

1. Les moteurs de recherche

1.1. Introduction

Internet nous donne accès à une masse considérable d'informations, des dizaines de millions de pages. Les documents publiés embrassent à peu près tous les sujets imaginables et sont diffusés aussi bien par des institutions que par des associations, des entreprises ou des individus. Ils sont de présentation et de qualité très inégales.

Ce système d'information et de communication est:

- **distribué**, l'information n'est pas centralisée, ni hiérarchisée,
- **hétérogène**, des ressources d'origine très différente coexistent,
- **non certifié**, l'information n'est pas toujours validée,
- **instable**, les sites d'information évoluent, naissent et disparaissent quotidiennement.

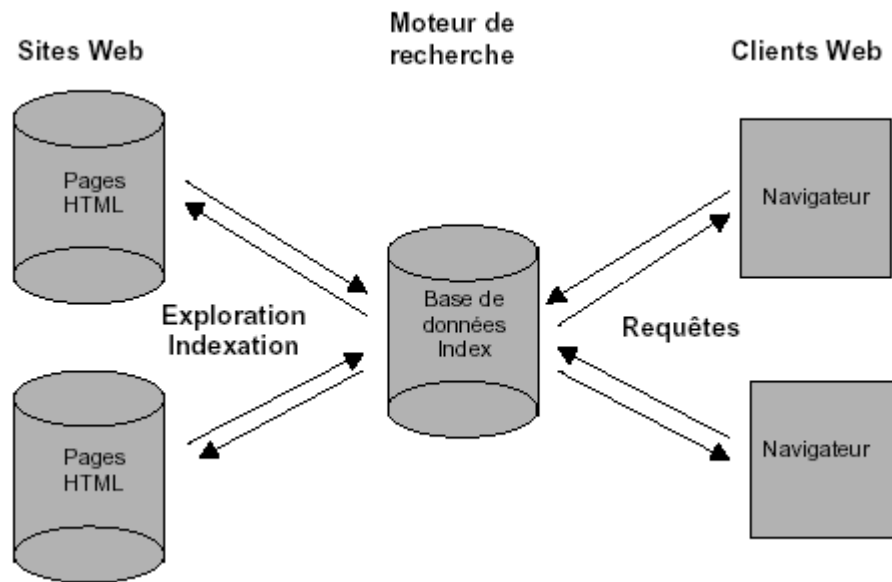
Attendez-vous donc à rencontrer le pire et le meilleur sur Internet. Pour y trouver de l'information intéressante et valable, vous devrez faire preuve d'ingéniosité dans vos recherches et de beaucoup de sens critique pour sélectionner les documents. Il est donc intéressant, non seulement de savoir comment rechercher de l'information, mais de savoir aussi comment la sélectionner.

1.2. Moteurs de recherche

Comment retrouver des documents dans cette masse (apparemment) inextricable? Il existe sur Internet une série d'outils permettant de faire des recherches. Une manière classique consiste à utiliser un (des) moteur(s) de recherche proposé(s) sur le Web. Dans son approche la plus simple, un moteur de recherche est une base de données qu'on peut interroger à distance. Plus précisément, un moteur de recherche est un serveur spécialisé dans la compilation et l'indexation des informations et qui, en outre, possède un module pour répondre aux requêtes (demandes de recherche) des utilisateurs.

Les moteurs de recherche possèdent, en général, trois modules principaux.

- **Le module de mise à jour**: son rôle est d'explorer l'Internet à la recherche de nouveaux sites ou de nouvelles informations sur des sites déjà répertoriés. Cette exploration peut se faire **automatiquement** par des **programmes (robots)**, qui rapatrient des informations prédéterminées, ou **manuellement** par des personnes spécialisées. Dans ce deuxième cas, l'information recueillie est moins importante en quantité, mais elle est plus sûre.
- **Le module d'indexation**: c'est la phase de structuration et de classification de l'information rapatriée. En général, les informations répertoriées sont:
l'adresse du site ;
le titre des pages ;
les mots clés, voire même, l'intégralité des pages.
Les méthodes d'indexation varient très fort d'un moteur à l'autre et ceci influence donc les résultats que l'on peut obtenir lors d'une requête.
- **Le module de requête et de présentation des résultats**: il s'agit d'une part de l'interface qui est présentée à l'utilisateur pour établir sa requête et visualiser les résultats et d'autre part d'un programme qui interprète et applique la requête de l'utilisateur.



