

# 1. Internet

**Licence professionnelle  
Université de Caen**

**Jean Fromentin**

<mailto:jfroment@info.unicaen.fr>

<http://www.info.unicaen.fr/~jfroment>

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;
  - le courrier électronique (email, **courriel**) ;

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;
  - le courrier électronique (email, **courriel**) ;
  - messagerie instantanée (chat, tchat, **clavardage**) : MSN, IRC, ... ;

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;
  - le courrier électronique (email, **courriel**) ;
  - messagerie instantanée (chat, tchat, **clavardage**) : MSN, IRC, ... ;
  - le transfert de fichiers : FTP, Peer-to-Peer, ... ;

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;
  - le courrier électronique (email, **courriel**) ;
  - messagerie instantanée (chat, tchat, **clavardage**) : MSN, IRC, ... ;
  - le transfert de fichiers : FTP, Peer-to-Peer, ... ;
  - etc.

## Internet

- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;
  - le courrier électronique (email, **courriel**) ;
  - messagerie instantanée (chat, tchat, **clavardage**) : MSN, IRC, ... ;
  - le transfert de fichiers : FTP, Peer-to-Peer, ... ;
  - etc.
- Ses utilisateurs sont désignés par le néologisme **internaute**.

## Internet

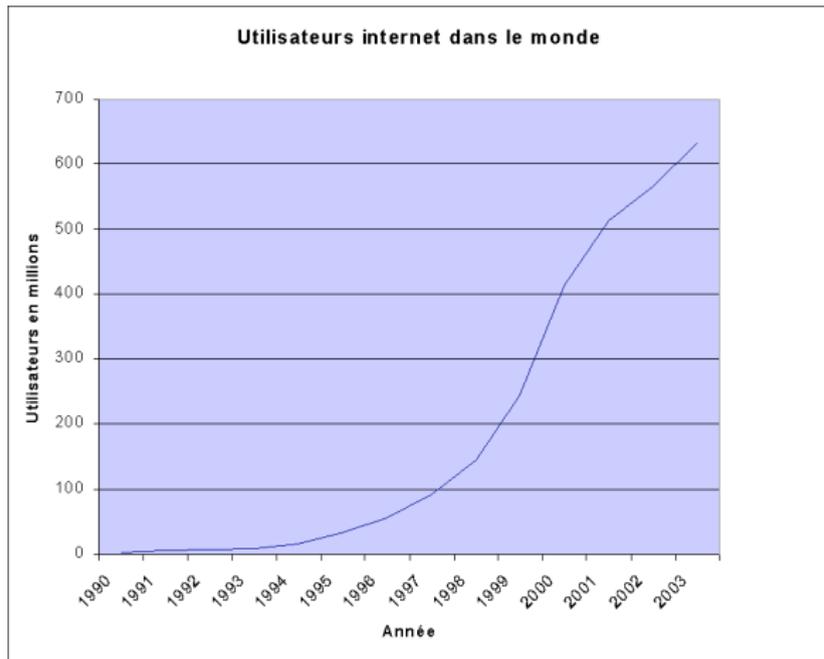
- **Internet** est le réseau informatique mondial qui rend accessible au public des services très variés comme :
  - le World Wide Web (WWW, Web, **toile d'araignée mondiale**) ;
  - le courrier électronique (email, **courriel**) ;
  - messagerie instantanée (chat, tchat, **clavardage**) : MSN, IRC, ... ;
  - le transfert de fichiers : FTP, Peer-to-Peer, ... ;
  - etc.
- Ses utilisateurs sont désignés par le néologisme **internaute**.
- Le mot **Internet** désigne un réseau informatique constitué de par l'interconnexion de plusieurs réseaux (LAN) : pour cette raison **Internet** est aussi appelé **réseau des réseaux**.

## Internet

- Internet ayant été popularisé par l'apparition du World Wide Web (vers 1993), les deux sont parfois confondus par le public non averti.

## Internet

- Internet ayant été popularisé par l'apparition du World Wide Web (vers 1993), les deux sont parfois confondus par le public non averti.



## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).

## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).
- Il a pour vocation de transférer les courriels d'un serveur à un autre.

## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).
- Il a pour vocation de transférer les courriels d'un serveur à un autre.
- L'utilisateur n'est jamais en contact direct avec un **serveur mail** mais utilise

## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).
- Il a pour vocation de transférer les courriels d'un serveur à un autre.
- L'utilisateur n'est jamais en contact direct avec un **serveur mail** mais utilise soit un **client de messagerie** (comme **Thunderbird**, **Outlook**),

## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).
- Il a pour vocation de transférer les courriels d'un serveur à un autre.
- L'utilisateur n'est jamais en contact direct avec un **serveur mail** mais utilise soit un **client de messagerie** (comme **Thunderbird**, **Outlook**), soit un **courrielleur web** ou **webmail**,

## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).
- Il a pour vocation de transférer les courriels d'un serveur à un autre.
- L'utilisateur n'est jamais en contact direct avec un **serveur mail** mais utilise soit un **client de messagerie** (comme **Thunderbird**, **Outlook**), soit un **courrielleur web** ou **webmail**, qui se charge de contacter le serveur pour envoyer ou recevoir des messages.

## Serveur mail

- Un **serveur mail** ou **serveur de messagerie électronique** est un logiciel serveur de **courriel**).
- Il a pour vocation de transférer les courriels d'un serveur à un autre.
- L'utilisateur n'est jamais en contact direct avec un **serveur mail** mais utilise soit un **client de messagerie** (comme **Thunderbird**, **Outlook**), soit un **courrielleur web** ou **webmail**, qui se charge de contacter le serveur pour envoyer ou recevoir des messages.
- La plupart des serveurs de messagerie ont les fonctions d'envoi et de réception, mais elles sont indépendantes et peuvent être dissociées physiquement en utilisant plusieurs serveurs.

## Serveur de messagerie électronique : protocole d'envoi

- Entre l'utilisateur et le serveur, l'envoi d'un courriel se déroule généralement via le protocole **SMTP**.

## Serveur de messagerie électronique : protocole d'envoi

- Entre l'utilisateur et le serveur, l'envoi d'un courriel se déroule généralement via le protocole **SMTP**.
- Puis c'est au serveur d'envoyer le message au serveur du destinataire, cette fonction est appelée **Mail Transfer Agent** en anglais, ou **MTA**.

## Serveur de messagerie électronique : protocole de réception

- Le serveur reçoit le message du serveur de l'expéditeur, il doit donc gérer des problèmes comme un disque plein ou une corruption de la boîte aux lettres et signaler au serveur expéditeur toute erreur dans la délivrance.

## Serveur de messagerie électronique : protocole de réception

- Le serveur reçoit le message du serveur de l'expéditeur, il doit donc gérer des problèmes comme un disque plein ou une corruption de la boîte aux lettres et signaler au serveur expéditeur toute erreur dans la délivrance.
- Il communique avec ce dernier par l'intermédiaire des canaux d'entrée-sortie standard ou par un protocole spécialisé comme **LMTP** (**Local Mail Transfer Protocol**).

## Serveur de messagerie électronique : protocole de réception

- Le serveur reçoit le message du serveur de l'expéditeur, il doit donc gérer des problèmes comme un disque plein ou une corruption de la boîte aux lettres et signaler au serveur expéditeur toute erreur dans la délivrance.
- Il communique avec ce dernier par l'intermédiaire des canaux d'entrée-sortie standard ou par un protocole spécialisé comme **LMTP** (**Local Mail Transfer Protocol**). Cette fonction de réception est appelée **Mail Delivery Agent** en anglais, ou **MDA**.

## Serveur de messagerie électronique : protocole de réception

- Le serveur reçoit le message du serveur de l'expéditeur, il doit donc gérer des problèmes comme un disque plein ou une corruption de la boîte aux lettres et signaler au serveur expéditeur toute erreur dans la délivrance.
- Il communique avec ce dernier par l'intermédiaire des canaux d'entrée-sortie standard ou par un protocole spécialisé comme **LMTP** (**Local Mail Transfer Protocol**). Cette fonction de réception est appelée **Mail Delivery Agent** en anglais, ou **MDA**.
- Finalement, lorsque le destinataire final désire accéder à ses messages, il lance une requête au serveur qui transmet les messages reçus généralement via le protocole **POP3** ou **IMAP**.

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.
  - 1990 : Le premier serveur Web est `nxoc01.cern.ch` ;

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.
  - 1990 : Le premier serveur Web est `nxoc01.cern.ch` ;  
Le premier navigateur web, appelé **Nexus**, est créé.

