République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud MAMMERI de Tizi-Ouzou
Faculté de Génie Electrique et d'Informatique
Département d'Informatique

Mémoire de Fin d'Etudes

En vue de l'obtention du diplôme Master 2 en informatique
***Option** : Conduite de projet informatique*

Thème

Filtrage en temps réel des microblogs exemple Twitter

Proposé et dirigé par :

Mr OUAMRANE Mohamed

Réalisé par :

LOUANCHI M^{ed} Amine

KEBIR Smail

Jury composé de :

Président :

Examinateurs:

Promotion 2011/2012

Résumé :

A l'heure où les sites de réseaux sociaux transforment les usages sur le Web, les échanges entre personnes deviennent de plus en plus faciles, ludiques et riches.

Le partage en temps réel de nouvelles, d'humeurs, et autres contenus (personnels ou personnellement sélectionnés) permet de tisser, de maintenir et de renforcer des liens sociaux entre personnes à des échelles encore inédites. Cependant, la quantité sans cesse croissante d'information circulant sur ces réseaux, souvent en temps réel, motive une régulation des signaux (ici appelées "interactions médiatisées"), de manière à réduire le temps nécessaire pour suivre ses réseaux sociaux, et modérer les interruptions induites, non favorables à une bonne productivité sur le traitement de tâches demandant une attention continue.

Dans le cadre de ce , nous avons développé un système de filtrage et de recommandation de ces signaux qui repose sur la similarité contextuelle entre utilisateurs, producteurs et consommateurs de ces signaux, pour évaluer leur pertinence. Notre approche consiste à agréger et interpréter les données de contexte sur les terminaux des utilisateurs, sous forme de mots-clés pondérés (tags), avant qu'elles ne puissent être exploitées par le serveur de recommandation, à la demande de l'utilisateur. Dans ce mémoire, nous présenterons un état de l'art couvrant la gestion de données contextuelles, les réseaux sociaux et leurs pratiques actuelles sur internet, et des techniques de recherche d'information. Ensuite, nous proposerons une formalisation de notre problématique de filtrage contextuel, l'implémentation d'une application de réseautage social d'entreprise, et nous discuterons les résultats expérimentaux obtenus auprès d'utilisateurs.

Mots-Clés: recherche d’informations, systèmes de recommandation, réseaux sociaux, surcharge informationnelle.

Abstract:

At a time when social networking sites revolutionize the usages on the Web, it has become rich, easy, and fun to share private or professional content. Sharing personal information in real-time (such as news, moods, etc…), supports the maintenance of social ties at a high scale. However, the information overload which emerged from the growing quantity of signals exchanged on these services, often in real-time, motivates a regulation of these signals (called "mediated interactions"), in order to reduce the temporal cost for maintaining social networks, and implied interruptions, which have a negative impact on productivity on tasks that require long-lasting attention.

In the frame of this thesis, we have developed a filtering and recommendation system that relies on contextual similarity between users that produce and consume social signals, as relevance criteria. In our approach, contextual information is aggregated and interpreted on users' terminal(s), before being submitted on-demand to a server in the form of a set of weighted tags. In this thesis, we present a broad state of the art on context-awareness, social networks and information retrieval, we propose a formalization of our filtering problem, and we implement and evaluate its application for enterprise social networking.

Keywords: information retrieval, systems, social networks, information overload

Remerciement

Tout d'abord, nous présentons nos remerciements les plus sincères à notre promoteur Mr. OUAMRANE Mohamed, pour nous avoir proposé un sujet aussi riche et pour nous avoir orientés durant la période de notre recherche.

Nous tenons aussi à remercier vivement les enseignants du département Informatique pour leur disponibilité et l'intérêt qu'ils ont eu pour notre travail.

Nous aimerions remercier également tous nos amis et camarades pour leur soutien et leur aide.

Introduction générale…….……11

Partie I……..13

Chapitre 1.....14

1. Filtrage d'information...…….……15
1.1. Définition……...……...15
1.2. Surveillance……..15
1.3. Filtrage en temps réel…….……...15
1.4. Filtrage et la géolocalisation…….16
1.4.1. Localisation par GSM……...16
1.4.2. Localisation par WIFI……...16
1.4.3. Localisation par IP……16
1.5. Que doit-on filtrer…….……16
1.5.1. La détection des contenus racistes……..……..17
1.5.2. Organisations criminelles……17
1.5.3. Chercheurs et enquêteurs……..……18
1.6. Technique DPI……..……18
1.6.1. Utilisation par les gouvernements……...…….19
1.6.2. Filtrage chez les FAI...…….……19
1.7. Serveur Proxy……...….21
1.7.1. Définition……...…21
1.7.2. Motif de conception de proxy…….……...…21
1.7.3. Proxy réseau……..……21
1.7.4. Accès internet……..…….22
1.7.5. Proxy de sécurité……..…….22
1.8. Pare-feu……..……..…...……..23
1.8.1. Définition……..23
1.8.2. Fonctionnement général……...…….23
1.8.3. Technologies utilisées……..…….……24
2. Les micros blogs…….25
2.1. Définition des blogs……..25
2.2. Mise en ligne du blog…….……...26
3. Les réseaux sociaux……26
3.1. Social……...…….26
3.2. Réseau……...……26
3.3. Réseau social……...……..26
3.4. Le Web 2.0……..……..27

Chapitre 2……29

1. Définition……29
1.1. Filtrage……29
1.2. Blocage……29
2. Filtrage……30
2.1. Efficacité d'un filtrage……30
2.1.1. Sous-blocage……30
2.1.2. Sur-blocage……30
2.1.3. Facilité de contournement……31
2.2. Coût d'un filtrage……31
2.3. Le filtrage par destination……31
2.3.1. Site web ou partie de site web……31
2.3.1.1. Filtrage IP……31
2.3.1.2. Filtrage par nom de domaine……32
2.3.1.3. Filtrage par URL……32
2.3.1.4. Filtrage dynamique (par contenu)……33
2.3.1.5. Filtrage hybride ……34
2.3.2. Messagerie Instantanée……34
2.3.2.1. Filtrage par IP……34
2.3.2.2. Filtrage protocolaire……35
2.3.2.3. Filtrage dynamique par zone géographique……36
2.3.3. Mail……36
2.3.3.1. Filtrage par IP……36
2.3.3.2. Filtrage dynamique (par contenu)……37
2.3.4. Usenet……38
2.3.4.1. Filtrage par groupe……38
2.3.4.2. Filtrage par hiérarchie……38
2.3.5. Paire à paire……39
2.3.5.1. Filtrage par protocole ……39
2.3.5.2. Filtrage dynamique……40
2.4. Filtrage par moyens……40
2.4.1. Introduction……40
2.4.2. Agir sur la source……41
2.4.3. Agir sur le milieu……41
2.4.3.1. Agir au niveau du cœur de réseau……41
2.4.3.2. Agir au niveau du Nœud de Raccordement Abonné……41
2.4.3.3. Agir au niveau des serveurs de noms(DNS) de l'opérateur……42
2.4.4. Agir à la destination……42
2.4.4.1. Sous la maitrise de l'utilisateur……42

2.4.4.2. Sous la maitrise d'un tiers……..……..42
2.5. Filtrage par type……..…..……43
2.5.1. Blocage sur les noms de domaine (DNS, DPI)……..……..……43
2.5.1.1. Objectif et terrain d'action……..43
2.5.1.2. Efficacité……..…43
2.5.2. Blocage sur les adresses IP (DNS, BGP)……..……44
2.5.2.1. Objectif et terrain d'action……..44
2.5.2.2. Efficacité……45
2.5.3. Filtrage sur les URL (DPI)……...45
2.5.3.1. Objectif et terrain d'action…….……45
2.5.3.2. Efficacité……45
2.5.4. Filtrage sur les contenus (DPI)……..…46
2.5.4.1. Objectif et terrain d'action…….……46
2.5.4.2. Efficacité……..…….……..46
2.5.5. Blocage sur les ports…….……46
2.5.5.1. Objectif et terrain d'action…….……46
2.5.5.2. Efficacité ……46
2.5.6. Filtrage hybride……..…47
2.5.6.1. Objectif et terrain d'action……..47
2.5.6.2. Efficacité……..…47
2.5.7. Filtrage sur les protocoles (DPI-traffic shaping)……..… 47
2.5.7.1. Objectif et terrain d'action……..47
2.5.7.2. Efficacité……..……47
3. Les enjeux……...…48
3.1. Enjeux de filtrage……..48
3.1.1. Le filtrage à l'intersection des techniques et des usages……...…48
3.1.2. Filtrage et responsabilité…….……..48
3.2. Filtrage par un tiers……...……48
3.2.1. Identification des tiers……...48
3.2.2. Une moindre responsabilité et une moindre maîtrise pour l'internaute……...…….48
3.2.3. Un respect de la légalité plus aisé.……50
3.2.4. Un enjeu éthique important……..…50
3.3. Filtrage sous la maitrise de l'internaute……...…….50
3.3.1. Une responsabilisation plus forte de l'internaute……...…..50
3.3.2. Un risque d'inégalité face au filtrage……...…….51
3.4. Filtrage et avenir d'Internet : les variables risques liées au filtrage ……..51

4. Conclusion……..……53
4.1. La question de la sécurité……..53
4.2. La question de l'innovation…….…….53

Chapitre 3.....55

1. Présentation de Twitter……..…….56
   1.1. Principe de fonctionnement…….……56
   1.2. Utilisation de Twitter…….57
2. Glossaire des termes clés…….……..58
   2.1. Tweet…….58
   2.2. Trending Topics…….…….……58
   2.3. Followers……..…….……..59
   2.4. Following…….……..…….……59
   2.5. Favorites…….…….……59
   2.6. Public timeline…….……..…….…..59
   2.7. Private Timeline…….…….……59
   2.8. Direct message……..…….…….60
   2.9. @reply…….60
3. Conditions d'utilisation de Twitter et charte de respect de la vie privée…...…60
   3.1. Conditions d'utilisation du service…….60
   3.2. Définition de vie privée…….……..61
   3.3. Charte de respect de la vie privée……..……..62
4. Risques liés à l'utilisation de Twitter……...……..……62
   4.1. Propagation de vers…….……62
   4.2. Indisponibilité de Twitter……63
   4.3. Usurpation d'identité……...63
   4.4. Tinyurl et bit.ly……63
   4.5. Spam……..……..63
   4.6. Relayer des informations non vérifiées……...63
   4.7. Contrôle de la diffusion de données personnelles……..64
5. Création d'un compte Twitter……..……..64
   5.1. Étape #0……...64
   5.2. Étape #1…….……..65
   5.3. Étape #2…….……..68
   5.4. Étape #3……..…….69
6. Suppression de messages et suppression de compte Twitter……..…70
   6.1. Effacer ses messages…….……..70
   6.2. Supprimer son compte Twitter……...…….71
   6.3. Procédure à suivre…….……..71

Partie II…….74

Chapitre 4…….75

1. Introduction……76
2. Objectif de notre application…….....76
3. Présentation de l'UML……...76
   3.1. Langage de modélisation……77
   3.2. Objectif de l'UML……...77
   3.3. Avantage de l'UML……78
4. Spécification……...78
   4.1. Diagramme de contexte……...78
   4.2. Diagramme des cas d'utilisations……78
   4.3. Diagramme de classe……...81
5. Conception……..83
   5.1. Diagramme de séquence……..83
6. Conclusion…….90

Chapitre 5…….91

1. Introduction……92
2. Langage de programmation utilisé…….92
3. Présentation de l'environnement de développement……..93
   3.1. Eclipse…….93
   3.2. Serveur de base de données PostgreSQL……93
      3.2.1. Définition……...93
      3.2.2. Quelques fonctionnalités de PostgreSQL……..94
      3.2.3. Triggers et contraintes……..94
      3.2.4. Les langages de procédure sous PostgreSQL……...94
      3.2.5. Portabilité……..95
4. Implémentation et description de notre Application……..95
   4.1. Coté des utilisateurs …….....95
      4.1.1. Espace administrateur ……...……96
5. Conclusion……101

Conclusion générale 102

Acronymes……...……104

Bibliographies……...105

Liste des figures :

Figure 1 : Principe du proxy réseau……..21

Figure 2 : Un pare-feu, représenté par un mur de briques, pour cloisonner le réseau privé…….23

Figure 3 : Pare-feu routeur, avec une zone DMZ…….24

Figure 4 : Espace de tweet……..…..56

Figure 5 : Interface utilisateur……..…58

Figure 6 : Page d'accueil de Twitter……...….65

Figure 7 : Page pour saisie les informations personnelles……..….67

Figure 8 : Interface pour se connecter à d'autre compte……..68

Figure 9 : Interface pour connecter à Twitter……...…69

Figure 10 : Interface pour sélectionner d'autre compte……...69

Figure 11 : Interface où on écrit les messages……..70

Figure 12 : Interface pour suprême un compte…….72

Figure 13 : Interface affiche les conséquences…….72

Figure 14 : Interface qui affiche la suppression de compte……..73

Figure 15 : Différent étapes de filtrage……...…….77

Figure 16 : Diagramme de contexte…….79

Figure 17 : Diagramme de cas d'utilisation générale……...…80

Figure 18 : Diagramme de cas d'utilisation des utilisateurs……80

Figure 19 : Diagramme Use Cases realatif à l'administrateur…….81

Figure 20 : Diagramme de cas d'utilisation détaillé «configuration »……82

Figure 21 : Diagramme de classe…….…83

Figure 22 : Diagramme de séquence « authentification »…….……85

Figure 23 : Diagramme de séquence de filtrage……...86

Figure 2 4 : Diagramme de séquence « ajout d'un mot-clé »……...……....87

Figure 25: Diagramme de séquence « Filtrage par profils »…….……..…….88

Figure 26: Diagramme de séquence « Filtrage par utilisateur »…….……89

Figure 27: Diagramme de séquence « Filtrage par pays »…….…….…….90

Figure 28: Page d'accueil administratif……96

Figure 29: Espace administratif……97

Figure 30: Espace pour vérifier l'émetteur…….98

Figure 31: Filtrage par IP……99

Figure 32: La suppression d'un profil…….99

Figure 33: L'affichage des profils……100

Figure 34: Ajout d'un mot clé……....101

Liste des tableaux :

Tableau 1 : Les informations demandées lors d'une création d'un compte…….67

Tableau 2 : Tableau des cas d'utilisations……81

Introduction générale :

« Quand les vents du changement soufflent,

certains construisent des abris, et d'autres des moulins. »

Proverbe chinois

Le succès des réseaux sociaux ne fait plus aucun doute et leurs taux d'activité ont atteint des niveaux sans précédent. Des centaines de millions d'internautes sont inscrits dans ces réseaux.

Ils échangent via des forums, maintiennent des blogs, racontent leurs dernières pensées, humeurs ou activités en quelques mots... Le développement des outils mobiles tels que les téléphones portables, permettant de contribuer à ces réseaux de n'importe quel endroit, a favorisé l'émergence de ces nouvelles pratiques. Twitter est l'un de ces réseaux. Il permet aux internautes de « microblogguer », c'est-à-dire d'envoyer des messages courts, des « tweets » de 140 caractères uniquement et de lire les messages d'autres utilisateurs. Les internautes restent ainsi connectés à leurs amis, familles ou collègues via le réseau (MILSTEIN 2008). En 2010, plus de 15 millions d'utilisateurs utilisent Twitter et plus de 6M de tweets sont produits chaque jour.

Dernièrement, Twitter a mis en place des APIs qui permettent de rechercher des messages selon différents critères et d'y associer des informations comme la date, des informations personnelles sur l'auteur du message issues de son profil (nom, bibliographie, goûts, etc.), sa situation géographique… Ces flux de documents représentent une incroyable richesse en termes d'exploration automatique.

Une récente étude conduite par le cabinet Burson-Marstellerx montre que « les deux tiers des 100 plus grandes entreprises dans le monde disposent d'au moins un compte sur le service de micro-blogging Twitter. Plus de la moitié des entreprises ont également au moins une page fan sur le réseau social Facebook, la moitié aussi sont dotées d'une chaîne sur la plate-forme vidéo de Google, YouTube. Les blogs ne sont plus utilisés que par un tiers de ces entreprises ». [Françoise Papa]

Twitter. Il fait état des motivations conduisant à la mise en œuvre du filtrage dans une société, des options techniques disponibles à cette fin.

Le présent mémoire propose une analyse détaillée de l'état actuel de la situation en matière de filtrage d'Internet et des observations concernant l'efficacité d'une telle

mesure et son impact, tant sur la lutte contre la cybercriminalité, que sur le maintien de la démocratie et de la sécurité des individus.

La question de savoir où se trouve le meilleur équilibre entre la protection de l'enfance et la protection des libertés démocratiques est très complexe, et doit être tranchée in fine à un niveau national, dans chaque pays, au terme d'un large débat entre les acteurs concernés, débat devant tenir compte des instruments internationaux contraignants, tels que la Convention européenne des droits de l'Homme (CEDH).

Selon les membres du Parlement européen, le libre accès à Internet, sans ingérence, est un droit d'une considérable importance. Internet est « une vaste plate-forme pour l'expression culturelle, l'accès à la connaissance et la participation démocratique à la créativité européenne, créant des ponts entre générations dans la société de l'information», dont l'accès est protégé par le droit à la liberté d'expression, même s'il n'est pas actuellement considéré comme un droit fondamental autonome.

De son côté, le web 2.0, via les réseaux sociaux, blogs et autres forums a élargi l'espace public à de nouvelles formes de prises de parole. Internet, plus qu'un média de communication et d'information, a renouvelé les possibilités de critique et d'action, devenant une forme politique à part entière.

De la campagne d'Obama aux révoltes arabes, de plus en plus d'événements politiques et sociétaux se sont initiés ou développés sur le Net.

Afin de présenter le travail qui nous a été assigné. Nous avons organisé notre travail en deux parties principales: La première partie concerne l'aspect théorique de notre projet, dans cette partie nous présenterons une synthèse sur le microblog Twitter et le filtrage.

La deuxième partie concerne l'aspect pratique de notre travail, cette partie sera divisée en deux chapitres.

Analyse et Conception : Dans ce chapitre nous allons présenter l'analyse de l'existant décisionnel.

Réalisation: Dans ce chapitre nous expliquerons la réalisation de notre programme de filtrage de contenu et les infrastructures logicielles utilisées, comme nous présenterons aussi quelques interfaces de notre application.

Nous avons créé un mini chat pour notre application.

Bien évidemment nous finirons notre travail par une conclusion générale.

Chapitre I

Filtrage et législation

Chapitre I

Filtrage et législation

1. Filtrage d'information:

1.1. Définition :

La quantité d'information produite de nos jours est devenue astronomique, et les supports de diffusion se multiplient, que ce soit le web, les chaînes télévisées, les radios, les téléphones cellulaires, etc. Les modalités d'accès à cette information évoluent donc elles aussi pour correspondre à ces nouveaux besoins.

Un système de filtrage d'information permet d'extraire, à partir d'un flot d'informations (documents), celles qui sont susceptibles d'intéresser un utilisateur ou un groupe d'utilisateurs ayant des besoins en information relativement stables appelés profils.

Le filtrage de l'information pourrait être défini comme étant le processus qui vise à filtrer les informations d'un flux pour ne faire parvenir que celles qui intéressent l'utilisateur.

Le filtrage est un ensemble de solutions techniques visant à limiter l'accès à certaine sites normalement accessibles sur le réseau Internet ayant pour objectifs le contrôle parental, la protection des enfants contre des contenus inappropriés, les restrictions d'un accès d'entreprise à un usage professionnel ou la protection des libertés individuelles. **[Réf. 14]**

1.2. Surveillance :

Une loi définit la surveillance comme « toute activité qui, opérée au moyen d'instruments techniques, consiste en l'observation, la collecte ou l'enregistrement de manière non occasionnelle des données à caractère personnel d'une ou de plusieurs personnes, relatives à des comportements, des mouvements, des communications ou à l'utilisation d'appareils électroniques et informatisés ». **[Réf. 6]**

1.3. Filtrage en temps réel :

Filtrage en temps réel se fait avant que le destinataire reçoit le message, puisque la messagerie instantanée se fait en temps réel donc c'est la même chose pour le filtrage afin que ce dernier ne bloc pas la discussion entre les internautes.

Un service de microblogage est à la fois un moyen de communication et un système de collaboration qui permet le partage et la diffusion de messages textuels. En comparaisons avec les autres réseaux sociaux sur le Web, les articles de microblog sont particulièrement courts et soumis en temps réel pour rapporter un événement récent. **[Réf. 2]**

Selon les statistiques officielles publiées en mars 2011, environ 140 millions de tweets sont chaque jour sur Twitter. Avec ce taux important, les données générées par les microblogs sont de plus en plus disponibles. Cependant, le filtrage doit se faire en temps réel pour arrivées aux meilleurs résultats.

1.4. Filtrage et la géo localisation :

1.4.1. Localisation par GSM :

GeoMobile.be est un service édité par la PME belge NETIKA Internet & Mobile Solution. Il est destiné uniquement aux entreprises, typiquement pour la gestion de flottes de véhicules ou de personnel mobile. **[Réf. 7]**

1.4.2. Localisation par WIFI :

HereCast est application de services géolocalisés basée sur les routeurs WIFI les plus proches. Chaque routeur WIFI émet un numéro d'identification unique. Comme les Access Point (AP) sont généralement fixes, le logiciel est capable de donner la position approximative de l'utilisateur.

La base de données des AP est mise à jours par les utilisateurs du logiciel. La position n'est pas donnée sous forme de coordonnées géographique mais sous forme d'une information textuelle comme par exemple « Sale de lecture informatique ». **[Réf. 7]**

1.4.3. Localisation par IP :

Windows Local Live est un service de carte sur Internet développé par Microsoft. Il possède un bouton « Locate Me » qui peut essayer de trouver la position d'un utilisateur à partir d'une liste d'Access Point (comme HereCast) ou bien à partir de son adresse IP (avec des bases similaires à ce que développe MaxWind). La carte est alors zoomée autour de la position de l'utilisateur.

Finalement, IPv6 intègre une gestion de la mobilité en cours de finalisation.**[http://local.live.com]**

1.5. Que doit-on filtrer ?

Le filtrage des contenus illégaux sur Internet peut être vu non seulement comme un moyen d'agir sur ceux qui transgressent la loi en rendant disponibles des contenus (les producteurs), mais aussi comme un outil de prévention pour empêcher les utilisateurs de télécharger des contenus illégaux (les consommateurs).

- Le producteur de contenu illégal ou le fournisseur de contenu illégal.
- Le consommateur de contenu illégal. **[Réf. 20]**

1.5.1. La détection des contenus racistes :

Le filtrage de contenus illicites sur internet est une problématique difficile. Le système de classification textuelle par apprentissage automatique nécessitant peu reçoit de d'interventions humaines. Ces techniques, traditionnellement utilisées avec des catégories définies par leur sujet (économique ou sport par exemple), sont fondés par la présence ou d'absence de mots. Nous présenterons une évaluation de ces techniques pour le filtrage de contenus racistes. Contrairement aux cas traditionnels, les documents ne doivent pas être catégorisés suivant le sujet suivant le point de vue énoncé raciste.

Le sens des mots clés est susceptible de changer, à la fois en synchronie(en fonction du contexte) et en diachronie (selon les époques). Ainsi, nègre reçoit des valeurs radicalement différentes selon qu'il s'agit de l'art nègre, de la tète de nègre, du nègre d'un écrivain tel que le mot était couramment usité au XVIIIème. En somme, l'insulte sale nègre n'est qu'un cas d'instanciation parmi d'autres.

Les phénomènes linguistiques de créativité lexicale rendent également la tache plus ardue. Par exemple, le verlan (bougnoule, gnou, rabzas), la modification de l'orthographe (nègre au lieu de naigre). Tous ces exemples montrent que les mots clés doivent aussi être mis à jour régulièrement. **[Réf. 6]**

1.5.2. Organisations criminelles :

Suivant leurs spécialités, elles vont essayer de vendre des objets volés ou illicites (drogues, armes, matière nucléaire), proposer et fournir des organes volés.

Elles vont viser des personnes précises pour faire partie de leur group ou pour obtenir quelque chose (argent, puissance, organes, filles, etc.).

Voici des cas sur Internet dans lesquels des organisations criminelles horizontales sont particulièrement impliqués au travers des systèmes à haute technologies:

- ✓ Ventes d'objets volés (œuvres d'art, armes, véhicules);

- ✓ Préparation d’attaques terroristes;
- ✓ Utilisation de cartes de crédit volées;
- ✓ Cyber blanchiment;
- ✓ Trafic de drogues ou achat de drogues;
- ✓ Production et dissémination d’images d’enfants abusés (pornographie infantile);
- ✓ Propagande sur mesure et communication de groupes extrémiste
- ✓ Filoutage et vol d’identité;
- ✓ Cyber terrorisme;
- ✓ Kidnapping;
- ✓ Piratage de logiciel;
- ✓ Fraude et contrefaçon;
- ✓ Propagation de "crimewares".

Vous pouvez également y trouver des gangs qui échangent des informations sur leurs mauvais coups. **[Réf. 3]**

1.5.3. Chercheurs et enquêteurs :

Les psychologues vont utiliser les profils pour classifier les personnes suivant différent modèles et aider les entreprises, les armées, les sectes, les agences gouvernementales et autres, à sélectionner le mail leur profil sans avoir la nécessité d’un entretien.

Les psychologues travaillant pour la police vont utiliser des profils psychologiques le cas d’investigation criminelle afin de trouver le suspect.

Les statisticiens vont utiliser les informations pour supporter les compagnes politiques ou publicitaires.

Les enquêteurs vont essayer de trouver des preuves dans le cas d’une investigation ou d’un procès. **[Réf. 6]**

1.6. Technique DPI :

En informatique, le Deep Packet Inspection (DPI), en français Inspection des Paquets en Profondeur, est l'activité pour un équipement d'infrastructure de réseau d'analyser le contenu (au-delà de l'en-tête) d'un paquet réseau (paquet IP le plus souvent) de façon à en tirer des statistiques, à filtrer ceux-

ci ou à détecter des intrusions, du spam ou tout autre contenu prédéfini. Le DPI peut servir notamment à la censure sur Internet ou dans le cadre de dispositifs de protection de la propriété intellectuelle.

Il s'oppose au *Stateful Packet Inspection*, qui ne concerne que l'analyse de l'en-tête des paquets. Le DPI peut provoquer un ralentissement sensible du trafic là où il est déployé.

1.6.1. Utilisation par les gouvernements :

En plus d'utiliser le DPI pour renforcer la sécurité de leurs réseaux, les gouvernements d'Amérique du Nord, d'Europe et d'Asie l'utilisent pour différents usages comme la surveillance et la censure. Ainsi, l'Iran utilise un tel système depuis 2008, fourni par Nokia Siemens Networks (NSN).
Les premières tentatives de contrôle des communications se sont traduites par la création d'un *Traffic Access Point* (TAP), un serveur tiers (proxy) connecté à un appareil de surveillance gouvernemental ; mais ces techniques ne sont plus d'actualité dans le cadre des nouveaux réseaux. Le DPI fait donc aujourd'hui partie des techniques de substitution qui remplissent des fonctions équivalentes, et peuvent être mises en œuvre par décision d'une cour de justice pour accéder aux flux de données d'un individu en particulier. Aux États-Unis, cet usage est soumis au CALEA (*Communications Assistance for Law Enforcement Act*).
Le dictateur Mouammar Kadhafi a utilisé le système Eagle de la société Amesys pour repérer et espionner ses opposants. Ce système de surveillance massive et d'interception de communications électroniques est également installé en France, sans que son usage soit clair à ce jour.

1.6.2. Filtrage chez les FAI :

Les fournisseurs d'accès à internet font d'avantage que fournir la bande passante. Ils sont ont leurs propres serveurs DNS et ils « routent » les messages des internautes. Ils ont donc le pouvoir technique. Le plus simple pour l'autorité de transmettre ses requêtes aux F.A.I. pour qu'ils « nettoient » le réseau.

Le DPI peut être mis en place par les FAI pour sécuriser leurs réseaux internes ; mais cette technologie peut aussi s'appliquer aux clients eux-mêmes, pour intercepter des communications illégales, pour mettre en place de la publicité ciblée, pour améliorer la qualité du service, pour offrir des services tiers, ou dans le cadre de la protection de la propriété intellectuelle.