Selon les moteurs, il y a actuellement deux approches de base:

Les répertoires validés.

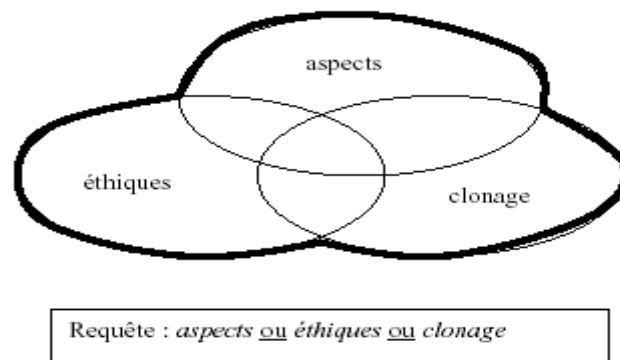
Il s'agit de constituer des répertoires organisés par **thèmes** ou **sujets** avec une validation, une classification humaine des ressources et une indexation automatique sur certaines zones. **YAHOO** est un exemple de moteur de ce type.

Les index.

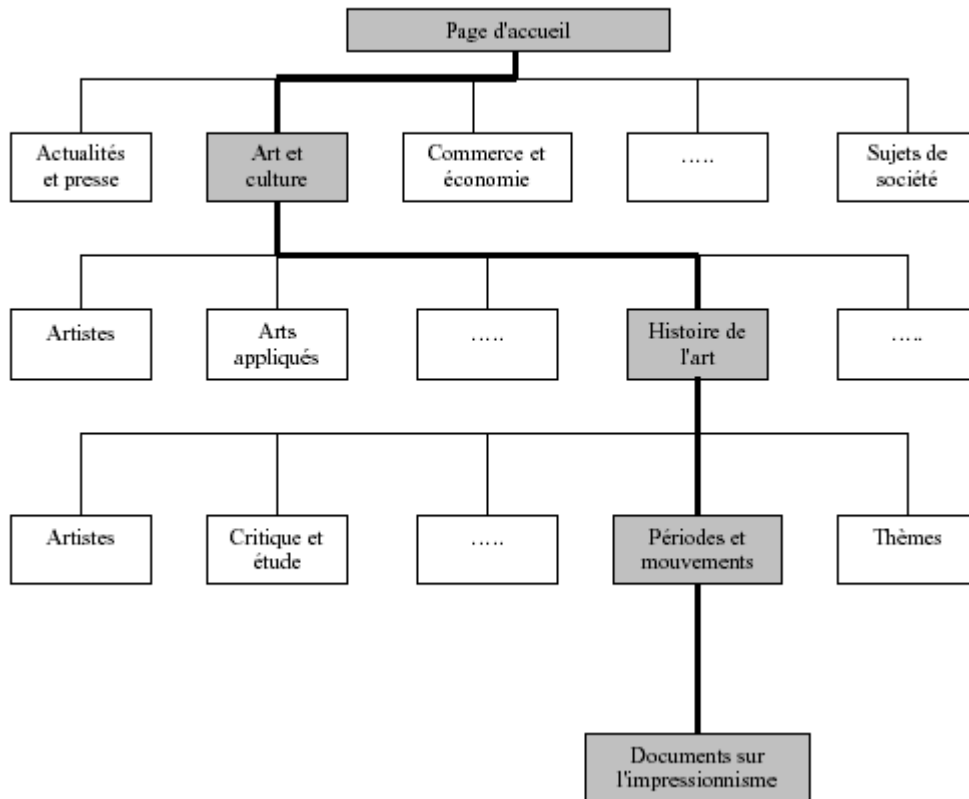
Il s'agit d'une collecte **automatique** de pages Web par un **robot**, suivie de l'indexation automatique des ressources. **ALTAVISTA** est un exemple de moteur de ce type.

1.3. Recherche d'informations

L'existence de deux types fondamentaux de moteurs de recherche nous amène assez naturellement à envisager deux formes de recherche d'informations: une recherche par thème basée sur l'exploitation des répertoires validés et une recherche par mot clé basée sur les index construits par les robots.



Arborescence d'une recherche par thème



2. Aspirer le web

Les pages de Web accessibles sur Internet représentent une quantité monstrueuse d'informations. Ceci ne signifie pas que toutes ces informations soient utiles et pertinentes.

