

Thème 2 – Le Web

Ouverture, page 41

194 milliards

C'est le nombre d'applications qui ont été téléchargées en 2018 dans le monde, soit 6 000 par seconde. En valeur, cela représente plus de 101 milliards de dollars dépensés sur les différents stores d'applications.

Source : Mobilemarketing.fr

Document c : L'araignée tisse sa Toile

Le Web (ou Toile) est un ensemble de ressources numériques reliées et accessibles par des liens cliquables, sur internet. À partir des années 2000, le Web « social » avec l'apparition de sites sociaux comme Facebook, Twitter, YouTube ou Flickr, permet le partage en tout genre.

Aujourd'hui le Web continue de se développer en devenant immersif et en interagissant avec le réel. Ainsi, les accès au Web se diversifient et se joignent à tout type de terminal (lecteur mp3, tablette, smartphone, montre connectée, etc.). Ces données collectées alimentent des bases de données dont l'exploitation et le traitement présentent un enjeu économique majeur.

Le Web démocratise les voix de l'information en se calquant sur le phénomène de mondialisation. Il gomme les frontières, mais reste tout de même façonné selon les lois et particularités des pays. Ceux qui créent le Web le font à leur image.

Unité 1 – Le fonctionnement du web

Document a page 42

Les liens hypertextes

Les pages Web sont constituées de diverses informations (textes, images, vidéos, animations, etc.) et de **liens hypertexte** qui permettent de naviguer de l'une à l'autre lorsqu'on clique dessus. Ces liens supportent soit des images, soit des morceaux de texte cliquables que les concepteurs de pages Web signalent, par exemple, en les soulignant. Au survol de la souris, le curseur se transforme en une petite main. Au clic sur le texte ou l'image, une nouvelle page se charge dans le **navigateur**.

Document b page 42

Le protocole de communication HTTP

Pour s'afficher dans le navigateur, un document doit avoir fait l'objet d'une demande de la part de l'utilisateur (une **requête**) : demande auprès du serveur qui héberge ce document de le lui envoyer. C'est cette relation client-serveur que désigne le **protocole http** (*HyperText Transfert Protocol*, « protocole de transfert hypertexte »). Pour que cela soit possible, il faut que le document en question puisse être localisé sur le Word Wide Web par une adresse unique. Celle-ci indique son emplacement sur un ordinateur serveur, adresse à partir de laquelle le navigateur va pouvoir envoyer sa requête. On appelle cette adresse une **URL** ou « *Uniform Resource Locator* », expression anglaise pouvant se traduire par « localisateur universel de ressources ».

Point info page 42

À sa création, le Web était destiné aux chercheurs. Le but était de faciliter l'échange d'informations entre les membres de la communauté scientifique en utilisant internet.

Document c page 43

Les extensions du domaine

Les **extensions de domaine ou noms de domaine** de premier niveau se divisent en deux catégories : celles associées à un pays ou un territoire (« .fr » ou « .uk », par exemple) et les celles associées à des activités (« .edu » pour l'éducation, « .com » pour le commerce, par exemple).

Document d page 43

Les noms de domaine

Sur le Web, un utilisateur ou une entreprise est libre de choisir le nom de domaine associé aux pages Web qu'il crée, sous réserve qu'il soit disponible. Ce choix est stratégique sur le plan de la communication car il est l'une des portes d'entrée de l'identité numérique.

Les noms de domaines ne peuvent pas être achetés à vie, mais loués sur une durée déterminée. Pour en louer un, les utilisateurs doivent se rapprocher de distributeurs agréés par l'**ICANN** (*Internet Corporation for Assigned Names and Numbers*), l'autorité de régulation d'internet.

Pour qu'une adresse soit accessible via son nom de domaine, ce dernier doit être enregistré dans deux DNS (Domain Name System) au minimum. Le DNS fait le lien entre le Web et internet.

Unité 2 – Langages d'une page Web

Document a page 44

Constituer une page Web

Depuis sa création en 1990, le langage informatique HTML (*HyperText Markup Language*) est utilisé pour créer des pages Web. C'est l'une des trois inventions issues du W3C avec le protocole HTTP et les adresses Web. Son acronyme signifie « langage de balisage d'hypertexte » car il permet en effet de réaliser de l'hypertexte à base d'une structure de balisage.

Le CSS – *Cascading Style Sheets* ou « feuille de styles en cascade » – s'est imposé dans les années 2000. Comme son nom l'indique, il permet de styler les textes d'une page Web : définir leur taille, leur couleur ou l'alignement d'un paragraphe.

