

# INF344 – Données du Web

## Introduction : le Web, les navigateurs

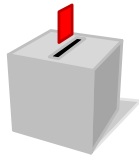
---

Antoine Amarilli



Le Web a été inventé...

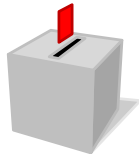
- **A:** À peu près en même temps qu'Internet
- **B:** 2 ans après Internet
- **C:** 5 ans après Internet
- **D:** Plus de 5 ans après Internet



# SONDAGE : Introduction du World Wide Web

Le Web a été inventé...

- **A:** À peu près en même temps qu'Internet
- **B:** 2 ans après Internet
- **C:** 5 ans après Internet
- **D: Plus de 5 ans après Internet**



- 1969** ARPANET, ancêtre d'Internet, est lancé
- 1974** Introduction de TCP
- 1990** Introduction du World Wide Web, HTTP, HTML
- 1994** Yahoo! est fondé
- 1995** Amazon.com, Ebay, AltaVista sont fondés
- 1998** Google est fondé
- 2001** Wikipédia est lancé

# SONDAGE : Nombre de domaines Internet

Combien y a-t-il de noms de domaines sur Internet?

- **A:** 3 millions
- **B:** 30 millions
- **C:** 300 millions
- **D:** 3 milliards



# SONDAGE : Nombre de domaines Internet

Combien y a-t-il de noms de domaines sur Internet?

- **A:** 3 millions
- **B:** 30 millions
- **C: 300 millions**
- **D:** 3 milliards



- Plus de 360 millions de **domaines**, dont 145 millions dans **.com**<sup>1</sup>
- **54%** du contenu est en **anglais**, **4%** en français<sup>2</sup>
- Google connaît plus d'un **trillion** ( $10^{12}$ ) d'URLs uniques<sup>3</sup>
  - Noter que le **même contenu** peut avoir plusieurs URLs différentes
  - Une large partie du Web n'est pas indexable : le **Web caché**

---

1. [https://www.verisign.com/en\\_US/domain-names/dnib/index.xhtml](https://www.verisign.com/en_US/domain-names/dnib/index.xhtml)

2. [https://w3techs.com/technologies/overview/content\\_language/all](https://w3techs.com/technologies/overview/content_language/all)

3. <https://googleblog.blogspot.fr/2008/07/we-knew-web-was-big.html>

# SONDAGE : Utilisateurs d'Internet

Quelle proportion de la population mondiale utilise Internet ?

- **A:** Moins d'un tiers
- **B:** Un peu moins de la moitié
- **C:** Un peu plus de la moitié
- **D:** Plus de deux tiers





# SONDAGE : Utilisateurs d'Internet

Quelle proportion de la population mondiale utilise Internet ?

- **A:** Moins d'un tiers
- **B:** Un peu moins de la moitié
- **C: Un peu plus de la moitié**
- **D:** Plus de deux tiers



- 54% de la population mondiale utilise Internet<sup>4</sup>
  - Disparité de **genre** : 48% des femmes et 58% des hommes
  - Disparité d'**âge** : 71% des gens ayant 15-24 ans<sup>5</sup>
- Pourtant, la connectivité **existe** :
  - 93% de la population a accès à un réseau mobile
  - 82% a accès à de la 4G

---

4. <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf>

5. <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf>

# Table des matières

Introduction

Navigateurs Web

Autres clients Web

URLs

# Comment les gens vont sur le Web ?

- On utilise un **navigateur Web**
- Dépend du **système d'exploitation sous-jacent**
- Fonctionne sur un **téléphone** ou sur un **ordinateur**

# Navigateurs Web historiques

**Mosaic** Premier navigateur graphique répandu, 1993-1997

- De 80% en 1994 à < 10% en 1996

# Navigateurs Web historiques

**Mosaic** Premier navigateur graphique répandu, 1993-1997

- De 80% en 1994 à < 10% en 1996

**Netscape** Lancé en 1994, basé sur Mosaic

- Propriétaire, usage non-commercial gratuit
- 80% d'usage en 1996, < 10% en 2001
- N'est plus maintenu depuis 2008

# Navigateurs Web historiques

**Mosaic** Premier navigateur graphique répandu, 1993-1997

- De 80% en 1994 à < 10% en 1996

**Netscape** Lancé en 1994, basé sur Mosaic

- Propriétaire, usage non-commercial gratuit
- 80% d'usage en 1996, < 10% en 2001
- N'est plus maintenu depuis 2008

**Internet Explorer** Lancé en 1995, basé sur Spyglass Mosaic

- Propriétaire, fourni avec Windows 95
- IE 6 lancé en 2001 atteint 80% du marché
- Mène au procès antitrust aux USA, 1998-2001