Si on s'intéresse exclusivement aux techniques de capture de ces informations, on peut donner quelques lignes directrices intéressantes en ce qui concerne les textes et les images, ceci en vue de les intégrer dans d'autres documents, et essentiellement des documents réalisés avec un traitement de texte.

2.1. Du texte

La technique du copier-coller est bien connue. Elle s'applique aussi aux navigateurs. La sélection du texte se fait dans la fenêtre d'affichage. La commande coller ayant pour but d'envoyer la sélection dans le presse-papiers, seuls les caractères sont pris en considération.

2.2. Des images

Deux techniques sont à prendre en considération:

- le copier-coller
- l'enregistrement de l'image dans un format choisi

2.2.1. Copier-coller

Cette technique est la plus simple. Toutefois, sa conséquence est que l'image capturée, quel que soit son format initial, risque d'être intégrée dans le nouveau document dans un format qui ne permet plus beaucoup de transformations par la suite (format simplifié).

2.2.2. Enregistrer l'image sous...

La technique de l'enregistrement est plus intéressante en ce sens qu'elle permet à d'autres outils logiciels de les retravailler au besoin.

Les formats les plus courants pour des images du Web sont

- **GIF** (Graphic Interface Format)
- et **JPEG** (Joint Photographic Experts Group)

Qui présentent l'avantage d'une compression efficace. Mais le format BMP(bitmap) est aussi possible (bien que peu recommandé à cause de la taille).

Il est possible d'enregistrer l'image au format référencé dans le document, mais aussi au format **bitmap**, ce qui facilite transformations par la suite (avec un outil simple de dessin par exemple).

Dans un cas comme dans l'autre, les informations, une fois capturées, peuvent être collées ou insérées dans un document géré par un logiciel qui les accepte.

3. Le courrier électronique (E-mail)

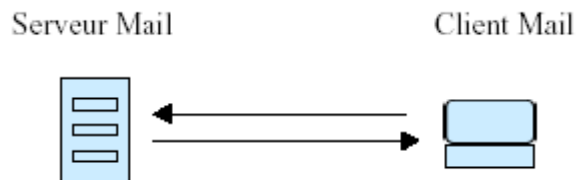
3.1. Introduction

Le courrier électronique est le service le plus utilisé sur l'Internet, après s'être répandu dans le monde professionnel, il est devenu l'un des outils favoris des particuliers. Quoi de plus normal, puisque ce service permet de communiquer en différé facilement et rapidement. De plus, il offre la possibilité **d'envoyer** et de **recevoir** des documents attachés au message, ce qui en fait un outil d'échange performant et intéressant. Ajoutons à cela qu'on peut tout aussi facilement échanger des messages entre 2 personnes qu'à l'intérieur de tout un groupe.

Nous allons essayer de comprendre le fonctionnement et l'utilisation du courrier électronique. Nous verrons que ce service fonctionne suivant le modèle client-serveur déjà abordé par ailleurs. Il faudra découvrir les particularités du dialogue de ces 2 partenaires pour bien comprendre ce service. Une bonne compréhension de ce fonctionnement devrait permettre une meilleure maîtrise des logiciels-clients associés à ce service.

3.2. Principes de base

Les applications de courrier électronique fonctionnent, elles aussi, selon le modèle client-serveur déjà abordé dans les paragraphes précédents.



Les 2 partenaires sont donc bien un **client** d'un côté, représenté par un logiciel-client, et un **serveur** de l'autre, représenté par un logiciel-serveur.

Ils sont reliés, comprenez plutôt connectés, par l'intermédiaire de l'Internet. ils peuvent donc dialoguer, bien sûr dans leur jargon (appelé plus académiquement **protocole**).

Rappelons-nous bien que le terme serveur-mail désigne la machine mais aussi, et surtout, le programme qui gère le service de courrier électronique et qui est installé (qui tourne) sur cette machine.

De même, le terme client-mail désigne la machine, mais aussi et surtout, le programme qui tourne sur la machine d'un utilisateur et qui permet à ce dernier de gérer son courrier électronique.

Pour accéder au service de courrier électronique, l'utilisateur doit, la plupart du temps, avoir un accès à l'Internet via un fournisseur.

Le grand principe de fonctionnement du courrier électronique ressemble à celui d'un autre service que nous utilisons quotidiennement, la **banque**.

