

Explorer l'espace logique*

Michel Volle†

30 septembre 2006

« Le Web est un espace construit autour de langages et de protocoles obéissant à des spécifications formelles. Les pages et les liens ayant cependant été créés par des êtres humains, leurs interactions forment sur le Web des structures macroscopiques¹. » (Tim Berners-Lee, Wendy Hall, James Hendler, Nigel Shadbolt, Daniel J. Weitzner, « Creating a Science of the Web », *Science*, 11 août 2006 [18]).

« Associez l'Internet, les satellites, les fibres optiques, la progression par bonds de la puissance de calcul (grâce à des circuits pas plus grands que quelques atomes), le progrès qualitatif du raccordement à large bande (qui transmet toujours plus de données vers les logements et les bureaux via des réseaux à fibre optique et des constellations de satellites), la carte du génome humain et des outils pour sélectionner puis assembler des gènes et même des molécules – et cela donne un magasin immense, global, en temps réel, offrant un choix et des possibilités presque infinis². » (Robert Reich, *The Future of Success, Working and Living in the New Economy*, Knopf 2000 [10]).

Résumé

L'espace géographique est pour chacun l'objet d'une expérience familière. La téléphonie a cependant procuré l'ubiquité au signal vocal, l'Internet la procure à la ressource informatique (textes, images, sons et données).

L'arrivée des téléphones mobiles, puis des terminaux mobiles, a rendu l'ubiquité absolue: le corps de l'utilisateur étant équipé, celui-ci n'a plus besoin pour accéder au service de se trouver dans un local équipé d'un terminal fixe.

L'ubiquité du service ouvre un espace nouveau, celui des signaux et des ressources informatiques dotées d'ubiquité; nous l'appellerons « espace logique ». Cet espace comporte des limites et des distances, mais elles diffèrent de celles qui existent dans l'espace géographique.

On peut doter l'espace logique d'outils qui permettent à l'utilisateur de s'y repérer. Ces outils s'articulent selon une architecture en couches; ils s'appuient sur la correspondance

* Source : <http://www.volle.com/travaux/espacelogique.htm>

† ©Michel Volle 2006, *GNU Free Documentation License*, <http://www.fsf.org/copyleft/fdl.txt>

1. « The Web is an engineered space created through formally specified languages and protocols. However, because humans are the creators of Web pages and links between them, their interactions form emergent patterns in the Web at a macroscopic scale. »

2. « Combine the Internet, wireless satellites, and fiber optics, great leaps in computing power (through circuits not wider than a few atoms), a quantum expansion of broadband connection (transmitting more and faster digital data into homes and offices through networks of fiber-optic cables and constellations of satellites), a map of the human genome and tools to select and combine genes and even molecules – and you have a giant, real-time, global bazaar of almost infinite choice and possibility. »

qu'un traitement statistique fait apparaître entre segments d'utilisateurs et segments de ressources. Enfin, la rencontre de l'espace logique et de l'espace géographique est l'occasion d'une offre de services spécifique.

1 Faut-il dire *mobilité* ou *ubiquité* ?

Depuis 1876 il n'est plus indispensable, pour converser avec une personne, de se trouver à portée de voix : le téléphone a conféré l'*ubiquité* au signal vocal. Lorsque seuls des téléphones fixes étaient disponibles, il fallait cependant pour pouvoir téléphoner se trouver près d'un poste téléphonique (dans l'appartement, au bureau, dans une cabine). À l'ubiquité du signal, le téléphone mobile a ajouté l'*ubiquité de l'accès* : *il équipe le corps de l'utilisateur* et non plus un local. L'ubiquité de la téléphonie est ainsi devenue totale, aux zones blanches du réseau près.

Depuis 1991 l'utilisateur du Web accède à des serveurs dont la localisation lui importe peu : que le serveur soit proche ou à l'autre bout du monde, le délai d'affichage est analogue. Le Web confère ainsi l'ubiquité aux *ressources informatiques*³ [1], celles-ci étant également accessibles à partir de tout ordinateur connecté au réseau. Lorsque l'ordinateur sera devenu lui aussi mobile, lorsqu'il équipera le corps de l'utilisateur⁴ l'informatique bénéficiera elle aussi de l'*ubiquité totale*.

Pour décrire cette évolution on utilise par analogie avec la téléphonie mobile le terme de *mobilité*. Mieux vaut dire *ubiquité* : l'utilisateur, qu'il soit mobile ou non, se trouve en effet plongé dans un *espace logique* où la distance géographique n'existe pas. Dans cette expression l'épithète « logique » ne renvoie pas au raisonnement mais au *langage* (*λογός* signifie à la fois *parole* et *raison*). Quant au mot « espace », il prend son sens car on peut, parmi les ressources, définir des limites et une « distance » (non géographique).

Les routes, les rues des villes, les portes des immeubles et des appartements délimitent, dans l'espace géographique, des voies d'accès licites. Il en est de même dans l'espace logique où les habilitations délimitent les ressources auxquelles un utilisateur particulier a accès et jouent donc le même rôle que les murs, portes et serrures de l'espace géographique⁵. Entre un utilisateur et une ressource on peut en outre définir une distance : celle de l'*intelligibilité* de cette ressource pour cet utilisateur, ou encore (et cela revient au fond au même, car nous ne comprenons que ce qui nous intéresse), celle de l'*intérêt* qu'elle présente pour lui⁶. On retrouve ainsi dans l'espace logique, sous des formes certes différentes, les *limites*, la *distance* qui nous sont familières dans l'espace géographique.