- Parce qu'ils acheminent tout le trafic de leurs clients, les FAI peuvent en effet surveiller leurs habitudes de navigation de manière très détaillée et connaître ainsi leurs centres d'intérêt

(puis revendre ces informations à des entreprises spécialisées dans la publicité ciblée comme Phorm, NebuAd et Front Porch).

- L'usage du DPI a aussi été envisagé par le Parlement néerlandais dans un rapport de 2009 en tant que mesure qui aurait été ouverte aux parties tiers pour renforcer la surveillance du respect de la propriété intellectuelle, visant plus particulièrement à réprimer le téléchargement de contenu protégé sous copyright. Suite à des critiques émanant d'ONGs, cette proposition devrait être abandonnée. En France, après l'arrivée de l'HADOPI et de la mise en route de l'analyse des échanges peer to peer, le DPI est évalué, mais l'HADOPI ne montre pas (pour l'instant du moins) la volonté de l'utiliser dans le cadre de son activité.
- Des fournisseurs d'accès affirment que les échanges de type *peer-to-peer* (P2P) posent un problème de trafic ; typiquement, dans le cadre du partage de fichiers (musique, vidéos, documents), la large taille des fichiers transférés nécessite une capacité accrue des réseaux. Le DPI leur permet de vendre l'idée d'une répartition plus juste de la bande passante, et d'éviter les congestions du réseau... En complément, la priorité peut être accordée à des services comme la VoIP ou les appels en vidéo-conférences, qui nécessitent un temps de latence moindre. Cette approche est privilégiée pour une attribution dynamique de la bande passante.
- Le recours au DPI par les FAI pose le problème du respect d'un certain niveau de service (service level agreement) dû aux clients (le contrôle des données ralentissant les débits entrant/sortant), et celui du respect de la vie privée (le DPI permet de connaître le contenu de tous les paquets transférés, des e-mails envoyés ou reçus aux sites web visités, en passant par les partages de musique, de vidéo ou de logiciels ; il permet aussi d'interdire les connexions à certaines adresses IP ou l'usage de certains protocoles, d'identifier certains usages ou le recours à certaines applications).

1.7. Serveur Proxy (Serveur mandataire) :

1.7.1. Définition :

Proxy est un terme informatique général qui désigne un composant logiciel qui se place entre deux autres pour faciliter ou surveiller leurs échanges.

Dans le cadre plus précis des réseaux informatiques, un **proxy** est alors un programme servant d'intermédiaire pour accéder à un autre réseau, généralement internet.

Par extension, on appelle aussi **proxy** un matériel (un serveur par exemple) mis en place pour assurer le fonctionnement de tels services.

1.7.2. Motif de conception proxy :

Dans son sens générique, un proxy est un motif de conception (design pattern) utilisé dans n'importe quel programme informatique.

Pour prendre une analogie du monde réel, si deux personnes qui ne parlent pas la même langue veulent se parler, elles vont demander l'aide d'un interprète. C'est un exemple de proxy. En informatique, un cas similaire pourra être celui de deux programmes utilisant des technologies différentes qui devraient communiquer entre eux. On pourrait alors utiliser un proxy pour traduire leurs échanges dans les deux sens. **[Réf. 4]**

1.7.3. Proxy réseau :

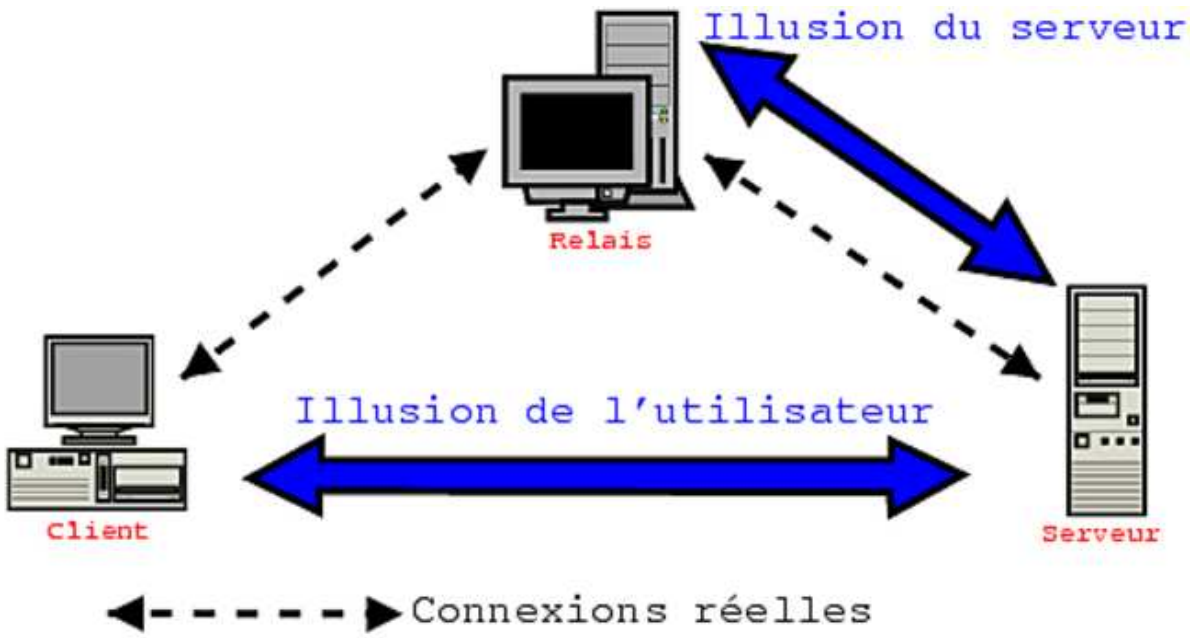


Figure 1 : Principe du proxy réseau

Dans l'environnement plus particulier des réseaux, un **serveur mandataire** ou ***proxy*** (de l'anglais) est ainsi une application informatique client-serveur qui a pour fonction de relayer des requêtes entre une application cliente et une application serveur (couches 5 à 7 du modèle OSI). Les serveurs mandataires sont notamment utilisés pour assurer les fonctions suivantes :

- Accélération de la navigation : mémoire cache, compression des données, filtrage des publicités ou des contenus lourds (java, flash);
- La journalisation des requêtes (logging) ;
- La sécurité du réseau local ;
- Le filtrage et l'anonymat. **[Réf. 4]**

1.7.4. Accès internet :

Les fournisseurs d'accès à internet (FAI) peuvent utiliser des proxys pour la connexion de leurs abonnés. Il faut pour cela que l'abonné paramètre correctement son système (via un logiciel d'installation fourni par le FAI). Mais il est également possible que le fournisseur d'accès utilise un proxy transparent (sans configuration par l'utilisateur).

Ce proxy permet aux FAI de connaître les habitudes de navigation de leurs abonnés.

En Algérie :

Le Cerist est l'hébergeur le plus important en Algérie. Il est historiquement le premier fournisseur d'accès Internet en Algérie. Selon Djamel Benradame, journaliste à Algérie-interface : « les autorités vont perfectionner leurs techniques de contrôle au fur et à mesure de la libéralisation des télécom ».

Des emails qui transitent par le Cerist sont déjà, de plusieurs heures à plusieurs jours en retard, ou disparaissent du réseau. Pour l'instant, nous mettons cela sur le compte d'un débit médiocre et de structures obsolètes du fournisseur d'accès public. La loi régissant l'information diffusée sur l'immense réseau n'est qu'à ses débuts d'où on trouve un vide juridique. La loi du 16 août 2009 portant règles particulière relatives à la prévention et à la lutte contre les infractions liées aux technologies de l'information et de la communication

1.7.5. Proxy de sécurité

L'utilité des serveurs mandataires est importante, notamment dans le cadre de la sécurisation des systèmes d'information.

Par exemple, il est presque systématique en entreprise ou dans les établissements scolaires que l'accès internet se fasse à travers un proxy. L'internaute ne voit pas la différence, sauf quand il tente de naviguer sur un site interdit, auquel cas il pourra recevoir un message d'erreur. Il se peut aussi qu'une boite de dialogue s'ouvre et demande un identifiant et un mot de passe avant de pouvoir surfer sur internet.

1.8. Pare-feu :

1.8.1. Définition :

Un **pare-feu**, ou **firewall** (de l'anglais), est un logiciel et/ou un matériel, permettant de faire respecter la politique de sécurité du réseau, celle-ci définissant quels sont les types de communication autorisés sur ce réseau informatique. Il mesure la prévention des applications et des paquets.

[Réf. 21]

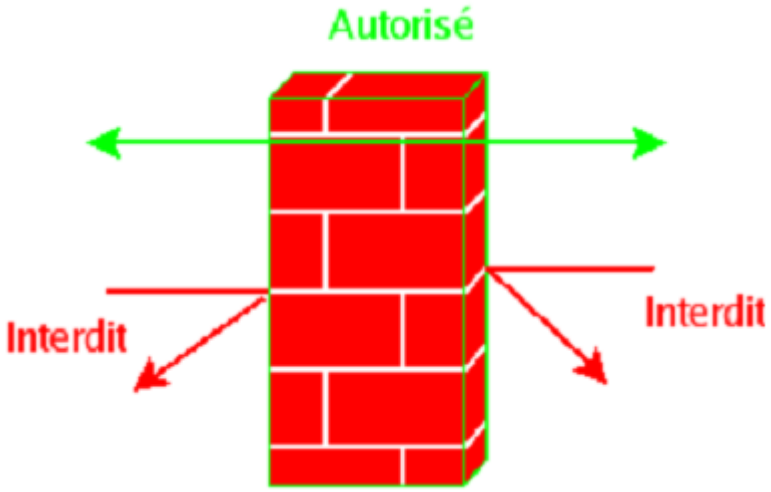


Figure 2 : Un pare-feu, représenté par un mur de briques, pour cloisonner le réseau privé.

1.8.2. Fonctionnement général :

Le pare-feu était jusqu'à ces dernières années considéré comme une des pierres angulaires de la sécurité d'un réseau informatique (il perd en importance au fur et à mesure que les communications basculent vers le HTTP sur SSL, court-circuitant tout filtrage). Il permet d'appliquer une politique d'accès aux ressources réseau (serveurs).

Il a pour principale tâche de contrôler le trafic entre différentes zones de confiance, en filtrant les flux de données qui y transitent. Généralement, les zones de confiance incluent Internet (une zone dont la confiance est nulle) et au moins un réseau interne (une zone dont la confiance est plus importante).

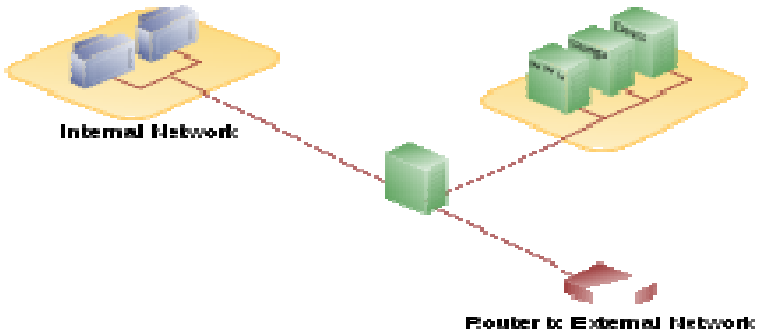


Figure 3 : Pare-feu routeur, avec une zone DMZ

Le but est de fournir une connectivité contrôlée et maîtrisée entre des zones de différents niveaux de confiance, grâce à l'application de la politique de sécurité et d'un modèle de connexion basé sur le principe du moindre privilège.
Le filtrage se fait selon divers critères. Les plus courants sont :

- L'origine ou la destination des paquets (adresse IP, ports TCP ou UDP, interface réseau, etc.) ;
- Les options contenues dans les données (fragmentation, validité, etc.) ;
- Les données elles-mêmes (taille, correspondance à un motif, etc.) ;
- les utilisateurs pour les plus récents.

Un pare-feu fait souvent office de routeur et permet ainsi d'isoler le réseau en plusieurs zones de sécurité appelées zones démilitarisées ou DMZ. Ces zones sont séparées suivant le niveau de confiance qu'on leur porte. **[Réf. 18]**

Enfin, le pare-feu est également souvent extrémité de tunnel IPsec ou SSL. L'intégration du filtrage de flux et de la gestion du tunnel est en effet nécessaire pour pouvoir à la fois protéger le trafic en confidentialité et intégrité et filtrer ce qui passe dans le tunnel. C'est le cas notamment de plusieurs produits du commerce nommés dans la liste ci-dessous.

1.8.3. Technologies utilisées :

Les pare-feu récents embarquent de plus en plus de fonctionnalités, parmi lesquelles on peut citer :

- Filtrage sur adresses IP / protocole,
- Inspection *stateful* et applicative,
- Intelligence artificielle pour détecter le trafic anormal,
- Filtrage applicatif :
  - HTTP (restriction des URL accessibles),
  - Courriel (Anti-pourriel),
  - Logiciel antivirus, anti-logiciel malveillant
- Traduction d'adresse réseau,
- Tunnels IPsec, PPTP, L2TP,
- Identification des connexions,
- Serveurs de protocoles de connexion (telnet, SSH), de protocoles de transfert de fichier (SCP),
- Clients de protocoles de transfert de fichier (TFTP),

- Serveur Web pour offrir une interface de configuration agréable,
- Serveur mandataire (« *proxy* » en anglais),
- Système de détection d'intrusion (« IDS » en anglais)
- Système de prévention d'intrusion (« IPS » en anglais) **[Réf. 18]**

2. Les micros blogs :

Un service de *microblogage* est à la fois un moyen de communication et un système de collaboration qui permet le partage et la diffusion de message textuels. En comparaison avec les autres réseaux sociaux sur le Web, les articles de microblog sont particulièrement courts et soumis en temps réel pour rapporter un événement récent. Dans ce mémoire, nous nous intéressons au service de microblogage de Twitter, étant le service le plus populaire et le plus largement utilisé. Twitter se distingue de sites similaires par certaines caractéristiques et fonctionnalités. Une des caractéristiques importantes est la présence de la relation sociale d'abonnement. **[Réf. 19]**

2.1. Définition des blogs :

Un **blog**, ou **blogue** ou encore **cybercarnet** est un type de site web, ou une partie d'un site Web. Comme son étymologie l'indique (*web log* signifie *journal de bord sur le web* en anglais), un blog est censé contenir régulièrement de nouveaux *billets*, c'est-à-dire des notes ou des articles agglomérés au fil du temps sur un sujet donné.

Début 2011 étaient dénombrés au moins 156 millions de blogs, publiant à la cadence d'un million de nouveaux billets par jour. Le nombre de blogs inactifs est élevé ; rares sont les blogs qui affichent une grande longévité, et l'écrasante majorité des blogs a été abandonnée par leurs auteurs.

Un *blogueur* a aujourd'hui loisir de mélanger textes, hypertexte et éléments multimédias (image, son, vidéo, applet) dans ses billets ; il peut aussi répondre aux questions et commentaires des lecteurs car chaque visiteur d'un blog peut ou bien laisser des commentaires sur le blog lui-même, ou bien contacter le blogueur par courrier électronique. **[Réf. 8]**

2.2. Mise en ligne du blog :

Pour que les internautes puissent consulter le blog, celui-ci doit être hébergé. L'hébergement peut se faire directement sur l'ordinateur de l'éditeur ou plus souvent, auprès d'un fournisseur d'hébergement. Une adresse d'accès doit ensuite être déterminée. Elle ne doit pas porter atteinte aux droits de la personnalité et plus particulièrement au nom de famille, au droit sur les signes distinctifs, au droit d'auteur et à l'ordre public. Du fait de cette mise en ligne, le blogueur reçoit la qualification d'éditeur de services de communication au public en ligne.

3. Les réseaux sociaux :

3.1. Social :

Dans une définition large de la notion du social, on peut l'entendre comme étant l'expression de l'existence de relations entre les vivants :

- Qui concerne la vie en société (ordre social, milieu social).
- Relatif à la vie des hommes en société. Etude des processus d'interaction : Interactions entre les individus, entre l'individu et les groupes, entre les groupes eux-mêmes.
- Qui se caractérise par le fait de vivre en société.
- Relatif aux, critères, comportements, réalités concrètes engendrés par société divisée et hiérarchisée. Ascension, différence, échelle, prestige, promotion, réussite sociale.

3.2. Réseau :

INFORMATIQUE : Réseau informatique. Interconnexion de un ou plusieurs ordinateurs avec plusieurs terminaux distants par l'intermédiaire des voies de transmission.

PSYCHO-SOCIOL : Réseau de communication. L'étude des réseaux de communication est au fond une psycho-sociologie écologique de la communication, elle porte sur les conditions de milieu dans lesquelles la communication s'exerce.

3.3. Réseau social :

Un réseau social est un ensemble d'entités sociales telles que des individus ou des organisations sociales reliées entre elles par des liens créés lors de leurs interactions mutuelles. Des réseaux sociaux peuvent être créés stratégiquement pour agrandir ou rendre plus efficient son propre réseau social (professionnel, amical, etc.).

Il existe des sites (applications) Internet aidant à se créer un cercle d'amis, à trouver des partenaires commerciaux, un emploi ou identifier des personnes ayant des intérêts communs. Ces sites sont regroupés sous l'appellation « réseaux sociaux » **[Réf. 12**]

Ce terme désigne un site Internet permettant à l'internaute de s'inscrire d'y créer une carte d'identité virtuelle appelée le plus souvent « Profil ». Le réseau est dit social en ce qu'il permet d'échanger avec les autres membres inscrits sur le même réseau : Des messages publics ou privés, des liens hypertexte, des vidéos, des photos, des jeux…

L'ingrédient fondamental du réseau social reste cependant la possibilité d'ajouter des amis (connaissances), et de gérer ainsi une liste de contacts.

L'émulation des réseaux sociaux fonctionne ensuite sur deux principes que l'on peut résumer ainsi :

- « Les amis de mes amis sont mes amis ».

- « Les personnes qui partagent les mêmes centres d'intérêts que moi sont mes amis ».

3.4. Web 2.0 :

Le terme « web 2.0 » s'impose vraiment en 2007 après avoir été dévoilé au grand public par Tom O'Reilly en 2005 dans son article extrait de la conférence fondatrice.

Par définition, Wikipédia nous dit que l'on qualifie de Web 2.0 les interfaces permettant aux internautes d'interagir à la fois avec le continu des pages mais aussi entre eux ; faisant ainsi du Web 2.0 le web communautaire interactif par opposition aux pages statiques que l'on a pu voir autour des années 2000. Maintenant, nous avons des sitcoms comme youtube, Widler, dailymotion, myspace, twitter technoraty ou encore facebook.

Se sont de nouvelles plateformes de forme Web 2.0.

L'avantage est leur gratuité. **[Réf. 19]**

Chapitre II

Technique de filtrage

1. Définitions :

Le filtrage et le blocage sont des notions distinctes, qu'il convient de définir.

1.1. Filtrage :

Le filtrage se définit comme une limitation sélective d'accès au Réseau.

Si le blocage permet d'arrêter tout trafic, sans distinction, le filtrage permet de laisser passer certaines informations, mais seulement après analyses de ces dernières. Ce sont ces analyses qui permettront de déterminer quelles informations sont autorisées à arriver jusqu'à leurs destinataires.

Tout comme pour le blocage, le filtrage peut être opéré de différentes manières, par différentes entités, selon le but recherché.

Néanmoins, des moyens supérieurs à ceux mis en œuvre pour le blocage seront mobilisés, car il s'agit d'une technique plus délicate, dans la mesure où certaines informations vont être acceptées et d'autres vont être expurgées. En effet, il va falloir s'interroger sur les critères permettant de différencier les éléments et installer un mécanisme sophistiqué pour comprendre la nature du trafic. Ce sont les informations se trouvant dans le cœur même de l'infrastructure d'accès au réseau qui seront examinées.

Ainsi, toute question de filtrage ou de blocage se pose forcément en termes de finalité – destination du filtrage – de moyens mis en œuvre pour atteindre cette finalité – efficacité et limites – et de responsabilités de la prise de décision au regard de la destination et des moyens – subsidiarité, qui va prendre la décision de filtrer, et proportionnalité, comment l'entité ayant décidé de filtrer va procéder. Ayant établi que le blocage n'était qu'un sous-ensemble du filtrage, c'est ce dernier qui sera privilégié. **[Réf. 11]**

1.2. Le blocage :

Le blocage consiste à interdire le passage d'information d'un point A à un point B. Sur Internet cela peut se traduire par l'acte de rendre impossible toute de connexion d'une machine vers un site Internet, ou bien d'empêcher un contenu d'atteindre un ordinateur personnel ou un poste informatique. L'exemple le plus célèbre de blocage sur Internet est Youtube. En effet, ce site de partage de vidéos a fait l'objet de différents blocages dans différents pays : Bangladesh, Chine, Maroc, Thaïlande, Tunisie ou encore la Turquie.

Le blocage est une action permettant d'arrêter le trafic. Cela revient à poser un barrage sur une route. Le blocage peut se faire de différentes manières et peut être opéré par différentes personnes, les techniques utilisées sont fonction de la personne qui souhaitera opérer un blocage ainsi que du but recherché par cette dernière. **[Réf. 11]**

2. Filtrage :

2.1. Efficacité d’un filtrage :

Il s’agit ici de présenter les bases qui vont nous permettre de décider si un filtrage (une technique de filtrage appliquée à un contexte) est efficace ou pas. Il ne s’agit donc pas de déterminer si un filtrage est dangereux (pour les infrastructures réseau par exemple), non opportun, ou d’apporter un quelconque jugement sur celui-ci : nous souhaitons mettre en lumière son efficience.

Habituellement, l’efficacité d’un processus se définit par la capacité à atteindre l’objectif pour lequel le processus a été défini.

Dans le cas du filtrage on ne peut se contenter de cette seule définition.

En effet, lorsque l’on se prononce sur l’efficacité d’un filtrage il faut aussi prendre en considération les effets collatéraux. Ainsi le filtrage d’un site Web ne peut être considéré comme efficace si ce filtrage implique, qu’en plus du site Web que l’on désire filtrer, on empêche l’accès à d’autres sites qui n’étaient pas censés être la cible de ce filtrage.

Lorsque l’on se prononce sur l’efficacité d’un filtrage, il ne faut pas non plus négliger la facilité de contournement. Un filtrage n’est pas efficace si n’importe qui peut facilement le contourner.

L’efficacité d’un filtrage s’évalue donc en fonction de 3 critères :

- Le sous-blocage.
- Le sur-blocage.
- La facilité de contournement.

2.1.1. Sous-blocage :

On parle de sous-blocage lorsqu’un filtrage n’arrive pas à atteindre son objectif. Par exemple, on souhaite filtrer un ensemble de fichiers circulant sur un réseau pair-à-pair et on n’arrive qu’à en bloquer la moitié.

Plus le sous-blocage est important, moins le filtrage est efficace.

2.1.2. Sur-blocage :

On parle de sur-blocage lorsqu’un filtrage filtre plus que ce qu’il devrait. Par exemple, on souhaite filtrer un ensemble de fichiers circulant sur un réseau pair-à-pair et on bloque la circulation de l’intégralité des fichiers circulant sur ce réseau.

Plus le sur-blocage est important, moins le filtrage est efficace.

2.1.3. Facilité de contournement :

On parle de facilité de contournement lorsqu'un filtrage peut être contourné. Par exemple on souhaite filtrer un site Web par l'IP de son serveur, mais il existe un site miroir dont le serveur possède une autre IP et par conséquent le site est toujours accessible.

Plus la facilité de contournement est importante, moins le filtrage est efficace.

2.2. Coût d'un filtrage :

Le coût financier d'un filtrage peut difficilement s'estimer.

Car au-delà du coût financier direct qui peut s'évaluer par :

- Le coût d'installation d'équipements/logiciels assurant le filtrage
- Le coût de découverte maintenance et mise à jour des éléments à filtrer

Il y a aussi les coûts financiers indirects difficilement estimables (ex : impact sur l'économie d'un réseau ralenti par une technique de filtrage).

2.3. Le filtrage par destination :

2.3.1. Site web ou partie de site web :

Le but est, par exemple, d'empêcher l'accès à http://www.w3c.org. On retiendra les 5 critères de filtrage suivants :

2.3.1.1. Filtrage IP :

Le filtrage par IP consiste à interdire l'accès et les connexions au serveur identifié par les IP incriminées.

Efficacité :

La facilité de contournement est existante : le site web peut changer d'adresse en quelques minutes et l'utilisateur pourrait utiliser un proxy situé dans un pays non concerné par le filtrage du site Web en question.

Le risque de sous-blocage est existant : le site web peut utiliser plusieurs adresses sans que celles-ci puissent être toutes infailliblement connues.

Le risque de sur-blocage est très important : il est impossible de connaitre précisément la liste des sites Web ou services Internet abrités par une même adresse IP, potentiellement plusieurs centaines voire des milliers selon la solution d'hébergement utilisée.

Conclusion :

Le blocage d’un site Web par adresses IP n'est que faiblement efficace.

Coût :

Faible. Le blocage d’une connexion en fonction de son IP d’origine est quelque chose de couramment employé de nos jours. C’est le principe même d’un pare-feu.

2.3.1.2. Filtrage par nom de domaine :

Le filtrage par nom de domaine consiste à interdire l’accès et les connexions aux serveurs identifiés par le nom de domaine incriminé.

Efficacité :

La facilité de contournement est existante : un tel blocage n'empêche aucunement l'accès au site par les adresses IP directement. De plus, changer de nom de domaine est une opération rapide et quasi triviale. Les moyens de communication du nouveau nom sont ensuite nombreux (réseaux sociaux, tchat, forums spécialisés...).

Le risque de sous-blocage est existant : un tel blocage n'empêche aucunement l'accès au site par d'autres noms de domaine.

Le risque de sur-blocage est important : plusieurs milliers de sites Web pourraient être bloqués en même temps que le site visé, par exemple le seul blocage de ‘wordpress.com’ entrainerait le blocage des centaines de milliers de blogs.

Conclusion :

Le blocage d’un site Web par nom de domaine du site n'est que faiblement efficace

Coût :

Faible.

2.3.1.3. Filtrage par URL :

Le filtrage par URL consiste à inspecter les requêtes HTTP par une technique reposant sur l’utilisation de DPI et de filtrer toutes les requêtes vers l’URL incriminée.

L’utilisation d’un proxy par lequel transiteraient toutes les requêtes HTTP permettrait de faciliter un tel filtrage.

Un exemple de ce type de filtrage pourrait être : sur un même site www.site.dz autoriser la page correspondant à l'URL www.site.dz/page1.html et bloquer la page correspondant à l'URL www.site.dz/page2.html.

Efficacité :

La facilité de contournement est existante : le filtrage par URL ne peut notamment s'appliquer aux sites auxquels on accède en HTTPS (à moins de casser la sécurité du protocole de chiffrement SSL).

Le risque de sous-blocage est important : il sera difficile de référencer l'intégralité des URL à filtrer.

Le risque de sur-blocage est minime.

Conclusion :

Le blocage d'un site Web par URL est relativement efficace.

Coût :

Coût d'un DPI généralisé, soit extrêmement couteux (voir DPI). Réduction du coût possible par l'utilisation d'un proxy.

2.3.1.4. Filtrage dynamique (par contenu) :

Le filtrage dynamique consiste à inspecter lors de l'accès au site Web, le contenu de ses pages et de filtrer selon ce que ces pages contiennent (ex : bloquer toutes les pages qui contiennent le mot « lapin »). Ce type de filtrage repose sur l'utilisation de DPI.

Efficacité :

La facilité de contournement est existante : le filtrage dynamique ne peut notamment s'appliquer aux sites auxquels on accède en HTTPS (à moins de casser la sécurité du protocole de chiffrement SSL).

Le risque de sous-blocage est important : un tel blocage nécessite la constitution et la mise à jour d'une base de contenu (ou d'identifiant de contenus) et il sera impossible de référencer tout ce que l'on souhaite filtrer.

Le risque de sur-blocage varie de minime à important : bloquer un mot ou un terme ou une expression bloquera l'ensemble des sites comportant ce mot, ce terme ou cette expression, même s'ils ne sont pas directement visés par le filtrage, le sur-blocage sera alors important.

Si le filtrage se fait sur un identifiant unique du contenu (une signature ou un hash d'un fichier par exemple) le sur-blocage sera minime.

Conclusion :

Le blocage dynamique d'un site Web est moyennement efficace.

Coût :

Coût d'un DPI généralisé, soit extrêmement couteux.