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.
  - 1990 : Le premier serveur Web est `nxoc01.cern.ch` ;  
Le premier navigateur web, appelé **Nexus**, est créé.  
La première page web est créée ;

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.
  - 1990 : Le premier serveur Web est `nxoc01.cern.ch` ;  
Le premier navigateur web, appelé **Nexus**, est créé.  
La première page web est créée ;
  - 1991 : Tim Berners-Lee rend le projet public ;

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.
  - 1990 : Le premier serveur Web est `nxoc01.cern.ch` ;  
Le premier navigateur web, appelé **Nexus**, est créé.  
La première page web est créée ;
  - 1991 : Tim Berners-Lee rend le projet public ;
  - 1993 : Le CERN met les logiciels du Web dans le domaine public.

## World Wide Web

- Le **World Wide Web** ou **Web** est un système hypertexte public fonctionnant sur **Internet** et qui permet de consulter, avec un **navigateur**, des **pages** mises en lignes dans des **sites**.
- L'image de la toile d'araignée provient des hyperliens reliant les pages web entre elles.
- Bref historique :
  - 1989 : Tim Berners-Lee, du CERN, propose un système hypertexte améliorant la diffusion des informations internes.
  - 1990 : Le premier serveur Web est `nxoc01.cern.ch` ;  
Le premier navigateur web, appelé **Nexus**, est créé.  
La première page web est créée ;
  - 1991 : Tim Berners-Lee rend le projet public ;
  - 1993 : Le CERN met les logiciels du Web dans le domaine public.

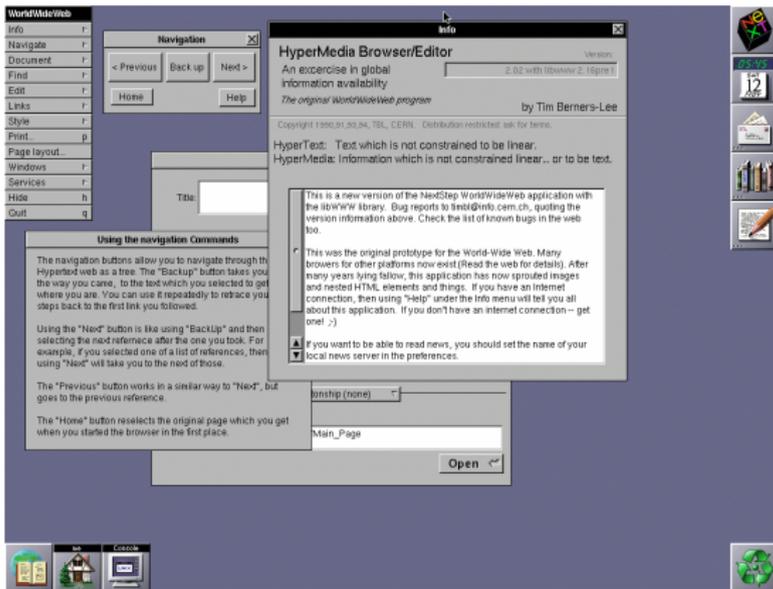


## Navigateurs

- Un **navigateur Web** (ou **butineur**, **brouteur**) est un logiciel conçu pour consulter le World Wide Web.

## Navigateurs

- Un **navigateur Web** (ou **butineur**, **brouteur**) est un logiciel conçu pour consulter le World Wide Web. Le terme navigateur web (ou navigateur Internet) est inspiré de **Netscape Navigator**.



# La guerre des navigateurs

Principaux navigateurs :

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

Moteur de rendu :

## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

Moteur de rendu :

- Trident (I.E.) ;

## La guerre des navigateurs

### Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

### Moteur de rendu :

- Trident (I.E.) ;
- Gecko (Mozilla, Firefox, Nets) ;

## La guerre des navigateurs

### Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

### Moteur de rendu :

- Trident (I.E.) ;
- Gecko (Mozilla, Firefox, Nets) ;
- Presto (Opera) ;

## La guerre des navigateurs

### Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

### Moteur de rendu :

- Trident (I.E.) ;
- Gecko (Mozilla, Firefox, Nets) ;
- Presto (Opera) ;
- KHTML (Safari ; Google Chrome) ;

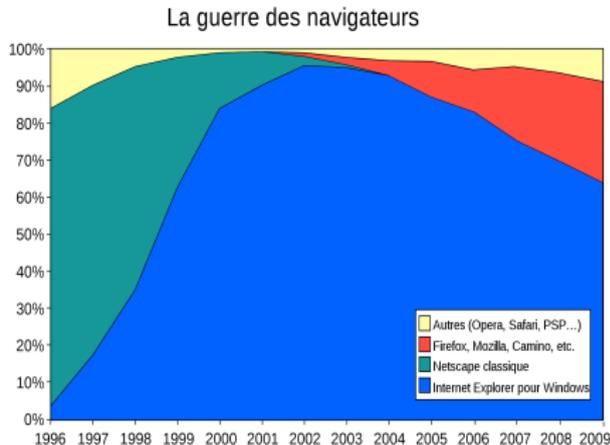
## La guerre des navigateurs

Principaux navigateurs :

- Netscape Navigator (1994) ;
- Internet Explorer (1995) ;
- Opera (1997) ;
- Mozilla (2002) ;
- Safari (2003) ;
- Firefox (2004) ;
- Google Chrome (2008).

