

Naviguer et collecter des informations sur le Web

Ce document est un support de cours dédié à l'apprentissage des principales fonctions d'un navigateur. Il explique notamment comment récupérer les informations sur une page Web. Pour l'utiliser, lancez votre navigateur et testez les différentes fonctions présentées.

Les actions à effectuer sont précédées d'une flèche (⇒).

1. Principes des base

1.1. Glossaire

Lisez ces quelques définitions. Même si ces notions vous paraissent peu claires au premier abord, vous pourrez y revenir au fur et à mesure de votre progression.

Internet : Contrairement aux idées reçues, le mot Internet ne désigne qu'un ensemble de « Tuyaux » reliant entre eux des millions d'ordinateurs dans le monde entier. A travers ces tuyaux, de nombreuses informations circulent telles le courrier électronique, le transfert FTP et surtout le Web dont nous allons traiter ici.

Web : Contraction de World Wide Web (www), ce qui signifie « Toile d'araignée mondiale ». Le Web désigne un ensemble mondial de documents écrits en langage HTML. Ces documents se présentent sous forme de textes et images et sont reliés entre eux à l'aide de liens dits « Hypertexte ». On accède au Web au moyen d'un logiciel appelé « Navigateur ».

Hypertexte : Un document est dit hypertexte lorsqu'il permet d'accéder à d'autres documents au moyen d'un simple « Clic » de souris sur une partie du texte. L'exemple le plus simple de l'hypertexte est l'aide de Windows : un clic sur certains mots appelle le glossaire ou le menu correspondant.

Navigateur : (syn. « butineur » ou « browser ») C'est un logiciel qui permet d'afficher les pages écrites en langage HTML. Si vous êtes connecté à l'Internet, il vous permettra de « naviguer » sur le Web. Il faut savoir que les logiciels gratuits *Netscape Navigator* et *Microsoft Internet Explorer* se partagent en gros chacun la moitié du marché.

HTML : Les documents disponibles sur le Web sont écrits en langage HTML (Hyper Text Markup Language). Lorsqu'on charge une page Web, on charge un fichier HTML (.htm) qui contient le texte, la mise en page et les liens du document. Tous les autres éléments tels les images sont d'autres fichiers auxquels le fichier HTML fait référence et que le navigateur charge séparément.

Page Web : Lorsque l'on charge un document HTML sur le navigateur, la page visualisée est appelée « page Web »

Site Web : Généralement lorsqu'un individu ou une entité diffuse des informations sur le Web, il le fait à l'aide de plusieurs pages situées sur le même serveur. Cet ensemble de pages constitue ce qu'on appelle un « site Web ».

Moteurs de recherche : Les moteurs de recherche sont des serveurs qui indexent un certain nombre de sites Web et vous permettent généralement de rechercher les documents qui vous intéressent à l'aide de mots clé. Ce sont en quelque sorte les « Annuaire » du Web.

