

# Bases de connaissances à propos de l'internet et du web version 0.5

Jean-Philippe GEORGET  
jean-philippe.georget@orleans-tours.iufm.fr

4 octobre 2002

## Table des matières

<b>1</b>	<b>Préambule</b>	<b>1</b>
<b>2</b>	<b>Historique rapide</b>	<b>2</b>
<b>3</b>	<b>Philosophie de l'internet – navigateurs – monopole</b>	<b>3</b>
<b>4</b>	<b>Des erreurs en surfant sur le web</b>	<b>4</b>
<b>5</b>	<b>Quelques liens</b>	<b>5</b>
<b>6</b>	<b>Recherche des informations sur le web</b>	<b>5</b>
6.1	Annuaire . . . . .	7
6.2	Moteurs de recherche . . . . .	7
6.3	Méta-moteurs . . . . .	7
<b>7</b>	<b>Conclusion</b>	<b>8</b>

## 1 Préambule

Ce document est un support de cours. Il donne quelques bases pour permettre de mieux comprendre l'évolution et le fonctionnement du web. Il détaille notamment divers points techniques qui permettent de mieux se l'approprier.

Un bon complément de ce cours est l'URL <http://cui.unige.ch/eao/www/Internet/Plan.html> qui est un cours en ligne mis à disposition par son auteur Bertrand Ibrahim de l'université de Genève.

## 2 Historique rapide

La consultation du lien suivant permet de se faire une idée de l'évolution de l'internet et du web.

– <http://cui.unige.ch/eao/www/Internet/Historique.html>

Pour résumer :

- *internet* est la concaténation de *inter* – comme en français – et de *net* – réseau. C'est l'interconnexion de réseaux d'ordinateurs de diverses natures qui peuvent fonctionner avec des systèmes d'exploitation différents<sup>1</sup> ;
- son existence débute à l'époque de la guerre froide dans les années 60 aux États-Unis. L'armée souhaitait créer un réseau informatique supposé résister à une attaque nucléaire. Il a été ensuite utilisé par les universités pour faciliter leurs échanges aux États-Unis puis dans d'autres pays. Les sociétés privées et le grand public y ont aujourd'hui accès ;
- le principe était – il l'est encore partiellement – que si une liaison était coupée entre deux bases, alors les communications pouvaient facilement passer par une troisième ou bien par une succession de bases afin que tous les messages arrivent à bon port. Le protocole qui permet cette opération et qui reste d'actualité s'appelle le protocole IP<sup>2</sup>.
- *internet* permet d'utiliser plusieurs *services* tels que la messagerie électronique, le web, les forums de discussions Usenet, le transfert de fichiers, etc. ;
- chaque service de l'internet demande, en général, un logiciel particulier pour être exploité : un navigateur pour le web, un logiciel de messagerie, etc. Certains logiciels proposent d'utiliser plusieurs de ces services par le biais de modules dépendants ;
- l'accès à une ressource située sur l'internet s'effectue généralement par le biais de son *URL* – *Uniform Resource Locator*. Prenons l'URL `ftp://dino.education.fr/divers/sommaire.zip`.
- `ftp://` est le service pour les transferts de fichiers<sup>3</sup> ;

---

<sup>1</sup>Un système d'exploitation permet d'utiliser de manière basique un ordinateur. Il gère l'affichage à l'écran, la saisie au clavier, les fichiers, etc. Voici quelques exemples : Microsoft Windows, MacOS, GNU/Linux. On peut ensuite utiliser des logiciels, préalablement installés ou à installer, pour aller au-delà de ces fonctionnalités basiques.

<sup>2</sup>Internet Protocol

<sup>3</sup>FTP – *File Transfert Protocol* – protocole de transfert de fichiers.

- `dino.education.fr` est le nom de l'ordinateur qui héberge cette ressource ;
- `/divers/sommaire.zip` est le chemin pour y accéder. Ici, c'est un fichier *ZIP*, c'est à dire un fichier compressé qui est stocké dans le répertoire `divers`.