Moteur de rendu :

- Trident (I.E.) ;
- Gecko (Mozilla, Firefox, Nets) ;
- Presto (Opera) ;
- KHTML (Safari ; Google Chrome) ;



## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format  
**HTML : Hyper Text Markup Language**

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

**HTML : Hyper Text Markup Language**

stockés sur un serveur, en utilisant le protocole

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

**HTML** : Hyper Text Markup Language

stockés sur un serveur, en utilisant le protocole

**HTTP** : Hyper Text Transfert Protocole.

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

**HTML** : **H**yper **T**ext **M**arkup **L**anguage

stockés sur un serveur, en utilisant le protocole

**HTTP** : **H**yper **T**ext **T**ransfert **P**rotocole.

- Les documents sont accessibles par une

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

**HTML** : **H**yper **T**ext **M**arkup **L**anguage

stockés sur un serveur, en utilisant le protocole

**HTTP** : **H**yper **T**ext **T**ransfert **P**rotocole.

- Les documents sont accessibles par une

**URL** : **U**niform **R**essource **L**ocator

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

**HTML** : Hyper Text Markup Language

stockés sur un serveur, en utilisant le protocole

**HTTP** : Hyper Text Transfert Protocole.

- Les documents sont accessibles par une

**URL** : Uniform Ressource Locator

comportant le nom du serveur **http** contenant le document, le chemin d'accès au document et le nom de celui-ci.

## World Wide Web : Technologies

- Le **World Wide Web** permet d'accéder à des documents au format

**HTML** : **H**yper **T**ext **M**arkup **L**anguage

stockés sur un serveur, en utilisant le protocole

**HTTP** : **H**yper **T**ext **T**ransfert **P**rotocole.

- Les documents sont accessibles par une

**URL** : **U**niform **R**essource **L**ocator

comportant le nom du serveur **http** contenant le document, le chemin d'accès au document et le nom de celui-ci.

- Exemple : <http://www.info.unicaen.fr/~jfroment/index.html>

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;`

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;` ;
  - 2. `//` pour les protocoles nécessitant un chemin d'accès ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : **http, https ftp, mailto** ;
    - 1.2. Le caractère de séparation : **;**
  - 2. **//** pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;` ;
  - 2. `//` pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de
    - 3.1. nom d'utilisateur ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;` ;
  - 2. `//` pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de
    - 3.1. nom d'utilisateur ;
    - 3.2. Le caractère : puis le mot de passe (facultatif) ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;`
  - 2. `//` pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de
    - 3.1. nom d'utilisateur ;
    - 3.2. Le caractère : puis le mot de passe (facultatif) ;
    - 3.3. Le caractère `@` ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : **http, https ftp, mailto** ;
    - 1.2. Le caractère de séparation : **;**
  - 2. **//** pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de
    - 3.1. nom d'utilisateur ;
    - 3.2. Le caractère : puis le mot de passe (facultatif) ;
    - 3.3. Le caractère **@** ;
  - 4. Nom de domaine du service ou son adresse IP : **www.google.com** ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;` ;
  - 2. `//` pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de
    - 3.1. nom d'utilisateur ;
    - 3.2. Le caractère : `:` puis le mot de passe (facultatif) ;
    - 3.3. Le caractère `@` ;
  - 4. Nom de domaine du service ou son adresse IP : `www.google.com` ;
  - 5. Le caractère `:` puis un numéro de port (facultatif) ;

## URL

- Une **URL** permet d'indiquer comment accéder à une ressource d'**Internet** indépendamment du contexte où elle est utilisée.
- La syntaxe générale est la suivante :
  - 1. Le schéma de représentation (normalement obligatoire) constitué de
    - 1.1 Protocole de communication : `http`, `https`, `ftp`, `mailto` ;
    - 1.2. Le caractère de séparation : `;` ;
  - 2. `//` pour les protocoles nécessitant un chemin d'accès ;
  - 3. Identifiant (**facultatif et déconseillé**) constitué de
    - 3.1. nom d'utilisateur ;
    - 3.2. Le caractère `:` puis le mot de passe (facultatif) ;
    - 3.3. Le caractère `@` ;
  - 4. Nom de domaine du service ou son adresse IP : `www.google.com` ;
  - 5. Le caractère `:` puis un numéro de port (facultatif) ;
  - 6. Nom complet de la ressource à demander une fois connecté.