# Navigateurs Web historiques

**Mosaic** Premier navigateur graphique répandu, 1993-1997

- De 80% en 1994 à < 10% en 1996

**Netscape** Lancé en 1994, basé sur Mosaic

- Propriétaire, usage non-commercial gratuit
- 80% d'usage en 1996, < 10% en 2001
- N'est plus maintenu depuis 2008

**Internet Explorer** Lancé en 1995, basé sur Spyglass Mosaic

- Propriétaire, fourni avec Windows 95
- IE 6 lancé en 2001 atteint 80% du marché
- Mène au procès antitrust aux USA, 1998-2001

**Firefox** Lancé en 2002 (Netscape libéré en 1998)

- Libre et gratuit
- Navigation à onglets
- Attaque le monopole de IE 6



# Navigateurs Web actuels

**IE** Développement de **IE 7** annoncé en 2005, lancé en 2006.  
Meilleure **conformité aux standards**

**Firefox** Toujours **dans la course**

**Safari** Lancé en 2003, navigateur par défaut de **Mac OS X**  
(succédant à Netscape puis IE sous Mac OS)

**Opera** Lancé en **1996**, mais **payant** jusqu'en 2000 et financé par la **publicité** jusqu'en 2005. Maintenant **gratuit** mais propriétaire

**Chrome** Lancé en **2008** par **Google**. Propriétaire mais version libre également disponible

**Mobile** Important! **55%** d'usage d'après [gs.statcounter.com](http://gs.statcounter.com)  
• **Safari** (iOS), **Android browser** ou **Chrome** (Android)  
• Firefox Mobile, Blackberry, Opera, UC Browser...

Différentes **versions**, cf [browsershots.org](http://browsershots.org) pour la compatibilité

## SONDAGE : Parts de marché des navigateurs (1/3)

Quel navigateur est dominant actuellement?

- **A:** Internet Explorer / Edge
- **B:** Mozilla Firefox
- **C:** Google Chrome
- **D:** Apple Safari



## SONDAGE : Parts de marché des navigateurs (1/3)

Quel navigateur est dominant actuellement?

- **A:** Internet Explorer / Edge
- **B:** Mozilla Firefox
- **C: Google Chrome**
- **D:** Apple Safari



## SONDAGE : Parts de marché des navigateurs (2/3)

Quel est son principal concurrent ?

- **A:** Internet Explorer / Edge
- **B:** Mozilla Firefox
- **C:** Apple Safari
- **D:** Un navigateur plus obscur ?



## SONDAGE : Parts de marché des navigateurs (2/3)

Quel est son principal concurrent ?

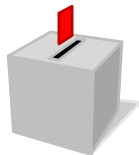
- **A:** Internet Explorer / Edge
- **B:** Mozilla Firefox
- **C: Apple Safari**
- **D:** Un navigateur plus obscur ?



## SONDAGE : Parts de marché des navigateurs (3/3)

Quel est la part de marché de Safari, le principal challenger ?

- **A:** 10%
- **B:** 20%
- **C:** 30%
- **D:** 40%



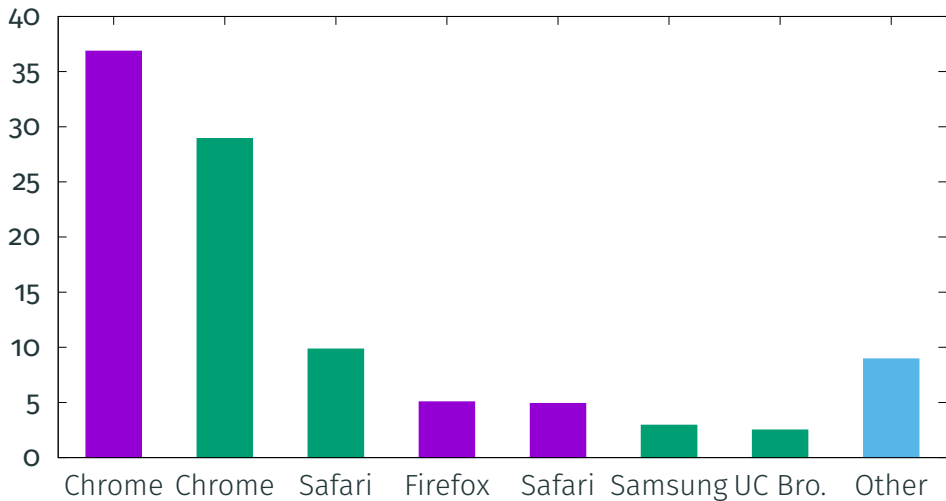
## SONDAGE : Parts de marché des navigateurs (3/3)

Quel est la part de marché de Safari, le principal challenger ?