L'utilisation de feuilles de style au format CSS repose sur l'idée de séparation du contenu et de la mise en forme.

Document c page 44

Code source d'une page Web

Pour accéder au code source d'une page Web, il faut commencer par effectuer un clic droit sur la page. Dans le menu contextuel, il faut ensuite sélectionner la commande « Examiner l'élément » avec Firefox, « Inspecter » avec Chrome ou « Inspecter l'élément » avec Edge, par exemple.

Point info page 45

Les deux langages JavaScript et PHP permettent d'exécuter des actions dans les pages Web demandées par l'internaute et les rendent dynamiques. Ces langages ont des bases différentes : JavaScript est un langage exécuté côté client (sauf Node.js) tandis que PHP est un langage exécuté côté serveur.

Document e page 45

Un Web unique et uniformisé

Le **W3C** ou World Wide Consortium est un organisme international à but non lucratif dont la mission première est de mettre en place et de certifier les standards du Web au fur et à mesure de l'évolution technologique (HTML, CSS, XML, JavaScript...), afin d'en uniformiser le contenu. Il met à disposition des outils de validation du code des pages HTML, des feuilles de styles et la validité des liens, ce qui permet d'optimiser le référencement des sites Web. Grâce au respect des règles qu'il impose, il permet l'affichage de manière uniforme d'un site Web sur n'importe quel support à jour.

Unité 3 – Mécanisme des échanges sur le Web

Vocabulaire page 46

Bande passante : débit binaire ou quantité d'informations pouvant être transmises simultanément sur une voie de transmission.

Document b page 46

Sécurité dans le protocole HTTP

L'*HyperText Transfer Protocol Secure* ou **HTTPS** est la version sécurisée de la connexion HTTP. Elle protège les données échangées en ne les rendant compréhensibles que par le destinataire, grâce au protocole TLS (Transport Layer Security) qui chiffre les données. Pour s'assurer que le serveur destinataire est bien celui qu'il prétend, le HTTPS authentifie la cible grâce aux certificats appelés communément « Certificats SSL (Secure Socket Layer) ».

Parole d'expert page 46

HTTP est un protocole de la « couche application » d'internet dont les données transitent via le protocole TCP. Il permet de récupérer des ressources telles que des documents HTML.

Le Web fonctionne sur le principe d'une architecture de type client-serveur/requête-réponse. Les clients et serveurs communiquent par l'échange de messages individuels. Ce type de structure permet de centraliser les ressources sur un même lieu et d'assurer leur accessibilité : l'administration, la sécurité, les pannes et l'évolution sont gérées au niveau du serveur. Si les liens pointant vers la ressource sont inactifs, il indique le message « *404 not found* ».

Le serveur est le maillon faible de la structure : en cas de panne ou de « piratage », les ressources ne sont plus disponibles. De plus, les coûts d'exploitation sont élevés : nécessité d'ordinateurs puissants et besoin élevé en bande passante.

Document c page 47

Intégration d'une ressource externe dans une page Web

Le code *iFrame* ou inline Frame est un document HTML qui sert à intégrer d'autres pages Web (source externe) dans une page Web. Des sites comme Youtube proposent des codes iFrames pour afficher directement les vidéos sur des sites Web tiers. Des régies publicitaires comme Google AdSense les proposent pour afficher des bannières publicitaires sur d'autres sites.

Document d page 47

Fonctionnement de l'iFrame

Une page Web complète pourra être construite à partir de différents sous-documents qui sont récupérés sur différents serveurs. Par exemple : du texte, des descriptions de mise en page, des images, des vidéos, des scripts et bien plus.

Unité 4 – Les moteurs de recherche

Document a page 48

Recherche des ressources

Des « robots » ou « crawlers » (programmes informatiques) parcourent le Web et indexent les pages à partir des mots qu'elles contiennent et en lien avec leur adresse URL. Par exemple, l'index peut indiquer que le mot « Louvre » est utilisé sur les pages 10, 27 et 157. Cela permet de gagner du temps de réponse face à la requête du visiteur. Toutes les pages ne sont pas sauvegardées. Certaines pages provenant de sites illégaux sont tout simplement blacklistées.

Point info page 48

« Je ne suis pas un robot ! » Les captcha servent à différencier les robots et les humains dans le cas où les pages se feraient espionner par un logiciel malveillant. Ils servent aussi de « machine training » pour entraîner les algorithmes à reconnaître les images.