## Exemples d'URLs

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

`http://users.info.unicaen.fr/~jfroment/enseignements.html`

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

`http://users.info.unicaen.fr/~jfroment/enseignements.html`

- 3. URL d'un fichier sur un site FTP :

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

`http://users.info.unicaen.fr/~jfroment/enseignements.html`

- 3. URL d'un fichier sur un site FTP :

`ftp://cdimage.ubuntu.com/cdimage/dvd/20090930/  
karmic-dvd-amd64.iso`

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

`http://users.info.unicaen.fr/~jfroment/enseignements.html`

- 3. URL d'un fichier sur un site FTP :

`ftp://cdimage.ubuntu.com/cdimage/dvd/20090930/  
karmic-dvd-amd64.iso`

- 4. URL d'un lien (mailto) vers une adresse courriel :

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

`http://users.info.unicaen.fr/~jfroment/enseignements.html`

- 3. URL d'un fichier sur un site FTP :

`ftp://cdimage.ubuntu.com/cdimage/dvd/20090930/  
karmic-dvd-amd64.iso`

- 4. URL d'un lien (mailto) vers une adresse courriel :

`mailto:Jean.Fromentin@info.unicaen.fr`

## Exemples d'URLs

- 1. URL du département informatique de l'université de Caen :

`http://www.info.unicaen.fr`

- 2. URL vers ma page web enseignements :

`http://users.info.unicaen.fr/~jfroment/enseignements.html`

- 3. URL d'un fichier sur un site FTP :

`ftp://cdimage.ubuntu.com/cdimage/dvd/20090930/karmic-dvd-amd64.iso`

- 4. URL d'un lien (mailto) vers une adresse courriel :

`mailto:Jean.Fromentin@info.unicaen.fr`

- **Attention** : les URLs sont sensibles à la casse (sauf le nom de domaine).

# URL et TLD

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - **com** : commerce ;
    - **net** : gestionnaire de réseaux ;

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...
  - 2. Les **ccTLD** (country code) :

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...
  - 2. Les **ccTLD** (country code) :
    - fr : pour la France ;

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...
  - 2. Les **ccTLD** (country code) :
    - fr : pour la France ;
    - us : pour l'Amérique ;

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...
  - 2. Les **ccTLD** (country code) :
    - fr : pour la France ;
    - us : pour l'Amérique ;
    - uk : pour la Grande Bretagne

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
  
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...
  - 2. Les **ccTLD** (country code) :
    - fr : pour la France ;
    - us : pour l'Amérique ;
    - uk : pour la Grande Bretagne
    - ...

## URL et TLD

- Dans un nom de domaine, le label le dernier point est appelé **TLD**, pour **Top level domain**.
- On distingue les
  - 1. Les **gTLD** (generic) :
    - com : commerce ;
    - net : gestionnaire de réseaux ;
    - org : organisations à but non commercial ;
    - edu : enseignement supérieur américain ;
    - ...
  - 2. Les **ccTLD** (country code) :
    - fr : pour la France ;
    - us : pour l'Amérique ;
    - uk : pour la Grande Bretagne
    - ...
  - ...

## URL : Exercices

- Donner le nom de domaine et son TLD des URLs suivantes :
  - 1. <http://www.info.unicaen.fr>
  - 2. <http://fr.wikipedia.org/wiki/.org>
  - 3. <http://www.info.unicaen.fr/index.php>
  - 4. <ftp://cdimage.ubuntu.com/cdimage/dvd/20090930/karmic-dvd-amd64.iso>
  - 5. <mailto:Jean.Fromentin@info.unicaen.fr>

## URL et TLD : suite

- L'**ICANN** (Internet **C**orporation for **A**ssigned **N**ames and **N**umbers) valide au niveau mondial le nom et la nature des **TLD**.

## URL et TLD : suite

- L'**ICANN** (Internet **C**orporation for **A**ssigned **N**ames and **N**umbers) valide au niveau mondial le nom et la nature des **TLD**.
- Les **registres** (registries) exploitent techniquement les **TLD**.