Pour `http://dino.education.fr/accueil/sommaire.htm/#meteo`, `http://` signifie le service web<sup>4</sup> situé sur le même ordinateur, et désigne un endroit repéré par une *ancree* placée dans le fichier `sommaire.htm`. Lorsque l'on demande juste de `http://dino.education.fr/`, on aura en fait la page d'accueil du site web en question<sup>5</sup>.

### 3 Philosophie de l'internet – navigateurs – monopole

En dehors des sites commerciaux, l'empreinte des universitaires est encore très présente sur le web et dans les forums. Le principe général est "j'ai créé des documents, je n'en perd pas l'usage si je les partage". Ce principe est repris notamment par les partisans des *logiciels libres*<sup>6</sup>.

Pour le web comme pour d'autres services, les différentes normes de l'*HTML* n'ont pas été mises au point en un jour. Plusieurs entreprises regroupées au sein du consortium *W3C*<sup>7</sup> y ont oeuvré. Le premier navigateur, *Mosaïc* de *NSCA*<sup>8</sup>, fut mis à la disposition du public gracieusement. Ensuite, *Netscape*, entreprise privée, développa *Netscape Navigator* et participa activement au développement de la norme. Ce navigateur a équipé 80% des machines connectées durant quelques années. *Microsoft* arrive dans les années 90 et tentent d'imposer un réseau "parallèle" : *MSNet*. Il développe alors le navigateur *Internet Explorer*. Le projet de réseau indépendant tourne court et l'entreprise développe ensuite son projet à l'intérieur de l'internet. Le fait que le navigateur fut ensuite installé d'office avec le système d'exploitation *Windows* explique le fait qu'aujourd'hui, c'est *Microsoft* qui détient 80%<sup>9</sup> du marché des navigateurs. L'icône *Internet* du bureau de *Microsoft Windows* lance en fait une connexion à l'internet et lance le navigateur *Internet Explorer* qui s'ouvre tout naturellement sur la page d'accueil de l'entreprise *Microsoft*. Le problème est que cette entreprise cherche à capter le marché en imposant ses normes sans vouloir les normaliser. Le navigateur *Opera* propose quelques fonctionnalités originales. A l'inverse de *Netscape Navigator* et d'*Internet Explorer*, celui-ci est payant mais

<sup>4</sup>*HTTP – Hyper Text Markup Langage* – langage hypertexte à balises

<sup>5</sup>En général, elle correspond au fichier `index.html` ou `index.htm`.

<sup>6</sup>Voir le site de l'AFUL, <http://www.aful.org/>.

<sup>7</sup><http://www.w3c.com/>.

<sup>8</sup>*National Center for Supercomputing Applications*

<sup>9</sup>Pourcentage approximatif, aucune étude ne permettant de fixer ce chiffre avec précision.

on peut l'utiliser gratuitement si on accepte un petit bandeau publicitaire en haut de la fenêtre du logiciel. Il fonctionne sur beaucoup de systèmes différents à l'inverse d'*Internet Explorer*.

*Mozilla*, logiciel libre et gratuit, est disponible sur plusieurs plate-formes et peut être téléchargé sur <http://frenchmozilla.sourceforge.net/>.

Concernant les différents logiciels, leur fonctionnement est similaire et chacun a ses points forts et faibles. Le tout est de savoir si en tant que citoyen, nous souhaitons qu'une seule entreprise fournisse le seul logiciel qui nous permettra de rechercher des informations, de gérer nos comptes bancaires, d'envoyer et de recevoir des messages, nos déclarations d'impôts, etc. ou si nous souhaitons garder des possibilités de choix. Il est clair qu'il vaut mieux y être sensibilisé avant que cela se produise.

## 4 Des erreurs en surfant sur le web

Lorsque l'on utilise un navigateur, l'adresse sur laquelle on clique ou celle que l'on tape peut ne pas fonctionner. Il peut y avoir au moins deux raisons. La première est que le nom de l'ordinateur utilisé est erroné. La deuxième est que le chemin de la ressource est erroné et donc qu'on ne peut pas la trouver sur l'ordinateur en question. En général, le navigateur donne une page d'explication en anglais.