On peut aussi y définir d'autres distances : une distance entre utilisateurs, d'autant plus grande que leurs centres d'intérêt sont plus éloignés ; une distance entre les ressources, d'autant plus grande que leur contenu est plus différent. C'est (cf. annexe) après avoir

3. Les « ressources informatiques », que nous appellerons « ressources » tout court, sont composées (1) de textes, données, sons et images ; (2) des logiciels utilisés pour les classer, trouver, traiter ; (3) de la mémoire et de la puissance des serveurs.

4. C'est l'affaire de quelques années avec l'extension des réseaux cellulaires à haut débit et la miniaturisation : certains terminaux, comme le Treo de Handspring, conjuguent déjà les fonctions de l'ordinateur et celles du téléphone mobile.

5. La gestion des droits d'accès sur le réseau téléphonique est assurée par le HLR (*Home Locator Register*), outil déjà ancien construit dans les années 90. Les opérateurs cherchent à le faire évoluer pour tenir compte de l'adressage IPv6.

6. Les forums constituent une masse où s'accumule le commentaire du commentaire et où l'on se perd sans trouver de point saillant. Anne Cauquelin [3], se référant à la philosophie stoïcienne, a parlé des incorporels, du vide qui entoure l'objet et lui permet d'exister devant l'attention du lecteur : les mots sont ainsi entourés d'un espace dans lequel ils peuvent signifier quelque chose. Le forum, étant trop plein, ne laisse pas jouer l'espace vide qui rendrait les mots expressifs. Produire le point saillant (une synthèse, une orientation) permettrait de valoriser le stock de textes.

défini ces deux distances que l'on peut évaluer, en s'appuyant sur l'observation statistique, la distance entre un utilisateur et une ressource.

L'ubiquité logique que procure l'Internet est *absolue* (sous réserve des droits d'accès). L'*ubiquité physique* qu'il procure est par contre, comme en téléphonie, limitée aux signaux : le réseau peut transmettre le signal vocal, ou l'image d'une personne, ou encore le codage de la forme géométrique d'un objet physique que le récepteur pourra reproduire, mais il ne peut pas transporter des objets physiques.

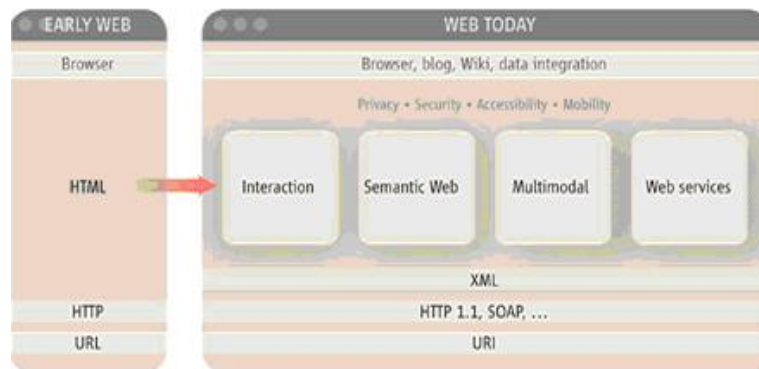


FIG. 1 – *Le Web hier et aujourd'hui (Source : World Wide Web Consortium)*

Lors de sa création en 1991 le Web comportait des textes codés en HTML (*Hypertext Markup Language*), transmis via le protocole HTTP (*Hypertext Transfer Protocol*) et consultables en utilisant un navigateur pointant vers une adresse URL (*Universal Resource Locator*).

Les utilisateurs disposent aujourd'hui d'une panoplie plus large : les SVG (*Scalable Vector Graphics*^a), le Web sémantique, des outils multimédia (navigateurs vocaux par exemple), des descriptions de services. Les ressources sont codées en XML (*eXtensible Markup Language*), transmises par les protocoles HTTP 1.1 ou SOAP (*Simple Object Access Protocol*) et leur adresse est un URI (*Uniform Resource Identifier*).

Il suffit de prolonger par la pensée cette évolution du Web pour anticiper les outils dont nous disposerons demain ([18] et [12]).

^a Format de fichier fondé sur XML et qui permet de décrire des graphiques vectoriels.

2 Comment cela va-t-il se passer pratiquement ?

Projetons-nous dans un avenir point trop lointain, mais suffisamment éloigné pour que les conséquences des tendances actuelles aient pu se déployer : disons donc dans dix ou quinze ans, en 2016 ou 2021 [9].

Dans les pays riches, les utilisateurs accèdent à des réseaux mobiles à haut débit (de l'ordre de 10 Mbit/s) et, dans leur entreprise comme à domicile, à des réseaux à plus haut débit encore (de l'ordre de 100 Mbit/s). Ils disposent d'un *terminal mobile*, successeur du téléphone mobile d'aujourd'hui, qu'ils portent sur eux et qui leur permet d'accéder à leurs ressources à tout moment et où qu'ils soient, et de *terminaux fixes*⁷, équipés d'un écran de

⁷. Sans doute aura-t-on entre temps trouvé pour ces équipements des dénominations plus séduisantes, plus familières que « terminal fixe » et « terminal mobile ».

grande dimension à haute définition (ou de plusieurs écrans) et d'un clavier, permettant d'accéder aux services et aux ressources selon une excellente ergonomie.

Les utilisateurs ont ainsi accès à des ressources informatiques publiques, personnelles et professionnelles (configuration de l'interface homme machine, fichiers, applications) résidant sur des serveurs sécurisés⁸. Lorsqu'ils se connectent les utilisateurs s'identifient et authentifient leur identification, après quoi ils accèdent aux ressources dans la limite de leurs habilitations.

