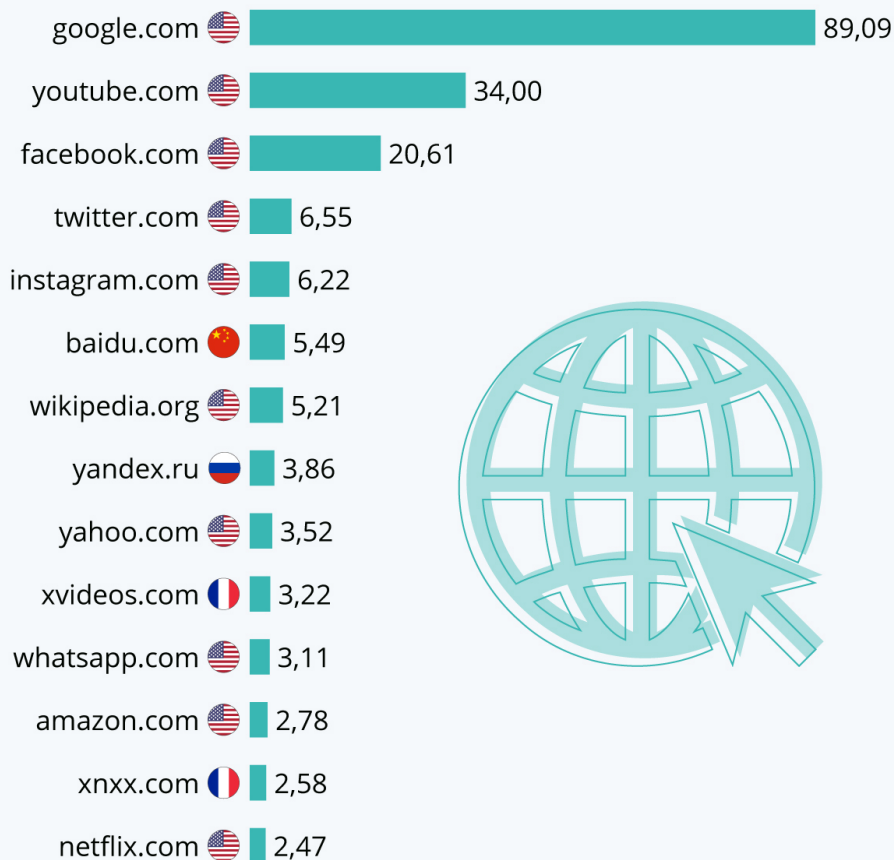
- **2004**, **Flickr** permet de diffuser gratuitement des photographies.
- **2005**, **YouTube** permet de partager des vidéos sur le Web, alors que le coût de l'hébergement web est encore élevé.



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

Les sites les plus visités

Sites internet les plus consultés dans le monde en novembre 2021, en milliards de visites

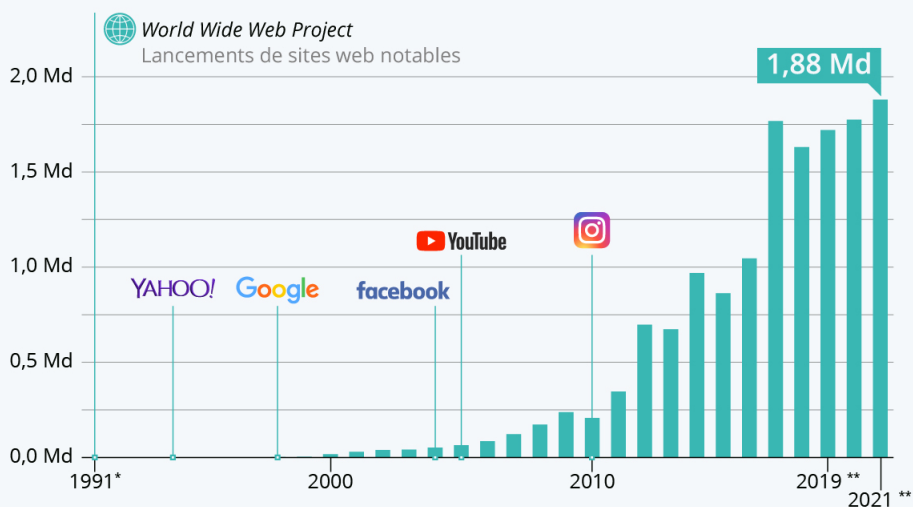


Source : SimilarWeb



Combien y a-t-il de sites Internet ?