- **A:** 10%
- **B: 20%**
- **C:** 30%
- **D:** 40%



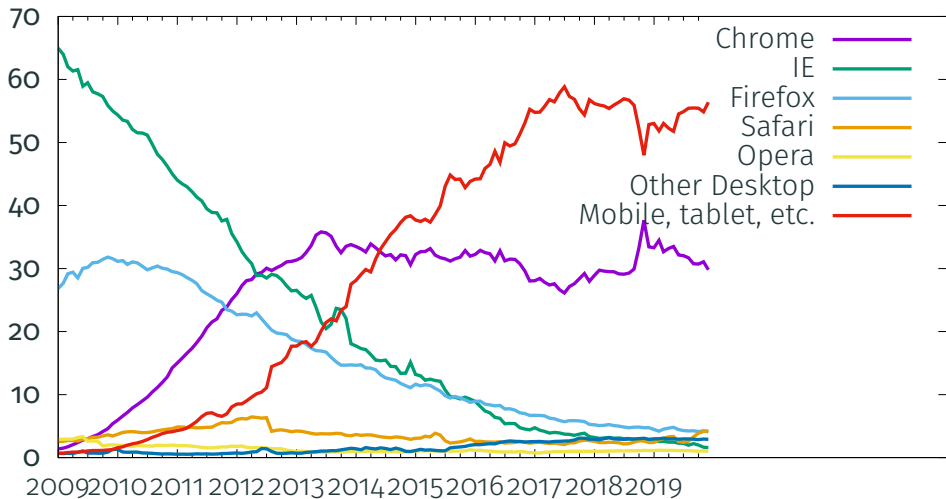
## Répartition récente



Source : [gs.statcounter.com](https://gs.statcounter.com) (décembre 2019)



# Évolution



Source : [gs.statcounter.com](https://gs.statcounter.com)

Quel proportion du trafic Web est sur mobile ?

- **A:** Moins d'un tiers
- **B:** Un peu moins de la moitié
- **C:** Un peu plus de la moitié
- **D:** Plus d'un tiers

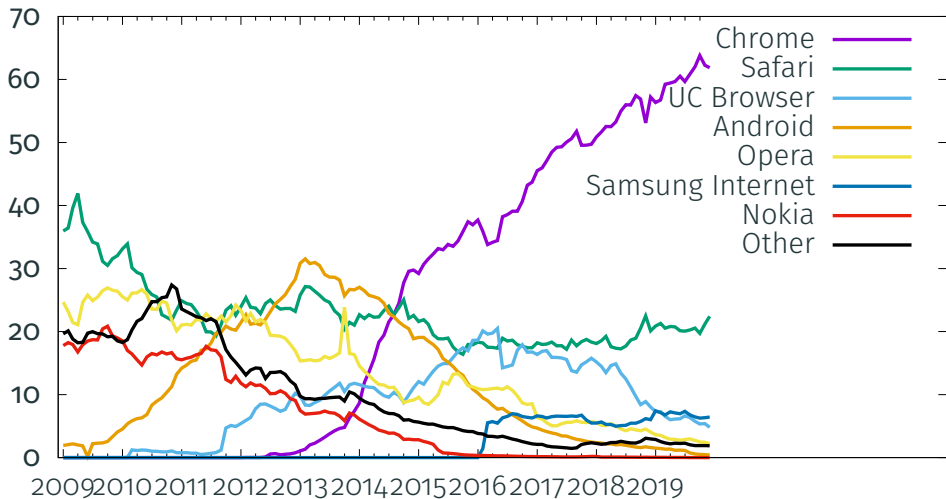


Quel proportion du trafic Web est sur mobile ?

- **A:** Moins d'un tiers
- **B:** Un peu moins de la moitié
- **C: Un peu plus de la moitié**
- **D:** Plus d'un tiers



# Évolution (mobile)



Source : [gs.statcounter.com](https://gs.statcounter.com)

# Moteurs de rendu

**Firefox** Gecko, et Servo (en cours de développement, utilise Rust)

**Safari** moteur Webkit

**Chrome** Blink (fork de Webkit, avril 2013)

**IE** D'abord Trident, puis EdgeHTML, puis Chromium depuis janvier 2020

**Opera** D'abord Presto, puis Blink

**Autres** Dillo, KHTML, autres moteurs exotiques, vieux, minimalistes...

# Résumé et perspectives

- **Domination** de Webkit/Blink et Chrome/Chromium
- Seuls concurrents : **Safari** (surtout sur mobile) et **Firefox**
- Blink est du **logiciel open-source** mais **contrôlé par Google**
- Beaucoup d'**autres navigateurs** utilisent ce moteur :
  - Certains **minimalistes**, comme `uzbl`
  - Des navigateurs **innovants** : Vivaldi, Brave

# Nouvelles questions autour des navigateurs (1)

- Écosystème d'**extensions** : Chrome Web Store, Mozilla Store
  - Firefox : **signature obligatoire** des extensions, et **compatibilité**

# Nouvelles questions autour des navigateurs (1)

- Écosystème d'**extensions** : Chrome Web Store, Mozilla Store
  - Firefox : **signature obligatoire** des extensions, et **compatibilité**
- **Bloqueurs de pub** (27% des gens<sup>6</sup>), jeu du chat et de la souris
  - Extension **uBlock Origin** basée sur **Easylist**  
`https://easylist.to/easylist/easylist.txt`
  - Aussi : **bloqueurs Javascript**, e.g., uMatrix, NoScript, etc.