## URL et TLD : suite

- L'**ICANN** (Internet **C**orporation for **A**ssigned **N**ames and **N**umbers) valide au niveau mondial le nom et la nature des **TLD**.
- Les **registres** (registries) exploitent techniquement les **TLD**.
- Les **registrars** vendent les **TLD**. Tout concepteur de site doit s'adresser à un registrar. Ils reversent une partie de ce qu'ils perçoivent au registre.

## URL et TLD : suite

- L'**ICANN** (Internet Corporation for Assigned Names and Numbers) valide au niveau mondial le nom et la nature des **TLD**.
- Les **registres** (registries) exploitent techniquement les **TLD**.
- Les **registrars** vendent les **TLD**. Tout concepteur de site doit s'adresser à un registrar. Ils reversent une partie de ce qu'ils perçoivent au registre.
- Pour retrouver des informations sur le propriétaire d'un site, utiliser le service **whois** du TLD concerné : l'**AFNIC** pour les **.fr** le **NIC** pour **.com**, **.us**, **.net**, **.org**, ...

## URL et TLD : suite

- L'**ICANN** (Internet Corporation for Assigned Names and Numbers) valide au niveau mondial le nom et la nature des **TLD**.
- Les **registres** (registries) exploitent techniquement les **TLD**.
- Les **registrars** vendent les **TLD**. Tout concepteur de site doit s'adresser à un registrar. Ils reversent une partie de ce qu'ils perçoivent au registre.
- Pour retrouver des informations sur le propriétaire d'un site, utiliser le service **whois** du TLD concerné : l'**AFNIC** pour les **.fr** le **NIC** pour **.com**, **.us**, **.net**, **.org**, ...
- Ex. : – <http://www.afnic.fr/outils/whois> pour **.fr** ;

## URL et TLD : suite

- L'**ICANN** (Internet Corporation for Assigned Names and Numbers) valide au niveau mondial le nom et la nature des **TLD**.
- Les **registres** (registries) exploitent techniquement les **TLD**.
- Les **registrars** vendent les **TLD**. Tout concepteur de site doit s'adresser à un registrar. Ils reversent une partie de ce qu'ils perçoivent au registre.
- Pour retrouver des informations sur le propriétaire d'un site, utiliser le service **whois** du TLD concerné : l'**AFNIC** pour les **.fr** le **NIC** pour **.com**, **.us**, **.net**, **.org**, ...
- Ex. : – <http://www.afnic.fr/outils/whois> pour **.fr** ;  
– <http://www.pir.org/> pour **.org** ;

## URL et TLD : suite

- L'**ICANN** (Internet Corporation for Assigned Names and Numbers) valide au niveau mondial le nom et la nature des **TLD**.
- Les **registres** (registries) exploitent techniquement les **TLD**.
- Les **registrars** vendent les **TLD**. Tout concepteur de site doit s'adresser à un registrar. Ils reversent une partie de ce qu'ils perçoivent au registre.
- Pour retrouver des informations sur le propriétaire d'un site, utiliser le service **whois** du TLD concerné : l'**AFNIC** pour les **.fr** le **NIC** pour **.com**, **.us**, **.net**, **.org**, ...
- Ex. : – <http://www.afnic.fr/outils/whois> pour **.fr** ;  
– <http://www.pir.org/> pour **.org** ;  
– <http://registrar.verisign-grs.com/whois/> pour **.com**.

Internet

○○

Serveur mail

○○○

World Wide Web

○○○○○

URL

○○○○

Outils de recherche sur Internet

●○○○○○

- Il y a principalement trois outils de recherche de page Webs :

- Il y a principalement trois outils de recherche de page Webs :
  - 1. Les annuaires ;

- Il y a principalement trois outils de recherche de page Webs :
  - 1. Les annuaires ;
  - 2. Les sélections ;

- Il y a principalement trois outils de recherche de page Webs :
  - 1. Les annuaires ;
  - 2. Les sélections ;
  - 3. Les moteurs de recherche.

- Il y a principalement trois outils de recherche de page Webs :
  - 1. Les annuaires ;
  - 2. Les sélections ;
  - 3. Les moteurs de recherche.
- Pour la recherche d'information, on peut utiliser :

- Il y a principalement trois outils de recherche de page Webs :
  - 1. Les annuaires ;
  - 2. Les sélections ;
  - 3. Les moteurs de recherche.
- Pour la recherche d'information, on peut utiliser :
  - 1. Les forums ;

- Il y a principalement trois outils de recherche de page Webs :
  - 1. Les annuaires ;
  - 2. Les sélections ;
  - 3. Les moteurs de recherche.
- Pour la recherche d'information, on peut utiliser :
  - 1. Les forums ;
  - 2. Les wikis (le plus célèbre est Wikipédia).