Les terminaux fixes sont installés au domicile, au lieu de travail et dans des lieux publics. Ils rassemblent les fonctions aujourd'hui remplies par l'ordinateur, le téléviseur et la chaîne à haute fidélité⁹. Si l'utilisateur s'installe en face d'un terminal fixe et demande la connexion, son terminal mobile émet vers le terminal fixe les informations nécessaires à l'identification : le terminal fixe donne alors, après saisie d'un mot de passe, accès aux ressources de l'utilisateur.

Réseaux et terminaux donnent accès à une plate-forme téléphonique, audiovisuelle et informatique sur laquelle sont offerts des services diversifiés, gratuits ou payants. L'accès aux ressources audiovisuelles, par exemple, est outillé de fonctions de recherche et de tri qui ont transformé les conditions pratiques de la programmation individuelle¹⁰ ([8] et [2]). La plate-forme comporte des outils sécurisés de gestion des comptes bancaires. Divers équipements (automobile, distributeur automatique de billets, portes à ouverture automatique, caisses des magasins etc.) sont eux aussi communicants et obéissent aux ordres de l'utilisateur.

Par rapport à la situation présente, voici les changements que cela implique :

- le téléphone mobile est devenu un terminal mobile qui incorpore, outre les fonctions du téléphone et de l'ordinateur, celles du GPS, du magnétophone, de l'appareil de photo, de la caméra, du téléviseur ; s'il est raccordé à des capteurs, il peut collecter l'électrocardiogramme, la tension etc. et envoyer des alertes à un centre de télésurveillance médicale. Il est doté d'un écran lisible¹¹. Le clavier est soit dépliable (des claviers dépliables équipent déjà certains Palmtops), soit une image projetée par un laser sur une surface plane etc. ;

- les ressources informatiques ne résident pas sur le disque dur du terminal mais sur des serveurs sécurisés (chiffrement, *back up*) accessibles via l'Internet, exploités par des opérateurs spécialisés et situés n'importe où dans le monde¹². Une entreprise peut soit exploiter elle-même les serveurs sur lesquels résident ses ressources informatiques, soit confier cette exploitation à des opérateurs ;

- les terminaux fixes et les terminaux mobiles sont en principe des terminaux « bêtes », équipés seulement du logiciel nécessaire pour la connexion à la ressource informatique et l'affichage de l'interface. Cela permet un démarrage rapide et l'affichage pratiquement instantané de l'interface de l'utilisateur. Ils peuvent toutefois, en option, être équipés comme les ordinateurs d'aujourd'hui du logiciel et de la mémoire de masse nécessaires pour une utilisation hors réseau. Les réseaux mobiles à haut débit et les réseaux fixes à très haut

8. Les ressources d'un même utilisateur peuvent résider sur plusieurs serveurs : les serveurs, étant en réseau, communiquent pour lui présenter des ressources « sans couture ».

9. HP vient de lancer un ordinateur doté d'une télécommande et qui remplit les fonctions du téléviseur.

10. Ces fonctions sont aujourd'hui disponibles, mais peu ergonomiques et donc peu utilisées. Les ressources graphiques et sonores ont besoin d'un titre et d'un commentaire, alors que les ressources textuelles peuvent s'indexer elles-mêmes. On peut prévoir des traitements de type textuel qui aident à construire un contexte d'interaction : à tout texte, le lecteur associe un paratexte [4], une enveloppe de significations : « je suis en train de lire un roman policier, je sais ce que l'on attend de la lecture d'un roman policier ».

11. Les recherches en cours s'orientent vers diverses solutions : écran à haute définition de taille réduite (comme celui que possède aujourd'hui l'iPod), projection de l'écran sur les lunettes, écran souple dépliable comme un journal etc.

12. Le service MediaMax, qui utilise Streamload [16], offre aujourd'hui sur l'Internet à chaque utilisateur un espace de stockage gratuit de 25 Go.

débit apparaissent alors comme complémentaires.

3 Architecture de l'espace logique

L'architecture de l'espace logique peut se représenter selon un modèle en couches (figure 2).

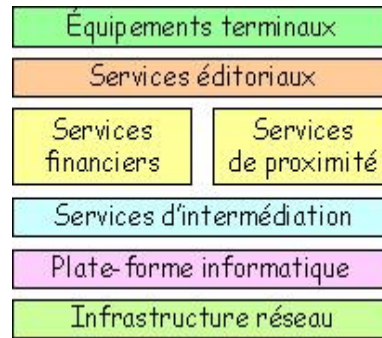


FIG. 2 – Couches de l'espace logique

- *infrastructure du réseau* : deux réseaux d'accès (mobile à haut débit, fixe à très haut débit) et un réseau de transport, l'Internet, avec les outils informatiques de gestion et supervision (adressage, réplication, dimensionnement¹³ etc.) ;

- *services de plate-forme informatique* : serveurs (processeurs, systèmes d'exploitation, outils de supervision), mémoires, sécurité (chiffrement, *back up*, protection contre les intrusions, antivirus etc.). Ces services traitent de la *physique informatique* et assurent la fiabilité et la disponibilité du service ainsi que la protection des ressources contre les indiscretions et les attaques ;

- *services d'intermédiation* : aide à la recherche des ressources qui correspondent aux besoins de l'utilisateur (moteurs de recherche, classification documentaire, dissémination sélective, *publish and subscribe* etc. [11]). L'intermédiation est associée, pour les ressources payantes, à des services financiers ; pour les produits matériels (livres, équipements etc.) à des services de proximité ;

- *services financiers* : traitement automatique des effets de commerce (dettes, créances, ordres de virement, répartition entre ayants droit) échangés par les utilisateurs et les fournisseurs ;

- *services de proximité* : livraison, installation, formation des utilisateurs, entretien, dépannage etc.