Pour profiter des services offerts par une banque, il faut avoir un **compte** dans cette banque et pour réaliser ses opérations bancaires, du moins avec une carte de banque, il faut un **code secret**, qui protège le client..

De même, pour accéder au service de courrier électronique sur un serveur, l'utilisateur doit posséder Un **compte** de courrier électronique sur ce serveur. Ce compte est assorti d'un **mot de passe** et donne à l'utilisateur la possibilité de **recevoir** et **d'envoyer** du courrier et met à sa disposition une **boîte aux lettres**, pour le courrier entrant et le courrier sortant.

Remarque :

Il faut bien distinguer ce compte de courrier électronique de l'éventuel compte d'utilisateur sur le serveur. Pour rappel ce compte permet à l'utilisateur de se connecter au serveur et via cette connexion d'accéder, par exemple, à l'Internet. De même, il ne faut pas confondre les serveurs d'accès et de courrier électronique. Il peut s'agir de 2 machines différentes. Ceci n'a d'ailleurs pas d'implications pour l'utilisateur, qui ne doit pas se soucier de savoir quel est son serveur d'accès, ni quel est son serveur de courrier électronique.

Le compte de courrier électronique aura la forme : **compte@serveur** (par ex.: **cheval_y@yahoo.fr**).

Tout utilisateur, possédant un compte de courrier électronique de ce type se verra attribuer une adresse de courrier électronique (appelée aussi adresse E-mail) plus explicite du type **alias@domaine**.

C'est cette adresse qui apparaît dans le courrier électronique et qui doit être utilisée par une personne désirant envoyer un courrier à cet utilisateur. A ce compte est aussi associé un mot de passe, permettant à l'utilisateur d'avoir accès à son courrier entrant. C'est une manière de préserver la confidentialité du courrier.

Remarque :

Il est important de noter que l'adresse E-mail comporte un nom de domaine, alors que le compte de courrier électronique comporte un nom de serveur.

Une fois en possession de son compte de courrier électronique et après avoir installé, configuré et lancé son logiciel de courrier électronique, l'utilisateur pourra **composer** des messages et les **envoyer**, **recevoir** des messages, les **lire** et les **classer**. Ce sont là les principales fonctions (primitives) du logiciel-client de courrier électronique.

Voyons maintenant comment s'effectue l'échange de courrier entre 2 personnes, possédant chacune un compte de courrier électronique.

Supposons que ces 2 comptes se trouvent sur 2 serveurs différents, ce qui n'est pas obligatoire, mais rend l'explication plus claire.

L'utilisateur **A** va composer un message avec son logiciel de courrier électronique. Pour le moment cet utilisateur ne doit pas être connecté sur l'Internet, il compose simplement son message comme il taperait n'importe quel texte avec son logiciel de traitement de texte. Il doit aussi indiquer l'adresse E-mail de son correspondant, pour le message puisse lui parvenir.

Maintenant que l'utilisateur a composé son message, il désire l'envoyer à son correspondant, pour cela il faut qu'il soit connecté sur le serveur de son fournisseur d'accès. Soit il

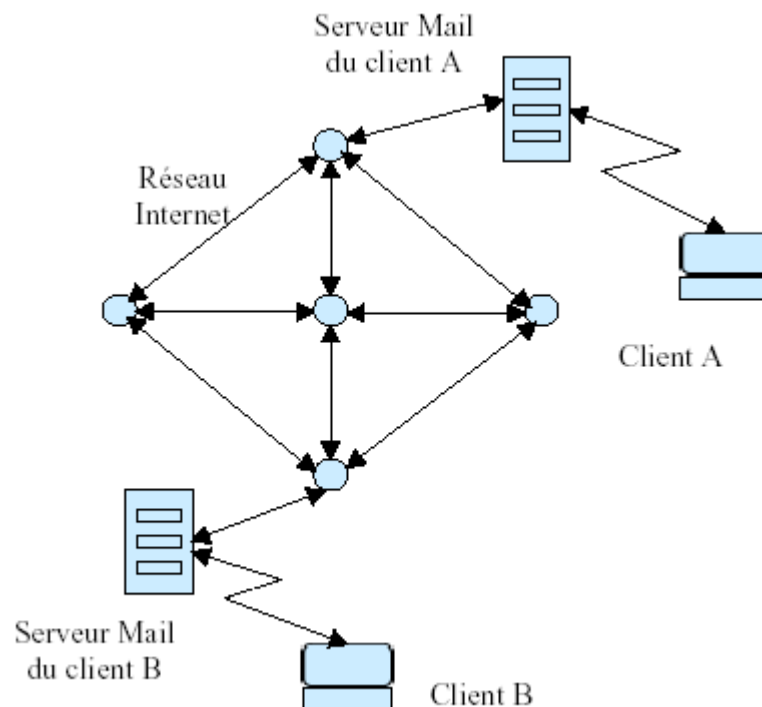
possède une connexion permanente, auquel cas il n'a rien à faire, soit il ne possède pas cette connexion et il doit l'établir, en général, via le réseau téléphonique.