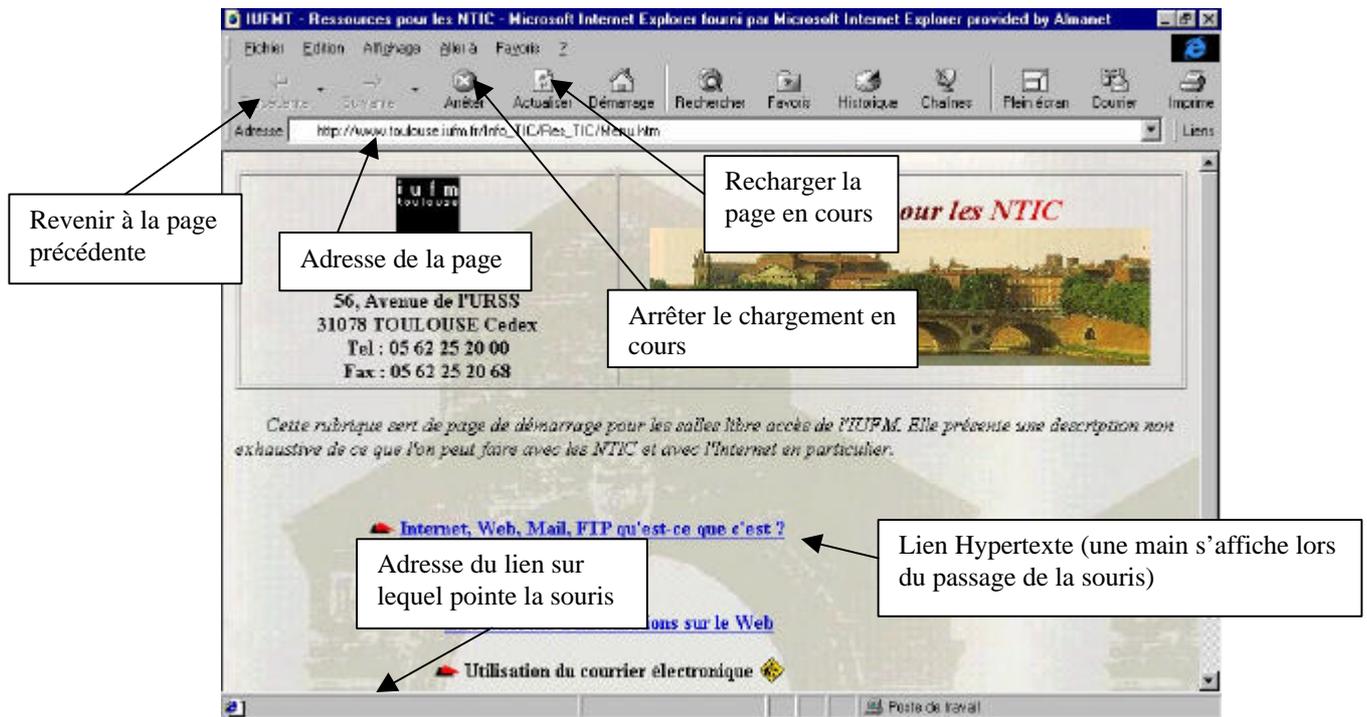
Kilo-Octet (Ko) : La taille des fichiers se mesure en octets. Pour votre culture sachez qu'en informatique Kilo ne veut pas dire 1000 mais $1024 = 2^{10}$, (de même Mega = $1024 \times 1024 = 2^{20}$). Pour avoir un ordre d'idée, sachez qu'une page de texte simple (.txt) occupe de l'ordre de 2Ko.

Bits par seconde (bps) : La vitesse des appareils de communication type Modem se mesure en Bits (0 ou 1) par seconde (et non en Baud !). Sachant qu'il faut au moins 8 bits pour transmettre un octet et en ajoutant les bits de contrôle, on peut compter approximativement 10 bits par octet. Concrètement un Modem 33600bps permet de communiquer à 3,3Ko/s.

Cookies : Parfois lors de votre navigation le navigateur vous demande si vous acceptez les « Cookies », il s'agit d'informations texte (<1Ko) enregistrées par un site sur votre disque à des fins statistiques. Contrairement à la rumeur il n'y a aucun risque de virus ! Par le biais d'un cookie un site peut savoir par exemple combien de fois vous êtes venu le consulter, quelles rubriques vous fréquentez le plus souvent ...

1.2. Principes de base de la navigation

Pour commencer, lancez votre navigateur (Ici, *Internet Explorer*) :



Seules les fonctions essentielles sont indiquées ci-dessus. Pour pouvoir les tester, placez-vous d'abord « Manuellement » sur une page en tapant l'adresse directement au clavier (après avoir cliqué sur la zone adresse à l'aide de la souris). Demandons par exemple <http://www.toulouse.iufm.fr>, ce qui indique de charger la page d'accueil du serveur de l'IUFM de Toulouse (notez que le « http:// » peut être omis, le navigateur l'ajoutera automatiquement).