2.3.1.5. Filtrage hybride :

Le filtrage hybride est un mélange d'un filtrage par IP puis d'un filtrage sur le contenu. Il évite de devoir effectuer du DPI sur la totalité des communications, mais nécessite un ancrage fort dans le réseau du FAI.

Efficacité :

La facilité de contournement est existante : le site web peut changer d'adresse IP en quelques minutes et l'utilisateur pourrait utiliser un proxy situé dans un pays non concerné par le filtrage du site Web en question.

Les risques de sous-blocage sont ceux du filtrage par IP soit existant.

Les risques de sur-blocage sont ceux du filtrage dynamique soit variant de minime à important selon le critère de décision (mots clé, signature de fichiers, etc.).

Conclusion :

Le blocage dynamique d'un site Web est moyennement efficace.

Coût :

Coût d'un DPI sélectif, moins couteux qu'un DPI généralisé mais couteux quand même.

2.3.2. Messagerie Instantanée :

Le but est, par exemple, de bloquer une messagerie instantanée sur une zone géographique prédéfinie. Il existe 3 critères de filtrage qui peuvent être utilisés pour cela.

2.3.2.1. Filtrage par IP :

Le filtrage par IP consiste à interdire l'accès et les connexions au serveur identifié par les IP incriminées.

Efficacité :

La facilité de contournement est existante : l’utilisateur pourrait utiliser un proxy situé dans un pays non concerné par le filtrage du système de messagerie en question.

Le risque de sous-blocage est minime.

Le risque de sur-blocage est important : bloquer par adresse IP revient à interdire totalement l'accès au système de messagerie instantanée concernée.

Conclusion :

Le blocage d’un service de messagerie instantanée par adresses IP n'est que faiblement efficace.

Coût :

Faible.

2.3.2.2. Filtrage protocolaire :

Le filtrage protocolaire consiste à interdire l’usage du ou des protocoles réseau utilisés par le service de messagerie instantanée.

Le risque de sur-blocage est important : la plupart des systèmes de messagerie instantanée utilisent le protocole HTTP, hors c’est le protocole sur lequel repose la majeure partie du Web. Interdire le protocole HTTP revient à bloquer une majeure partie du Web.

Conclusion :

Le blocage d’un service de messagerie instantanée par adresses IP n'est que faiblement efficace.

Coût :

Moyen

2.3.2.3. Filtrage dynamique par zone géographique :

Le filtrage par contenu par zone géographique utilise des technologies DPI dans des endroits ciblés du réseau. Il permettrait de filtrer uniquement certains messages en fonction de leur provenance géographique et de leur contenu.

La facilité de contournement est existante : l'utilisateur pourrait utiliser un proxy situé dans un pays non concerné par le filtrage du système de messagerie en question.

Le risque de sous-blocage est important : les modes d'accès aux messageries instantanées sont différents (Web, application mobile...) et les réseaux hétérogènes (nationaux, internationaux, roaming, VPN...) si bien qu'il serait difficile prendre en compte tout aspect pour réussir à filtrer ce que l'on souhaite filtrer.

Le risque de sur-blocage est important : pour les mêmes raisons que ci-dessus, il serait difficile de ne pas filtrer autre chose que ce que l'on souhaite filtrer.

Conclusion :

Le blocage d'un service de messagerie instantanée par adresses IP n'est que faiblement efficace.

Coût :

Coût d'un DPI généralisé, soit extrêmement couteux.

2.3.3. Mail :

2.3.3.1. Filtrage par IP :

Le filtrage par IP consiste à bloquer les mails provenant des serveurs identifiés par les IP incriminées.

Efficacité :

La facilité de contournement est faible : le changement d'IP pour un serveur d'envoi de mail n'est pas aisé. Il ne s'agit pas comme dans le cas d'un site Web de déplacer un contenu sur un autre serveur.

Le risque de sous-blocage est existant : si tous les mails provenant des serveurs identifiés par les IP sont bien bloqués, il est possible que d'autres serveurs envoient les mails que l'on désirait bloquer.

Le risque de sur-blocage est très important : tous les mails provenant des serveurs identifiés par les IP seront bloqués y compris ceux en dehors du périmètre de filtrage.

Conclusion :

Le blocage de mails par les adresses IP des serveurs d’expédition n'est que faiblement efficace.

Coût :

Faible.

2.3.3.2. Filtrage dynamique (par contenu) :

Le filtrage dynamique consiste à inspecter le contenu des mails et de filtrer selon ce que le mail contient (ex : bloquer toutes les mails qui contiennent le mot « lapin »). Suivant l’endroit où il est réalisé, ce type de filtrage peut reposer sur l’utilisation de DPI.

Efficacité :

La facilité de contournement est faible : le moyen de contournement le plus simple consisterait à chiffrer le contenu des mails, mais cela supposerait que seuls les destinataires aient les moyens de déchiffrer ce contenu (utilisation de crypto-asymétrique, échange sur canal privé, etc..). Ce moyen n’est pas envisageable à grande échelle.

Le risque de sous-blocage est existant : un tel blocage nécessite la constitution et la mise à jour d’une base de contenu (ou d’identifiant de contenus) et il sera impossible de référencer tout ce que l’on souhaite filtrer.

Le risque de sur-blocage varie de minime à important : bloquer un mot ou un terme ou une expression bloquera l'ensemble des mails comportant ce mot, ce terme ou cette expression, même s'ils ne sont pas directement visés par le filtrage, le sur-blocage sera alors important.

Si le filtrage se fait sur un identifiant unique du contenu (une signature ou un hash de contenu du mail par exemple) le sur-blocage sera minime.

Conclusion :

Le blocage dynamique de mails est moyennement efficace.

Coût :

Suivant la technique utilisée, le coût varie de moyen à élevé.

2.3.4. Usenet :

2.3.4.1. Filtrage par groupe :

Les informations et contenus sur Usenet sont organisés en groupes. On peut imaginer Usenet comme un arbre dont chaque feuille est un groupe. A l'origine, l'intégralité de l'arborescence est répliquée sur chaque serveur Usenet.

Chaque groupe a un nom qui reflète le type supposé des contenus et informations présents dans le groupe. Il est donc envisageable sur un serveur Usenet d'interdire l'accès à (ou de ne pas y copier) un groupe que l'on voudrait filtrer.

La facilité de contournement est existante : chaque groupe étant en théorie répliqué sur la totalité des serveurs Usenet existants, interdire ou supprimer un groupe revient à l'interdire et le supprimer sur tous les serveurs. Ainsi si un serveur ne joue pas le jeu, le groupe restera accessible via ce serveur et n'importe quel internaute pourra s'y connecter (via un canal chiffré pour éviter toute tentative de filtrage complémentaire par DPI).

Le risque de sous-blocage est existant : les utilisateurs peuvent dissimuler du contenu que l'on souhaiterait interdire dans des groupes ayant des noms ne laissant pas à penser qu'il faille les filtrer.

Le risque de sur-blocage est existant (à moins de considérer la totalité des contenus du groupe comme à filtrer) : ce filtrage risque d'empêcher l'accès à des contenus n'étant pas dans le périmètre de filtrage.

Conclusion :

Le blocage d'une partie de Usenet par le blocage d'un groupe est moyennement efficace.

Coût :

Faible.

2.3.4.2. Filtrage par hiérarchie :

Comme expliqué précédemment on peut imaginer Usenet comme un arbre dont chaque feuille est un groupe.

S'il est possible sur un serveur Usenet d'interdire l'accès à une feuille de cet arbre (un groupe), il est aussi possible d'interdire l'accès ou de supprimer toute une branche, que l'on appelle une hiérarchie, de cet arbre et ainsi tout un ensemble de groupes.

La facilité de contournement est existante : c'est la même facilité de contournement que pour le filtrage par groupe.

Le risque de sous-blocage est minime : en empêchant l'accès à toute une hiérarchie, on a de fortes chances d'empêcher l'accès aux contenus ciblés.

Le risque de sur-blocage est important (à moins de considérer la totalité des contenus d'une hiérarchie comme étant à filtrer) : ce filtrage empêchera l'accès à des contenus n'étant pas dans le périmètre de filtrage.

Conclusion :

Le blocage d'une partie de Usenet par le blocage d'un groupe est moyennement efficace.

Coût :

Faible.

2.3.5. Paire à paire (et autres protocoles applicatifs d'échange de contenus) :

2.3.5.1. Filtrage par protocoles :

Il est généralement possible de reconnaître qu'un flux réseau est engendré par un logiciel pair-à-pair. Le filtrage par protocoles revient à identifier les flux réseau engendrés par des logiciels pair-à-pair et à les interdire. Cette technique peut nécessiter l'utilisation de DPI.

La facilité de contournement est faible : de nos jours, l'identification de flux est une technique maîtrisée. Même si de nos jours certains protocoles P2P essayent de brouiller et de dissimuler leurs flux, il existe des techniques heuristiques permettant la reconnaissance en dépit du brouillage.

Le risque de sous-blocage est minime : filtrer un protocole équivaut à prévenir la circulation de tous les contenus (et par conséquent ceux que l'on souhaite filtrer).

Le risque de sur-blocage est important (à moins de considérer la totalité des contenus circulant sur le réseau P2P comme étant à filtrer) : ce filtrage empêchera la circulation de contenus n'étant pas dans le périmètre de filtrage.

Conclusion :

Le filtrage d'un réseau P2P par le blocage du protocole qu'il utilise est faiblement efficace.

Coût :

Elevé (nécessite du DPI)

2.3.5.2. Filtrage dynamique (par contenu) :

Le filtrage dynamique consiste à inspecter le contenu circulant sur le réseau P2P et de filtrer ceux-ci (ex : bloquer toutes les vidéos sur les « lapins »). Ce type de filtrage repose sur l'utilisation de DPI.

Efficacité :

La facilité de contournement est existante : le filtrage dynamique ne peut notamment s'appliquer aux sites auxquels on accède en HTTPS (à moins de casser la sécurité du protocole de chiffrement SSL).

Le risque de sous-blocage est important : un tel blocage nécessite la constitution et la mise à jour d'une base de contenu (ou d'identifiant de contenus) et il sera impossible de référencer tout ce que l'on souhaite filtrer.

Le risque de sur-blocage est minime : bloquer un contenu clairement identifié n'entrainera pas le blocage des autres contenus.

Conclusion :

Le filtrage dynamique d'un réseau P2P est moyennement efficace.

Coût :

Coût d'un DPI ciblé sur certains types de flux, très couteux.

2.4. Filtrage par moyens

2.4.1. Introduction :

Lorsqu'une information circule, il y a trois acteurs qui interviennent :

- La source d'information.
- Le milieu par lequel l'information se propage.
- Le destinataire de l'information.

Dans le cas qui nous concerne, la source est généralement le diffuseur d'un contenu sur Internet, le milieu se trouve être le réseau (Internet) et la destination l'utilisateur final (l'internaute).

Afin d'exercer un filtrage sur Internet, le filtrage consistant à bloquer la circulation de certains contenus (site Web, fichier, etc.), il est possible d'agir sur chacun des

trois acteurs : source, milieu, destinataire et nous allons voir dans la suite de cette partie comment cela est possible et avec quelles conséquences.

2.4.2. Agir sur la source :

On ne parle généralement pas de filtrage ou de blocage dans ce cas. Il est question de retirer le contenu ou le service purement et simplement.

2.4.3. Agir sur le milieu :

Il s'agit de tenter de résoudre un problème en s'attaquant au milieu par lequel le problème est propagé. On agit uniquement sur celui qui permet, par la mise à disposition de ses « tuyaux » de « solliciter » ou de « constater » l'infraction en faisant circuler ce que l'on ne veut pas voir utilisé. On ne réprime alors pas l'usage (et on ne risque pas d'entrer dans une pédagogie de responsabilisation) on tente une épuration du milieu.

Pour agir sur le milieu, il y a deux possibilités, que nous allons décrire ici.

2.4.3.1. Agir au niveau du cœur de réseau :

Le problème c'est que l'Internet n'a pas vraiment de cœur, de début ou de fin. On ne peut pas créer, à l'instar de l'eau, des stations d'épuration qui se retrouveraient forcément sur le passage de la collecte des eaux usées ou de la distribution des eaux propres. L'Internet est un réseau a-centré.

Par conséquent, il n'y a que deux manières de tenter cette épuration :

- On dispose des mécanismes d'épuration partout ou du moins à chaque croisement possible. Ceci n'est pas imaginable de par le nombre de croisements.
- On réduit le nombre de croisements en recentrant Internet. A l'extrême, on crée de gros points de concentration qui nous permettront d'épurer plus facilement. Ceci se fait au détriment de la décentralisation, entrainant donc une baisse des performances et de la résilience, ou bien à un coût (humain et technique) très élevé de par le nombre d'équipements et de configurations à mettre en production et à maintenir. Cette recentralisation entrainerait Internet loin de ses principes initiaux.

2.4.3.2. Agir au niveau du Nœud de Raccordement Abonné (DSLAM ou autre) :

Le Nœud de Raccordement de l'Abonné (ou NRO dans le cas de la fibre) est l'endroit dans lequel aboutissent les lignes de communication des abonnés (fibres, cuivre, hertzien, ...).

C'est généralement le premier endroit ou le trafic est susceptible de prendre ou d'arriver depuis plusieurs voies d'accès. Dans le cas très majoritaire de la France, ces NRA sont équipés de DSLAM (expliquer l'acronyme) permettant de collecter le trafic des lignes ADSL".

Tous les DSLAM en service ne sont cependant pas forcément capables de traiter les paquets IP. Beaucoup ne sont en effet dédiés qu'au simple transport depuis le DSLAM vers un autre équipement et n'embarquent aucune capacité d'analyse, ne serait-ce que de la source ou de la destination des paquets. De manière générale, leur mission ne concerne que le transport de paquets et leurs capacités de traitement sont orientées vers cette mission.

L'ajout d'équipements permettant ce genre de traitement serait envisageable, mais il faut garder à l'esprit qu'il existe environ 13 000 centraux téléphoniques. Cela revient à l'ampleur du travail qui a dû être effectué pour déployer l'ADSL ces dix dernières années.

2.4.3.3. Agir au niveau des serveurs de noms (DNS) de l'opérateur :

C'est un cas particulier du filtrage qui intervient sur un service et non sur l'infrastructure elle-même. Mais en l'espèce, sans DNS, la plupart des services de l'Internet, utilisés au quotidien ne fonctionneraient plus. Cela n'empêche pas de consulter le contenu, mais limite son accessibilité.

L'analogie la plus proche est celle du réseau routier sur lequel on retire les panneaux indicateurs : vous pouvez toujours vous rendre à votre destination si vous connaissez déjà le chemin. Il est aussi possible d'orienter différemment les utilisateurs en indiquant d'autres directions.

La plupart des utilisateurs utilisent le mécanisme de résolution d'adresses de leur opérateur. Ils seraient alors facilement « trompables ». Mais il serait alors possible d'utiliser d'autres DNS, qui eux, diraient la vérité (et ne filtreraient plus en remplaçant une destination par une autre).

2.4.4. Agir à la destination :

2.4.4.1. Sous la maitrise de l'utilisateur :

C'est un autre moyen de filtrer et la popularité des logiciels de contrôle parental est là pour nous confirmer la faisabilité technique d'une telle solution. Ces logiciels sont installés à la discrétion des utilisateurs et même s'il existe des paramétrages par défaut c'est toujours l'utilisateur qui peut choisir les paramètres de filtrage appliqués par ces logiciels.

Avec la multiplication des écrans « connectés » domestiques (Télévision, ordinateurs, consoles de jeu, Smartphones … voire demain quasiment tout objet ayant un intérêt à remonter de l'information ou que l'on doit « actionner »), il devient totalement illusoire de penser le « contrôle parental » comme on le pensait avant. Il devient impossible de protéger objet par objet … et d'ailleurs de quoi, de qui et pour qui ? (que se passe-t-il si nous sommes plusieurs à utiliser le même objet ?)

Tout comme, il est absurde de protéger du cambriolage son logement en n'empêchant pas le malfaiteur d'entrer, mais en disposant des cadenas sur chacun des objets contenus présents dans le logement (la première protection utile étant de penser à fermer la porte du logement), il est inconcevable d'imaginer un « contrôle parental » installé sur chacun des équipements connectés.

Si on poursuit le raisonnement avec l'analogie de la porte, se posent d'autres questions :

- Qui (et quoi) sera la porte blindée de notre « chez nous » numérique ?
- Qui la contrôlera ?
- Au-delà d'une grande responsabilité, ne s'agit-il pas d'un grand pouvoir qui se dessine ?

C'est en tous les cas le point de complexité majeure du sujet.

2.4.4.2. Sous la maitrise d'un tiers :

Une sécurisation de l'accès à Internet réalisée « à l'insu du plein gré » de l'utilisateur s'apparente à un appartement qui nous est livré avec une porte blindée, mais dont d'autres ont la clef.

Un système « d'épuration », sans la participation de l'utilisateur est une erreur si on considère que l'on ne pourra jamais tout filtrer et que par conséquent le premier rempart doit être l'utilisateur lui-même et non reposer sur l'autre, qu'il s'appelle un parent ou un firewall.

2.5. Le filtrage par type : (Quelles techniques employer ?)

2.5.1. Blocage sur les noms de domaine (DNS, DPI) :

2.5.1.1. Objectif et terrain d'action :

Il s'agit d'empêcher, à un niveau ou à un autre, la conversion du nom d'un site en adresse IP pour empêcher les machines des utilisateurs de connaitre cette IP et donc d'accéder au site.

Cette technique se met généralement en place au niveau du DNS des fournisseurs d'accès; sachant que s'il existe une poignée de très gros représentants, il faut également compter sur plusieurs centaines de petites structures n'ayant parfois qu'une dizaine d'utilisateurs.

2.5.1.2. Efficacité :

Le blocage complet d'un nom de domaine entraine un sur-blocage potentiel, par exemple dans le cas où plusieurs sous-domaines ou sous-sites sont hébergés à la même adresse (exemple: site.com/blog1 et site.com/blog2). On peut reprendre l'exemple présenté précédemment : le seul blocage de « wordpress.com » entrainerait le blocage des centaines de milliers de blogs.

Il est possible pour n'importe quel utilisateur de se servir d'autres serveurs DNS que ceux de son FAI, souvent même en dehors du pays (Google en propose, par exemple) rendant le filtrage sur les DNS du FAI inopérant.

Le filtrage DNS peut alors être couplé avec un filtrage protocolaire n'autorisant les requêtes DNS que sur les serveurs du fournisseur d'accès effectuant le filtrage. Ce filtrage peut, en l'état, être considéré comme une atteinte à la neutralité et peut être contourné au moyen de services DNS utilisant d'autres ports ou protocoles, ou encore en passant par un VPN.

Cette technique peut aussi faire l'objet de DPI, avec tous les travers que cela comporte, directement en cœur de réseau, permettant cette fois d'intercepter la totalité des requêtes DNS effectuées.

De manière générale, le filtrage DNS peut aisément être contourné au moyen d'un VPN permettant de simuler une connexion depuis un autre opérateur, par exemple à l'étranger.

2.5.2. Blocage sur les adresses IP (DNS, BGP) :

2.5.2.1. Objectif et terrain d'action :

Il s'agit d'empêcher le trafic en provenance ou à destination d'une adresse IP prédéfinie. Concrètement, sur le réseau d'un opérateur, cette mesure peut être implémentée au niveau du DNS, en l'empêchant de répondre l'adresse IP cible du blocage à ses clients. On retrouve les mêmes problématiques de contournement que sur le blocage des noms.

Il peut également être effectué au moyen d'injection BGP, technique consistant à donner, pour l'adresse IP cible, une fausse information à l'ensemble des routeurs du réseau afin que les demandes de connexions ne puissent atteindre le serveur réel.

L'emploi de technologies type VPN permet de contourner très aisément ce type de filtrage, en employant d'autres réseaux non soumis aux mesures de blocage.

Il peut enfin être mis en place en inspectant les paquets et en interrompant la connexion une fois qu'elle a été établie. L'inspection de paquets permettant plus de flexibilité, elle ne s'arrête généralement pas à un simple blocage des adresses IP, puisqu'elle permet beaucoup plus.

2.5.2.2. Efficacité :

Ce type de filtrage permet donc effectivement de bloquer l'accès à la cible définie depuis le réseau où est effectué le filtrage, mais le sur-blocage est énorme et non estimable au moment où l'ordre de blocage est donné et il est, en fonction de la méthode employée, aisé à contourner ou difficile à mettre en place.

Par ailleurs, il n’est pas dit que la stabilité d’Internet résiste à un grand nombre d’injections d’annonce BGP et que finalement, pour protéger leur réseau, les opérateurs ne se protègent pas de ce genre d’annonces (route flap dampening).

2.5.3. Filtrage sur les URL (DPI) :

2.5.3.1. Objectif et terrain d’action :

Il s'agit de contrôler précisément les URL auxquelles on laisse accès. C'est un filtrage applicable uniquement au Web non chiffré.

Il s'agit, via l'inspection de paquets, d'interrompre une connexion préétablie lorsqu'une URL "interdite" est détectée.

2.5.3.2. Efficacité :

Ce type de filtrage présente tous les problèmes du DPI et peut être facilement contourné par l'emploi de connexions chiffrées (HTTPS), par changement d'URL du contenu ou par l'utilisation de VPN.

L'utilisation de ces technologies sur HTTPS est théoriquement possible en ayant des certificats valides pour des sites que l'on ne possède pas. Il suffit pour cela de posséder une autorité de certification reconnue par les navigateurs. C'est le cas de nombreux états ou d'entreprises privées pouvant, le cas échéant, être soumises par un état.

2.5.4. Filtrage sur les contenus (DPI) :

2.5.4.1. Objectif et terrain d’application :

Suivant le même principe que le filtrage sur des URL, le filtrage sur le contenu inspecte les paquets transmis à la recherche d'un contenu spécifique : texte, image, son, vidéo, pour en empêcher la transmission.

2.5.4.2. Efficacité :

Il ne présente pas le problème du filtrage par URL qui peut être contourné en publiant le même contenu à une autre adresse, mais il oblige alors à un contrôle de l'intégralité du trafic.

La nécessité d'inspecter la totalité du trafic implique de mettre en place des équipements d'inspection de paquets dimensionnés pour supporter l'intégralité de la charge induite sur le réseau par les abonnés. Cela nécessite aussi de centraliser le trafic ou bien de démultiplier les équipements pour en placer plus près des abonnés.

2.5.5. Blocage sur les ports :

2.5.5.1. Objectif et terrain d’application :

Pour simplifier, considérons que chaque type d'application utilisant Internet le fait par le biais d'un numéro prédéfini pour chacune. Par exemple, le Web utilise massivement le port 80 ou, pour les sites sécurisés, le 443. L'envoi d'email se fait quant à lui par le port 25.

2.5.5.2. Efficacité :

Un filtrage par numéro de ports est simple à mettre en place. Presque tous les équipements réseau savent le faire. C'est le principe de base du fonctionnement des firewalls. On pourrait, par exemple, décider que les serveurs Usenet utilisant le protocole NNTP sur le port 119 sont une menace pour le respect du droit d'auteur. Il conviendrait donc de fermer le port 119 pour régler le problème.

Si on met de côté le fait que NNTP n'est qu'un protocole et ne détermine donc pas la licéité des contenus qui y transitent, les personnes souhaitant l'utiliser auront tôt fait d'utiliser un autre port que le 119.

Certaines applications sont mêmes auto-adaptatives, comme beaucoup de logiciels peer-to-peer qui vont tester la connexion de l'internaute à la recherche de ports bloqués et utiliser les premiers disponibles qu'ils trouveront.

Le blocage de ports est donc l'un des plus simples, mais aussi l'un des moins efficaces.

2.5.6. Filtrage hybride :

2.5.6.1. Objectif et terrain d’action :

Il s'agit d'un mélange du filtrage utilisant BGP et inspection de paquets.

Concrètement, pour éviter le sur-blocage entraîné par l'interdiction totale d'accéder à une IP et pour éviter le coût et les problèmes de redondance engendrés par le DPI, le filtrage hybride propose de ne faire passer au travers du filtre DPI que le trafic à destination des adresses IP distribuant le contenu à bloquer.

2.5.6.2. Efficacité :

L'effet sur le reste du trafic est supposé être nul, l'impact sur la globalité du réseau également. C'est bien entendu en supposant que les contenus à filtrer ne font pas partie de grosses plateformes de distribution de contenus. Car si ce genre de plateforme se retrouve soudain à devoir filtrer l'ensemble du trafic venant d'un site comme MegaUpload, elle aurait tôt fait d'être saturée et d'agir, finalement, comme un filtre simple interdisant l'ensemble du trafic.

Tout comme quand il est utilisé dans le cas du filtrage IP, l'utilisation détournée de BGP n'est pas un acte anodin. De nombreux administrateurs chevronnés commettent souvent des erreurs de configuration ayant plus ou moins d'impact sur le réseau tout en étant des actions de pure routine, on envisage aisément que des cas comme YouTube/Pakistan Telecom pourraient se produire régulièrement si l'utilisation du filtrage via BGP était institutionnalisée.

2.5.7. Filtrage sur les protocoles (DPI-traffic shaping) :

2.5.7.1. Objectif et terrain d’action :

Le filtrage sur les protocoles inspecte les paquets transmis à la recherche d'une signature de protocole. Une signature de protocole permet d’identifier qu’une suite de paquets transmis entre une source et une destination correspond bien à un type de protocole réseau (et donc si on simplifie : à un usage). Une fois le protocole identifié comme étant à filtrer, la connexion est interrompue.

2.5.7.2. Efficacité :

Ce filtrage oblige alors à un contrôle de l'intégralité du trafic.

La nécessité d'inspecter la totalité du trafic implique de mettre en place des équipements d'inspection de paquets dimensionnés ou de « traffic shapping » pour supporter l'intégralité de la charge induite sur le réseau par les abonnés. Cela nécessite aussi de centraliser le trafic ou bien de démultiplier les équipements pour en placer au plus près des abonnés.

3. Les enjeux :

3.1. Enjeux de filtrage :

3.1.1. Le filtrage à l'intersection des techniques et des usages :

Parler des problèmes engendrés par Internet est un abus de langage, car Internet n'est pas mauvais en soi, tout comme les technologies – le P2P (le pair à pair), le streaming – ne sont pas mauvaises par nature, ce qui n'est potentiellement pas le cas de l'usage que nous en faisons.

Ainsi la thématique du filtrage d'Internet ne peut être réduite à une problématique uniquement technique. Elle concerne aussi les usages des internautes. Les usages sont à comprendre comme une combinaison entre un possible technique et des utilisateurs qui se sont emparés des moyens pour en déterminer une utilisation. En surplus des réglementations, il existe également des habitudes et des usages particuliers, souvent regroupés sous l'appellation de Netiquette.

Internet est comme un objet qui peut se révéler indispensable à notre quotidien, mais en même temps imprévisible : mal utilisé, Internet, entraîne sur un terrain, non pas de non-droit, mais d'illégalité où généralement toutes les peines ont déjà été largement pensées et prévues.

Au-delà de la question des usages, se pose également la question de l'espace et du temps d'Internet. Internet impose un devoir de vigilance tant par l'aspect polymorphe de ses contenus que par leur vitesse de diffusion et les conditions de leur accessibilité. La même chose peut être mise en évidence pour les modalités d'accès à un contenu selon la localisation de l'internaute et du contenu. En effet, un usage peut être légal sur un territoire donné et être illégal sur un autre. Il peut également être illégal pendant un laps de temps et devenir légal par la suite. Les jeux d'argent en ligne semblent illustrer cette problématique.

En effet, dans le cas de la France, sur le plan de la légalité les jeux d'argent en ligne étaient considérés comme illicites. Suite à des changements législatifs, la mise en place de l'ARJEL et la coopération des sites concernés, les jeux d'argent en ligne sont progressivement entrés dans la légalité au regard de la loi française.

Internet ne peut pas être réduit au Web ou à un service précis, ce n'est pas non plus uniquement un réseau de diffusion. Il peut être utilisé pour cela, mais il n'a pas besoin de diffuser pour exister puisqu'il existait avant ces usages. Internet est un réseau d'échanges.

L'expérience du passé et l'humilité de reconnaître que l'on se trompe souvent lorsque l'on tente de prévoir le futur de ces technologies, nous enseignent qu'il est

préférable qu'il soit neutre et symétrique et si possible, au plus haut débit envisageable.