Dans le deuxième cas, on peut en général retrouver rapidement la ressource recherchée. Il suffit d'effacer dans la barre d'adresse le nom du fichier puis de valider. En cas d'échec, effacer le nom du dernier dossier et de valider, et ainsi de suite. Au pire, on obtiendra la page d'accueil du site en question.

Par exemple, si je cherche

<http://www.education.fr/accueil/profs/index.htm> et que je tape

<http://www.education.fr/accueil/profs/inde.htm>, j'obtiens une erreur. Je peux ensuite tenter

<http://www.education.fr/accueil/profs/>. En cas d'échec, j'essaie

<http://www.education.fr/accueil/>

et enfin

<http://www.education.fr/> qui me donne la page d'accueil du portail de l'éducation en France du Ministère de l'Éducation Nationale.

## 5 Quelques liens

- <http://www.education.fr> – le portail des professionnels de l'éducation, MEN ;
- <http://www.education.gouv.fr> – Site du ministère de l'Éducation Nationale ;

- <http://nte37.micro-video.fr> – le CARM37 – Centre d’Appui et de Ressources Multimédia d’Indre et Loire – assure le suivi et une aide de proximité pour les enseignants du département 37;
- <http://www.educnet.education.fr/> – Educnet, site ressource du MEN, fiches juridiques à l’intention des enseignants, classement par thèmes disciplinaires ;
- <http://www.educasource.education.fr/> – ressources électroniques pour les enseignants par thèmes disciplinaires ;
- <http://listes.educnet.education.fr/B2i/> – B2I – Brevet Informatique et Internet – un brevet qui n’est pas seulement du ressort du professeur de technologie ;
- [http://www.educnet.education.fr/listes\\_educnet/listes.htm](http://www.educnet.education.fr/listes_educnet/listes.htm) – listes de diffusion hébergées par Educnet ;
- <http://www.cafepedagogique.net/> – le *Café pédagogique* et sa liste de diffusion qui permet de se tenir informé facilement des sites qui peuvent intéresser chaque discipline ou des thèmes transdisciplinaires ;
- <http://www.epi.asso.fr/epinet.htm> – Epi.net, liste de diffusion de l’EPI, Asso Enseignement Public et Informatique. La liste de diffusion a un intérêt similaire à celle du *Café pédagogique* ;
- <http://www.francopholistes.com/> – catalogue de listes de diffusion ;
- <http://www.bnf.fr> – Bibliothèque Nationale de France ;
- <http://www.louvre.fr> – le Louvre.

## 6 Recherche des informations sur le web

Notons pour commencer que le bouton *Rechercher* d’un navigateur aboutit en général à la page de recherche du concepteur du logiciel et que ces systèmes sont peu performants.

Pour trouver des informations sur le web, on peut utiliser différents systèmes de recherche d’informations.

Voyons quelques éléments qui permettent de chercher une information sur le web. Le lecteur pourra se reporter à <http://www.abondance.com/> pour de multiples compléments pertinents sur le sujet et notamment sa FAQ<sup>10</sup>.

Certains systèmes proposent deux types de recherche : simple ou avancée. L’aide est proposée au moins sur la page d’accueil<sup>11</sup>.

Il faut distinguer la langue de l’interface – des informations qui s’affichent lors de la recherche – de la langue des sites recherchés. L’interface peut-être en langue

<sup>10</sup>[http://docs.abondance.com/faq\\_recherche.html](http://docs.abondance.com/faq_recherche.html)

<sup>11</sup>Mots-clés : tips, help, advanced research, FAQ – Frequently Asked Questions, Foire Aux Questions.

anglaise et me permettre de rechercher des sites en français ou d'autres langues.

Quelques éléments de syntaxe sont *relativement* généralisés.