Document b page 48

Indexation des ressources

Une fois les données collectées par les robots, un algorithme va les classer en fonction de plusieurs critères, comme le nombre de liens pointant vers une page. Le principe de fonctionnement est fondé sur le fait que plus un site est cité par d'autres sites, plus il sera considéré comme pertinent et donc plus son score sera élevé. Un bon score garantit une place de choix au site dans la page des résultats, c'est le **référencement naturel**.

Document d page 49

Le Web profond ou Web invisible

Toutes les pages Web, bien qu'accessibles avec un navigateur internet, ne sont pas référencées par les moteurs de recherche. Ces pages ont des particularités : elles sont dynamiques, protégées par un mot de passe et contiennent des ressources volumineuses, entre autres. Ces ressources non indexées par les moteurs de recherche composent le **Web profond** ou le « **Deep Web** » (96 % du Web). Le Dark Web en représente la partie illégale. Les ressources indexées, quant à elles, composent le **Web de surface**.

Point info page 49

Le moteur de recherche Google traite 1 milliard de requêtes chaque jour ! À l'heure actuelle, Google indexe près de 10^{12} pages, contre 10^9 en 2000. On est encore loin du 10^{100} : le fameux nombre « gogol » qui a inspiré son nom à la société. Google existe en 80 langues, y compris en Klingon, une langue qui existe uniquement dans la série télévisée *Star Trek* !

Unité 5 – Risques et sécurité sur le Web

Document a page 50

Historique d'un navigateur

Lorsque vous surfez sur le Web, les navigateurs enregistrent chronologiquement votre activité : sites visités, cookies, préférences, etc. Toutes ces informations, sauvegardées dans un fichier, composent l'historique. L'intérêt premier est de simplifier la navigation de l'utilisateur : retrouver un site qui a été consulté, garder en mémoire les mots de passe de sites visités fréquemment, pré-remplir les formulaires avec ses coordonnées personnelles, etc. Ces données restent cependant accessibles aux sites connectés.

Document b page 50

Cookies publicitaires

Lorsqu'un utilisateur accède à certains sites Web, un témoin de connexion, appelé **cookie**, est stocké dans le navigateur de l'utilisateur. Les cookies fournissent, au service qui les dépose, des données concernant la navigation de l'utilisateur (sites visités, requêtes dans les moteurs de recherche, géolocalisation, adresse IP, etc.). Ils facilitent ainsi la navigation et/ou apportent des fonctionnalités supplémentaires comme la gestion de l'identification. Ils peuvent être utilisés pour présenter à l'utilisateur sur un site ou en dehors de celui-ci des publicités ciblées sur ses centres d'intérêt.

Source : blog.octave.biz

Parole d'expert page 50

La navigation privée ne rend pas invisible. Tant que le navigateur est sollicité pour visiter des sites, les cookies continuent d'être déposés et donc potentiellement de transmettre en temps réel des informations personnelles. Il en est de même pour votre historique de navigation ou encore les mots de passe enregistrés qui ne s'effacent qu'une fois que votre navigateur aura été fermé. De plus, le fournisseur d'accès internet a l'obligation de garder les fichiers de connexion de votre navigation pendant une durée d'un an.

Programmes malveillants, page 51

Il existe plusieurs types de logiciels malveillants. Le **ransomware** ou rançongiciel prend en otage les données personnelles de l'utilisateur en les chiffrant et réclame de l'argent pour obtenir la clé de déchiffrement. Le spyware ou logiciel espion investit votre ordinateur incognito pour divulguer des informations sur son contenu.

Le **phishing** ou hameçonnage prend généralement la forme d'un mail copiant l'interface d'un service officiel, pour collecter des données permettant d'usurper l'identité de l'utilisateur grâce aux informations qu'il aura renseignées.

Document d page 51

Extensions malveillantes sous un navigateur

Un **plug-in** est une extension de navigateur qui ajoute certaines capacités et fonctionnalités. En l'installant, le développeur peut modifier l'interface utilisateur, ajouter des fonctionnalités de service Web, comme bloquer des publicités ou traduire des pages. Tous les navigateurs proposent des centaines de ces mises à jour pour améliorer la productivité, la personnalisation, les achats, les jeux, etc.

Le site AdGuard a repéré quatre fausses extensions de bloqueur de publicité. En utilisant des scripts cachés, ils peuvent récolter les données des utilisateurs durant leur navigation sur le site.