Après validation, votre ordinateur se connecte (via Internet) à notre serveur qui va retourner un fichier HTML que votre navigateur va interpréter. Suivant votre vitesse de connexion texte et images vont alors s'afficher progressivement.

⇒ Testez les effets de chacune des fonctions présentées ci-dessus.

Les liens hypertextes peuvent être indifféremment internes ou externes, c'est à dire qu'un clic de souris peut appeler indifféremment un fichier situé sur le même disque dur que sur le disque dur d'un autre serveur dans le monde entier.

⇒ Testez la rubrique « Liens » du serveur <http://webdev.toulouse.iufm.fr/pha/> et connectez-vous à différents sites suivant vos aspirations.

Les quelques pages que vous avez regardées font partie de ce qu'on appelle le Web, contraction de World Wide Web (ou www). La navigation ne serait pas très drôle s'il fallait connaître toutes les adresses et les rentrer « Manuellement ». Des outils de sélection ont heureusement été développés : Les moteurs de recherche. Ce sont des serveurs qui indexent un certain nombre de sites Web et vous permettent généralement de rechercher les documents qui vous intéressent à l'aide de mots clé.

⇒ Placez-vous sur Yahoo France à l'adresse <http://www.yahoo.fr> et testez différentes recherches selon vos aspirations en choisissant judicieusement vos mots clé.

⇒ Lorsque vous aurez trouvé un site intéressant, utilisez le menu « Favoris » de votre logiciel de navigation (ou « Signets » ou « Bookmarks ») et trouvez par vous même son utilité.

Vous avez maintenant compris les principes de base de la navigation : c'est extrêmement simple et aucune connaissance informatique n'est vraiment requise ! Ayant accès à une véritable bibliothèque mondiale, vous voudrez sûrement dès lors pouvoir récupérer les informations que vous avez consultées : c'est ce que nous allons voir maintenant.

2. Récupérer des informations

2.1. Légalité

Attention, même s'il règne sur le Web un esprit assez libertaire, la loi est la même que pour toute diffusion : Il faut respecter la propriété intellectuelle d'autrui.

Les copies ne sont légales que si elles sont explicitement autorisées, vous ne devez pas vous approprier ni même utiliser des textes ou des images sans l'accord explicite de leur auteur.

Heureusement cependant, l'éthique du réseau (la Netiquette comme disent certains) est tournée vers le partage : la raison d'être du Web c'est l'échange et la plupart des responsables de sites Web vous autorisent explicitement à diffuser gratuitement leurs productions pour peu que vous conserviez l'intégrité des documents et/ou que vous signaliez leur origine.

En bref : **Ne copiez rien qui ne soit explicitement autorisé !**

2.2. Récupérer un document diffusé sur une page Web

⇒ **Avant de tester les différentes méthodes de récupération de documents, créez un répertoire temporaire (par exemple c:\aeffacer\)** dans lequel vous enregistrerez vos fichiers (utilisez pour cela le gestionnaire de fichiers)

La plupart des ressources pédagogiques diffusées sur le Web le sont sous forme de fichiers (.doc, .pdf, .zip ...). Il suffit généralement de cliquer sur un lien pour télécharger le document désiré. Après avoir cliqué sur le lien, deux cas de figure se présentent :

- Soit votre navigateur vous demande s'il faut l'enregistrer sur le disque, il suffit alors de choisir le répertoire dans lequel vous voulez rapatrier le fichier.
- Soit votre navigateur appelle automatiquement le programme permettant de visionner le fichier. Si vous voulez seulement rapatrier le fichier il vous faudra cliquer sur le bouton droit de la souris puis « Enregistrer la cible/le lien sous »

2.2.1. Taille des documents et vitesse de téléchargement

Avant même de commencer à télécharger, faites toujours attention à la taille des fichiers et à la durée prévisible de téléchargement. La taille est généralement indiquée en Kilo-octets (Ko). Ne tentez pas de rapatrier un fichier de 15Mo (≈15000Ko) avec votre connexion par Modem !