Une fois la connexion établie, l'utilisateur envoie son message qui aboutit dans la boîte aux lettres de courrier sortant sur son serveur de courrier. De là, ce message sera envoyé sur le serveur de courrier du destinataire (utilisateur **B**), plus précisément dans sa boîte aux lettres de courrier entrant.

Ce message va rester sur le serveur jusqu'au moment où le destinataire (**B**) lance son logiciel-client de courrier électronique, se connecte et "demande" au serveur s'il y a du courrier dans sa boîte de courrier entrant.

Si effectivement il y a du nouveau courrier, ce qui est le cas dans notre exemple, le serveur-mail du client **B** va envoyer une copie du (des) nouveau(x) message(s) au logiciel-client de **B**.

Selon le paramétrage fait par l'utilisateur **B**, le logiciel-client **B** enverra (ou n'enverra pas) un ordre d'effacement du message au logiciel-serveur. Ceci a une conséquence très importante : si l'ordre d'effacement est envoyé, seule la machine-client **B** possédera une copie de ce message et celui-ci ne pourra plus être lu à partir d'une autre machine-client, même par le propriétaire du compte de courrier.



L'utilisateur peut aussi paramétrer son logiciel pour qu'une copie des messages reste sur la machine-serveur. Dans ce cas, l'utilisateur pourra consulter ses messages à partir d'autant de machines différentes qu'il le désire, à condition que les logiciels-clients de chaque machine soit configurés de cette manière et avec ses paramètres personnels.

Du fonctionnement global du service de courrier électronique, on retiendra une chose importante, il s'agit d'un service asynchrone, c'est-à-dire qui ne demande pas la présence simultanée des correspondants, un petit peu comme l'est le service postal par opposition au service téléphonique.

4. Configuration d'un logiciel de courrier électronique

La description du fonctionnement du service de courrier électronique, qui précède, met en évidence les différents paramètres de configuration d'un logiciel-client de courrier électronique.

Une bonne compréhension de ce fonctionnement est nécessaire pour configurer correctement votre logiciel-client de courrier électronique, selon l'utilisation que vous voulez en faire. Il ressort que l'utilisateur devra fournir au logiciel-client les renseignements suivants:

4.1. Paramètres utilisateur

Ces paramètres concernent le compte que l'utilisateur possède.

- Le compte de courrier électronique, appelé aussi compte **POP** (ou **POP account** dans les versions anglaises).
- l'adresse de courrier électronique (adresse E-mail), appelé aussi adresse de retour (ou **return address** dans les versions anglaises), par exemple `cheval_y@yahoo.fr`.

4.2. Paramètres serveur

Ces renseignements concernent la machine-serveur sur laquelle l'utilisateur possède son compte de courrier électronique.

- L'adresse du serveur **POP** (protocole pour le courrier entrant), par exemple **pop.menara.ma**.
- L'adresse du serveur **SMTP** (protocole pour le courrier sortant), par exemple **smtp.menara.ma**.

4.3. Paramètres de connexion

Pour configurer complètement le logiciel de courrier électronique, il faudra encore préciser le type de connexion au serveur, soit par l'intermédiaire d'un réseau local, soit par l'intermédiaire d'une ligne téléphonique.

Remarque : le logiciel-client de connexion est un autre logiciel, il suffit de préciser ici quel est le type de connexion et non pas comment cette connexion doit s'établir.

4.4. Paramètres du courrier

Concernant les messages, le paramètre le plus important est celui de la conservation du courrier sur le serveur.

Il existe 2 possibilités:

- **Laisser une copie** des messages sur le serveur, éventuellement pendant **x jours**.
- **Supprimer** les messages sur le serveur quand ils sont copiés sur le client.