D'après Blogdumoderateur.com

Document e page 51

Le déni de service

Le 28 février 2018, la plateforme GitHub a subi une attaque par déni de service distribuée (DDOS) d'une rare violence. Après cela, il lui était impossible de traiter les demandes reçues ; elle a donc dû se mettre hors service.

Unité 6 – Enjeux éthiques et sociétaux du Web

Document a page 52

Accès généralisé à la publication et à la diffusion

Le Web 2.0, ou **Web social** permet aux internautes de publier et d'interagir sur le Web avec d'autres internautes. Pour cela ils n'ont pas besoin de connaissances techniques, mais ils doivent s'inscrire et se connecter. Chacun peut à sa manière publier ce que bon lui semble pour le meilleur comme pour le pire.

Wikipédia est une encyclopédie collaborative très populaire. En janvier 2016, les 17 000 contributeurs francophones ont effectué environ 760 000 modifications sur Wikipédia. Mais souvent cible de trolleurs qui disséminent des erreurs. De nombreux employés traquent ces dissonances qui peuvent parfois avoir des conséquences néfastes sur l'image des personnes concernées par ces fausses affirmations.

Vocabulaire page 52

« Troll » : personne qui participe à un débat dans le but de nourrir artificiellement une polémique et de perturber l'équilibre de la communauté concernée.

Document b page 52

Législation encadrant les publications sur le Web

La loi sur la liberté de la presse du 29 juillet 1881 indique les limites de la liberté d'expression. La loi de confiance dans l'économie numérique de 2002 oblige les sites à présenter des mentions légales. Les atteintes à la vie privée et au droit d'auteur sont aussi condamnables sur le Web.

Document c page 52

Personnalisation de l'expérience utilisateur

Le Web 3.0 ou Web sémantique est lié à l'essor des objets connectés et au développement de l'intelligence artificielle. Il permet une adaptation aux modes de pensée des utilisateurs et à l'intégration des objets connectés dans leur environnement. C'est l'avenir du Web.

Document d page 53

Comment retenir l'attention des internautes

L'offre de services numériques, médiatiques, télévisuels est devenue surabondante. Se démarquer devient plus difficile, c'est pourquoi on se focalise désormais sur la captation d'une ressource qui tend à devenir rare : l'attention des internautes. Le moteur Google présente plusieurs widgets pour attirer l'attention de l'utilisateur, comme les *Google doodles*.

Vocabulaire page 53

L'**infinite scroll** est une fonctionnalité qui donne un aspect infini à la page. Des articles s'ajoutent à chaque fois que l'on descend sur la page, afin d'inciter l'internaute à chercher la suite.

Les **notifications « push »** sont des alertes sur les « smartphones » qui informent de l'arrivée d'une nouveauté.

Le **binge-watching** est le fait de regarder une série télévisée d'une seule traite. Comportement accentué par les services de streaming et de vidéo à la demande.

Document e page 53

2019, les 30 ans de l'invention du Web

Tim Berners-Lee rappelle ses idéaux de l'époque : « Il y a 30 ans, j'avais imaginé le Web comme une plateforme ouverte qui permettrait à quiconque, partout, de partager des informations, d'accéder à des opportunités et de collaborer par-delà les frontières géographiques et culturelles ». Bien que le Web ait répondu à cette vision, trois tendances actuelles sont sources d'inquiétudes croissantes pour lui : la perte de contrôle sur nos données personnelles en échange d'accès gratuit à des services, la facilité de dissémination de la désinformation par des personnes malintentionnées et la transparence de la publicité politique en ligne pour influencer un électorat.

Au-delà de ce constat, il propose des pistes de développement pour permettre aux utilisateurs de maîtriser l'utilisation de leurs données. Il envisage d'ailleurs la création d'un « contrat Web » ambitionnant d'établir les principes d'une gouvernance en ligne pour assurer la véracité des informations sur le Web.

Le Mag' des SNT, pages 54-55

Grand angle

Le « darknet » ou plutôt les « darknets » (internet clandestin) sont un ensemble de réseaux anonymes et chiffrés utilisant des protocoles de communication spécifiques. Ces réseaux sont souvent confondus avec le « Deep Web » (toile profonde) qui est la partie du Web non accessible via les moteurs de recherche. Les darknets permettent aux internautes de naviguer sur le Web sans laisser de trace.