Nombre de sites Internet en ligne de 1991 à 2021



* En date du 1er août 1991.

** Dernières données disponibles 2019 : 28 octobre, 2020 : 2 juin, 2021 : 6 août.

Source : Internet Live Stats



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. 7]

En 1990, le premier site web `info.cern.ch` voit le jour au CERN. Quelles technologies ont mises au point Tim Berners-Lee et Robert Cailliau pour y parvenir ?

- Le protocole HTTP
- Un navigateur web
- Un serveur web
- 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. 10]

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

- 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.
- 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.
- 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. 13]

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 ?

- de l'ordre d'une seconde
- de l'ordre de dix secondes
- de l'ordre d'une minute
- 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. 17]

Classer dans leur ordre de création les navigateurs suivants :

WorldWideWeb	NSCA Mosaic	Netscape Navigator	Opera	Internet Explorer	Firefox
Safari	Chrome				

- 🔍 *Des butineurs aux navigateurs web (1990-2008) (cf. p.18)*

NB : première version de Opera et Internet Explorer selon Wikipédia

- Opera : avril 1995
- IE : août 1995

Solution n°5

[exercice p. 21]

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

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

Solution n°6

[exercice p. 24]

Un moteur de recherche intègre nécessairement les fonctions suivantes :

- Un programme qui parcourt le Web pour référencer les pages existantes (*crawler*)
- Un programme qui décrit les pages référencées dans une base de données (*indexation*)
- Une interface qui permet à l'utilisateur d'exprimer une recherche (*formulaire de recherche*)
- Une présentation ordonnée des résultats de recherche (*pages des résultats*)
- 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

Solution n°7

[exercice p. 27]

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

- Amazon
Amazon ouvre en 1995, entre en bourse en 1997 et réalise ses premiers bénéfices en 2001.
- Wikimedia Foundation (qui gère Wikipédia)
La Wikimedia Foundation est une organisation à but non lucratif.
- 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)
- Iliad (maison mère de Free)
Free est créé en 1999, Iliad 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. 29]

Exercice

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

- Environ 2.000
- Environ 20.000
- Environ 200.000
- 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 ?

- Environ 60.000
- Environ 600.000
- Environ 6.000.000
- Environ 60.000.000

Exercice

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

2001

Abréviations



WCMS : (ou WMS) Web Content Management System

Bibliographie



[**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](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>.

[**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 v1.0⁴) — 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.org⁵

Répartition des internautes dans le monde en 2013 p. 5

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. 8

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

Page www.lemonde.fr du 22 juin 1997 p. 12

<https://web.archive.org/web/19970622152113/http://www.lemonde.fr/>

NCSA Mosaic (page de Robert Cailliau au Cern) p. 19

Universel - Transfert dans le Domaine Public - Selbst Erstellt · <https://commons.wikimedia.org/w/index.php?curid=75617889>

Mozilla (1999) p. 19

web.archive.org⁶

Évolution de la part respective des navigateurs entre 1996 et 2009 p. 20

Attribution - Partage dans les Mêmes Conditions

Yahoo! en 1996 (depuis Internet Archive) p. 25

Yahoo! home page · web.archive.org⁷

Wired.com invents the web banner ad (Oct. 27, 1994) p. 28

wired.com⁸

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>

7. <https://web.archive.org/web/19961017235908/http://www2.yahoo.com/>

8. <https://www.wired.com/2010/10/1027hotwired-banner-ads/>

Site de l'IUMA en 1996 (depuis Internet Archive) p. 30

Licence : Domaine Public - IUMA · <https://web.archive.org>

Nombre de sites Internet entre 1991 et 2021 p. 32

Statista / Live Stats · fr.statista.com⁹

Nombre de sites web entre 1991 et 2018 p. 32

InternetLiveStats¹⁰

⁹. <https://fr.statista.com/infographie/19063/nombre-de-sites-web-internet-dans-le-monde/>

¹⁰. <https://www.internetlvestats.com/>