3.1.2. Filtrage et responsabilité :

Le principe de liberté, souvent confondu avec celui de l'arbitraire dans le langage commun, inclut nécessairement celui de la responsabilité, définie comme le fait d'agir en toute connaissance de cause, en toute conscience. Selon l'entité qui opère le filtrage et selon les moyens utilisés pour le mettre en place, la notion de responsabilité se trouve modifiée.

L'internaute serait placé dans une situation de responsabilité moindre si le filtrage d'Internet lui échappait ou s'il n'en avait pas l'initiative. Il serait alors dans une situation de dépendance avec un tiers pour la mise en place d'un système de filtrage. Examinons les enjeux de ces deux possibilités : filtrage par un tiers, filtrage à l'initiative de l'internaute.

3.2. Filtrage par un tiers :

3.2.1. Identification des tiers :

En effet, il y a différentes entités qui peuvent opérer un filtrage d'Internet : l'autorité judiciaire, l'opérateur de réseau, l'opérateur de services, l'hébergeur d'un site Web, l'employeur, l'administrateur réseau et l'internaute. Ainsi l'autorité judiciaire va par exemple opérer un filtrage afin de répondre à un besoin légitime, notamment celui du respect de la légalité. Dans le cas de l'opérateur de réseau, l'opérateur de service, de l'hébergeur, de l'administrateur réseau et de l'employeur, il s'agit également de s'assurer du respect de la loi ainsi que de préserver les ressources du système afin de ne pas créer un déséquilibre dans une architecture réseau. En effet, les ressources du système sont partagées entre plusieurs machines: si l'une des machines consomme, de manière excessive, certaines ressources, les autres machines rencontreront des dysfonctionnements, créant une réaction en chaîne. Cet impératif de préservation pourrait justifier qu'un administrateur filtre une partie des éléments arrivant sur les machines composant son réseau afin de ne pas déséquilibrer l'ensemble de la structure.

3.2.2. Une moindre responsabilité et une moindre maîtrise pour l'internaute :

Dans l'hypothèse où un filtrage est opéré par un tiers, l'internaute est davantage déresponsabilisé. Il pourrait donc penser que les contenus qu'il consulte sont conformes avec la légalité. Il pourrait donc s'exonérer, au sens juridique du terme, de sa responsabilité. Lorsque les techniques de filtrage ne sont pas mises en place par l'internaute lui-même, il peut y avoir une perte de maîtrise de la part de

l'internaute, une perte d'initiative quant aux choix des contenus désirés. Cette perte de moyens pourrait aboutir à une fracture numérique, à savoir un Internet à plusieurs niveaux, constitué d'une hiérarchie fondée sur la possibilité donnée aux individus de contourner les processus de filtrage. Sur le plan technique, lorsque l'internaute est placé dans une situation de consultation de contenus préalablement filtrés par un système non transparent, il n'est pas créateur. Il ne va pas contribuer à l'expansion du réseau, laissant ainsi la place pour un cercle d'initiés.

3.2.3. Un respect de la légalité plus aisé :

Toutefois, si le filtrage est à son initiative, le législateur pourra s'assurer plus aisément du respect de la loi et surtout évaluer pleinement la validité des intermédiaires afin de protéger non seulement l'utilisateur, mais aussi de garantir le respect des droits. **[Réf. 20]**

Notons néanmoins qu'un filtrage des échanges et des communications pourrait conduire à un excès de protection de la part des internautes, notamment par l'utilisation de technologies de chiffrement. Or, il pourrait y avoir une régulation de chiffrement et de cryptage, comme cela a été le cas en France jusqu'à la loi du 26 juillet 1996. Ce type de régulation est à même de créer des inégalités et peut potentiellement fausser la concurrence. On obtient alors un Internet à deux vitesses avec d'un coté ceux autorisés à utiliser la cryptographie « forte » et ceux n'en ayant pas l'autorisation (et qui par conséquent se trouvent désavantagés par rapport aux premiers).

3.2.4. Un enjeu éthique important :

Soulignons que les enjeux en termes d'éthiques et de vie privée sont majeurs dans le cas du filtrage opéré par un tiers sans l'intervention de l'internaute. En effet, dans l'hypothèse où l'internaute n'est pas informé des techniques de filtrage ni des filtres mis en avant, certaines informations le concernant pourraient être analysées sans que son consentement soit recherché.

3.3. Filtrage sous la maitrise de l'internaute :

Dans ce cas, l'internaute décide ce qu'il désire consulter et met en œuvre lui-même une solution de filtrage. Cette décision n'appartient qu'à lui.

3.3.1. Une responsabilisation plus forte de l'internaute :

L'internaute va donc chercher à se protéger d'éventuelles agressions, à protéger ses ressources systèmes et à éviter certains contenus, pour des raisons qui lui sont propres, et dont il aura l'initiative. C'est le sens des solutions centrées sur l'utilisateur qui placent l'initiative de ce dernier au cœur du processus de filtrage.

L’enjeu est d’initier une dynamique responsable quant à ce que l’on accepte ou non, dans le respect de la position de l’utilisateur. Or, en plaçant l’utilisateur dans une situation de responsabilité, cela permet de donner les moyens au plus grand nombre, de prendre part au monde numérique qui va baigner, faciliter et permettre la plupart des échanges. L’internaute va être créateur, initiateur, concepteur, hébergeur.

3.3.2. Un risque d’inégalité face au filtrage :

L’inconvénient de ce type de filtrage réside dans le fait que cela suppose une maîtrise technique que tous les internautes peuvent ne pas avoir. Cela implique donc un effort de pédagogie supplémentaire. Par ailleurs, placer entre les mains de l’internaute la question de la responsabilité du filtrage amène à une pleine conscience de cette responsabilité, à une nécessaire connaissance fine de la loi et des différentes réglementations que tout le monde n’a pas : on risquerait une mise en place différentielle du filtrage qui n’impacterait pas les utilisateurs d’une façon uniforme.

Par conséquent, il pourrait être nécessaire de définir des paramétrages de filtrage « par défaut » contournables et documentés dont le sens et la finalité seraient à même d’être appréhendés par tous quelque soit son niveau.

Par ailleurs lorsque l’internaute décide de lui-même de ce qu’il souhaite filtrer, selon des paramètres et des problématiques qui lui sont propres, il prend le risque de sur-bloquer inutilement. Par exemple, s’il a paramétré une liste de mots-clefs dans son moteur de recherche, certaines recherches qu’il mènera pourraient être biaisées du fait d’un filtrage volontairement ou involontairement trop poussé. Cependant, contrairement à un filtrage qui ne serait pas placé sous la maîtrise de l’utilisateur, un mauvais paramétrage n’entrainerait de conséquences que pour l’utilisateur.

3.4. Filtrage et avenir d’Internet : les variables risques liées au filtrage

Outre l’initiateur du filtrage, il convient d’évoquer les impacts globaux du filtrage en termes de variables-risques. La procédure n’est pas anodine et a des impacts durables sur la mise en place d’une politique numérique pour le pays. Ces risques variables peuvent assez facilement être étudiés dans des mesures d’impacts qui devraient toujours éclairer des décisions touchant à Internet. Il n’est pas le lieu d’étayer tous les sujets indiqués, mais simplement de les mentionner

La mise en place d’infrastructures – logiciels, réseaux, machines – performantes est un enjeu clair de compétitivité. La question n’est pas seulement de mettre à disposition des personnes un réseau, ni de donner un moyen de distraction

supplémentaire, il s'agit de donner les moyens, au plus grand nombre, de prendre part, au monde numérique qui va baigner, faciliter et permettre la plupart de nos échanges. Cette thématique touche directement l'un des principaux enjeux du principe d'accès libre au réseau.

Penser des mécanismes de filtrage va nécessairement entraîner un projet global de management des risques qui lui sont liés. En effet, le filtrage a des conséquences sur le réseau lui–même et sur la circulation des données :

- **Qualité et fluidité de nous réseaux :** Un réseau non filtré aura une meilleure qualité en termes de débit puisqu'il n'y aura aucun point de passage obligatoire à franchir.
- **Risque de perte compétitive :** Filtrée, l'infrastructure pourrait être moins performante Ne risquons-nous pas de décourager toute une population d'entrepreneurs, d'inventeurs et de faire fuir les meilleurs à l'échelle internationale ?
- **Risque de dénaturation et d'oubli :** Pour des raisons économiques et pragmatiques, il pourrait être décidé de re-centraliser à l'extrême les points d'échanges et de routage d'Internet. Au lieu d'avoir une myriade de points d'entrée et de sortie, il y aurait un point d'entrée unique et un point de sortie unique. Cette hypothèse conduit nécessairement à une saturation du réseau. Ce type de réseau ne serait dès lors plus réellement Internet et on ne voit pas nettement ce qui le distinguerait des précédents réseaux de diffusion, comme par exemple le Minitel. Les risques potentiels de fragilisations volontaires ou involontaires des architectures et des systèmes seraient alors beaucoup plus élevés qu'aujourd'hui. La conséquence serait alors une simplification des attaques et des paralysies complètes de ce type de systèmes. A l'heure actuelle, les attaques sont disséminées en différents points d'entrée et de sortie. Canaliser l'ensemble des données en une seule route comportant un point d'entrée et de sortie pourrait amener à la saturation.
- **La pérennité des contenus filtrés :** Enfin, il convient de souligner que le filtrage ne résout pas le problème du contenu. En effet, filtrer un contenu n'implique pas sa disparition du réseau. Le contenu litigieux existera toujours. Il sera simplement invisible pour certaines personnes. Par ailleurs, le fait de filtrer un contenu plutôt que de le supprimer n'exclut pas l'apparition de sites miroirs, qui vont relayer le contenu. L'exemple illustrant le mieux cette hypothèse est Wikileaks. Lorsque le site a été neutralisé, les internautes ont relayé les différents câbles, montrant ainsi que ce n'était pas parce qu'une information était bloquée ou filtrée qu'elle

n’existait plus. Cet exemple pourrait être la démonstration que le filtrage ne résout pas le problème initial : l’existence d’un contenu litigieux.

4. Conclusion :

La problématique de filtrage d’un contenu litigieux concerne également la notion de persistance des contenus sur Internet. Ceux-ci en effet, même filtrés persistent. Ils constituent une sorte de mémoire de l’Internet qu’il est finalement difficile d’effacer, mais à laquelle il est tout aussi difficile d’accéder. Dès lors, la question du filtrage intéresse aussi l’indexation des données filtrées et leur visibilité dans les moteurs de recherche. Autrement dit, le filtrage ne peut faire l’économie ultérieure d’une réflexion complexe sur ces derniers.

4.1. La question de la sécurité :

Avec la problématique de la persistance des données, se pose la question de la sécurité sur Internet, question d’autant plus essentielle lorsqu’elle est mise en rapport avec les publics dits fragiles, notamment les enfants. De manière naturelle, penser à recourir au filtrage afin de protéger ce public semble légitime.

Néanmoins il a été vu que les contenus filtrés continuent d’exister sur Internet. Ils ne sont pas effacés et restent accessibles aux personnes qui ont des compétences techniques particulières. Cela pose ainsi la question de l’égalité d’accès aux contenus sur Internet entre une sphère composée de personnes qui ont des connaissances techniques permettant de contourner les mesures de filtrages et d’accéder même illégalement à des contenus (films, musique…) et une autre sphère de personnes qui ne les posséderaient pas (et qui continueront donc à payer pour accéder à des contenus que d’autres obtiennent gratuitement). Risque-t-on de se diriger vers un filtrage à deux vitesses, dépendant de la maturité technologique des individus ? On a là une question importante qui touche la démocratisation de l’accès à Internet et l’antinomie possible entre le filtrage et la neutralité du net.

4.2. La question de l’innovation :

Indéniablement, le filtrage a pour conséquence de rendre Internet davantage centralisé car davantage dépendant de points de passages. D’où une moindre rapidité d’un Internet centralisé par rapport à d’autres Internets (afférents à d’autres pays) ralentissant alors certains protocoles ou rendant plus lents certains services.

Aujourd’hui, les usages ont évolué du fait des possibilités techniques. La fibre optique permet une avancée sans commune mesure avec ce que permettait le cuivre. On ne parle plus de temps de commutation en centaines de millisecondes, mais en nanosecondes. De quoi demain sera fait ? Certains y travaillent déjà15. N’allons-nous pas être de facto écartés de ces sujets, n’ayant plus les moyens

techniques de nous en rapprocher ? Ceci signifie que toute politique de promotion du filtrage va de pair avec un effort technologique important : le temps de l'innovation sur Internet est d'une rapidité importante. Une politique de filtrage doit donc être en phase avec ces évolutions sous peine de risquer une obsolescence forte et donc une in-opérabilité rapide.

Ce risque d'obsolescence est à mettre en rapport avec la proximité des stratégies de filtrage et de certaines infrastructures (réseaux, machines, etc.) : plus les stratégies de filtrage se rapprochent des infrastructures, plus elles appellent la prise en considération de multiples facteurs apparemment étrangers aux finalités de la stratégie initiale envisagée. C'est là qu'intervient la notion de « complexité » du filtrage : celui-ci n'est pas une grille de contrôle applicable délibérément, mais un système complexe au sens d'une mise en mouvement d'une série d'entités en interactions. Ces entités, malgré la précision de plus en plus poussée des ingénieries algorithmiques, ne sont pas à l'abri d'une certaine imprévisibilité (on l'a vu avec le phénomène du sur blocage).

Dès lors, toute stratégie de filtrage nécessite une délimitation des objectifs et une anticipation claire et rigoureuse des conséquences de sa mise en œuvre : c'était là un des objets de ce document, que de mettre en lumière les conséquences d'un choix d'un type de filtrage.

La prudence est donc de mise et tout déploiement d'une stratégie de filtrage est à envisager à bon escient. En effet, l'accumulation de stratégies désordonnées de filtrage pourrait provoquer des dysfonctionnements importants dans les réseaux, susceptibles de pénaliser la compétitivité numérique du pays, c'est-à-dire sa capacité à répondre aux exigences des entreprises, aux aspirations individuelles en matière d'accès au numérique et à la vitalité d'un modèle économique du numérique en pleine maturation. En la matière, c'est une dialectique féconde entre liberté et responsabilité qui est à rechercher.

Avançons alors l'idée que l'utilisateur puisse devenir non pas tant son propre FAI, mais son FAL : Fournisseur d'Accès Local. On considère ici qu'il convient de privilégier une autonomisation locale de l'accès, en même temps qu'un déplacement stratégique des entités centralisées (les grands FAI) vers la périphérie (le réseau domestique). Cette dynamique est intéressante car porteuse d'un double potentiel : la conviction d'une responsabilisation efficace de l'utilisateur en même temps qu'une potentialité d'autonomisation de l'individu par la technologie. »

Chapitre III

Présentation de Twitter

1. Présentation de Twitter :

Twitter (terme d'argot anglais pouvant signifier gazouiller est un service de type Web 2.0 / réseau social autour du thème du micro-blogging. La raison d'être de ce service est de permettre de rester connecté avec sa communauté en échangeant de courts messages textes qui ont la particularité d'être limités à 140 caractères. Twitter permet à chaque utilisateur de signaler à tous les membres de son réseau "ce qu'il est en train de faire" tel que résumé dans le texte introduisant la saisie des Tweets : Qu'est-ce que vous faites?

Une des différences entre Twitter et un blog traditionnel réside dans le fait que Twitter n'appelle pas directement une participation à commenter les messages postés.

- **Type :** Microblog.
- **Architecture :** Web 2.0
- **Date de création :** 2006
- **Auteur :** Jack Dorsey, Evan Williams et Biz Stone
- pour le moment gratuit et affranchi de publicité.

Figure 4 : Espace de tweet.

1.1. Principe de fonctionnement :

Les messages échangés, similaires à des SMS, sont appelés updates (mises à jour) ou Tweets (qui peut être traduit de l'anglais en gazouillis). Ils permettant de suivre depuis sa propre page d'accueil Twitter les actualités et les statuts de tous les membres de son réseau.

Les updates des membres de son réseau sont affichés sur la page publique du compte Twitter de son auteur (http://twitter.com/nom_utilisateur) et immédiatement relayés dans la private timeline (page d'accueil) de chacun des membres qui se sont abonnés pour suivre ces updates (appelés followers).

Par défaut, tous les tweets de tous les membres Twitter sont rassemblés au sein de la Public Timeline (http://twitter.com/public_timeline), une variation d'un blog public mis à jour en temps réel avec tous les tweets des membres.

Twitter permet d'envoyer et de recevoir des updates depuis son site Internet mais également par SMS ou par téléphones portables, BlackBerry, iPhone, et autres équipements de communication nomades.

La flexibilité est un atout de Twitter. Celle-ci est renforcée par la mise à disposition d'APIs maintenues par Twitter. Une API est une interface de programmation ouverte et documentée pouvant être utilisée par des développeurs pour mettre en œuvre facilement et rapidement de nouveaux services et applications s'appuyant sur la plateforme de communication Twitter. Ainsi, et bien que Twitter ne se limite à des messages textes, des applications permettent d'envoyer des photos via Twitter au travers de sites tiers.

1.2. Utilisation de Twitter :

Aujourd'hui Twitter est essentiellement utilisé par des internautes blogger friands de communication. Le site Internet a montré ses capacités à relayer très rapidement des informations sur le vif, comme lors des attentats de Bombay au mois de novembre 2008 ou lors de la cérémonie d'investiture du Président Barack Obama aux USA en janvier 2009. Certains professionnels, pour la plupart liés à l'industrie des médias, utilisent massivement et professionnellement Twitter. C'est ainsi le cas de certains journaux d'informations, comme CNN ou BBC qui utilisent des robots pour envoyer des flashs d'information via Twitter. Les candidats à la campagne présidentielle américaine de 2008, Barack Obama en tête, ont ainsi massivement utilisé cet outil de communication.

Un des aspects négatifs mis en avant par les détracteurs de Twitter est la vacuité de certains échanges et le manque de crédit et de véracité de certaines informations relayées. Le dessinateur François Cointe fait remarquer que si l'on devait utiliser Twitter pour décrire réellement ce qu'on est en train de faire, tout le monde devrait écrire qu'il est en train d'écrire sur Twitter.

Figure 5 : Interface utilisateur.

2. Glossaire des termes clés :

2.1. Tweet :

Court message texte limité à 140 caractères échangé via Twitter.

Twitter limite les messages échangés à 140 caractères afin de permettre la mise à jour de son statut de manière brève et spontanée.

Update :

Synonyme de Tweet.

2.2. Trending Topics :

Les Trending Topics sont les sujets les plus souvent abordés par la communauté des utilisateurs de Twitter. Ils sont présentés sous forme de mots-clés les plus souvent utilisés dans les Tweets.

2.3. Followers :

L'ensemble des comptes Twitter qui suivent les Tweets d'une personne donnée. On peut le traduire en français par « suiveurs ».

Les followers est la liste des personnes qui vous suivent sur Twitter. Cela implique que les updates sont diffusés à tous les followers.

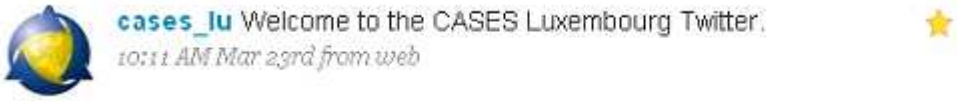
2.4. Following :

L'ensemble des comptes Twitter que l'on suit pour être informé de leurs updates. On peut le traduire en français par « suivre ».

Les following est la liste des personnes que vous suivez sur Twitter. Ceci implique que l'on reçoit tous les updates des membres que l'on suit.

2.5. Favorites :

Tweets que vous avez lu, qui vous ont plus et que vous avez marqués comme favoris (favorites). Ces derniers sont identifiés par  et toujours accessibles dans le menu Favorites de votre profil public (sauf si vous protégez vos updates).



A noter que vous pouvez ajouter comme favori n'importe quel Tweet, que vous soyez ou non un follower de son auteur.

2.6. Public timeline :

La public timeline est la section publique de Twitter qui regroupe tous les Tweets de tous les membres (qui ne protègent pas leurs updates).

La public timeline de Twitter est accessible par l'adresse http://twitter.com/public_timeline.

2.7. Private timeline :

La private timeline est la section privée de Twitter regroupant les propres Tweets de l'utilisateur mais également ceux des membres qu'il suit.

2.8. Direct message :

Twitter propose d'envoyer des messages directement à un membre précis. Le message ne s'affiche pas sur la public timeline ni sur la private timeline du destinataire mais dans le menu Direct Message.

2.9. @reply :

Un @ reply est un tweet public envoyé d'un membre à un autre. Ce qui le distingue d'un tweet habituel est le préfixe @nom_utilisateur avec le nom de l'utilisateur auquel on souhaite faire une réponse. La réponse est alors affichée dans un menu spécifique dans la configuration Twitter du destinataire du message.

À noter qu'il n'est pas nécessaire de suivre une personne pour faire un @ reply et ainsi lui envoyer un message. Il est toutefois possible de bloquer un utilisateur pour ne plus recevoir de @ reply.

3. Conditions d'utilisation de Twitter et charte de respect de la vie privée :

Twitter communique sur le site Internet les conditions d'utilisation du service (http://twitter.com/tos) et la charte de respect de la vie privée (http://twitter.com/privacy). Ces informations sont susceptibles d'être modifiées à tout moment et sans préavis par Twitter.

3.1. Conditions d'utilisation du service :

Twitter impose d'être âgé d'au moins 13 ans pour créer un compte et l'utiliser.

Il est interdit de se servir de Twitter pour des actions en violation des lois. Il est également interdit de créer des sites Internet ressemblant à Twitter (phishing). Il est également interdit d'utiliser Twitter pour diffuser des Spams, des virus ou vers informatiques.

L'internaute est responsable de toutes les actions qu'il effectue par son compte mais également de toutes les informations qu'il échange et communique par Twitter.

Il est de sa responsabilité de choisir et de conserver son mot de passe de manière sécurisée.

Le non respect des conditions d'utilisation peut entraîner la suppression du compte Twitter incriminé.

Le cybersquatting est interdit sur Twitter. Le site peut décider de réattribuer des comptes si des actes de cybersquatting ou assimilé sont mis en évidence.

Twitter assure ne pas s'approprier le droit d'auteur des communications de ses membres. Toutes les informations déposées par les membres sur Twitter restent leur propriété. Twitter se réserve le droit de retirer des contenus qui seraient contraire aux droits d'auteur ou qui violeraient les lois relatives aux copyrights.

3.2. Définition de vie privée :

Au sens classique ou historique, la vie privée a été définie comme « le droit de vivre en paix ». Au 21^{e} siècle, cependant, la vie privée a revêtu plusieurs dimensions. Pour certaines personnes, la vie privée signifie avoir droit à un espace privé, pouvoir effectuer des communications privées, être libre de toute surveillance et respecter le caractère sacré de la personne. (Vie, 2003)
Il n'existe pas de définition juridique précise de la vie privée. La vie privée est l'ensemble des activités d'une personne qui relève de son intimité par opposition à la vie publique. Les contours de la notion de vie privée sont relativement flous. L'atteinte à la vie privée peut résulter de la diffusion d'un écrit ou d'une image concernant la personne.
On peut considérer cependant comme privé :

- **L'intimité :** Etat de santé, opinions politiques et religieuses, appartenance ethnique, mœurs, relations personnelles, sociales, appartenance syndicale, vie professionnelle…..

- **La vie familiale,**
- **Le domicile,**
- **Les loisirs,**
- **Les circonstances de la mort,**
- **Le doit à l'image,**
- **La correspondance privée,**
- **Les atteintes à l'honneur et à la réputation.**

Le droit au respect de la vie privée est proclamé par la loi. [Badinter, 2007]

La vie privée ou plus précisément « le droit à l'intimité de la vie privée », fait partie des droits civils. Les composantes de la vie privée n'on pas fait l'objet d'une définition ou d'une énumération limitative afin d'éviter de limiter la protection aux seules prévisions légales. Les tribunaux ont appliqué le principe de cette protection au droit à la vie sentimentale et à la vie familiale, au secret relatif à la santé, la résidence et le domicile, et le droit à l'image ». **[Broudo, 2009]**

3.3. Charte de respect de la vie privée :

Twitter recommande de fournir le plus d'informations possibles sur son profil (biographie, ville, site Internet, etc.) mais précise bien que renseigner plus d'informations que celles requises pour l'inscription sur le service est totalement optionnel et sans impact pour le bon fonctionnement du compte.

Twitter se réserve le droit de communiquer à sa seule discrétion les informations que vous lui avez confiées aux autorités publiques ou à des entreprises privées dans le but de coopérer et de répondre à des procédures légales.

Twitter considère comme asset (bien informationnel) toutes informations personnelles de chacun de ses membres et se réserve le droit de vendre, transférer ou partager partie ou totalité de ces assets en relation avec une fusion, une acquisition, une réorganisation ou une fermeture de Twitter.

Twitter assure être impliqué dans la confidentialité des données personnelles de ses membres et déclare utiliser les moyens de protection organisationnelle, physique et logique pour protéger ces informations contre des accès non autorisés.

Twitter se réserve le droit d'utiliser les informations de contact renseignées par ses membres (numéro de téléphone, adresse e-mail, etc.) pour proposer des informations de nature commerciale sur ses produits et services. Le site se réserve également le droit de proposer des informations de nature commerciale pour des services et produits qui ne sont pas gérés par Twitter. Il est toutefois possible de se désinscrire de telles communications au sein de sa page de configuration.

4. Risques liés à l'utilisation de Twitter :

4.1. Propagation de vers :

Avec sa popularité grandissante, Twitter est confronté au même type de menaces que l'on retrouve sur les réseaux sociaux. Des vers informatiques sont conçus pour exploiter Twitter et se propager de comptes en comptes et diffuser des messages. Ceci s'est déjà passé en avril 2009 où quelques centaines de comptes Twitter ont été corrompus et plusieurs milliers de Tweets diffusés par un ver informatique mais sans porter atteinte aux informations personnelles.

Les premiers cas de propagation semblent encore avoir un impact limité, mais il est fort à parier que des vers utilisant Twitter seront probablement développés dans le but d'impacter beaucoup plus sérieusement la sécurité.

4.2. Indisponibilité de Twitter :

Twitter connaît un grand succès auprès des internautes mais la rançon de cette « gloire » est une indisponibilité ponctuelle du service. Une étude a montré que Twitter est un des sites de réseau social qui a subi le plus d'indisponibilités en 2008, ce qui représente un total de 84 heures, soit une moyenne de 7 heures par mois (ou près de 15 minutes par jour).

4.3. Usurpation d'identité :

Il n'y a pas de certitude sur l'identité de son interlocuteur sur Twitter et il est aisé d'y « emprunter » l'identité de son choix. Des exemples célèbres font état de l'usurpation de l'identité d'hommes ou de femmes politiques ou bien du vol de mots de passe pour des comptes Twitter célèbres (Britney Spears ou Barack Obama en 2008).

4.4. Tinyurl et bit.ly :

Puisque Twitter permet uniquement de diffuser des messages limités à 140 caractères, un problème survient lorsque l'on souhaite communiquer à ses followers une longue adresse Internet comme, par exemple,

http://www.cases.public.lu/fr/publications/fiches/pdf/droit_a_image.pdf (71 caractères). Il serait bien plus facile de diffuser une adresse telle que http://tinyurl.com/bv5ov4 ou http:///bit.ly/bv5ov4 pour accéder au même document.

Twitter permet de réaliser cette transformation à l'aide du site bit.ly qui permet de faire le lien entre une adresse Internet de son choix et une adresse du type http://bit.ly/tuv123 pour pouvoir l'intégrer facilement dans un tweet.

4.5. Spam :

Des comptes Twitter peuvent être utilisés pour diffuser du spam. Il n'est pas rare de recevoir une demande d'un follower ou de voir un follower suivre vos updates sans que vous ne le connaissiez. Ce type d'approche peut être un préalable pour envoyer des @ replies relayant des spams.

4.6. Relayer des informations non vérifiées :

Twitter fait parti du Web 2.0, où l'internaute est consommateur d'informations mais également acteur en en publiant.

Une des fonctionnalités les plus utilisées de Twitter est la transmission en temps réel d'informations sur des évènements d'actualité. Il faut se souvenir que bien

souvent ces informations sont envoyées de Twitter en Twitter et que peu de personnes en vérifient les sources et la vraisemblance. En fait, l'information initiale peut avoir été mal comprise puis mal relayée ce qui peut avoir des conséquences négatives sur la teneur et la véracité des propos. De même, de nombreux canulars (hoax) y circulent.