Voici un repère : Télécharger **100Ko** de données depuis un établissement (connecté généralement à 64Kbps soit 6 à 7 Ko/s) prend au **meilleur cas 15s**, comme le débit est partagé entre tous les utilisateurs et que les lignes sont souvent surchargées il faudra **généralement compter une bonne minute !** Pour des fichiers plus importants, appliquez la règle de trois.

2.2.2. Types des documents

On peut pratiquement trouver tous les type de fichier sur le Web, certains sont cependant privilégiés notamment en ce qui concerne les ressources pédagogiques :

Type	Caractéristiques
.DOC	Ce sont les documents Word, pour les visionner il faut disposer du logiciel payant ou éventuellement se procurer la visionneuse gratuitement distribuée par Microsoft. Ces documents peuvent contenir des Macrovirus, faites donc attention à les vérifier après téléchargement. Attention aussi à la version du logiciel : un document tapé sous Office 97 ne peut être lu sur Office 95. Comme les fichiers Word sont parfois volumineux, on les trouve plus souvent sous forme compressée (.Zip).
.XLS	Ce sont les fichiers du tableur Excel qui répondent aux mêmes contraintes que les fichiers .DOC
.PDF	Ce type de fichier est très répandu sur le Web, il s'agit de documents Adobe Acrobat qui peuvent contenir textes et images. Leur succès vient de leur taille : elle est très réduite de sorte qu'une notice technique de 20 pages peut tenir sur 50Ko. Autres intérêts majeurs : les documents PDF sont exempts de virus, ils peuvent être protégés (pour empêcher leur édition et leur impression) et la visionneuse Acrobat Reader est gratuite et largement distribuée.
.EXE	Certains document sont distribués sous forme exécutable autocompressée : une fois le fichier téléchargé, il suffit de lancer l'exécutable pour créer les fichiers document. L'avantage de ce type de fichier est leur taille réduite, de plus ils n'ont besoin d'aucun programmes additionnels comme Winzip pour être décompressé. L'inconvénient majeur est le fichier exécutable lui-même qui, dès qu'il est lancé peut contaminer la machine avec un virus.
.ZIP	C'est de loin le type de fichier le plus diffusé. Il s'agit de documents compressés au format Zip dont le principe est expliqué dans le paragraphe suivant.

⇒ **Trouvez et téléchargez dans votre répertoire temporaire différents types de documents.**

2.2.3. Les fichiers Zip

Etant donné la relative lenteur d'Internet, les documents en téléchargement sont très souvent fournis sous forme compressée au format ZIP, ce qui réduit d'autant leur volume donc le temps de chargement. Le principe de la compression est relativement simple : il consiste à compiler un ou plusieurs documents sur un seul fichier en "éliminant" les informations inutiles.

Principe de fonctionnement : Sur un fichier Word, la ligne : « XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX » (30 lettre X) occupe **30 Octets**. Si nous passons le fichier à travers une « moulinette » qui remplace la ligne par « 30*X », nous n'occupons plus sur le disque que **4 Octets**. Les algorithmes de compression sont plus complexes dans la réalité mais le principe est le même. Ainsi, on peut facilement obtenir des taux de compression de 25% (la taille est divisée par 4 !) sur un fichier texte.

Avantage : Le temps de chargement d'un fichier compressé est beaucoup plus court que celui du même fichier non compressé,

Inconvénient : Il faut disposer du logiciel de décompression adapté et savoir l'utiliser.

Il existe de nombreux logiciels de compression. Heureusement un standard s'est actuellement imposé sur le Web, il s'agit du format ZIP. Le logiciel le plus utilisé pour gérer les fichiers Zip est le shareware WinZip qui se révèle vite indispensable à tout internaute.

L'utilisation de WinZip est simple et intuitive, s'il est installé sur votre machine vous pouvez simplement décompresser les fichiers Zip téléchargés : il suffit de double-cliquer sur le fichier concerné et d'extraire les fichiers dans un répertoire quelconque.