---

6. <https://www.statista.com/topics/3201/ad-blocking/>



## Nouvelles questions autour des navigateurs (2)

- Filtrage des **robots** : robots exclusion standard, CAPTCHAs
  - **reCAPTCHA** : fait à présent faire du travail bénévole pour Google

## Nouvelles questions autour des navigateurs (2)

- Filtrage des **robots** : robots exclusion standard, CAPTCHAs
  - **reCAPTCHA** : fait à présent faire du travail bénévole pour Google
- **Sécurité** : isolation des sites Web, un processus par site
  - Disponible dans Chrome, en phase de tests dans Firefox (Project Fission)

## Nouvelles questions autour des navigateurs (2)

- Filtrage des **robots** : robots exclusion standard, CAPTCHAs
  - **reCAPTCHA** : fait à présent faire du travail bénévole pour Google
- **Sécurité** : isolation des sites Web, un processus par site
  - Disponible dans Chrome, en phase de tests dans Firefox (Project Fission)
- **Prise d'empreinte** des navigateurs Web :  
<https://panopticlick.eff.org/>

## Nouvelles questions autour des navigateurs (2)

- Filtrage des **robots** : robots exclusion standard, CAPTCHAs
  - **reCAPTCHA** : fait à présent faire du travail bénévole pour Google
- **Sécurité** : isolation des sites Web, un processus par site
  - Disponible dans Chrome, en phase de tests dans Firefox (Project Fission)
- **Prise d'empreinte** des navigateurs Web :  
<https://panopticlick.eff.org/>
- Minage de **cryptomonnaies** dans le navigateur : CoinHive (mais fermé en avril 2019)

## Nouvelles questions autour des navigateurs (2)

- Filtrage des **robots** : robots exclusion standard, CAPTCHAs
  - **reCAPTCHA** : fait à présent faire du travail bénévole pour Google
- **Sécurité** : isolation des sites Web, un processus par site
  - Disponible dans Chrome, en phase de tests dans Firefox (Project Fission)
- **Prise d'empreinte** des navigateurs Web :  
<https://panopticlick.eff.org/>
- Minage de **cryptomonnaies** dans le navigateur : CoinHive (mais fermé en avril 2019)
- **Tor** et **services cachés** sur Tor

# Table des matières

Introduction

Navigateurs Web

Autres clients Web

URLs

# Navigateurs textuels

```
sampi:~$ w3m 'http://en.wikipedia.org' (14:24:59)
Télécom ParisTech

From Wikipedia, the free encyclopedia
(Redirected from Telecom ParisTech)
Jump to: navigation, search
"ENST" redirects here. For the airport with this ICAO airport code, see Sandnessjøen Airport, Stokka.

Crystal Clear This article may need to be rewritten entirely to comply with Wikipedia's quality
app kedit.svg standards. You can help. The discussion page may contain suggestions. (May 2009)

Coordinates: 48°49′35″N 2°20′47″E﻿ / ﻿48.82639°N 2.34639°E﻿ / 48.82639; 2.34639

Télécom ParisTech

Logo telecomparisTech.png

Motto L'École au cœur de la Société de l'Information

Established 1878

Type French Grande École

President Yves Poilane

Admin. staff 340 (2006)

Students 1249 (2006)

Location Paris, France

Campus Paris, Sophia Antipolis

⌂ ↑ ↓ Viewing <Télécom ParisTech - Wikipedia, the free encyclopedia>
```

Aussi : **navigateurs auditifs** pour les malvoyants...

Nombreux **programmes automatisés** sur le Web :

- **Crawlers** pour moteurs de recherche (indexation) :  
Googlebot (Google), bingbot (Bing)...
- **Crawlers** plus obscurs : meanpathbot, ahrefsbot, ezoomsbot, Career-X bot, msnbot, news bot, seznambot, ChangeDetection bot? ...
- **Lecteurs RSS** et agrégateurs
- **Harvesters** d'adresses email
- Utilisateurs d'**API**



# Table des matières

Introduction

Navigateurs Web

Autres clients Web

URLs

- Structure générale :

`https://en.wikipedia.org/wiki/Telecom_Paris#History`  
protocole      machine                      chemin                      fragment

- Uniform Resource Locator
- Identifie une **ressource** sur le Web
- Barre d'adresses

- Matériel de cours inspiré de notes par Pierre Senellart