## Outils de recherche sur Internet : Annuaires

- Description :

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories.

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.

## Outils de recherche sur Internet : Annuaires

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :

## Outils de recherche sur Internet : Annuaires

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif;

## Outils de recherche sur Internet : Annuaires

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif ;
  - Non qualitatif ;

## Outils de recherche sur Internet : Annuaires

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif ;
  - Non qualitatif ;
  - Très catégorisé.

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif;
  - Non qualitatif;
  - Très catégorisé.
- **Utilisation** :

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif;
  - Non qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document dans un domaine particulier.

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif;
  - Non qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document dans un domaine particulier.
- **Exemple** :

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif;
  - Non qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document dans un domaine particulier.
- **Exemple** :
  - <http://www.recherche.lycos.fr/annuaire>

## Outils de recherche sur Internet : Annuaire

- **Description** : Répertoire des sites par catégories. Un site apparaît dans l'annuaire si son propriétaire l'a inscrit.
- **Propriétés** :
  - Non exhaustif;
  - Non qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document dans un domaine particulier.
- **Exemple** :
  - <http://www.recherche.lycos.fr/annuaire>
  - <http://fr.dir.yahoo.com>

## Outils de recherche sur Internet : Sélections

- Description :

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories.

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).
  
- **Propriétés** :

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).
  
- **Propriétés** :
  - Non exhaustif;

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).
  
- **Propriétés** :
  - Non exhaustif ;
  - Qualitatif ;

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).
  
- **Propriétés** :
  - Non exhaustif ;
  - Qualitatif ;
  - Très catégorisé.

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuoproclamés).
  
- **Propriétés** :
  - Non exhaustif;
  - Qualitatif;
  - Très catégorisé.
  
- **Utilisation** :

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).
  
- **Propriétés** :
  - Non exhaustif ;
  - Qualitatif ;
  - Très catégorisé.
  
- **Utilisation** : Chercher un document **fiable** dans un domaine particulier.

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuooproclamés).
- **Propriétés** :
  - Non exhaustif;
  - Qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document **fiable** dans un domaine particulier.
- **Exemple** :

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuoproclamés).
- **Propriétés** :
  - Non exhaustif;
  - Qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document **fiable** dans un domaine particulier.
- **Exemple** :
  - <http://www.lii.org>

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuoproclamés).
- **Propriétés** :
  - Non exhaustif;
  - Qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document **fiable** dans un domaine particulier.
- **Exemple** :
  - <http://www.lii.org>
  - <http://www.bonweb.fr>

## Outils de recherche sur Internet : Sélections

- **Description** : Répertoire des sites par catégories. Les classifications sont faites par des experts (reconnus ou atuoproclamés).
- **Propriétés** :
  - Non exhaustif;
  - Qualitatif;
  - Très catégorisé.
- **Utilisation** : Chercher un document **fiable** dans un domaine particulier.
- **Exemple** :
  - <http://www.lii.org>
  - <http://www.bonweb.fr>
  - <http://www.signets.bnf.fr>

## Outils de recherche sur Internet : Moteur de recherche

- Description :

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;
  - Non catégorisé.

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;
  - Non catégorisé.
- **Utilisation** :

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;
  - Non catégorisé.
- **Utilisation** : Chercher un document à l'aide de mots clés.

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;
  - Non catégorisé.
- **Utilisation** : Chercher un document à l'aide de mots clés.
- **Exemple** :

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;
  - Non catégorisé.
- **Utilisation** : Chercher un document à l'aide de mots clés.
- **Exemple** :
  - <http://www.altavista.com>

## Outils de recherche sur Internet : Moteur de recherche

- **Description** : Indexation des mots contenus dans les pages WEBS de manière automatique.
- **Propriétés** :
  - Tend à être exhaustif ;
  - Attribution de scores ;
  - Non catégorisé.
- **Utilisation** : Chercher un document à l'aide de mots clés.
- **Exemple** :
  - <http://www.altavista.com>
  - <http://www.google.fr>

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** :

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien
  - 2. l'**indexeur** :

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien
  - 2. l'**indexeur** : retient les mots contenus dans chaque page web en constituant un **index**.