- *services éditoriaux* : ressources gratuites ou payantes mises à disposition sur le réseau (logiciels, textes, musique, audiovisuel, données etc.) ;

- *équipements terminaux*, répartis en deux grandes familles (fixes et mobiles).

4 Changements par rapport à aujourd'hui

On retrouve dans ce monde certains services que nous connaissons déjà : le commerce électronique est en place avec les outils de paiement informatisés et il s'articule avec des services de livraison pour les produits pondéreux. Le futur apportera donc non pas de l'émer-

13. Le trafic sera plus aléatoire que sur le réseau téléphonique en raison de la diversité des services : il faudra dimensionner le réseau en conséquence.

gence de services entièrement nouveaux, mais l'exploitation systématique, plus puissante aussi parce que disposant de moyens nouveaux, des services qui existent déjà aujourd'hui.

Cette intensification a des conséquences pratiques : elle « change la vie » tout comme le téléphone mobile, qui lui aussi n'avait « rien de nouveau », l'a changée lorsqu'à partir de 1995 son prix a baissé, que sa couverture géographique s'est densifiée et qu'il est devenu un produit de masse.

La baisse du prix du haut débit, son extension aux réseaux mobiles, la généralisation de la tarification forfaitaire, les progrès dans l'ergonomie du terminal mobile et du terminal fixe auront élargi et modifié la relation entre les utilisateurs (personnes ou entreprises) et la ressource informatique. Des services dont le « business plan » aurait été auparavant peu crédible seront rentables. Des équipements qui n'avaient jamais été mis en réseau (systèmes d'arrosage, équipements ménagers, chauffage etc.) lui seront raccordés, ce qui facilitera leur télécommande, leur télémaintenance et le téléchargement de nouvelles versions de leur logiciel. L'informatisation des transactions facilitera l'accès des PME à des services de comptabilité, de gestion d'agenda etc., ainsi que le partage des centres d'appel.

La ressource informatique d'une personne contiendra son dossier médical, l'historique de ses comptes bancaires, de ses revenus et de ses dépenses, celui de son cursus scolaire et des formations qu'elle a suivies etc. L'accès à ces informations doit être sécurisé¹⁴ mais l'utilisateur pourra s'il le souhaite l'ouvrir de façon permanente ou temporaire à des personnes qu'il autorise (médecins, administration fiscale etc.)¹⁵.

Cependant tout ne sera pas automatisé : parallèlement à la croissance de la part prise par l'automate dans les services croîtra pour l'utilisateur le besoin d'une assistance et d'un dialogue avec un être humain, que ce soit par réseau (centre d'appel, visiophonie) ou en face à face, afin de traiter les problèmes qui ne sont pas du ressort de l'automate (expliquer, comprendre). Il en résultera un changement de la structure de l'emploi.

L'ubiquité de l'espace logique permettra des usages aujourd'hui inédits. Il se peut par exemple qu'une personne qui habilite l'administration fiscale à consulter automatiquement l'historique de ses revenus bénéficie d'une réduction de l'impôt sur le revenu en contrepartie de la fiabilité de l'information et de la commodité d'accès. Une personne qui consulte un médecin pourra, pendant la durée de la consultation, lui donner accès à son dossier médical, à l'historique des prescriptions et traitements etc.

5 Outils de l'espace logique

Les outils nécessaires à la conquête de l'ubiquité sont en cours de mise au point, voire déjà disponibles : réseaux à haut débit, outils de classement et de recherche, de chiffrement et de gestion des habilitations etc. L'outillage avance et la synergie des divers outils converge, mais sans toutefois semble-t-il que le point vers lequel s'oriente cette convergence soit clairement perçu : les réseaux à haut débit et l'UMTS se mettent en place sans que l'on sache exactement à quoi ils vont servir et il en est de même des outils de recherche. Il est d'autant plus intéressant d'explorer les conséquences prévisibles de cette synergie.

À chaque utilisateur, on pourra associer dans l'espace logique un *domaine* comportant l'ensemble des ressources susceptibles de l'intéresser. Ce domaine sera défini (1) par

14. Le chiffrement est un enjeu crucial : pour que les entreprises, les personnes, puissent déposer leurs ressources sur le réseau, il faut que la confidentialité soit inviolable fût-ce même pour un organisme qui, comme la NSA, dispose des outils les plus puissants.

15. Un rapide calcul montre que l'ensemble des conversations (horodatées et localisées) auxquelles un être humain participe pendant sa vie tient, après traitement par un logiciel de reconnaissance vocale et codage en mode caractère, sur un CD-Rom. Que cette possibilité soit utilisée ou non, elle illustre ce qu'une personne pourra indexer et stocker dans une mémoire de taille relativement modeste.

segmentation, le domaine étant commun à tous les utilisateurs appartenant à un même segment ; (2) par *personnalisation*, l'observation des requêtes et du comportement d'un utilisateur permettant de préciser l'information sur ses besoins que donne la segmentation. On pourra lui fournir des outils qui :

- facilitent la recherche des ressources intéressantes (moteurs de recherche, mais aussi synthèses, critiques, paratexte etc.) ;

- indiquent, de façon proactive, les changements survenus dans le contenu de son domaine (nouvelles ressources etc.) ;

- permettent de prendre connaissance du contenu de son domaine, de ses frontières, des relations qu'il entretient avec les domaines connexes, des extensions possibles de son champ d'intérêt.