Inversement, si vous voulez transférer des fichiers en occupant un minimum d'espace (par mail ou sur disquette), démarrez le logiciel Winzip et créez une nouvelle archive en ajoutant tous les fichiers que vous voulez inclure.

⇒ **Copiez un fichier texte (.TXT) quelconque dans votre répertoire et compressez le fichier à l'aide de WinZip, observez les tailles avant et après compression.**

⇒ **Si ce n'est déjà fait, récupérez un document zippé sur le Web et décompressez-le dans votre répertoire.**

2.3. Récupérer une image

Pour récupérer une image à partir de votre navigateur, placez-vous sur l'image désirée, cliquez sur le bouton droit de la souris et faites "Enregistrer l'image sous".

Les images disponibles sur le Web sont de deux types qui sont tous deux compressés :

Les images .GIF : Elles sont limitées à 256 couleurs mais permettent des effets de transparence et de petites animations (qui sont en fait plusieurs images apparaissant successivement). La compression GIF est non destructive et est très efficace avec les images "logiques" comme les schémas, boutons et icônes.

Les images .JPG : (prononcer J-Peg) Elles sont en 16 Millions de couleurs et permettent d'obtenir des taux de compression impressionnants, notamment avec les photos. Cette compression est destructive.

Pour retravailler les images récupérées, le logiciel le plus réputé est le Shareware Paint Shop Pro dont l'utilisation est relativement simple et intuitive.

Attention à la propriété des images, vous ne pouvez récupérer légalement que les images librement diffusables. Certains sites se sont d'ailleurs spécialisés dans la diffusion d'images libres de droits : consultez-les.

⇒ **Récupérez des images en les enregistrant dans votre répertoire, essayez ensuite de les éditer pour voir comment vous pourrez les réutiliser.**

2.4. Récupérer une page Web

Les pages Web sont en fait des fichiers HTML (extension .HTM ou .HTML), vous pouvez facilement les enregistrer pour récupérer les informations textuelles en faisant **Fichier/Enregistrer sous**.

Attention, le fichier HTML ne contient que le texte, les liens et la mise en page. Par conséquent vous n'enregistrerez pas les images! Vous pouvez ensuite relire les informations en local sur votre machine non connectée et récupérer les informations hors connexion.

Rq : Quand une page possède des « Frames », c'est à dire plusieurs cadres, cliquez d'abord dans le cadre désiré avant de procéder à l'enregistrement.

⇒ **Récupérez une page Web en l'enregistrant dans votre répertoire, double-cliquez ensuite sur le fichier depuis l'explorateur et testez un copier/coller vers le traitement de textes Write.**

Pour récupérer l'ensemble d'un site Web ou une page Web complète avec les images il existe des logiciels spécialisés appelés « aspirateurs » comme MemoWeb ou WebDown (ce dernier est disponible gratuitement). Ces outils sont très pratiques mais doivent être utilisés en prenant certaines précautions : Un téléchargement de niveau 2 (la page courante, les pages liées et les pages liées aux pages liées) peut facilement appeler 200 documents et encombrer ainsi rapidement la bande passante de l'établissement et votre disque dur.

⇒ **Si le temps le permet, installez et testez le logiciel WebDown avec un téléchargement de niveau 1**

⇒ **Avant d'éteindre la machine, n'oubliez pas d'effacer votre répertoire temporaire**

3. Retrouver ces informations

La plupart de ces informations sont disponibles sur le serveur de l'IUFM de Toulouse : <http://www.toulouse.iufm.fr> , vous pouvez laisser vos critiques ou commentaires à A.Gosselin à l'adresse : webmaster@toulouse.iufm.fr

Pour conclure, sachez que l'informatique ne peut s'apprendre qu'en « merdant ». Il est illusoire de croire qu'une formation est suffisante, seule une la pratique régulière vous permettra de progresser. Sinon vous aurez bien vite oublié les quelques notions présentées ici. Alors lancez vous ...