Pour y accéder, il est nécessaire d'utiliser un logiciel spécifique, comme The Onion Router ou TOR pour le plus connu. Une fois installé, ce type de navigateur permet deux types de navigation : anonyme sur le Web traditionnel et l'accès au « darkweb », Web des réseaux utilisant des protocoles spécifiques (sites en .onion par exemple).

Les caractéristiques techniques des « darknets » (anonymat et chiffrement) font d'eux le lieu idéal des activités illégales. Ils permettent aussi, pour les opposants politiques, de s'exprimer publiquement et de contourner la censure, et pour les journalistes d'enquêter et de communiquer avec leurs sources en toute discrétion. Ils permettent enfin, à tout internaute, de faire le choix de ne pas laisser de trace sur internet.

En conclusion, les « darknets » ne sont pas bons ou mauvais par nature, ce sont seulement des dispositifs techniques. C'est l'usage qu'en font les internautes qui leur confère leur aspect légitime ou illicite.

Lire ! *Millenium*

Dans cette saga policière, Lisbet Salander est une hackeuse qui pénètre par effraction dans les systèmes et les réseaux informatiques. Sous le pseudonyme de « Wasp » (guêpe), elle est une figure légendaire dans la communauté internationale des hackers. Elle utilise ses talents pour gagner sa vie et aider le journaliste Mikael Blomkvist dans ses enquêtes. Agissant librement sur le Dark Web, Lisbet ne pourra pas se cacher longtemps. **Jouer avec le Web, c'est exposer sa vie réelle.**

Et demain ?

Garantir la pérennité du Web, c'est le développer tout en encadrant ses revers. Le concept de « Privacy by Design » est de garantir l'intégration de la protection de la vie privée dès la conception d'une nouvelle application, produit ou service.

Concrètement, il oblige les entreprises et autres responsables du traitement des données personnelles à offrir à leurs utilisateurs ou clients le plus haut niveau possible de protection des données et de garantir, par défaut, le respect de la vie privée. Prenons le cas des boutons de partage social présents sur la plupart des pages Web. Certains de ces boutons contiennent du code qui enregistre les visiteurs de la page et envoie les données au réseau social, même s'ils ne partagent pas la page en question. Si l'on veut respecter le principe du « Privacy by Design », les boutons de partage doivent, par défaut, être inactifs ; ils ne sont activés que si le visiteur les utilise pour partager la page.

À l'heure du Big Data, l'application de ce concept semble incompatible avec la collecte massive des données. Une étude luxembourgo-canadienne démontre que la mise en œuvre des trois principes suivants permettrait de garantir la protection des

données personnelles sans freiner le développement du Big Data : la minimisation des données, la dépersonnalisation des données et l'utilisation de contrôles d'accès.

Métier : développeuse

« Mon client voulait améliorer son application de vente en ligne pour y intégrer de nouvelles fonctionnalités. Nous avons établi ensemble un cahier des charges précis. Ensuite, avec le graphiste qui s'est occupé du design et de l'ergonomie, j'ai réalisé les programmes informatiques et j'ai défini les algorithmes de traitement des données répondant au besoin du client.

Ce que j'aime dans ce métier, ce sont ses côtés créatif et relationnel. Chaque client a un besoin différent que je dois comprendre et analyser pour y répondre. Le choix des solutions techniques adaptées à la future application est également un enjeu majeur qui m'oblige constamment à étudier les nouveautés en matière de codage. Lors de mon recrutement, j'avais mis en avant mon cursus scolaire orienté maths/informatique et mes qualités de rigueur et de bonne communicante. »

En bref

1. Le Web a 30 ans

Né le 12 mars 1989 et inventé par Tim Berners-Lee, le Web est aujourd'hui composé d'1,4 milliard de sites et compte 4,12 milliards d'internautes (dont 53 % se connectant via leur mobile). En France, il y a 53 millions d'internautes, dont 34 millions de mobinautes.

2. Moteurs alternatifs

Qwant, DuckDuckGo, Ecosia, Lilo... Ces moteurs de recherche sont des alternatives à Google et ont pour spécificités de respecter la vie privée des internautes et/ou de s'inscrire dans des démarches écologiques et solidaires.

3. Lolcat

Un lolcat est l'association d'une image (ou d'une vidéo) de chat et d'une phrase (ou d'un mot) à visée humoristique. Le lolcat « Charlie Schmidt's Keyboard Cat! » a été vu plus de 54 millions de fois sur Youtube !