Des communautés d'utilisateurs pourront se former à l'exemple de ce qui se passe aujourd'hui sur Flickr où l'on stocke, documente et partage des photographies, l'index documentaire amorçant le fonctionnement communautaire en facilitant la recherche.

* *

Il est difficile de recenser *a priori* tous les usages possibles du terminal mobile, toutes les situations d'interaction, de même qu'il aurait été très difficile à Gutenberg de prévoir toutes les utilisations de l'imprimerie, des journaux aux livres et magazines jusqu'aux notices techniques et fiches que l'on trouve dans l'emballage des médicaments. C'est un thème de recherche en soi, qui peut tout au plus s'amorcer par l'observation et l'extrapolation raisonnée des pratiques actuelles.

Les possibilités ne seront pas immédiatement utilisées : si la ressource informatique personnelle semble par exemple le lieu naturel où l'on peut stocker un dossier médical, il faudra du temps pour que la corporation médicale, le milieu hospitalier, acceptent de s'y adapter. De façon générale les institutions, confrontées à une transformation profonde de leurs procédures, seront lentes à en tirer parti.

Certaines personnes vont résister, car l'informatisation de la vie personnelle ne sera pas bien vécue par tout le monde ; elle posera d'ailleurs, tout comme l'a fait le téléphone mobile, des problèmes de savoir-faire et de savoir-vivre nouveaux et irritants.

Les prédateurs chercheront les failles éventuelles des systèmes de chiffrement et de sécurité : l'espace logique devra donc, pour être habitable, être équipé de dispositions juridiques, surveillé par une gendarmerie vigilante, sanctionné par un appareil judiciaire qualifié.

Les entreprises, enfin, rencontreront dans l'espace logique de nouvelles formes de commerce et d'organisation de la force de travail auxquelles elles devront elles aussi s'adapter.

Les efforts d'organisation personnelle et collective, la transformation des institutions, prendront plus de temps que l'innovation technique.

* *

L'évocation de ce futur proche éveille inévitablement une question à la fois naïve et fondamentale : « en quoi cela nous rendra-t-il plus heureux, plus intelligents ? ». La réponse est implacable : ni le bonheur, ni l'intelligence ne dépendent des outils qui, comme le téléphone, l'automobile, l'avion, la télévision – et demain les réseaux à haut et très haut débit, le terminal mobile et son couplage au terminal fixe – ne font que modifier notre rapport au temps et à l'espace. Le mathématicien le plus savant, pour ne prendre que cet exemple, se contente d'un crayon et de papier. L'espace logique ne nous apporte ni le bonheur, ni l'intelligence : il pourra être utilisé par le barbare comme par le sage, et certains s'inquiètent déjà des moyens qu'il offrira aux terroristes.

Construire sans tarder les savoir-faire et les savoir-vivre que nécessite l'espace logique

apparaît alors comme un enjeu pour la civilisation. C'est pourquoi il importe de voir où conduit l'évolution actuelle: il ne s'agit pas de faire l'apologie béate de la technique, mais de se préparer à la vie dans l'espace logique.

6 Mise en perspective : temps et espace

Pour comprendre la nature et les enjeux de l'espace logique il est utile de se tourner vers l'histoire.

L'espace logique est plus ancien que le « cyberspace » qui s'est ouvert sur l'Internet. Nous nous trouvons en effet déjà dans l'espace logique pendant une conversation, au cinéma, au théâtre, devant notre téléviseur. Nous y accédons aussi lorsque nous entrons dans une bibliothèque: « la lecture de tous les bons livres est comme une conversation avec les plus honnêtes gens des siècles passés¹⁶ ».

Dans une bibliothèque par exemple, et sous la double contrainte de ses préférences et des limites du catalogue, le lecteur est affranchi des contraintes de l'espace géographique et du temps. Comme tous les monuments, un texte peut être détruit ou mutilé mais, alors qu'aucun monument ne peut rester intact, certains textes nous sont à travers les millénaires et les continents parvenus tels qu'ils étaient sortis des mains de leur auteur: si nous savons les lire, nous entrons en conversation avec lui. Rien n'est plus émouvant que ces échanges par delà la mort, par delà les frontières du langage et de la culture. La *political correctness* prétend cependant les proscrire sous prétexte que leurs auteurs seraient des morts blancs du sexe masculin, « *they are all white, male and dead*¹⁷ »...

On peut donc dire que l'espace logique dont nous parlons n'a rien de nouveau, et il est vrai qu'il ne faut pas s'exagérer sa nouveauté. Toutefois il n'est pas indifférent que l'on puisse y disposer grâce au Web d'une plus grande diversité de ressources, d'outils plus efficaces de recherche et de sélection.

L'espace logique, devenu sur l'Internet accessible à tout moment et en tous lieux, équipé d'outils de recherche et de classement, s'y manifeste d'une façon plus évidente et qualitativement plus complète qu'il ne le faisait auparavant dans les bibliothèques. C'est cet espace ainsi complété, achevé, que nous voulons explorer ici. Mais nous l'aborderons en gardant en mémoire les leçons que l'on peut tirer des pratiques immémoriales de la conversation, de la lecture, du spectacle.