Un signe plus accolé devant un mot signifie que l'on exige sa présence alors qu'un signe moins signifie que l'on ne souhaite pas voir apparaître ce mot. Par exemple, *+vanille +fraise chocolat -menthe* donnera généralement des liens vers des pages qui contiennent obligatoirement les mots *fraise* et *vanille*, éventuellement *chocolat* et jamais *menthe*.

On peut souvent utiliser les connecteurs logiques *AND*, *OR*, *NOT*. Par exemple, *vanille AND fraise* donnera généralement des liens vers des pages qui contiennent les deux mots ou éventuellement vers des pages qui n'en contiennent qu'un seul.

Certains systèmes ignorent les caractères accentués<sup>12</sup> du français alors que d'autres en tiennent compte. Il en est de même pour la casse<sup>13</sup> des lettres. En général, "langue française" – avec les guillemets – impose que l'expression soit complète alors que *langue française* – sans les guillemets – peut renvoyer des liens vers des pages où *langue* est situé en bas de la page alors que *française* est situé en haut de la page, ces deux termes n'ayant rien à voir l'un avec l'autre. Sur *Altavista*, *NEAR* demande à ce que deux mots ne soient pas séparés eux-mêmes par plus de 10 mots. Cette fonctionnalité n'est pas proposée en général.

Pour dresser une typologie des systèmes disponibles, on définit généralement trois types de systèmes : les annuaires ou catalogues thématiques, les moteurs de recherche, les méta-moteurs. Il arrive aussi que certains sites proposent les trois types mais il reste toujours une dominante.

Enfin pour expliquer que cette puissance de recherche est à notre disposition à titre gracieux, il faut fournir quelques explications. Vous trouverez en général sur ces sites des bandeaux publicitaires qui sont modifiés en fonction de votre équation de recherche, des services annexes fournis par des partenaires comme des sites de ventes de livres par correspondance. Ces systèmes sont aussi des vitrines de la technologie qu'ils exploitent et qui peut être vendue à des entreprises pour leurs réseaux internes.

## 6.1 Annuaires

Ce sont des sites où l'on trouve des liens classés par thèmes de manière plus ou moins hiérarchisés. Ces sites ont été indexés manuellement et l'on peut effectuer une recherche dans tout le catalogue ou limitée à un thème, une catégorie. On peut aussi parcourir les différentes catégories et lister les sites proposés dans chacune

---

<sup>12</sup>ç, à, é, ...

<sup>13</sup>Alternance des majuscules et minuscules.

d'elles. Étant donné que cette indexation est "manuelle", la liste des liens répertoriés est grande mais limitée et concerne des thèmes « grand public ». En général, seules les pages d'accueil des sites sont présentées.

Quelques exemples :

- <http://www.yahoo.fr/> – Yahoo France
- <http://www.voila.fr/> – Voilà , France Telecom

## 6.2 Moteurs de recherche

Des "robots logiciels" surfent sur le web et explorent les sites jusqu'à une certaine profondeur – dans la hiérarchie des dossiers. Les mots sont indexés automatiquement après avoir été éventuellement filtrés. Une recherche sur ce type de système donne énormément de liens – pas toujours vers la page d'accueil des différents sites. Le classement des réponses dépend du moteur utilisé. Il est très rare de n'avoir aucune réponse et très courant d'en avoir plusieurs centaines de milliers. Il est donc utile de se reporter à l'aide disponible à partir de la page d'accueil.

- <http://www.google.fr/> – Google (fr)
- <http://www.hotbot.com/> – Hotbot
- <http://www.altavista.fr/> – Altavista (fr)

## 6.3 Méta-moteurs

Ces systèmes interrogent les deux autres catégories (annuaires et moteurs). Il existe par exemple <http://www.wwsearch.com/>. Le logiciel *Copernic* existe dans une version gratuite et permet d'interroger d'autres moteurs. Il a ses adeptes mais nécessite d'être installé sur la machine alors que les autres systèmes de recherche sont simplement disponibles sur le web.

## 7 Conclusion

Ce document est en cours de rédaction. Le lecteur voudra bien pardonner les éventuelles erreurs qu'il contient.