4.7. Contrôle de la diffusion de données personnelles :

Twitter permet de gérer :

- Les Tweets que vous publiez,
- La liste des personnes qui vous suivent (Followers),
- La liste des personnes que vous suivez (Following),
- Vos Tweets favoris (Favorites),
- Les @ replies que vous envoyez,
- Les @ replies que vous recevez.

Le seul moyen de contrôler la diffusion de vos tweets est l'activation du paramètre Protect my updates via la configuration de votre profil.

Ceci a pour objectif :

- ✓ De ne donner accès à ses propres tweets seulement à destination des followers que vous choisissez,
- ✓ De contrôler la liste des followers,
- ✓ De ne pas diffuser ses tweets sur la public timeline.

5. Création d'un compte Twitter :

La sécurité d'un compte Twitter commence par le respect des bonnes pratiques lors de sa création.

5.1. Étape #0 :

La création d'un compte commence par

la connexion sur le site http://www.twitter.com.

Une fois sur la page d'accueil de Twitter, il suffit de cliquer sur le bouton

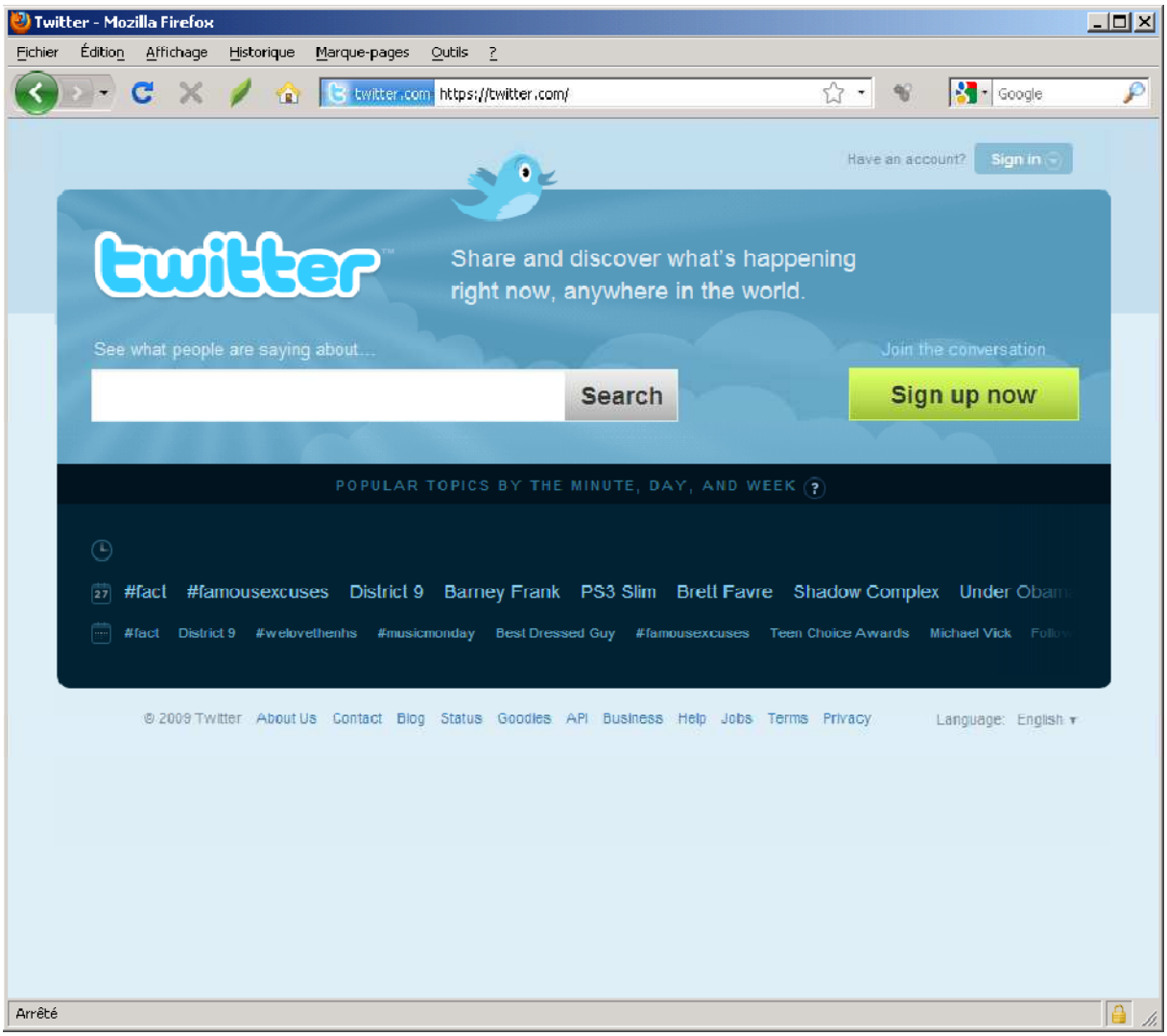


Figure 6 : Page d'accueil de Twitter.

5.2. Étape #1 :

De nombreuses informations sont demandées pour la création d'un compte, toutefois. Twitter permet de modifier ultérieurement l'ensemble de ces informations.

Il est important de noter que lors de la création d'un compte.

Noms des champs	**Descriptions**	**Recommandations CASES Luxembourg**
Full Name	Le nom d'utilisateur tel qu'il apparaîtra dans le profil public	Ne pas utiliser votre vrai prénom ni votre vrai nom de famille ni tout autre information révèlerait votre identité réelle Ne pas utiliser de surnom qui pourrait vous porter préjudice et ternir votre réputation
Username	Le nom d'utilisateur qui, combiné avec le mot de passe, permet de vous authentifier sur Twitter Le nom d'utilisateur qui sera renseigné est celui qui sera utilisé dans l'URL publique de votre profil Twitter (comme rappelé dans la fenêtre de saisie)	Ne pas utiliser votre vrai prénom ni votre vrai nom de famille ni tout autre information qui révèlerait votre identité réelle Ne pas utiliser de surnom qui pourrait vous porter préjudice et ternir votre réputation
Password	Le mot de passe, qui, combiné avec le Nom d'utilisateur, permet de vous authentifier sur Twitter pour rédiger vos *tweets*	Utiliser un mot de passe fiable en suivant les bonnes pratiques décrites dans la fichethématique et le dossier suivants : •http://www.cases.public.lu/fr/publications/fiches/pdf/Fich_MotsdePasse.pdf • http://www.cases.public.lu/fr/pratique/comportement/mot_de_passe/ À noter que Twitter ne demande pas de confirmation de la saisie du mot de passe. Il best donc recommandé de saisir son mot de passe très soigneusement afin d'éviter les fautes de frappe. Il sera toutefois possible de demander la réinitialisation du mot de passe en recevant la procédure à suivre sur l'adresse e-mail saisie lors de la création du compte
Email	L'e-mail qui sera utilisé par Twitter pour notifier des évènements liés à votre compte Twitter. Cette adresse e-mail est également Utilisée pour la procédure de réinitialisation de mot de passe	Utiliser une adresse e-mail dépersonnalisée de votre véritable identité pour ne pas donner d'informations personnelles vous concernant

☐ *I want the inside scoop-please me email updates*	En cochant cette case (décochée par défaut), vous autorisez Twitter à envoyer sur votre e-mail des informations, newsletters ou publicités sur le fonctionnement du service	Ne cocher cette case qu'en connaissance de cause, en sachant que vous pourrez recevoir de la publicité sur l'adresse e-mail que vous avez saisie précédemment
Type the words above	Le Captchat (*Type the words above*) pour s'assurer que la création du compte Twitter n'est pas réalisée par des robots	n/a

Tableau 1 : Les informations demandées lors d'une création d'un compte.

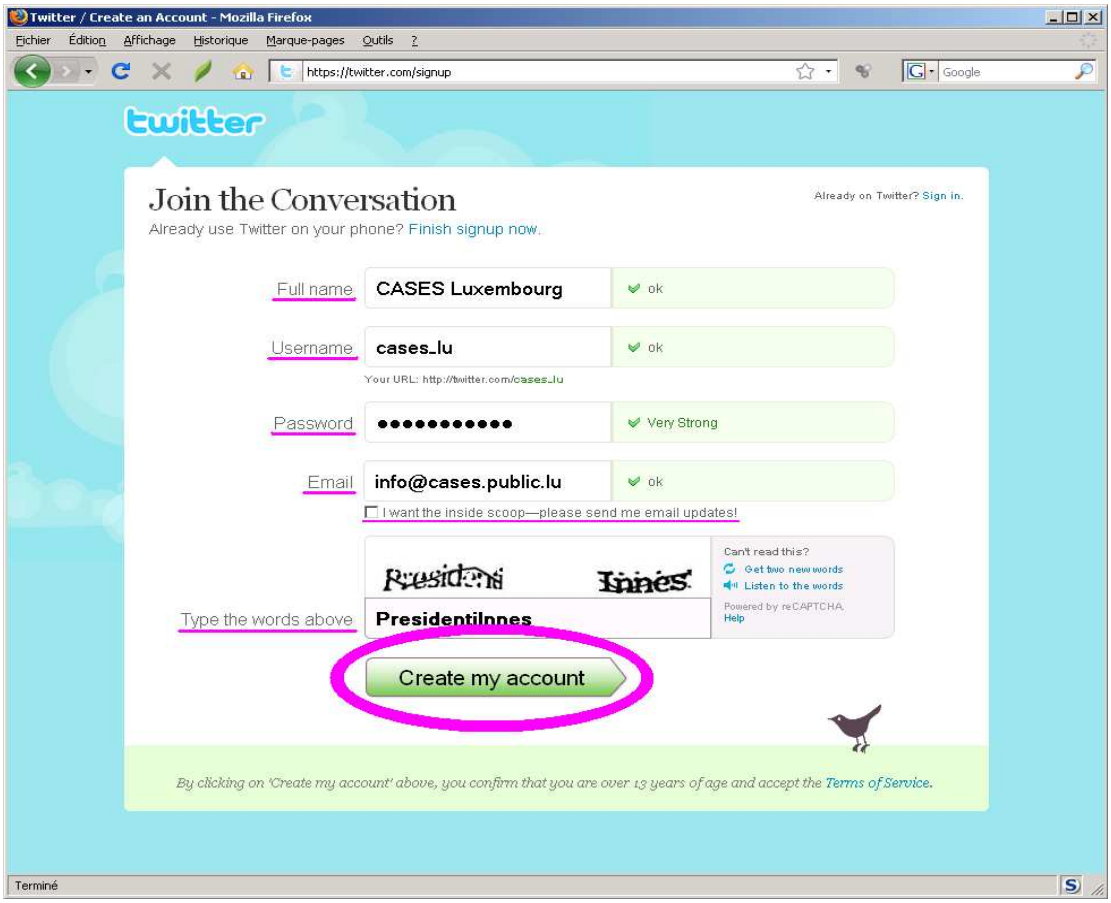
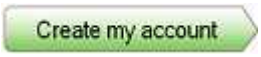


Figure 7 : Page pour saisie les informations personnelles.

Une fois les différents éléments renseignés, on passe à l'étape suivante en cliquant sur



5.3. Étape #2 :

Twitter propose de se connecter à votre place sur différents comptes de messagerie sur Internet pour collecter la liste des adresses e-mail de vos contacts pour ensuite les inviter à utiliser Twitter.

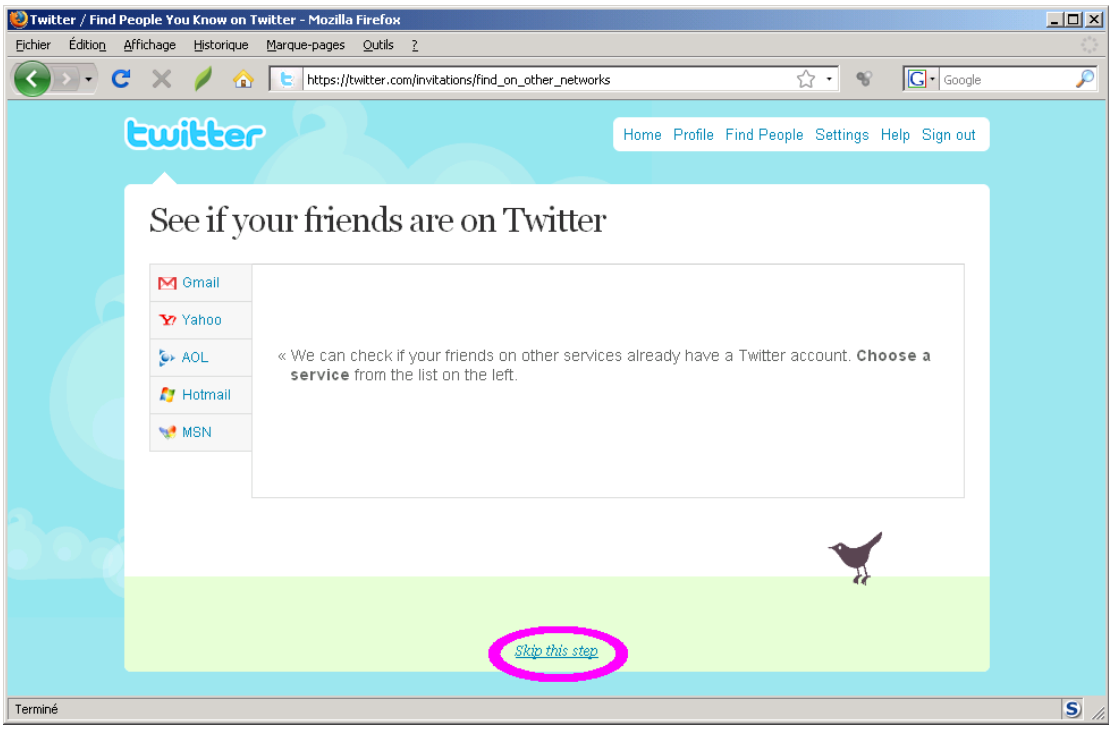


Figure 8 : Interface pour se connecter à d'autre compte.

Evidemment pour permettre à Twitter de se connecter à votre place sur vos différents comptes de messagerie sur Internet, vous devez lui confier vos identifiants et vos mots de passe comme cela est demandé dans la capture d'écran suivante :

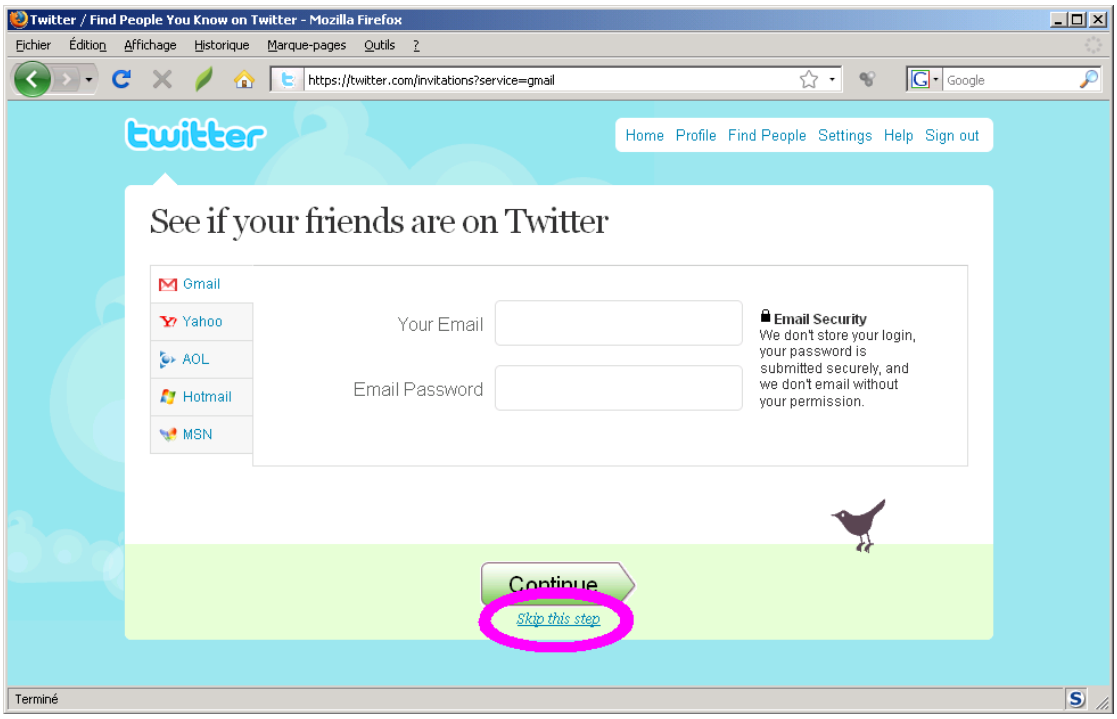


Figure 9 : Interface pour connecter à Twitter.

5.4. Étape #3 :

Twitter propose au cours de cette étape de suivre une sélection de comptes Twitter pour lesquels on devient immédiatement followers.

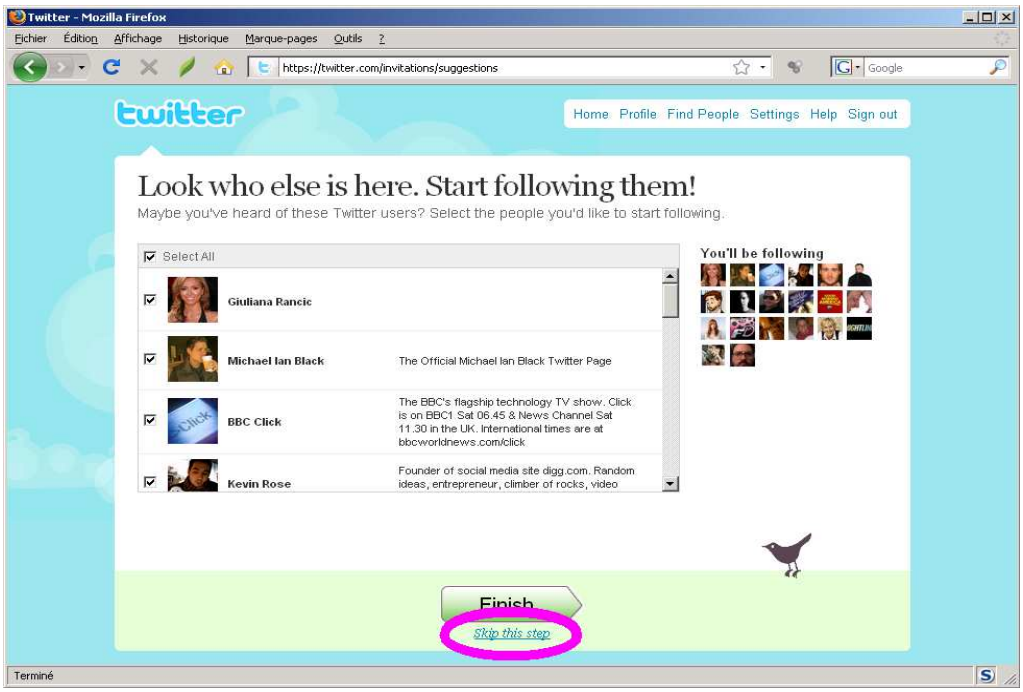


Figure 10 : Interface pour sélectionner d'autre compte.

Cette sélection de 20 comptes Twitter est constituée par le site lui-même et n'apporte aucun gage de qualité ou de pertinence des informations.

Ca y est, vous êtes connecté à Twitter et vous pouvez écrire votre premier tweet !

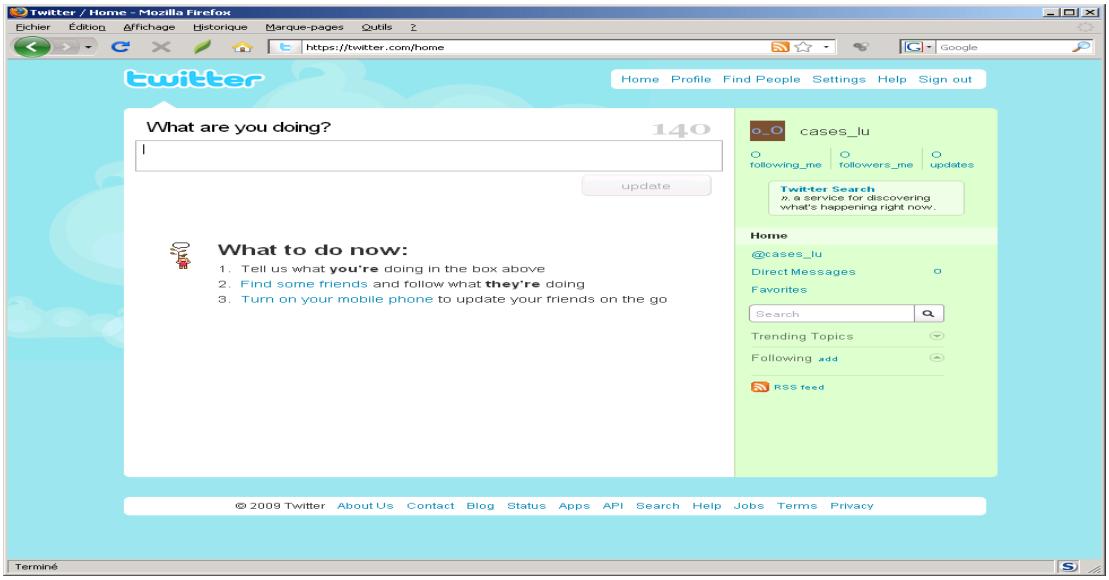


Figure 11: Interface où on écrit les messages.

6. Suppression de messages et suppression de compte Twitter :

6.1. Effacer ses messages :

Bien qu'il soit impossible de modifier un message une fois écrit, il est possible d'effacer les messages de son choix (tant que l'on en est l'auteur).

Pour effacer un message, il faut cliquer sur  .

Il est important de noter qu'un message effacé peut déjà avoir été publié dans la public timeline. Ceci veut dire que des traces de l'existence de ce message peuvent subsister sur Internet. De plus, bien qu'il soit possible de supprimer ses messages, il est impossible d'effacer des messages de following une fois publiés sur sa private timeline, même si le contenu est offensant, à moins de demander à son auteur de le supprimer (ce qui aura pour conséquence de supprimer ce messages de toutes les private timeline de chaque follower).

6.2. Supprimer son compte Twitter :

Twitter propose une option pour supprimer son compte Twitter. La suppression de son compte a les conséquences suivantes :

- Aucune restauration d'un compte supprimé n'est possible.
- Un compte supprimé reste visible sur Twitter un temps indéterminé après sa suppression et peut toujours apparaître dans des résultats de moteurs de recherche même après sa suppression (NB : ceci est hors du contrôle de Twitter).
- Il est impossible de créer un nouveau compte Twitter en utilisant une information (nom d'utilisateur, numéro de téléphone ou adresse e-mail) déjà utilisée dans un ancien compte, même si celui-ci a été supprimé.

A noter qu'il n'est pas nécessaire de supprimer un compte si l'on souhaite seulement changer son nom d'utilisateur (ceci est possible dans le menu Settings).

6.3. Procédure à suivre :

Pour supprimer son compte Twitter, il faut se rendre dans le menu et choisir le lien situé en bas de la page

- N'oubliez pas de modifier les informations liées à l'adresse e-mail utilisée pour ce compte Twitter si vous avez l'intention de créer ultérieurement un nouveau compte Twitter en utilisant cette même adresse e-mail.

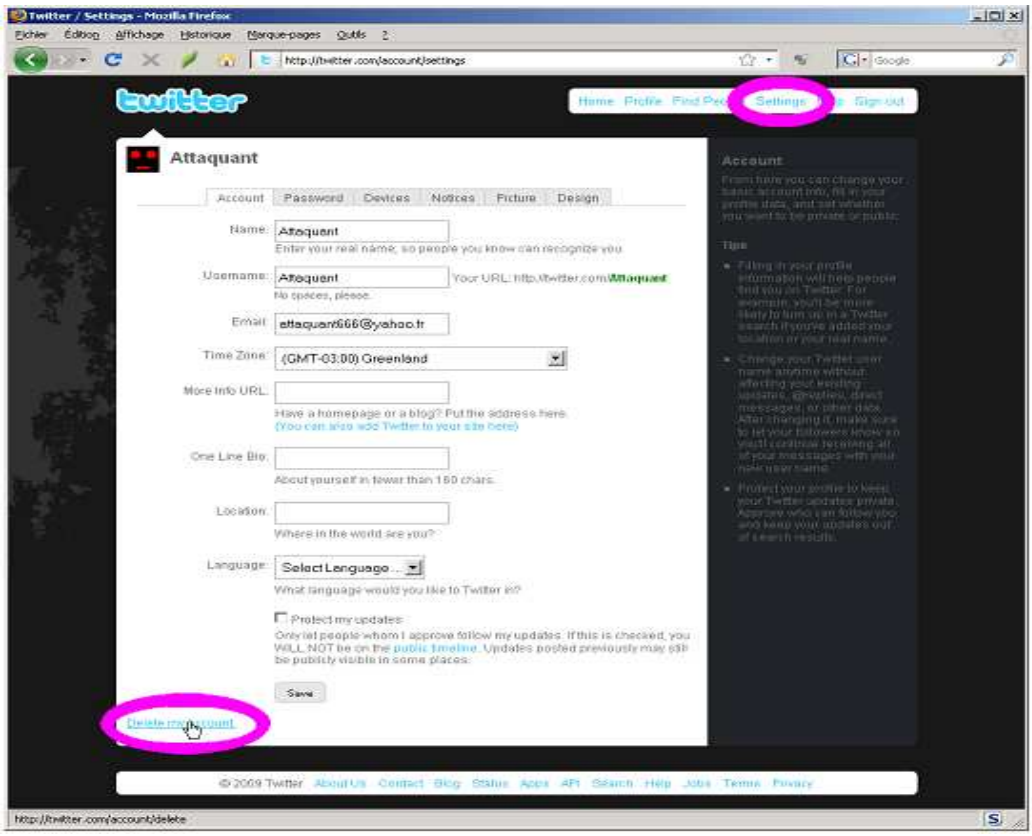


Figure 12 : Interface pour suprême un compte.

- La page suivante s'affiche et rappelle les conséquences de la suppression de son compte :

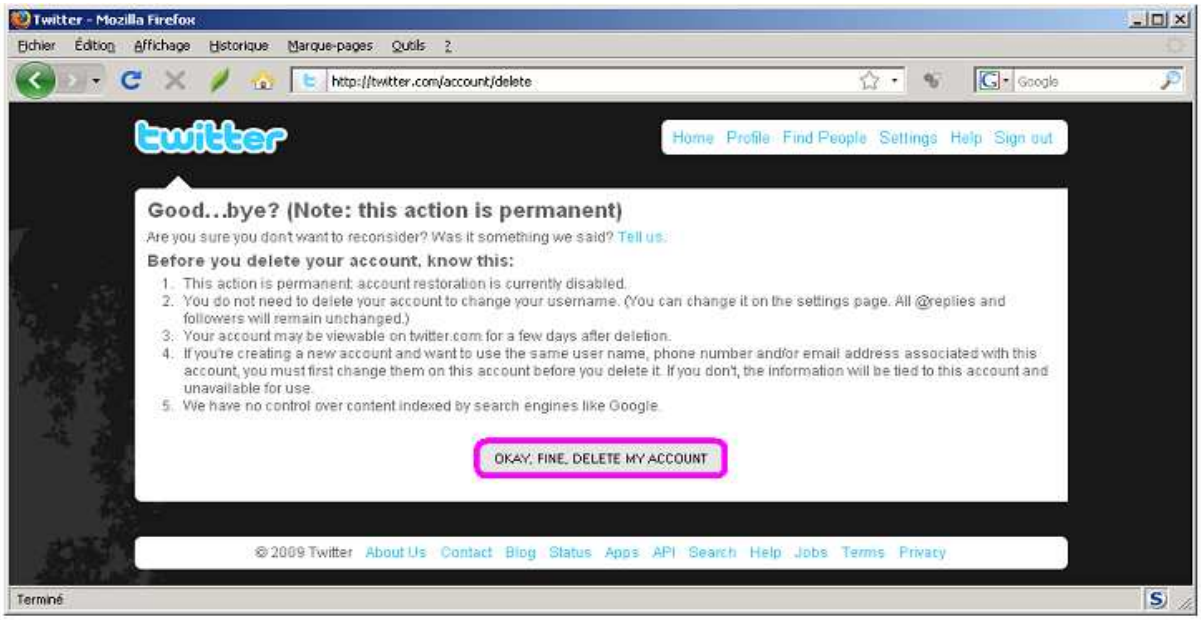


Figure 13 : Interface affiche les conséquences.