* *

L'espace logique est d'abord la négation de l'espace géographique. Déjà, dans les entreprises, l'Intranet a permis de combler (voire de retourner) l'écart de compétence professionnelle qui existait entre la direction générale et les directions régionales¹⁸. Une part des travaux de l'entreprise s'effectue dans l'espace logique: interprétation des textes, classement comptable, programmation informatique etc. Le télétravail sera donc mis en œuvre pour les activités pour lesquelles il est économiquement efficace (c'est-à-dire: non pas à 100 % ni pour toutes les activités, mais de façon significative tout de même).

16. René Descartes (1596-1650), Discours de la méthode (1637). C'est sans doute en s'interrogeant sur la qualité des conversations que l'on peut trouver la clé du choix de ses lectures comme de la façon de lire.

17. Cela rappelle le cri par lequel von Stahremberg galvanisait ses troupes: « *Wir wollen keine deutsche Kultur! Wir wollen keine französische Kultur! Wir wollen gar keine Kultur!* » (Nous ne voulons pas de la culture allemande! nous ne voulons pas de la culture française! nous ne voulons pas de culture du tout!) et la fameuse phrase de Goebbels « *Wenn ich das Wort Kultur höre, entsichere ich meinen Revolver* » (Quand j'entends parler de culture, j'arme mon revolver).

18. Les « gens des DR » sont des lecteurs plus attentifs de la documentation technique, désormais accessible et à jour sur l'Intranet, que ne le sont les « gens de la DG » qui comptent pour s'informer sur des conversations entre experts.

L'espace logique est aussi en relation avec l'espace géographique : le commerce des produits physiques sur le Web se boucle nécessairement par une livraison ; la localisation d'appel est à la base de plusieurs des services offerts sur les réseaux UMTS ; les automobilistes utilisent des systèmes de navigation qui s'appuient sur le GPS. Voici une exploration (non exhaustive) des services qui concrétisent l'articulation entre espace logique et espace géographique :

6.0.1 Localisation

La localisation, fournie de façon précise par le GPS incorporé dans le terminal mobile, est un des attributs de l'utilisateur et contribue à la définition des ressources qui l'intéressent : il s'intéresse potentiellement à sa région, sa commune, aux manifestations culturelles locales, aux commerçants du coin, aux services de proximité qui peuvent lui être offerts.

Il peut être utile aussi de communiquer aux personnes habilitées la localisation des personnes dépendantes (enfants, personnes âgées).

6.0.2 Déplacements

Le GPS permet à l'utilisateur de savoir où il se trouve pendant ses déplacements. Les systèmes de navigation peuvent équiper le piéton aussi bien que l'automobile.

Le calcul d'itinéraire peut se faire en temps réel, c'est-à-dire en tenant compte de la situation présente du trafic (chantiers, embouteillages, manifestations etc.).

Le piéton peut localiser les taxis et transports en commun disponibles : emplacement des stations et arrêts, durée d'attente, durée prévisible du voyage.

6.0.3 Ressources physiques

Le terminal mobile peut :

- informer l'utilisateur sur les ressources physiques proches et ouvertes (pharmacies, hôpitaux, distributeurs automatiques de billets, agences bancaires, bureaux de poste, restaurants, cinémas etc.) [7] ;
- afficher des vitrines virtuelles (quand on passe devant un immeuble, indique des magasins dont certains ne se voient pas depuis la rue) ;
- donner une information météorologique localisée ;
- signaler au touriste des monuments intéressants, des expositions etc., et lui donner des explications historiques sur les lieux qu'il parcourt ;
- indiquer les noms des habitants d'un immeuble, etc.

6.0.4 Système d'information géographique

L'utilisation locale d'un SIG fournit des informations diverses et à des échelles diverses ; elle permet de localiser les réseaux qui se trouvent en sous-sol, de visualiser le cadastre etc.

* *

7 Annexe : outils mathématiques

Les utilisateurs du Web disposent actuellement de moteurs de recherche dont Google est le plus fameux et d'encyclopédies en ligne dont Wikipedia est la plus connue. Les liens hypertexte leur permettent de « surfer » pour trouver des ressources inconnues et, d'aventure, intéressantes.

Les outils actuellement disponibles, aussi puissants et utiles qu'ils soient, ont des limites. Google indexe des chaînes de caractère et les résultats qu'il fournit sont altérés par des homonymies. Le *surf* est aléatoire. Les outils de dissémination sélective, censés apporter automatiquement à l'utilisateur des informations qui l'intéressent, sont dans les limbes.

Cependant les recherches en cours ont entrepris de corriger ces défauts et esquissent l'outillage de l'espace logique. Des outils sémantiques visent à supprimer les homonymies dans les moteurs de recherche [15]; le contenu du Web est soumis à des outils de classification automatique délimitant des *clusters* de textes [6]; des outils de *datamining* examinent les co-occurrences lexicales et détectent les tendances de l'évolution des contenus¹⁹ [14]; les annonceurs observent le comportement des utilisateurs de l'Internet [5].

* *

Entre un utilisateur et une ressource, on peut définir une distance qui est celle de l'*intelligibilité* ou, ce qui revient au fond au même, de l'*information* que ce texte peut apporter au lecteur en prenant « information » non au sens qu'a ce mot dans la théorie de Shannon [13] mais au sens étymologique : une information, c'est quelque chose qui vous *informe*, qui modifie ou complète la forme intérieure de votre représentation du monde, qui vous *forme* vous-même²⁰. L'information ainsi conçue a une *signification* : elle suscite une action de la part de celui qui la reçoit ou du moins elle modifie (*trans-forme*) les conditions de son action future. Plus un texte peut apporter d'information à un lecteur, plus il présente d'intérêt pour lui : *intelligible*, *intéressant*, *informatif* sont donc des synonymes.