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien
  - 2. l'**indexeur** : retient les mots contenus dans chaque page web en constituant un **index**.
  - 3. le **classificateur** ou **guichetier** :

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien
  - 2. l'**indexeur** : retient les mots contenus dans chaque page web en constituant un **index**.
  - 3. le **classificateur** ou **guichetier** : à chaque requête

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien
  - 2. l'**indexeur** : retient les mots contenus dans chaque page web en constituant un **index**.
  - 3. le **classificateur** ou **guichetier** : à chaque requête
    - 3.1. cherche dans l'**index** les pages contenant les mots clés recherchés et constitue un index inversé

## Moteur de recherche

- Un moteur de recherche est constitué de trois éléments majeurs :
  - 1. le **robot** ou **aspirateur**, **crawler**, **spider** : parcourt les pages du web, de lien en lien
  - 2. l'**indexeur** : retient les mots contenus dans chaque page web en constituant un **index**.
  - 3. le **classificateur** ou **guichetier** : à chaque requête
    - 3.1. cherche dans l'**index** les pages contenant les mots clés recherchés et constitue un index inversé
    - 3.2. classe les pages trouvées suivant une pondération propre au moteur de recherche (algorithmes de classement).

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :
  - 1. dont le contenu est dans un format impropre à l'indexation ;

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :
  - 1. dont le contenu est dans un format impropre à l'indexation ;
  - 2. issue de site dynamique ;

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :
  - 1. dont le contenu est dans un format impropre à l'indexation ;
  - 2. issue de site dynamique ;
  - 3. à accès restreint ;

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :
  - 1. dont le contenu est dans un format impropre à l'indexation ;
  - 2. issue de site dynamique ;
  - 3. à accès restreint ;
  - 4. non liée.

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :
  - 1. dont le contenu est dans un format impropre à l'indexation ;
  - 2. issue de site dynamique ;
  - 3. à accès restreint ;
  - 4. non liée.
- Le web auquel on ne peut pas accéder par des techniques classiques d'indexation est appelé **Web profond** (ou **Web invisible**).

## Web profond

- **Attention** : Les moteurs de recherche n'explorent pas tous le web !
- En particulier ils ne trouveront jamais de pages :
  - 1. dont le contenu est dans un format impropre à l'indexation ;
  - 2. issue de site dynamique ;
  - 3. à accès restreint ;
  - 4. non liée.
- Le web auquel on ne peut pas accéder par des techniques classiques d'indexation est appelé **Web profond** (ou **Web invisible**).
- La société BrightPlanet estimait en **2001** que sa taille était de **500** fois plus importante que le WEB indexé.

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.
- Une équipe de chercheurs allemands a étudié le comportement des robots d'indexation face à des sites contenant énormément de pages.

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.
- Une équipe de chercheurs allemands a étudié le comportement des robots d'indexation face à des sites contenant énormément de pages. Ils ont créé un site web composé de **2147483647** pages ( $2^{31} - 1$ ).

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.
- Une équipe de chercheurs allemands a étudié le comportement des robots d'indexation face à des sites contenant énormément de pages. Ils ont créé un site web composé de **2147483647** pages ( $2^{31} - 1$ ). Il faut au minimum **31** clics pour arriver à certaines pages.

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.
- Une équipe de chercheurs allemands a étudié le comportement des robots d'indexation face à des sites contenant énormément de pages. Ils ont créé un site web composé de **2147483647** pages ( $2^{31} - 1$ ). Il faut au minimum **31** clics pour arriver à certaines pages. Ils ont laissé ce site en ligne, sans le modifier, pendant une année.

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.
- Une équipe de chercheurs allemands a étudié le comportement des robots d'indexation face à des sites contenant énormément de pages. Ils ont créé un site web composé de **2147483647** pages ( $2^{31} - 1$ ). Il faut au minimum **31** clics pour arriver à certaines pages. Ils ont laissé ce site en ligne, sans le modifier, pendant une année.
- Les résultats montrent que le nombre de pages indexées pour ce site, dans le meilleur des cas, ne dépasse pas

## Web opaque

- Le **Web opaque** est la partie du Web qui pourrait être indexé mais qui ne l'est pas.
- Une équipe de chercheurs allemands a étudié le comportement des robots d'indexation face à des sites contenant énormément de pages. Ils ont créé un site web composé de **2147483647** pages ( $2^{31} - 1$ ). Il faut au minimum **31** clics pour arriver à certaines pages. Ils ont laissé ce site en ligne, sans le modifier, pendant une année.
- Les résultats montrent que le nombre de pages indexées pour ce site, dans le meilleur des cas, ne dépasse pas **0,0049%**.