Rappelons la conséquence importante de ce choix, si l'utilisateur choisit la première possibilité, il pourra lire et relire son courrier à partir de n'importe quelle machine connectée à l'Internet et configurée avec ses paramètres personnels. Si, au contraire, l'utilisateur choisit la deuxième possibilité, ceci ne sera plus possible.

5. Composition d'un message

Parmi les primitives de base d'un logiciel-client de courrier électronique, on trouve assez logiquement une fonction qui permet de créer (**composer**) un nouveau message. Notons bien que durant cette phase, il n'est pas nécessaire d'être connecté puisqu'il n'y a pas d'échange entre le client et le serveur.

Chaque message comporte 4 parties bien distinctes jouant chacune un rôle bien défini:

- **L'entête du message** qui permet d'indiquer quel(s) est (sont) le(s) destinataire(s) et quel est le sujet du message.
- **Le corps du message** qui contient le texte du message.

- **La signature du message** qui permet à l'auteur de se personnaliser.
- **Les objets attachés** (joint) au message, par exemple des documents qu'on veut envoyer.

5.1. L'entête du message

L'entête comporte:

- **Une zone "A" ou "TO"**, dans laquelle on introduit la (les) adresse(s) du (des) destinataire(s), très précisément l'adresse E-mail, ou un alias qui désigne l'adresse de courrier électronique d'une ou de plusieurs personnes. (voir. Carnet d'adresses)
- **Une zone "CC"** (copie carbone ou carbon copy) dans laquelle on peut introduire une ou plusieurs adresses de courrier électronique de personnes qui recevront aussi le message.
- **Une zone "CCI"** (copie carbone invisible) ou **"BCC"** (Blind carbon copy), dans laquelle on peut mettre une ou plusieurs adresses de courrier électronique qui recevront aussi le message, mais sans que les autres destinataires ne le sachent.
- **Une zone "OBJET" ou "SUBJECT"**, dans laquelle on peut écrire le sujet ou l'objet du message, ce qui permettra au destinataire de se faire une idée du contenu sans devoir le lire et aussi, par la suite, de classer ses messages par sujet.

5.2. Le corps du message

Il s'agit du texte écrit pour le destinataire, le message qu'on veut lui envoyer. Cette zone est "contrôlée" selon le logiciel de courrier électronique par un éditeur de texte, éventuellement par un "mini-traitement" de texte. Le contenu et la présentation du texte est libre et n'est soumise à aucune règle spécifique, Si ce n'est celle de la présentation, de la politesse, ...

5.3. La signature du message

L'utilisateur peut personnaliser la signature de son message et fournir ainsi des renseignements personnels plus précis que sa simple adresse de courrier électronique. L'insertion de la signature peut se faire automatiquement ou manuellement.

5.4. Les objets attaches (joint) au message

L'utilisateur a la possibilité, très intéressante et très utile, de joindre des fichiers à son message. Les fichiers peuvent être de n'importe quelle nature, documents provenant d'un traitement de texte, d'un tableur, une image, un fichier compressé, un exécutable, ...

L'utilisateur peut ainsi communiquer rapidement à son correspondant des documents. Bien sûr le destinataire doit posséder les logiciels nécessaires à la relecture de ces documents. Il est donc prudent de se renseigner sur les logiciels possédés par le destinataire.

6. Envoi et réception de courrier

Chaque logiciel-client de courrier électronique possède une primitive d'envoi et une primitive de rapatriement du courrier. Souvent ces 2 primitives sont combinées, l'envoi et le rapatriement se font en même temps.

Autant il n'était pas nécessaire d'être connecté pour rédiger son courrier, autant cette connexion est indispensable pour envoyer et "rapatrier" son courrier sur sa machine. En effet l'envoi et le rapatriement nécessitent un dialogue entre le logiciel-client de l'utilisateur et le logiciel-serveur. Avant d'utiliser les fonctions d'envoi et de rapatriement, il faut donc vérifier que la machine-client est bien connectée à l'Internet.

Rappelons une chose importante, l'envoi du courrier se fait sur la machine-serveur sur laquelle l'utilisateur possède un compte, de même le rapatriement des messages se fait à partir de la même machine-serveur et ces opérations n'ont rien à voir avec La machine-serveur de l'envoyeur (cas du Courrier entrant) ou du destinataire(cas du courrier sortant).