Pour pouvoir recevoir une information, il faut avoir été formé, et c'est en recevant de l'information que l'on se forme. L'amorce de ce cycle est enfouie dans les origines de la personne tout comme l'amorce du cycle de la poule et de l'œuf est enfouie dans les origines de la vie.

* *

Notons U l'ensemble des utilisateurs, R l'ensemble des ressources. Supposons définie une distance $d(u,r)$ entre l'utilisateur u et la ressource r , d'autant plus petite que cette ressource est plus intelligible pour cet utilisateur. Nous appellerons « domaine de l'utilisateur u » l'ensemble $D(u)$ des ressources dont la distance à l'utilisateur est inférieure à un seuil conventionnel s :

$$D(u) = \{r \mid d(u,r) < s\}.$$

Nous appellerons « lectorat de la ressource r » l'ensemble $L(r)$ des utilisateurs dont la distance à la ressource est inférieure à s :

$$L(r) = \{u \mid d(u,r) < s\}.$$

Considérons un sous-ensemble D de R . Nous noterons $L(D)$ l'ensemble des utilisateurs intéressés par toutes les ressources que contient D :

$$L(D) = \{u \mid D \subset D(u)\}.$$

19. Voir le moteur de recherche Clusty, qui propose des clusters de documents.

20. « Informer » et « instruire » sont proches : informer, c'est donner une forme ; instruire, c'est donner une structure.

Considérons un sous-ensemble L de U . Nous noterons $D(L)$ l'ensemble des ressources qui intéressent tous les utilisateurs appartenant à L :

$$D(L) = \left\{ \bigcap_u D(u) \mid u \in L \right\}.$$

Nous dirons que L et D sont *en correspondance* si $L[D(L)] = L$ et $D[L(D)] = D$.

Nous allons esquisser la démarche qui permet de construire deux segmentations (l'une des utilisateurs, l'autre des ressources) dont les segments mutuels sont statistiquement en correspondance.

* *

Les outils de segmentation en cours de mise au point utilisent des techniques statistiques connues et éprouvées (notamment les *Support Vector Machines*²¹), mais à une échelle et sur des volumes qui exigent des algorithmes et des processeurs puissants. Pour classer des textes, ils considèrent la taille, la structure formelle, les métadonnées et les co-occurrences lexicographiques. Pour classer des images et les enregistrements audio ou vidéo, ils considèrent les métadonnées (index etc.) et, plus difficilement, les formes représentées ainsi que les paramètres techniques de l'image ou du son.

Dans tous les cas, il faut définir une distance entre les ressources, puis une distance entre agrégats de ressources; on peut alors, par classification ascendante hiérarchique, construire une ultramétrie sur l'ensemble des ressources puis obtenir une segmentation en « coupant les branches les plus longues » de l'arbre qui représente l'ultramétrie.

Si l'on veut que le Web puisse signaler à l'utilisateur de façon proactive les ressources susceptibles d'intéresser celui-ci, il faut segmenter la population des utilisateurs en utilisant *d'abord* des données « intrinsèques » (c'est-à-dire indépendantes de l'usage qu'ils font de la ressource) observées par enquête auprès d'un échantillon représentatif : âge et sexe, catégorie socioprofessionnelle, lieu de résidence, lieu de travail, métier, fonction dans l'entreprise, statut matrimonial, nombre et âge des enfants, *hobbies* etc.

Puis on observe, toujours sur le même échantillon, les pratiques dans l'utilisation du Web et on demande éventuellement aux utilisateurs de noter l'intérêt des ressources qu'ils consultent. On établit un tableau des fréquentations en croisant la segmentation des utilisateurs et celle des ressources (figure 3).

Une analyse factorielle des correspondances [17], réalisée sur ce tableau, met en évidence la corrélation entre les deux segmentations (c'est parce que l'on recherche cette corrélation qu'il ne fallait pas prendre en compte de façon prématurée, pour classer les utilisateurs, les données concernant leurs consultations). L'ordre des classes sur le premier axe de l'analyse fournit un tableau aussi proche que possible de la forme diagonale (figure 4).

Il est alors possible d'associer statistiquement à chaque segment des utilisateurs une catégorie de ressources et inversement. Cela fournit une segmentation globale, définie à la fois selon les paramètres propres aux utilisateurs et selon ceux des ressources : les ensembles U et R sont ainsi mis en correspondance segment par segment. On peut de la sorte, ayant identifié les ressources qui intéressent le plus un segment d'utilisateurs, amorcer une politique de diffusion proactive.

On trouvera en dehors de la diagonale quelques cases bien remplies qui font exception à la logique d'ensemble. C'est là un des résultats les plus précieux de l'analyse : ces cases contiennent des utilisateurs dont le comportement s'écarte du comportement majoritaire,

21. Les SVM sont un outil d'analyse discriminante linéaire, mais dans un espace où l'on a introduit des fonctions puissance ou autre des données observées; on obtient ainsi dans l'espace d'origine une frontière qui n'est pas linéaire.