- Pour confirmer la suppression il faut cliquer sur le bouton

OKAY, FINE, DELETE MY ACCOUNT

- La suppression du compte dans les conditions décrites précédemment est confirmée par le message *Your ac-count will be deleted. Bye!* et par le retour à la page d'accueil de Twitter comme montré sur la capture d'écran suivante:
    - A noter que le message de confirmation écrit en anglais utilise le temps du futur, ce qui signifie que le compte Twitter n'est pas immédiatement supprimé mais le sera (dans un futur proche)

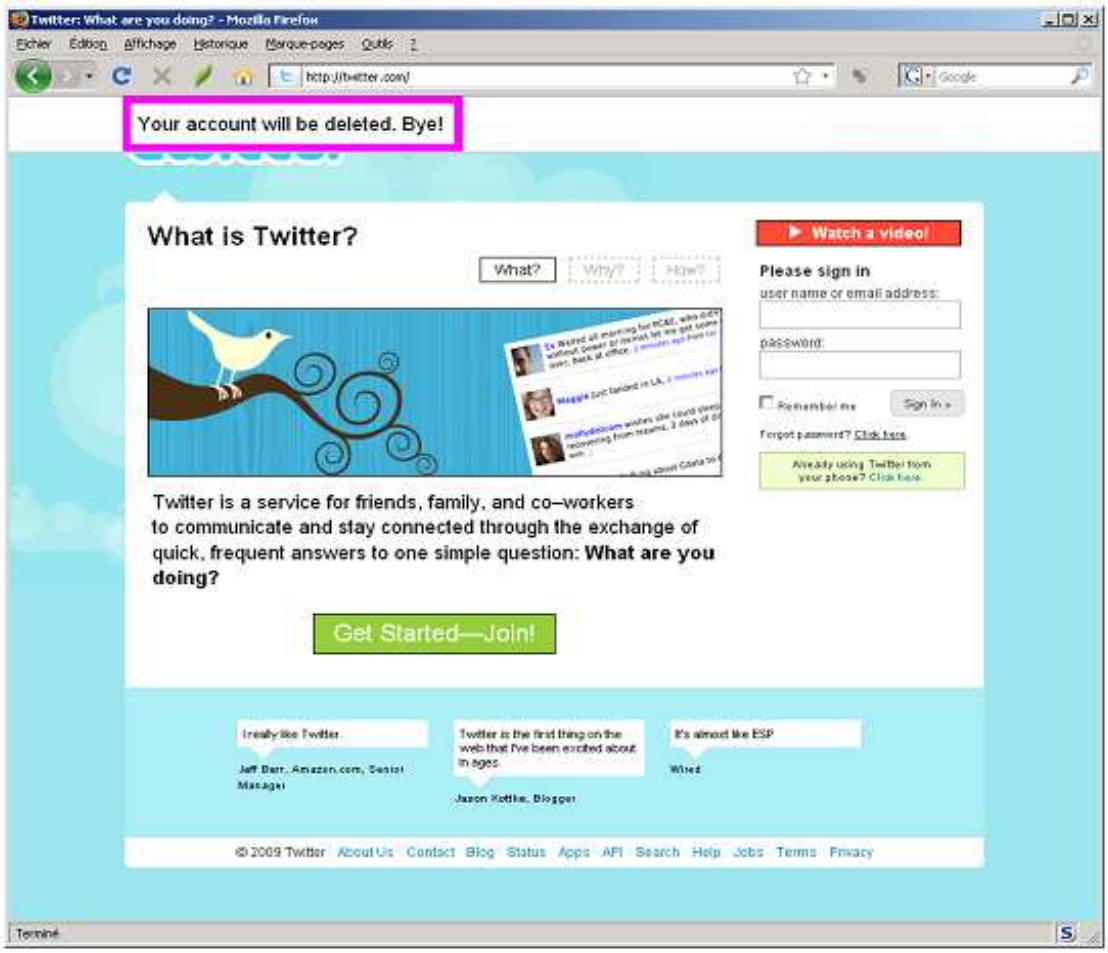


Figure 14 : Interface qui affiche la suppression de compte.

Partie II

Analyse et conception

Chapitre IV

Analyse et conception

1. Introduction :

Le recours à la modélisation est depuis longtemps une pratique indispensable au développement des logiciels, car un modèle sert à anticiper les résultats du codage, c'est en effet une représentation abstraite d'un système qui permet d'en faciliter l'étude, c'est un outil majeur de communication entre les divers intervenants au sein d'un projet. En outre, les systèmes devenant de plus en plus complexes, leur compréhension et leur maitrise globale dépassent les capacités d'un seul individu. La construction d'un modèle abstrait permet de remédier à ce problème. Le modèle a notamment l'atout de faciliter la schématisation du système, à savoir la possibilité de partir d'un de ses éléments et de suivre ses interactions et liens avec d'autres parties du modèle. Associé au processus (cycle) de développement, le modèle représente l'ensemble des vues sur une expression des besoins ou sur une solution technique. Le modèle sert donc d'objectifs différents selon le niveau de développement et sera construit à partir des points de vue de plus en plus détaillés. La modélisation objet consiste à créer une représentation informatique des éléments du monde réel auxquels on s'intéresse, sans se préoccuper de l'implémentation, ce qui signifie indépendamment d'un langage de programmation. Il s'agit donc de déterminer les objets présents et d'isoler leurs données et les fonctions qui les utilisent.

Pour le développement de notre application, nous avons opté pour une démarche de conception orientée objet, en nous basant sur la modélisation en UML.

2. Objectif de notre application :

L'objectif de notre application est de faire filtrer le contenu indésirables microblog par la méthode de FILTRAGE DE CONTENU, bloquer le tweet si c'est nécessaire.

- **Vérification de l'émetteur :** Filtrage par profil et par zone géographique
- **Analyse de contenu :** Filtrage de contenu.

Notre application est exécutée au niveau de serveur mais on peut l'appliqué aussi dans un :

- ✓ Pare-feu
- ✓ F.A.I.
- ✓ Proxy

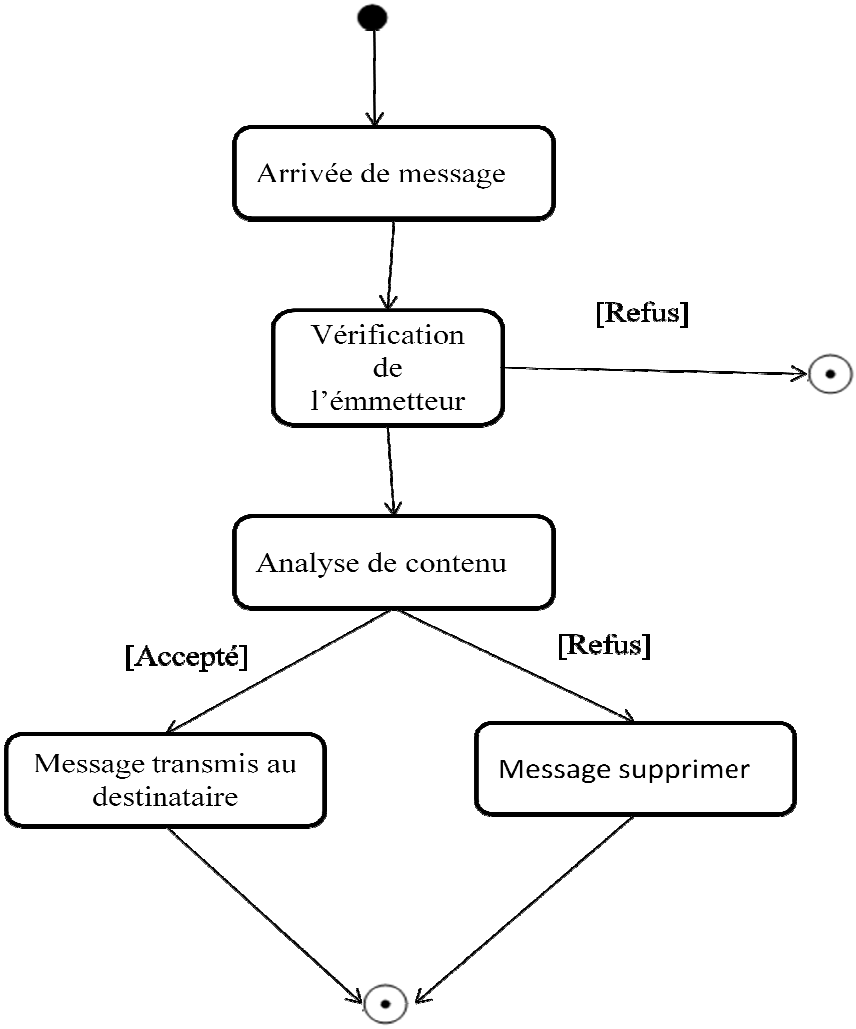


Figure 15 : Différent étapes de filtrage.

3. Présentation de l'UML :

Face à la diversité des méthodes d'analyse et de conception objet et en particulier aux différentes notations des mêmes concepts, L'UML (Unified Modeling Language) constitue une étape importante dans le domaine de la convergence des notions utilisées dans le domaine de l'analyse et de la conception. UML représente une synthèse de trois méthodes OMT (Objet Modeling Technic), Booch (Grady Booch) et OOSE (Objet Oriented Software Engenering).

L'UML permet de représenter la dynamique d'une application par la série de diagrammes qu'il offre. **[Réf. 25]**

3.1. Langage de modélisation UML :

L'UML (Unified Modeling Language) est un langage de spécification, de représentation graphique et de documentation d'un système orienté objet. UML a une notation graphique simple, précise et homogène. Il propose de plusieurs modèles qui sont des descriptions du système étudié et qui sont :

- **Les diagrammes d'activité :** représentation du comportement d'une opération en termes d'action.
- **Les diagrammes de cas d'utilisation :** représentation des fonctions du système du point de vue de l'utilisateur.
- **Les diagrammes de classes :** représentation de structure statique en termes de classes et de relations.
- **Les diagrammes de collaboration :** représentation spatiale des objets, des liens et des interactions.
- **Les diagrammes de déploiement :** représentation du déploiement des composants sur les dispositifs matériels.
- **Les diagrammes d'états transitions :** représentation du comportement d'une classe en terme d'état.
- **Les diagrammes d'objet :** représentations des objets et de leurs relations, correspond à un diagramme de collaboration simplifié, sans représentation des envois de message.
- **Les diagrammes de séquences :** représentation temporelle des objets et de leurs interactions.

Ces modèles sont élaborés par les utilisateurs au moyen des diagrammes. Un diagramme spécifie un aspect précis du modèle. UML offre une vue complète des aspects statiques et dynamiques d'un système en distinguant 09 diagramme.

3.2. Objectif de l'UML :

L'analyse objet, la conception objet et de l'implémentation du logiciel, il est destiné aux systèmes a fortes composantes logiciel, il a été utilisé avec succès dans les domaines tels que :

- Système informatique ;
- Les services bancaires et financiers ;
- La télécommunication ;
- Les transports ;
- Les services distribués basés sur le web ;

3.3. Avantages d'UML :

- Il permet de représenter l'aspect traitement du système aussi bien que l'aspect donné ;
- Il s'utilise sur l'ensemble de développement du logiciel ;
- Les principaux outils de modélisation permettent de travailler avec UML ;

4. Spécification :

4.1. Diagramme de contexte :

L'objectif de la messagerie instantanée étant en premier lieu de communiquer avec un ou plusieurs contacts, nous avons donc défini la fonction principale du système : communiquer. Il vient immédiatement le premier acteur du système : Filtrer.

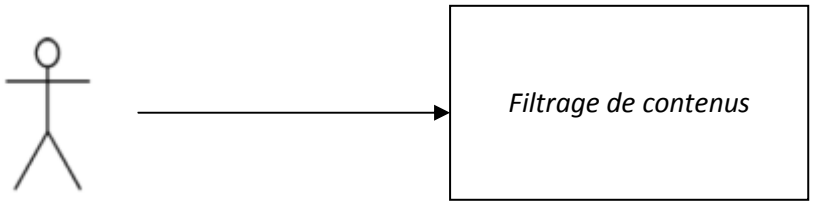


Figure 16 : Diagramme de contexte.

4.2. Diagramme des cas d'utilisations :

Les diagrammes de cas d'utilisation (use cases) représentent les cas d'utilisation, les acteurs et les relations entre les cas d'utilisation et les acteurs. Ils décrivent, sous la forme d'actions et de réactions, le comportement d'un système du point de vue d'un utilisateur. Ils permettent de définir les limites du système et les relations entre un système et l'environnement.

Un cas d'utilisation est une manière spécifique d'utiliser un système. C'est l'image d'une fonctionnalité du système, déclenchée en réponse à la simulation d'un acteur externe.

Dans notre cas, nous avons les cas d'utilisation suivants :

- Cas d'utilisation relatif à l'utilisateur
- Cas d'utilisation relatif à l'administrateur.
-

4.2.1. Cas d'utilisation :

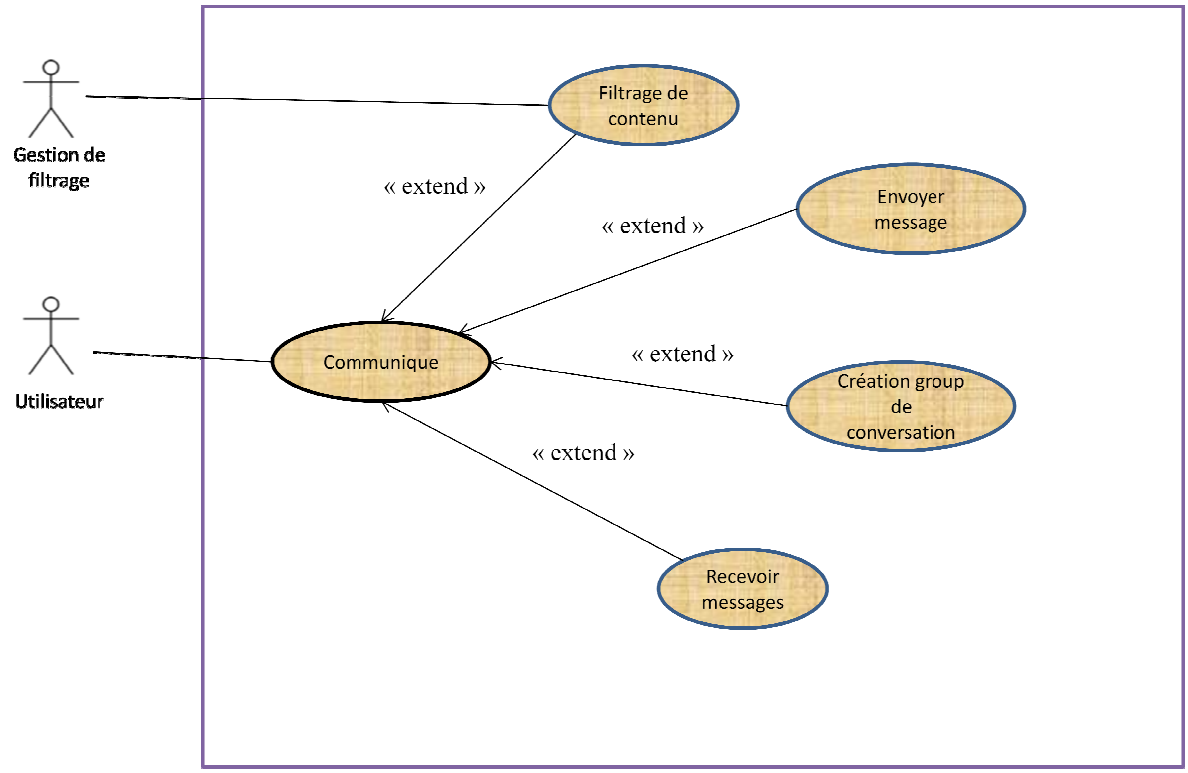


Figure 17 : Diagramme de cas d'utilisation générale

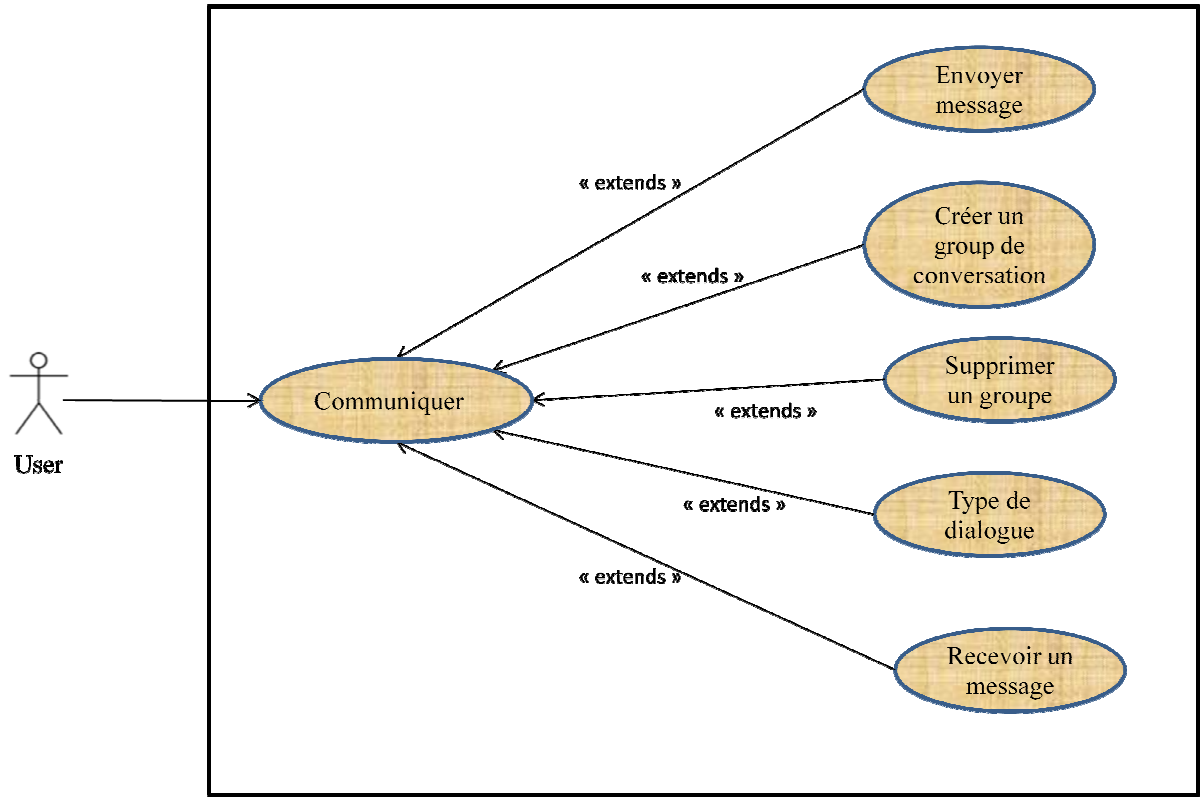


Figure 18 : Diagramme de cas d'utilisation des utilisateurs.

4.2.2. Les cas d'utilisation :

Acteurs	Cas d'utilisation	Sous cas d'utilisation
Gestion de filtrage	S'authentifier	
	Configuration (Filtrage par mots-clés)	Ajouter un mot-clé Supprimer un mot-clé
	Configuration (Filtrage des liens)	Ajouter un lien Supprimer un lien
	Configuration (Filtrage par profil et mots-clés)	Ajouter un mot-clé Ajouter un profil Supprimer un mot-clé Supprimer un profil
UTIL	S'authentifier	
	Accéder à la page réserve pour l'utilisateur	Sélectionner le mode de filtrage
	Connecter à l'internet	Saisie un site de Twitter
	Quitter l'application	

4.2.3. Cas d'utilisation relatif au gestionnaire de filtrage :

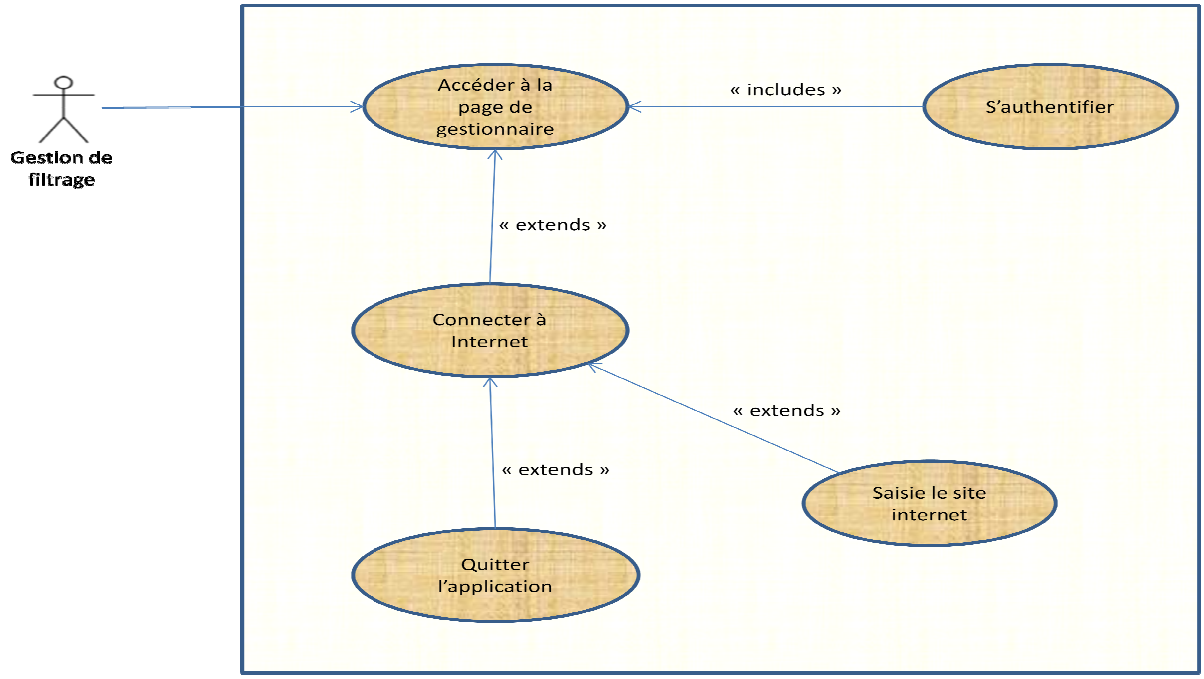


Figure 19 : Diagramme Use Cases realatif à l'administrateur

4.2.4. Diagramme de cas d'utilisation détaillé « configuration »

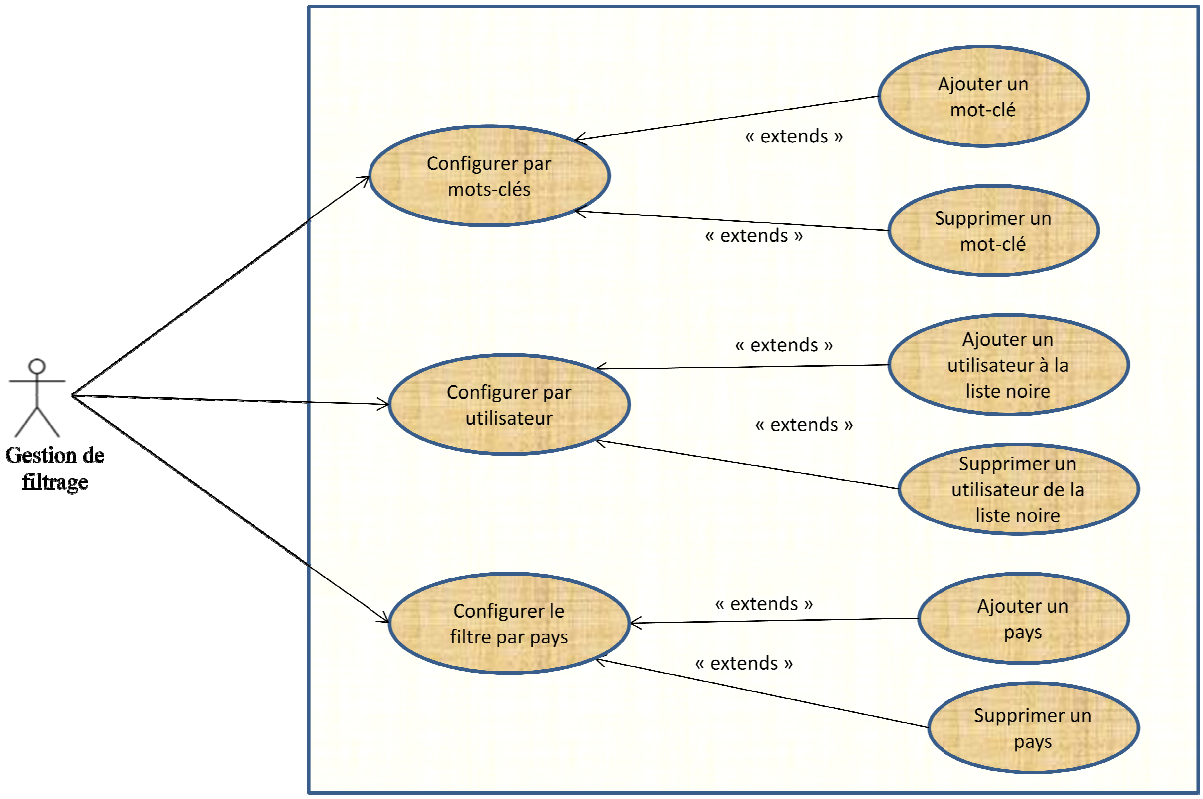


Figure 20 : Diagramme de cas d'utilisation détaillé «configuration »

4.3. Diagramme de classe :

On va décrire le modèle général du système. Ceci est possible à travers des diagrammes de classes qui offrent une vue statique du système, en représentant les classes et les relations entre les classes.

Une classe est un ensemble d'objet ayant les mêmes caractéristiques. En présentera quelques diagrammes de classes correspondant aux cas d'utilisation déjà décrits :

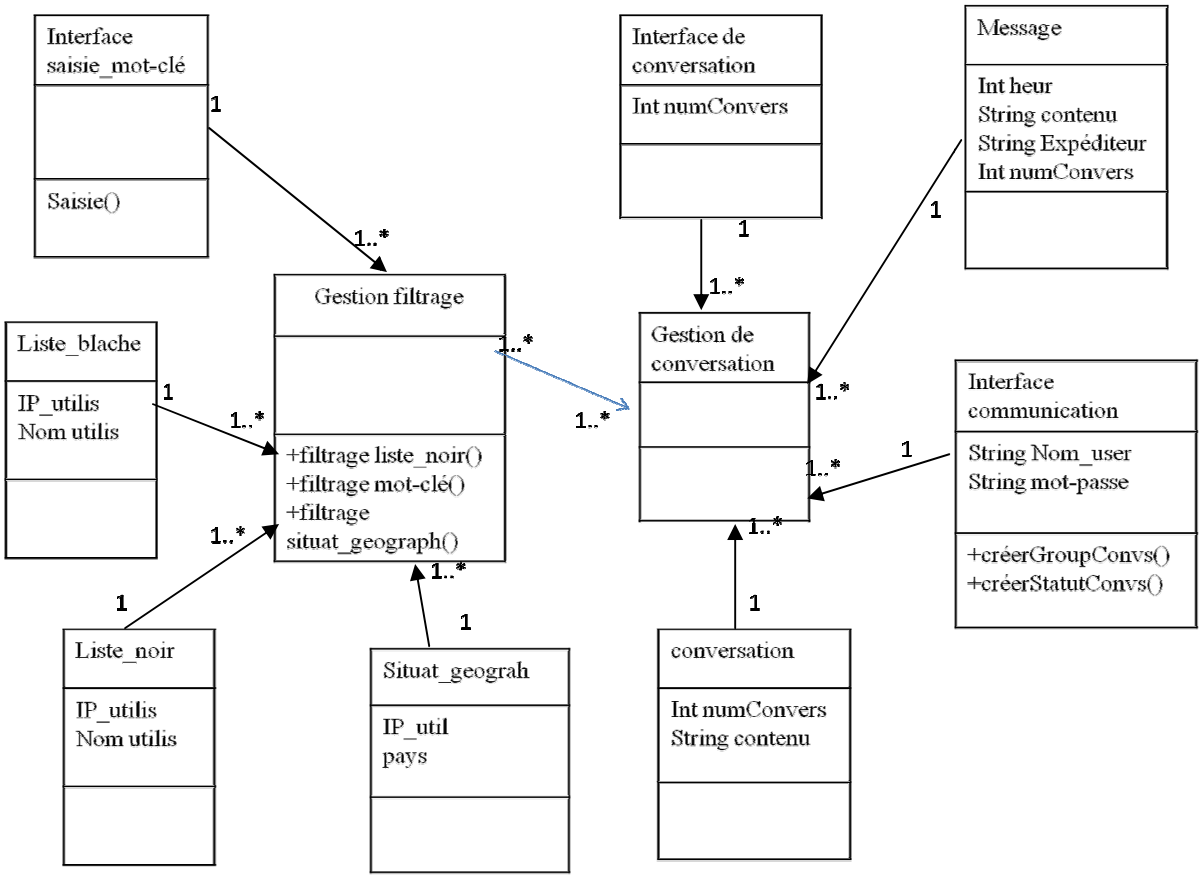


Figure 21 : Diagramme de classe.