7. Gestion de son courrier électronique

7.1. Boîtes de courrier

Tous les logiciels de courrier électronique possèdent une boîte de réception, une boîte d'envoi et une poubelle, comme un(e) secrétaire bien organisée. La boîte de réception contient tous les messages qui sont entrés (qui ont été lus sur le serveur). Un signe distinctif permet de reconnaître les messages lus par l'utilisateur des messages non lus.

L'utilisateur peut créer des boîtes, en fait il s'agit de dossiers, dans lesquelles il classera son courrier en fonction des sujets ou de tout autre critère spécifique. Après la création des boîtes, la technique du glisser-coller est utilisée pour le classement.

Parmi tous les messages reçus, certains ne doivent pas être gardés, d'autres sont inintéressants, il faut donc pouvoir s'en débarrasser. La poubelle est là pour ça.

7.2. Carnet d'adresses

Faut-il que je retienne ou que je note (sur un bout de papier) les adresses électroniques de tous mes correspondants ? Question angoissante, qui, heureusement, a une réponse plutôt rassurante : le carnet d'adresses.

Les logiciels-clients de courrier électronique possèdent une fonction, appelée très souvent carnet d'adresses, qui permet de mémoriser les adresses E mail des correspondants en les associant avec un nom clair(**alias**). On peut alors utiliser ce nouvel alias pour envoyer un message.

Mieux encore ! On peut créer, dans le carnet d'adresses, un alias qui correspond à un groupe de personnes, c'est-à-dire lui associé plusieurs adresses électroniques.

7.3. Répondre et rediriger

Les logiciels-clients de courrier électronique possèdent encore 2 autres facilités intéressantes : **répondre** à un message et **rediriger** un message. Le terme facilité est utilisé à dessein, car ces fonctionnalités peuvent être réalisées à partir des primitives déjà vues, essentiellement "envoyer du courrier" et des primitives de base de l'environnement, comme par exemple copier-coller.

Première situation : j'ai reçu un message auquel je voudrais répondre point par point, dois-je créer un nouveau message, copier-coller chaque point du message original, taper ma réponse et l'envoyer?

Non, il suffit d'utiliser la fonction "répondre" qui créera automatiquement un nouveau message, placera l'adresse du destinataire et le texte du message original. Il suffira d'insérer les réponses ou commentaires à chaque point du message.

Deuxième situation : j'ai reçu un message qui ne me concernait pas directement ou je voudrais faire prendre connaissance de ce message à quelqu'un d'autre, dois-je créer un nouveau message, copier-coller le message, inscrire l'adresse E-mail du destinataire et l'envoyer?

Non, il suffit de rediriger le message. L'appel à cette fonction créera automatiquement un nouveau message contenant déjà le texte. Il suffira d'ajouter l'adresse du destinataire ou mieux de la reprendre dans le carnet d'adresses. Le destinataire final saura que le message a été redirigé.

Ces 2 exemples sont assez éloquentes sur la manière dont les logiciels fonctionnent. Il existe des primitives (fonctions de base) qui peuvent être regroupées d'une certaine manière pour obtenir une fonction (action) plus élaborée.

8. les listes de diffusion

Dans un des paragraphes précédents, nous avons abordé une fonction assez utile, le carnet d'adresses. Celui-ci permet, rappelons-le, d'associer à un alias une série d'adresses E-mail, c'est-à-dire de se créer une facilité pour travailler dans un groupe d'échange. Il subsiste cependant une lourdeur et une limitation à ce système :

Chaque membre du groupe doit encoder l'alias et les adresses associées, de plus les membres extérieurs à ce groupe ne savent pas qu'il existe.

Un certain nombre de serveurs offre un service appelé **liste de diffusions**, qui centralise des groupes de discussion, autour d'un thème défini. Les utilisateurs peuvent s'inscrire (et se désinscrire) à une liste, de manière plus ou moins automatique. Chaque membre de la liste de diffusion peut **déposer** un message, qui sera diffusé automatiquement à tous les autres membres du groupe.

La différence avec l'utilisation du carnet d'adresses d'un logiciel-client de courrier électronique, c'est que la gestion des inscriptions et de la diffusion des messages qui est gérée par le logiciel-serveur de liste de diffusion. La constitution du groupe peut évoluer dynamiquement sans intervention des membres du groupe.