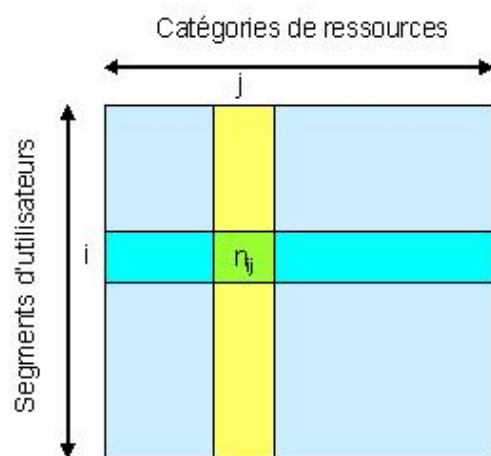


FIG. 3 – *Croisement des segmentations*

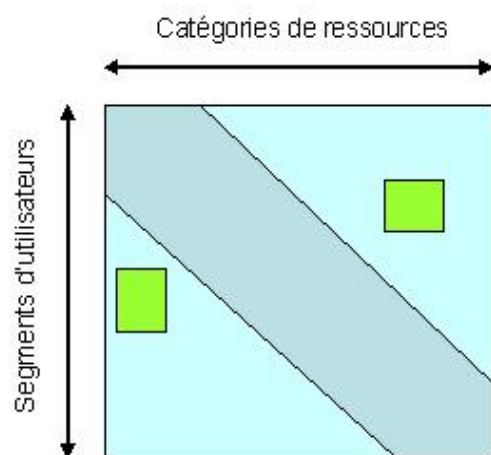


FIG. 4 – *Diagonalisation du tableau*

qu'ils soient « en avance » ou « en retard » dans la consultation des ressources ou pour tout autre raison. Les examiner donne un aperçu sur la dynamique de la demande.

Lorsque l'outillage est ainsi amorcé, on peut mettre les segmentations à jour en recherchant, parmi les combinaisons linéaires des données intrinsèques et des ressources consultées par un utilisateur, celles qui sont les mieux corrélées : il faudra ici recourir aux techniques de l'analyse canonique [17]. Cette analyse permet de :

- placer un nouvel utilisateur dans l'espace documentaire, une fois connu le segment auquel il appartient (par la suite, la personnalisation permettra de faire évoluer ce qu'on lui propose) ;

- signaler aux utilisateurs d'un segment les nouveaux documents qui font partie du segment documentaire qui *a priori* les intéresse ;

- fournir à chacun une vue de l'espace logique, de la façon dont il est structuré, et de l'endroit où on le situe lui-même (il peut ainsi organiser des « voyages » dans cet espace) ;

- identifier les segments « voisins », potentiellement intéressants les ceux des utilisateurs d'un segment donné qui souhaitent élargir leur horizon.

Références

- [1] Francis André. *Libre accès aux savoirs : Open Access to Knowledge*. Futuribles, 2005.
- [2] Sandra Braman. Transformations of the research enterprise. *Educause ReviewEducause ReviewEducause Review*, août 2006.
Voir <http://www.educause.edu/ir/library/pdf/ERM0641.pdf>.
- [3] Anne Cauquelin. *Fréquenter les incorporels : contribution à une théorie de l'art contemporain*. PUF, 2006.
- [4] Gérard Genette. *Palimpsestes*. Seuil, 1982.
- [5] Saul Hansel. Advertisers trace paths users leave on internet. *The New York Times*, 15 août 2006.
- [6] Thorsten Joachims. *Learning to Classify Text using Support Vector Machines*. Kluwer Academic Publishers, 2002.
- [7] Vassilis Kostakos Eamonn O'Neil. Designing urban pervasive systems. *Computer*, septembre 2006.
- [8] Sergey Brin Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International World Wide Web Conference*. Elsevier Science, 1998.
L'article qui décrit les origines de Google.
- [9] Pew Internet & American Life Project. The future of the internet, 24 septembre 2006.
Voir http://www.pewinternet.org/pdfs/PIP_Future_of_Internet.pdf.
- [10] Robert Reich. *The Future of Success, Working and Living in the New Economy*. Knopf, 2000.
- [11] Weiguo Fan Linda Wallace Stephanie Rich. Tapping the power of text mining. *Communications of the ACM*, septembre 2006.
- [12] Michael M. Roberts. Lessons for the future internet: Learning from the past. *Educause Review*, août 2006.
Voir <http://www.educause.edu/apps/er/erm06/erm0640.asp?bhcp=1>.
- [13] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, juillet - octobre 1948.
- [14] Colin Stewart. « data miners » at uci moving beyond google. *The Orange County Register*, 8 août 2006.
Voir http://www.ocregister.com/ocregister/money/columns/article_1236329.php.

- [15] Michael Stutz. Open source search technology goes beyond keywords. *NewsForge*, 25 septembre 2006.
Voir <http://software.newsforge.com/software/06/09/19/1531258.shtml?tid=130>.
- [16] Jean-Baptiste Su. Streamload : L'avenir du stockage de données privées appartient à internet. *L'Expansion*, 8 septembre 2006.
Voir <http://www.lexpansion.com/NLTech/4272.581.146624.html>.
- [17] Michel Volle. *Analyse des données*. Economica, quatrième édition, 1997.
Voir <http://www.volle.com/ouvrages/andon.htm>.
- [18] Tim Berners-Lee Wendy Hall James Hendler Nigel Shadbolt Daniel J. Weitzner. Creating a science of the web. *Science*, août 2006.

Pour lire un peu plus :

- Marketing des NTIC : prospective
<http://www.volle.com/travaux/marketingntic.htm>
- Marketing et connaissance des clients
<http://www.volle.com/travaux/marketing3.htm>
- Science du marketing
<http://www.volle.com/travaux/marketing2.htm>
- L'usage des TIC dans les entreprises
<http://www.volle.com/rapports/diact.htm>