5. Conception :

On va donner les diagrammes de séquence et le diagramme de classe.

5.1. Diagramme de séquence :

Il s'agit de représenter les différentes interactions possibles entre les objets de notre système selon un point de vue temporel.

Les symboles qui sont utilisés :

Les objets interface : Un objet d'interface représente l'interface entre l'acteur et le système tel des pages web ou les écrans de saisie.

L'icône :

Les objets entité : ils sont des objets décrits un cas d'utilisation et qui se trouvent dans d'autres cas d'utilisation tels le visiteur en ligne.

L'icône :

Les objets contrôlent : ils représentent les activités des processus du système. Ils d'érigent les activités des objets entité et interface.ces objets sont obtenus en extrayant les verbes des cas d'utilisation.

L'icône :

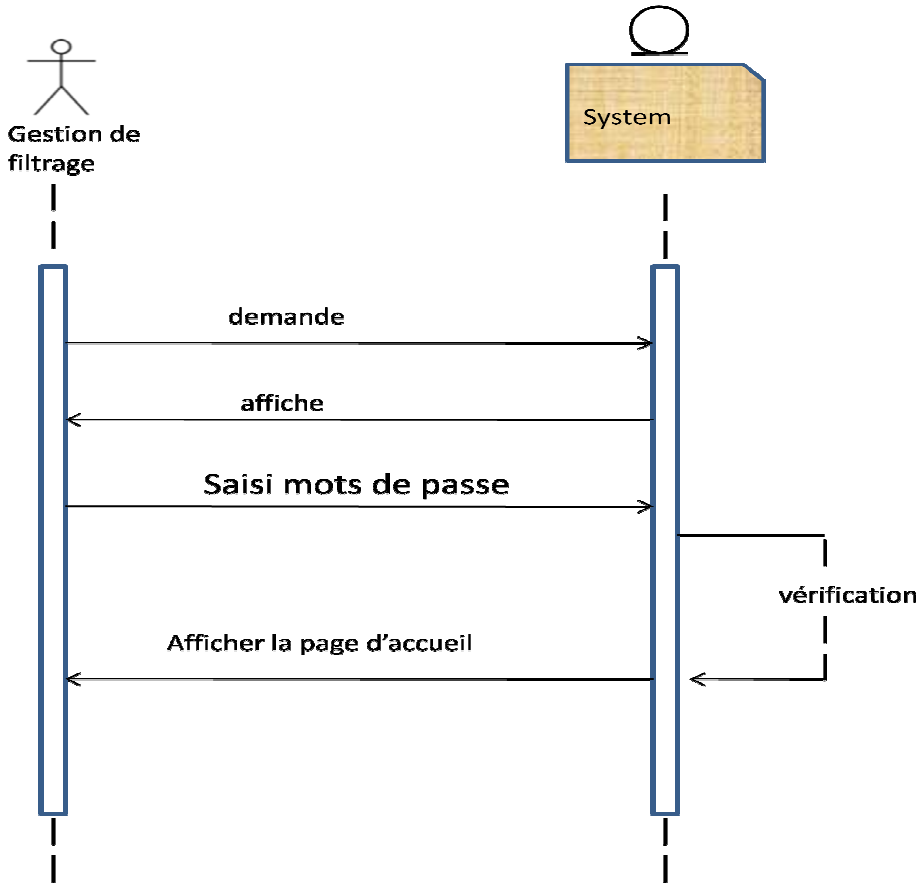


Figure 22 : Diagramme de séquence « authentification ».

1- L'administrateur demande le formulaire de l'authentification.
2- L'administrateur saisie le Login et le mot de passe.
3- Le system vérifie la validité de Login et le mot de passe.
4- L'application affiche la page d'accueil.

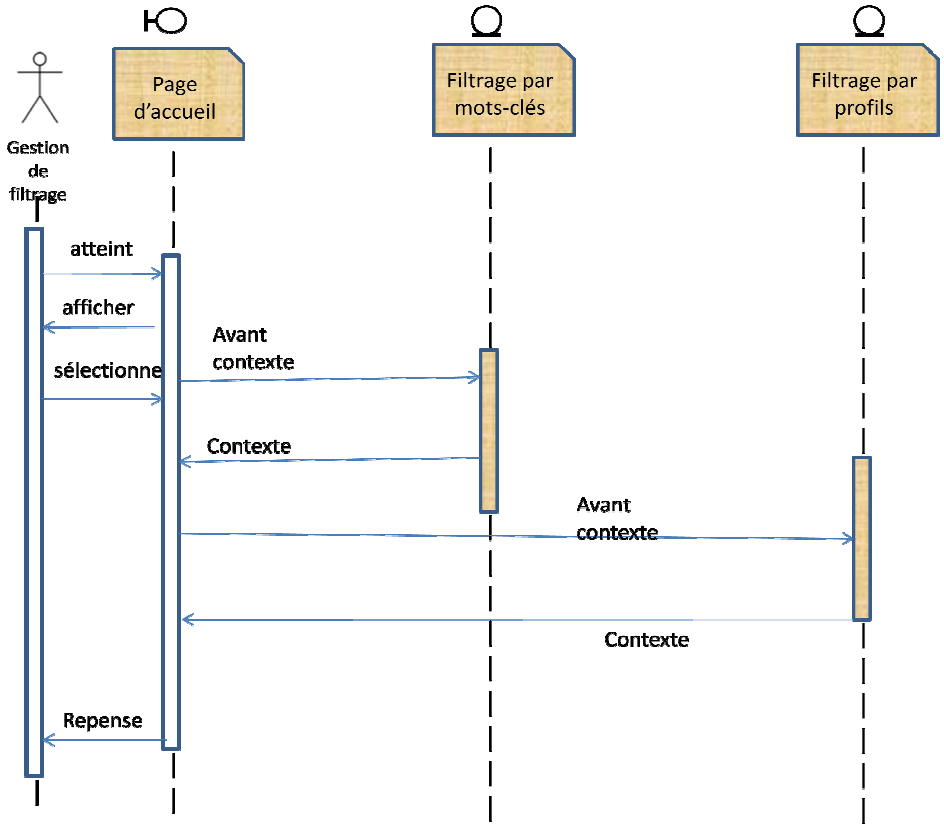


Figure 23:Diagramme de séquence de filtrage

1- L'administrateur s'authentifie pour accéder à son espace.
2- L'administrateur demande la page de filtrage par mots-clés.
3- L'administrateur demande la page de filtrage par profils.

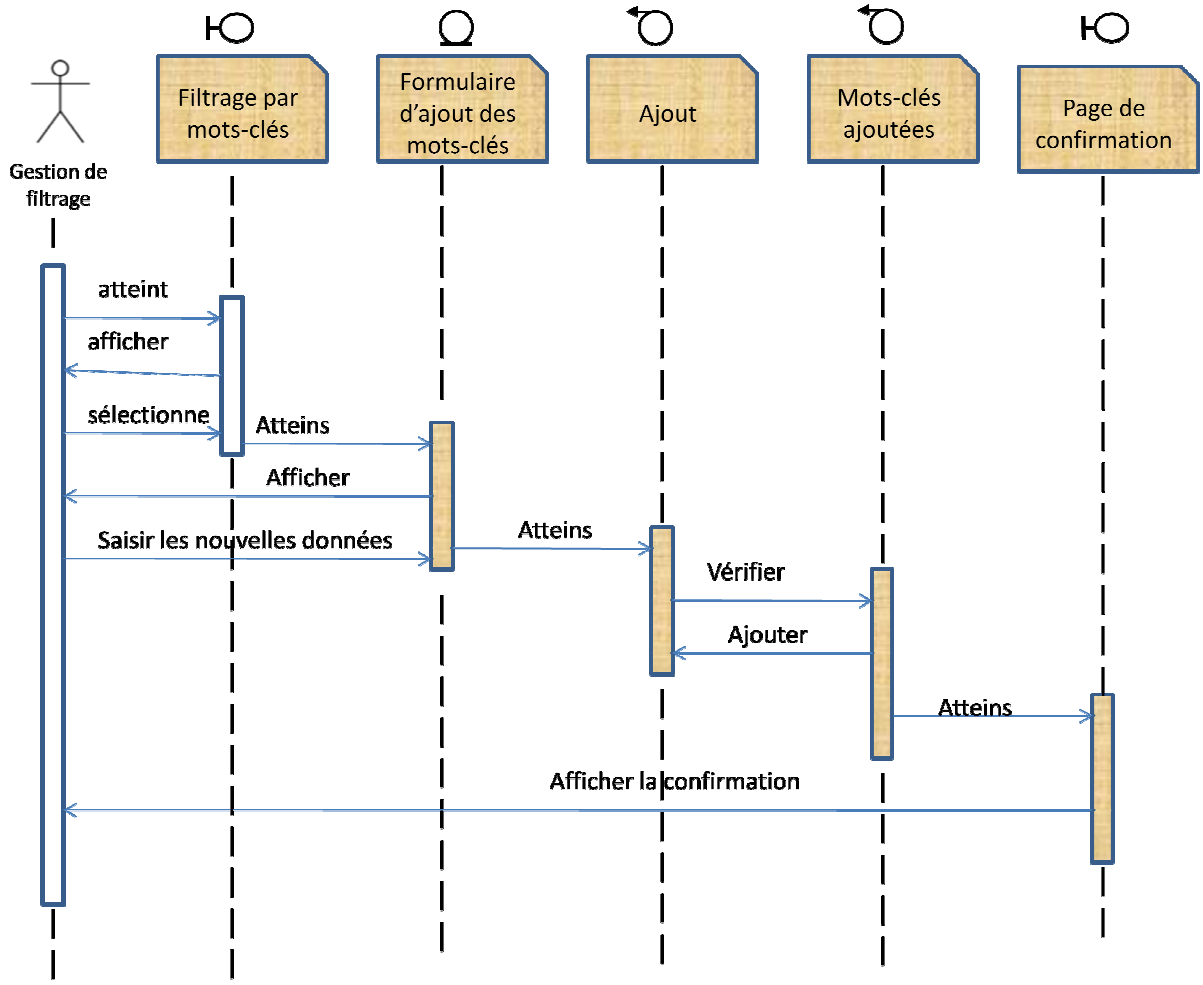


Figure 24 : Diagramme de séquence « ajout d'un mot-clé »

4- L'administrateur s'authentifie pour accéder à son espace.
5- L'administrateur demande le formulaire d'ajout.
6- L'administrateur saisit les données.
7- L'application envoi la requête.
8- L'application stocke les données au niveau de la base de données.
9- L'application confirme l'enregistrement.

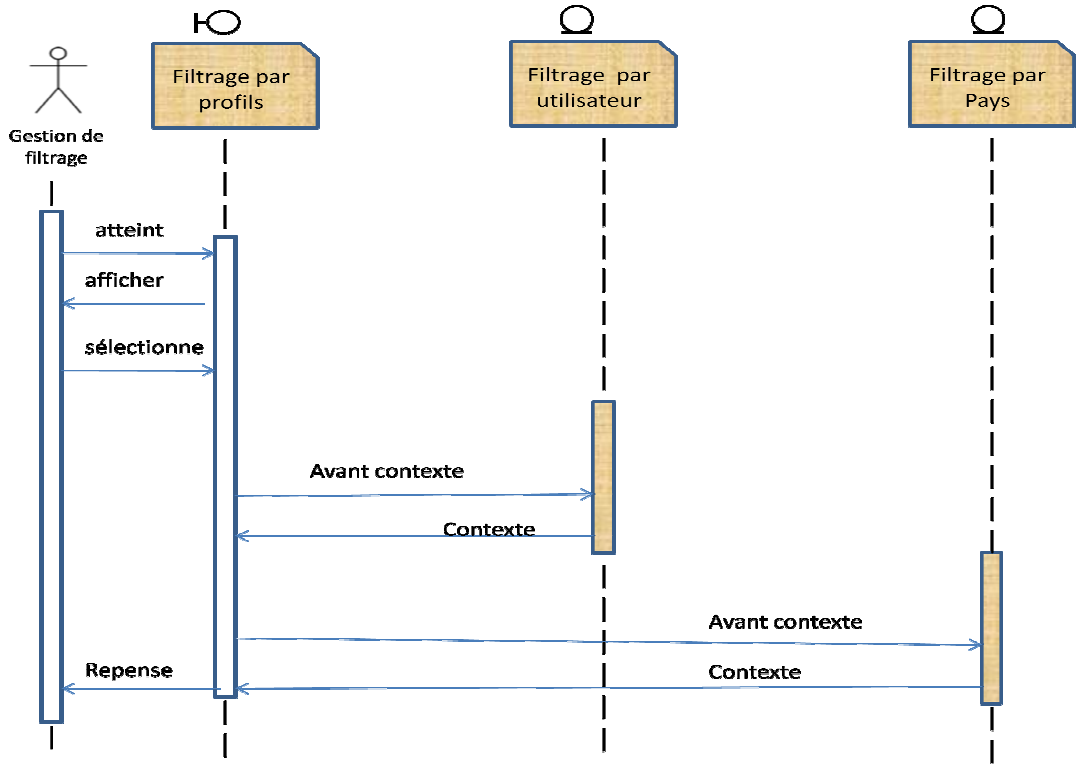


Figure 25 : Diagramme de séquence « Filtrage par profils»

1- L'administrateur s'authentifie pour accéder à son espace.
2- L'administrateur demande la page de filtrage par pays.
3- L'administrateur demande la page de filtrage par utilisateur.

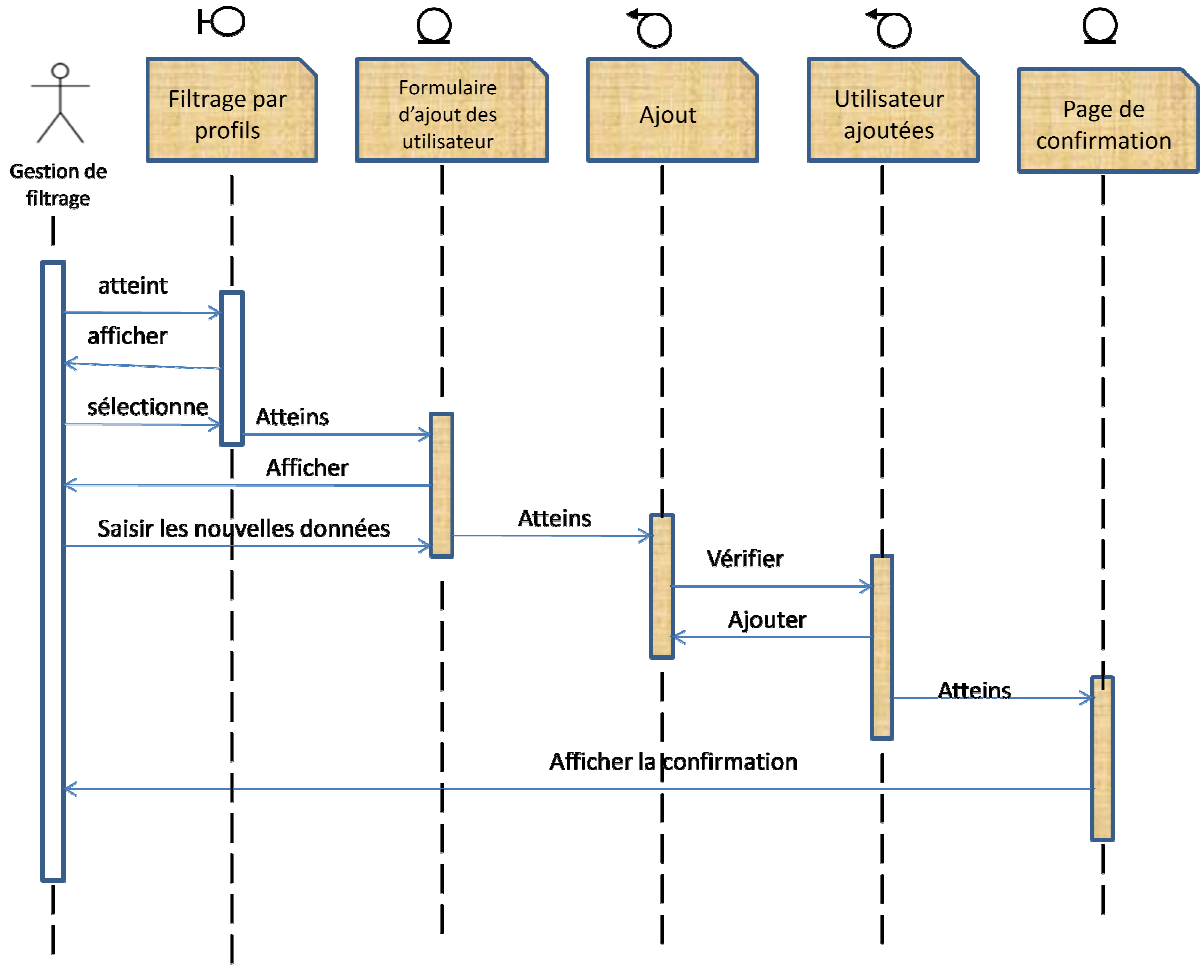


Figure 26 : Diagramme de séquence « Filtrage par utilisateur»

1- L'administrateur s'authentifie pour accéder à son espace.
2- L'administrateur demande le formulaire d'ajout.
3- L'administrateur saisit les données.
4- L'application envoi la requête.
5- L'application stocke les données au niveau de la base de données.
6- L'application confirme l'enregistrement.

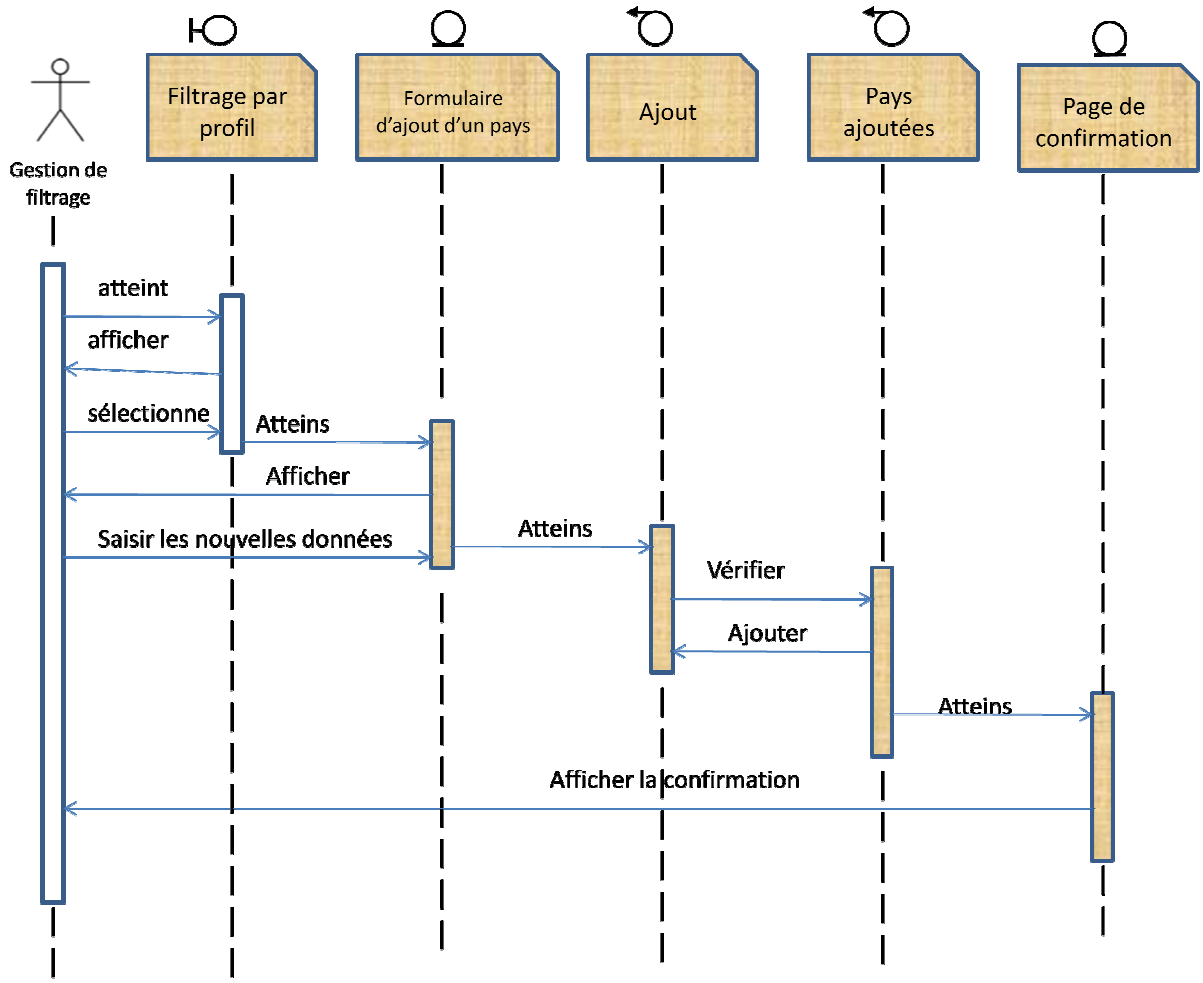


Figure 27 : Diagramme de séquence « Filtrage par pays»

1- L'administrateur s'authentifie pour accéder à son espace.
2- L'administrateur demande le formulaire d'ajout.
3- L'administrateur saisit les données.
4- L'application envoi la requête.
5- L'application stocke les données au niveau de la base de données.
6- L'application confirme l'enregistrement.

6. Conclusion :

Dans ce chapitre, nous avons proposé une démarche de modélisation pour développer notre application, cette démarche est basée sur la méthode UML , en commençant par définir les objectifs recherchés de notre application, ensuite nous avons défini les différents cas d'utilisations, diagrammes de séquences et le diagramme de classe général dans la phase de spécification des besoins, puis la conception nous avons élaboré le diagrammes de classe et le modèle logique. Une fois la conception terminée, nous passons à la réalisation qui fera l'objet du chapitre suivant, dans lequel nous présenterons les outils et environnement utilisés pour développer l'application, son fonctionnement général ainsi que toutes les interfaces utilisées.

Chapitre V

Réalisation

1. Introduction :

Après avoir présenté dans le chapitre précédent les différentes étapes : l'Analyse et Conception de notre système, nous allons présenter dans ce chapitre, le l'éternuement qui nous a servi d'appui pour le développement de notre application.

2. Langage de programmation utilisé :

Java est un langage de programmation orienté objet mis au point en 1991 par la firme SUN. Il est caractérisé par les points suivants :

- **Java est indépendant de toute plate-forme** : Une application développée en java fonctionne (sans aucune modification, même pas une recompilation) dans n'importe quel environnement disposant d'une MJV (Machine Virtuelle Java).
- **Java est un langage orienté objet** : Dans ce type de programmation, on ne manipule pas des fonctions et des procédures, mais des objets qui s'échangent des messages. Le principal avantage, outre le fait que l'on peut créer des objets de toutes natures représentant les véritables objets du problème à traiter, est chaque objet peut être mis aux point séparément.
- **Java est extensible à l'infini :** Java est écrit en java. Idéalement, toutes les catégories d'objets (appelées classes) existant en java sont définies par extension d'autres classes, en partant de la classe de base la plus générale : la classe Object. Pour étendre le langage, il suffit donc de développer de nouvelles classes. Ainsi, tous les composants écrits pour traiter un problème particulier peuvent être ajoutés au langage et utilisés pour résoudre de nouveaux problèmes comme s'agissait d'objets Standards.
- **Java est un langage à haute sécurité :** Contrairement à C++, Java a été développé dans un souci de sécurité maximale. L'idée maitresse est qu'un programme comportant des erreurs ne doit pas pouvoir être compilé. Ainsi, les erreurs ne risquent pas d'échapper au programmeur et de passer les procédures de tests. En détectant les erreurs à la source, on évite qu'elles se propagent en s'amplifiant.
- **Java est un langage compilé :** Java est un langage compilé, c'est-à-dire qu'avant d'être exécuté, il doit être traduit dans le langage de la machine sur laquelle il doit fonctionner. Cependant, contrairement à de nombreux compilateurs, java traduit le code source dans le langage d'une machine virtuelle, appelée JVM (Java Virtual Machine). Le code produit, appelé bytecode, ne peut pas être exécuté directement par le processeur de la machine. Le bytecode est ensuite confié à un interpréteur qui le lit et exécute.

3. Présentation de l’environnement de développement :

3.1. Eclipse :

Eclipse est un environnement de développement intégré Open Source extensible, universel et polyvalent, permettant de créer des projets de développement mettant en œuvre n'importe quel langage de programmation. Eclipse est principalement écrit en Java et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions. La spécificité d'Eclipse vient du fait de son architecture totalement développée autour de la notion de plug-in. Les participants à cette formation seront familiarisés au développement avec Eclipse et au débogage des applications Java et seront autonomes pour configurer le produit.

L'espace de travail permet de voir des perspectives qui assurent une vision particulière d'un projet. Chaque perspective contient des vues et des éditeurs qui permettent de travailler sur une entité.

3.2. Serveur de base de données PostgreSQL :

3.2.1. Définition :

PostgreSQL est ce que l'on appelle un SGDB (système de gestion de base de données), est un logiciel capable d'enregistrer et conserver des informations (numérique) et de les restituer à la demande d'un utilisateur. Les concurrents de PostgreSQL les plus courants sont oracle, sybase, informix, inter base etc., mais l'avantage de PostgreSQL par rapport aux autres SGDB est sa gratuité, alors que des systèmes tel qu'oracle ne sont abordable que par de très grosse société. PostgreSQL est toutefois un peu plus rudimentaire que ses compagnons qui possèdent énormément d'outils d'aide à l'utilisation ou développement. Il possède toutefois leurs caractéristiques principales et essentielles.

PostgreSQL est un serveur de base de données (DBMS) transactionnel très puissant, PostgreSQL peut stocker plus de types de données que les types traditionnels entiers, caractères, etc. L'utilisateur peut créer des types, des fonctions, utiliser l'héritage de type etc.

PostgreSQL est largement reconnu pour son comportement stable, proche de Oracle. Mais aussi pour ses possibilités de programmation étendues, directement dans le moteur de la base de données. Le traitement interne des données peut aussi être couplé à d'autres modules externes compilés dans d'autres langages.

3.2.2. Quelques fonctionnalités de PostgreSQL :

3.2.2.1. Triggers et contraintes :

Une procédure stockée (fonction) peut être appelée dans le cadre d'un trigger (aussi appelé déclencheur). Un trigger est une opération qui sera appelée lors d'un accès à une table. Un trigger peut être appelé dans le cadre d'un SELECT et/ou INSERT, et/ou UPDATE. L'appel peut être exécuté soit pour chacune des lignes manipulées, soit une fois par ordre SQL exécuté. De plus n trigger peut intervenir avant et/ou après que la manipulation des données soit réellement éjective.

PostgreSQL gère également les contraintes. Celles-ci sont déclarées lors de la création de la table. Nous avons bien sûr les contraintes classiques : NULL/NO NULL, UNIQUE, PRIMARY KEY, et REFENCES (clef étrangère) mais également CHECK qui est moins courante. Cette contrainte permet d'écrire une expression booléenne qui acceptera ou n'acceptera pas les données.

En déclarant une contrainte de clé étrangère, on peut également spécifier le comportement en cas de suppression de la donnée référencée (refuser, suppression en cascade, passage de la clé étrangère à NULL ou bien à la valeur par défaut définie pour cette colonne).

3.2.2.2. Les langages de procédure sous PostgreSQL :

De nombreux systèmes de base de données proposent des langages intégrés pour l'écriture de procédures (les procédures sont des fonctions directement stockées dans la base de données).

Oracle dispose ainsi du langage PL/SQL. PostgreSQL possède également un langage d'écriture des procédures, semblable au PL/SQL d'Oracle : PL/pgSQL.

Tout comme PL/SQL, PL/pgSQL est un langage orienté bloc supportant la déclaration de variables, les boucles, des constructions logiques et une gestion des erreurs avancée.

PostgreSQL supporte la surcharge de fonction. Il n'est pas possible avec PL/SQL de surcharger une fonction.

3.2.3. Portabilité :

PostgreSQL est le serveur de bases de données le plus standardisé des logiciels open source.

- Compatible ANSI SQL
- Possibilité d'ajouter des extensions pour SHA1, MD5, XML et autres fonctionnalités
- Outils pour générer du code SQL portable à partager avec d'autres systèmes compatibles SQL
- Fonctions pour simplifier la transition vers d'autres bases de données relationnelles respectant moins le langage SQL.

4. Implémentation et description de notre Application :

4.1. Coté de des utilisateurs :

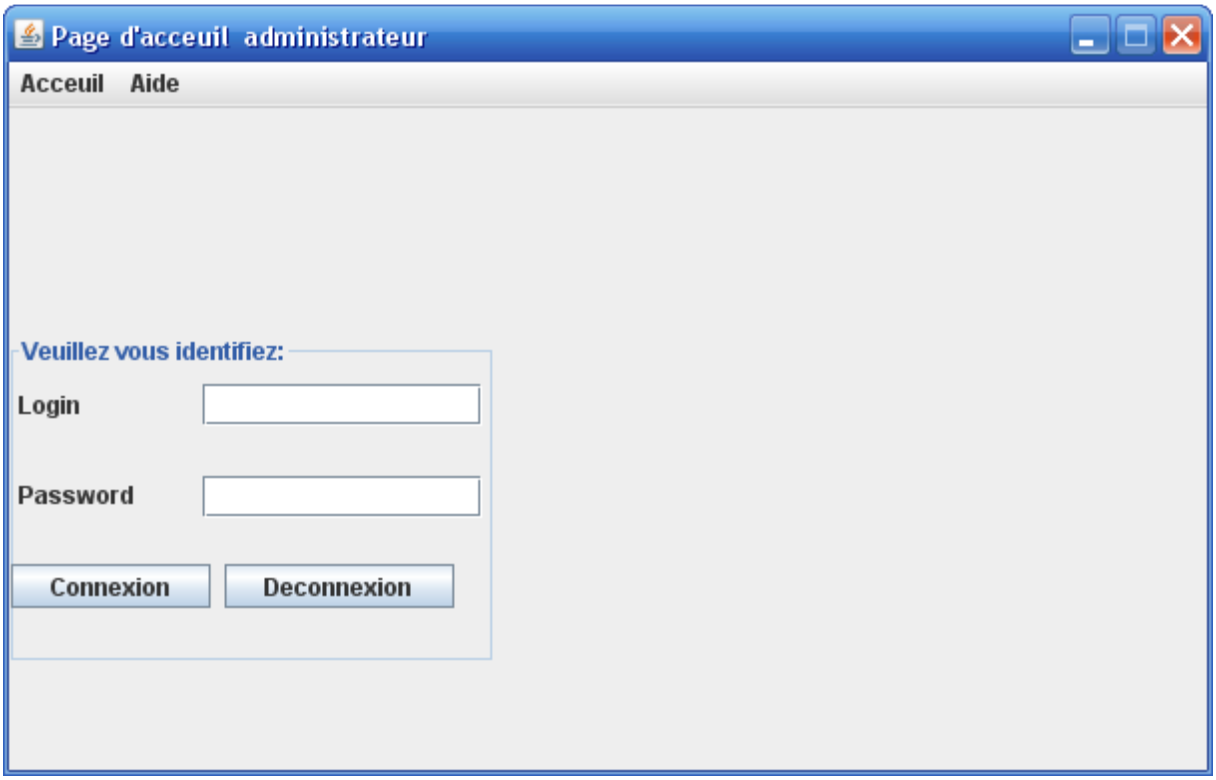


Figure 28 : Page d'accueil administratif.

Dans cette page l'administrateur entre son login es son mon de passe, via ce formulaire d'authentification pour accéder a son propre espace. Dans le cas ou le mot de passe est erroné, l'administrateur doit refaire son authentification.

4.1.1. Espace administrateur :

Lorsque l'administrateur authentifie la fenêtre suivante apparaitre :

Figure 29 : Espace administratif.

L'administrateur click sur le bouton « Gérer les profils » la fenêtre suivante apparait :

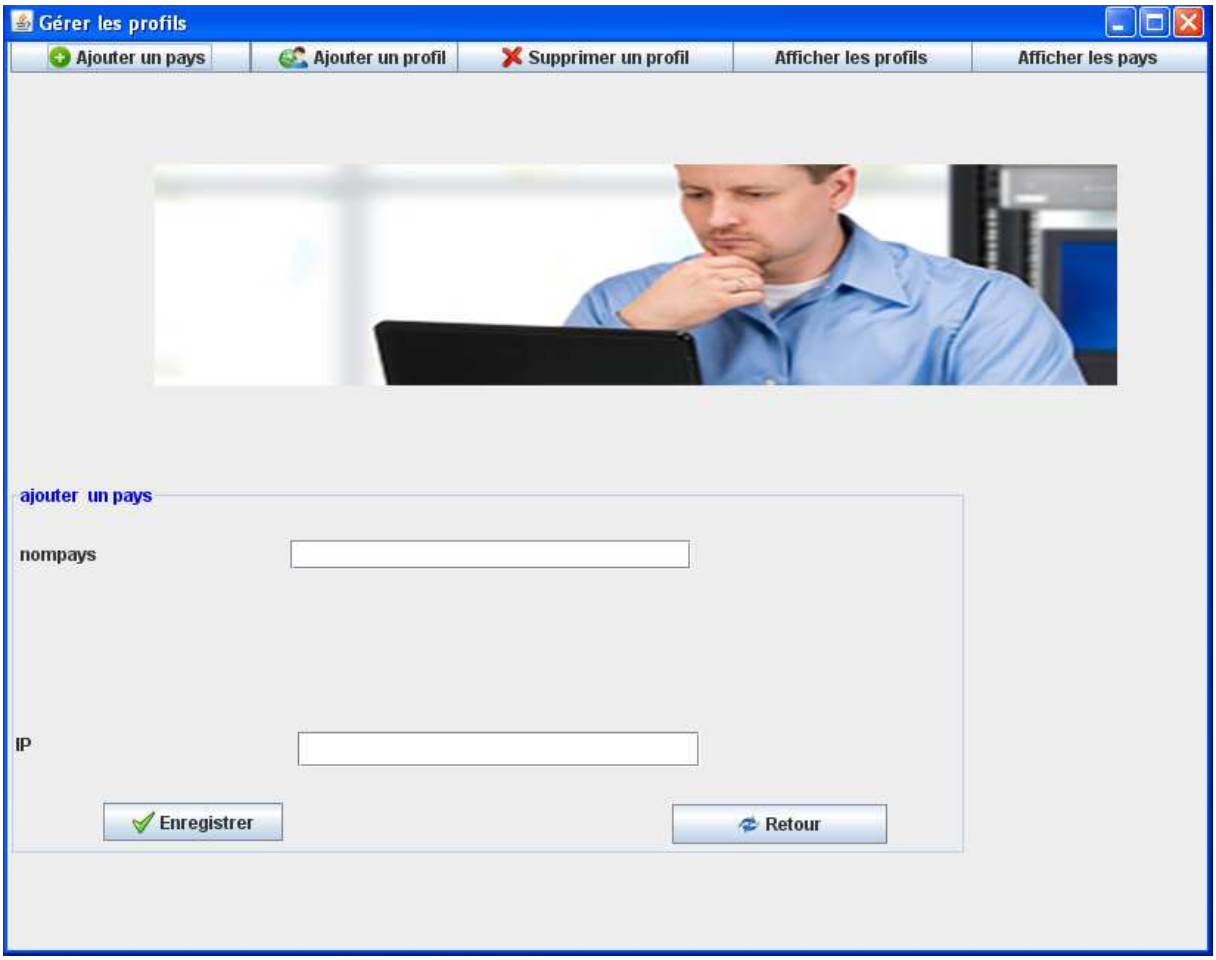


Figure 30 : Espace pour vérifier l'émetteur.

1. **Filtrage par pays :** dans ce menu l'administrateur permet :

✓ Ajouter un nom de pays.

✓ L'ajout de l'IP de pays.

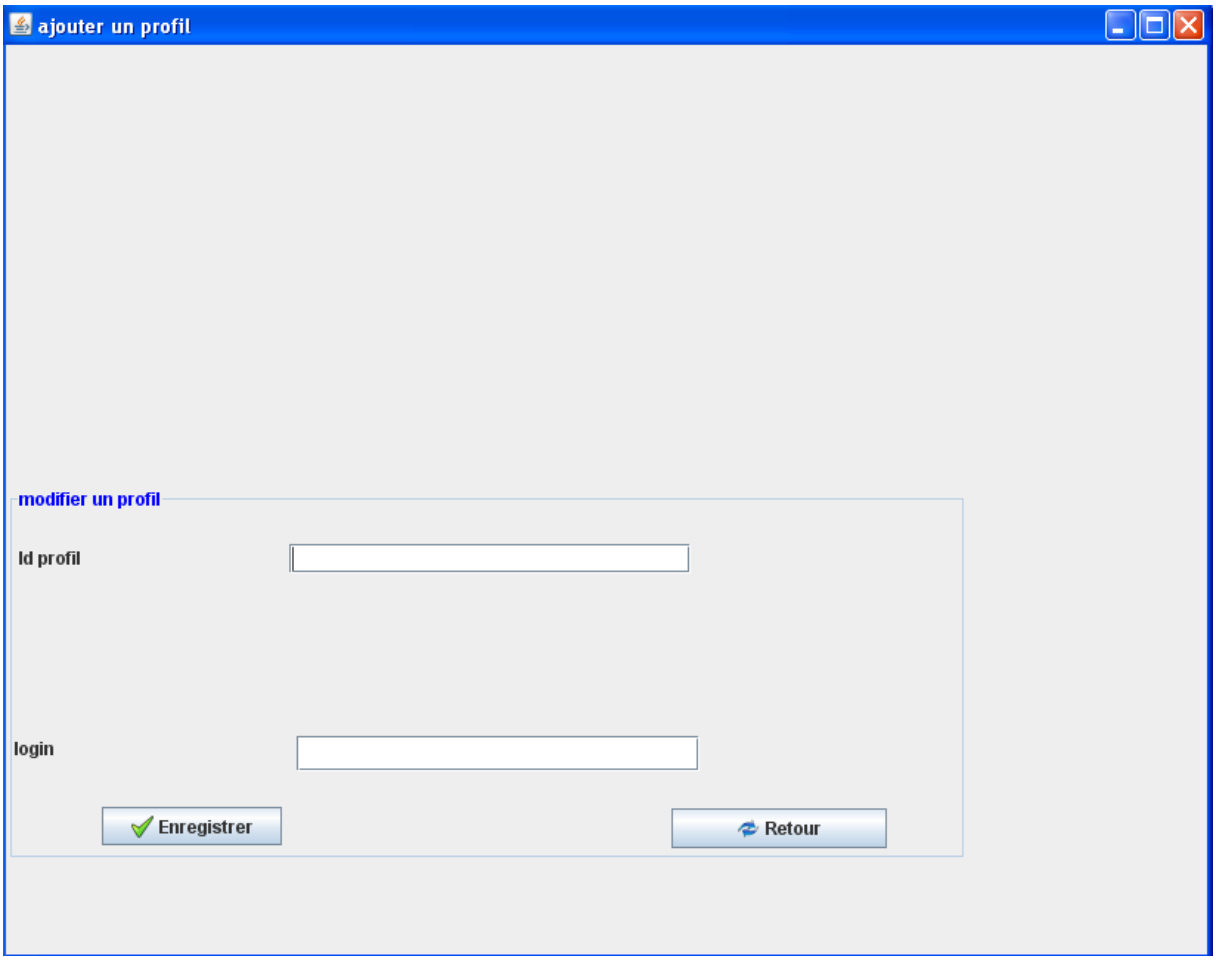


Figure 31 : Filtrage par IP.

2. **Filtrage par profil :** L'administrateur dans cette interface permet :
   - ✓ Ajout d'un Id profil.
   - ✓ Ajout de login.

Figure 32 : La suppression d'un profil.

3. **La suppression d'un profil :** L'administrateur permet aussi de :
   - ✓ Ajout d'un Id.

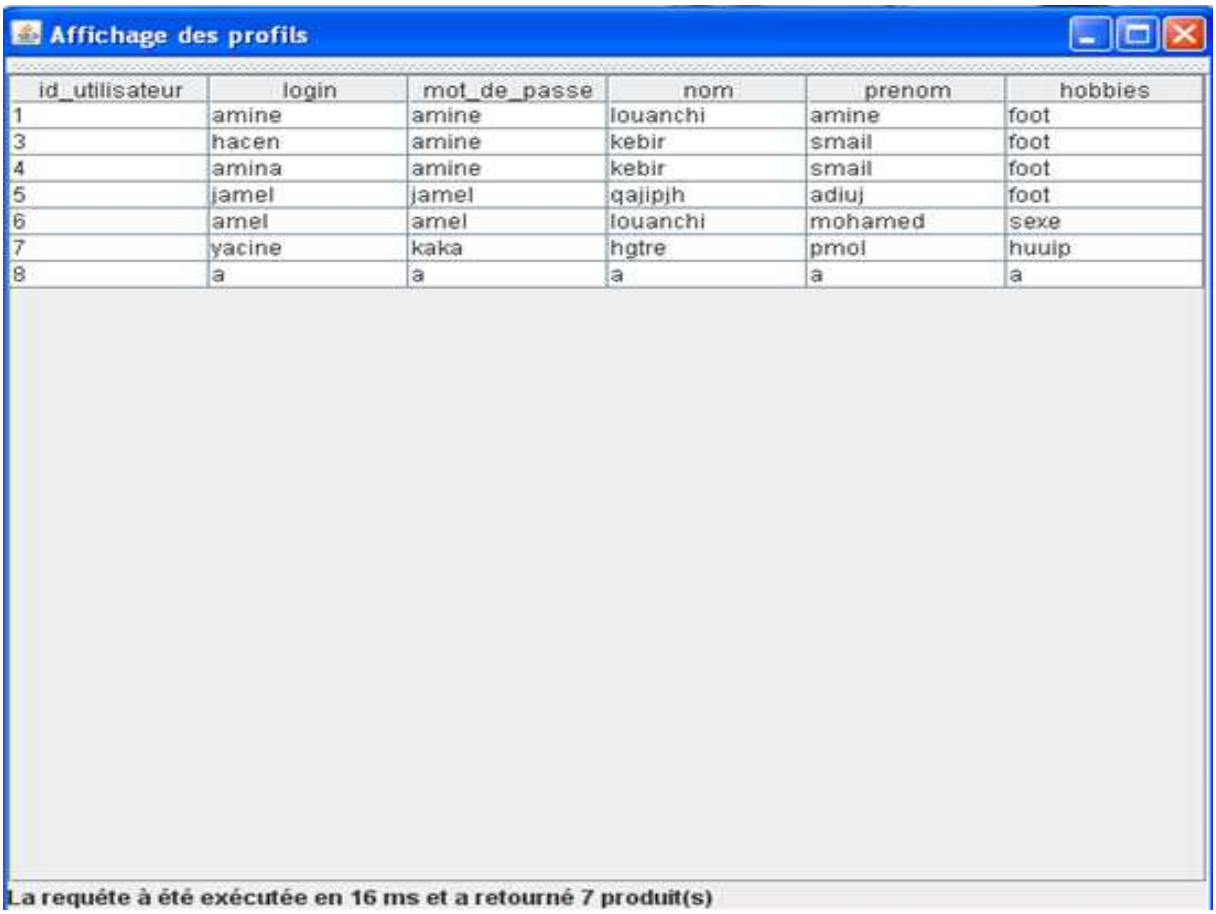


Figure 33 : L'affichage des profils.

4. **L'affichage des profils :** L'administrateur permet aussi de :

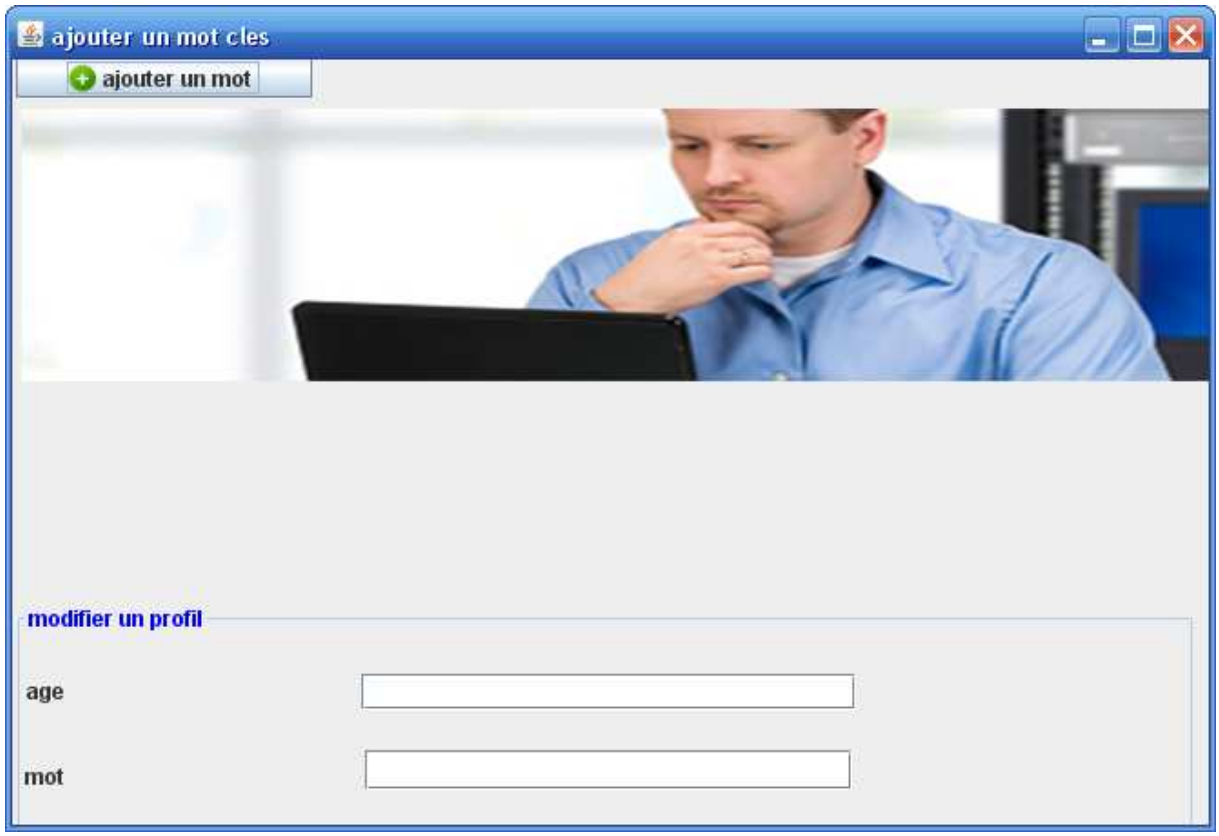


Figure 34 : Ajout d'un mot clé.

5. **Filtrage par mots-clés :** L'administrateur dans cette interface permet :
   - ✓ Ajouter un mot-clé

6. Conclusion :

Dans ce chapitre nous avons en premier lieu présenté l’environnement et les outils utilisés pour implémenter notre projet. Puis nous avons présenté des exemples réels de cas d’utilisation de notre application.

Conclusion générale :

Les systèmes de filtrage d'information ont connu une avancée significative ces dix dernières années, depuis les premiers systèmes collaboratifs classiques à ceux à base de contenu ou hybrides. Ils ont été largement investis dans divers domaines tels que le commerce électronique (livres, cinéma, musique, voyages, restauration), etc.

Les filtres actuellement disponibles pour la protection des mineurs laissent toujours à désirer en ce qui concerne leur performance générale. L'avenir est dans la mise au point des mécanismes de classification qui tiennent d'avantage compte des informations contextuelles, et dans la combinaison intelligente de différents concepts de filtrage, tels que le filtrage sémantique et filtrage en temps réel.

Le filtrage sémantique concerne le sens des mots , La demande pressante des institutions en matière de protection des usagers contre les contenus illicites ou préjudiciables sur Internet (racisme, xénophobie, pédophilie) invite à dépasser les systèmes de filtrage automatique conventionnels bases sur des listes de mots clés ou des annuaires d'adresses préétablies, peu efficaces.
Le « **real -time web** » (rtw) est un ensemble de technologies et de pratiques dans lequel l'utilisateur reçoit des informations au moment où elles sont publiées, il se distingue donc des paradigmes dans lesquels c'est l'usager ou l'une de ses applications qui va vérifier régulièrement les mises à jour d'une source basée sur le web.

Acronymes :

BGP : Border Gateway Protocol. C'est un protocole d'échange d'informations d'accessibilité de réseaux, conçu pour prendre en charge un volume important de données. Ce protocole a permis la décentralisation du routage d'Internet.

DNS Domaine Name System (système de noms de domaine). Il s'agit d'un système décentralisé permettant de la correspondance entre une adresse IP et un nom de domaine, dont l'exécution est distribuée sur plusieurs machines. Il est constitué de serveurs.

DPI Deep Packet Inspection – Inspection profonde de paquets. Technologie permettant d'analyser le contenu d'un paquet.

FAI Fournisseur d'Accès Internet. Sont regroupés dans cette catégorie les opérateurs de réseaux et les opérateurs de services.

FTP File Transfer Protocol : protocole de transfert de fichiers. Protocole de communication destiné à l'échange de fichiers sur un réseau. Permet de les copier afin d'alimenter un site Web.

Hébergeur Entité ayant pour vocation de mettre à disposition des internautes des contenus et gérés par des tiers.

ICANN Internet Corporation for Assigned Names and Numbers. Société pour l'attribution des noms de domaine et des numéros sur Internet. Depuis 1998, elle est chargée de superviser les règles d'attribution des adresses et des domaines sur Internet. Site officiel : HTTP://www.icann.org/

P2P Peer to Peer, en français, le pair à pair. Il s'agit d'un modèle de réseau informatique sur Internet. Par abus de langage le pair à pair est aussi parfois employé pour des applications réseaux (réseaux organisés en modèle pair à pair) et utilisées pour l'échange des fichiers.

PGP Pretty Good Privacy (confidentialité plutôt bonne). Logiciel gratuit de chiffrement et de signature de données développée par Philip Zimmermann**.**

Port Désigne un point d'accès à un ordinateur.

RSA Algorithme de cryptographie asymétrique permettant de chiffrer les communications par deux clefs : l'une publique, l'autre privée.

Spam Désigne des messages non sollicités qui encombrent les boîtes aux lettres électroniques.

URL Uniform Resource Locator. Adresse Internet exclusive permettant d'atteindre un site précis depuis n'importe quel endroit dans le monde.

VPN Virtual Private Network ou réseau virtuel. C'est un mécanisme qui permet de réaliser un échange d'informations avec chiffrement de la communication.

Bibliographies:

[Réf. 1] Manuscrit auteur, publié dans "Médias011 : Y a-t-il une richesse des réseaux ?, Aix en provence : France (2011)" par Camille Alloing.

[Réf. 2] Mohamed Boughanem, Lynda T. et Lamjed Ben J. « Intégration des facteurs temps et autorité sociale dans un modèle bayésien de recherche de tweets », IRIT, Université Paul Sabatier.

[Réf. 3] Cormac C., Marco G., Estelle De M. et Hein Dries-Ziekenheiner « Filtrage d'Internet. Equilibrer les réponses à la cybercriminalité dans une société démocratique ». Rapport dans le cadre d'un financement de l'Open Society Institute. Traduction française : Estelle De Marco le 11 mai 2010.

[Réf. 4] Centre d'Expertise Gouvernemental de Réponse et de Traitement des Attaques informatiques « Filtrage et pare-feux ».

[Réf. 5] COMITÉ D'EXPERTS SUR LES NOUVEAUX MÉDIAS « Projet de recommandation CM/Rec(2012)…du Comité des Ministres aux Etats membres sur la protection des droits de l'homme dans le cadre des services de réseaux sociaux ».

[Réf. 6] Laurent COLLÉE « Sécurité et vie privée sur les réseaux sociaux » Université de LUXEMBOURG.

[Réf. 7] Patrick DESSALLE « Conception et réalisation d'une plateforme de deploiement de services géolocalisés » Université Libre de Bruxelles 2006.

[Réf. 8] Firas Damak , Karen Pinel-Sauvagnat et Guillaume Cabanac « Recherche de microblogs : quels critères pour raffine résultats des moteurs usuels de RI? » Université de Toulouse – IRIT UMR 5505 CNRS.

[Réf. 9] Françoise Papa, Université de Grenoble, France

[Réf. 10] Guillaume Valadon et Yves-Alexis Perez, « Architecture DNS sécurisée ». ANSSI 51bd. De La Tour-Maubourg 75700 Paris Cedex 07 France

[Réf. 11] Lab Réseaux et Techniques, « Livre vert sur les techniques de filtrage ». V2, le Mercredi 14 décembre 2011.

[Réf. 12] Houda OUFAIDA et Omar NOUALI « Le filtrage collaboratif et le web 2.0. Etat de l'art », Lavoisier | Document numérique 2008/1 - Volume 11, ISSN 1279-5127 | ISBN 2-7462-2318-9 | pages 13 à 35

[Réf. 13] site de zero « Analyser le réseau et filtrer le trafic avec un firewall »

[Réf. 14] samir YAHIAOUI « Structuration des métadonnées en vue d'un filtrage D'information sur le web » ESI. 2008

[Réf. 15] Benjamin Rosoor(*), Laurent S.(*), Sandra B.(**,***), Pascal P.(***) et Mathieu R.(***). « Quand un Tweet détecte un catastrophe naturelle ». (*) Web Report, France (**) Dépt. MIAP, Université Montpellier 3, France (***) LIRMM, CNRS, Université Montpellier 2, France

[Réf. 16] Leila Kosseim, Guy Lapalme « Une expérience en extraction d'information bilingue ». T.A.L., vol. n°1, pp. 1-22 Laboratoire RLI, Université Montréal.

[Réf. 17] Amine BENHAMZA « Identification des leaders d'opinion en politique, Dans l'écosystème Twitter ». NFE204 – CNAM 2010.

[Réf. 18] Ibra Seye « Détection d'anomalies dans les configurations de firewalls De dernière génération ». TELECOM Bretagne.

[Réf. 19] **«** Les médias Sociaux » Travaux mené par les membres de l'IAB France. Publication Novembre 2010.

[Réf. 20] Charlotte BOGSZ « Le régime juridique applicable aux réseaux sociaux » Septembre 2009 -Panthéon - Sorbonne- Université Paris 1.

[Réf. 21] Sharevb « Sécurité GNU/Linux. Iptable, principe de base ».

[Réf. 22] B. Starynkévitch(*), M. Daoudi(**) «Architecture du système POESIA de filtrage de contenu Internet ». 2003
(*) Commissariat à l'énergie Atomique – DRT/LIST/DTSI/SLA
(**) MIIRE Telecom Lille 1.

[Réf. 23] Romain Vinot, Natalia Grabar, Mathieu Valette «Application d'algorithmes de classification automatique pour la détection des contenus racistes sur l'internet ». Juin 2003
Université Paris 6.

[Réf. 24] Michel PLAISENT, Nabil BENKIRANE « La surveillance des employés Branchés : timidité ou éthique ? » Revue Francophone de de @management N° 10 – février 2004.
Université du Québec à Monttréal

[Réf. 25] Josef Gabay et David Gabay « Analyse et conception » ebooks-land.net Paris 2008.

Site internet:

www.twitter.com

www.guide-twitter.com -V2

www.wikipédia.com

www.siteduzero.com

http://www.bulletins-electroniques.com/actualites/66759.htm
http://www.controle-parental.net

http://wwww.controleparent.com/10/comment-utiliser-un-logiciel-de-controle-parental.html

http://themetricsystem.rjmetrics.com

http://delegation.internet.gouv.fr/mineurs/enquete.htm

http://www.poesia-filter.org

www.lesiteduzero.com

www.javafr.com