

## LET'S DICE<sup>1</sup>

# Bibliothèque numérique, apprentissage automatique et interface 3D

Éric BRIYS

Co-fondateur, Cyberlibris et Professeur-Associé, Ceregmia - Université des Antilles et de la Guyane

Richard NOCK

Ceregmia - Université des Antilles et de la Guyane

Pour la bonne compréhension de cette article et de ses illustrations, la version couleur est exceptionnellement disponible dès maintenant sur notre site à l'adresse : <[http://www.abd-bvd.be/cah/papers/2012-2\\_Briys-Nock.pdf](http://www.abd-bvd.be/cah/papers/2012-2_Briys-Nock.pdf)>

Voor het goed verstaan van dit artikel en van haar illustraties is de kleurenversie uitzonderlijk beschikbaar vanaf nu op onze site op volgend webadres: <[http://www.abd-bvd.be/cah/papers/2012-2\\_Briys-Nock.pdf](http://www.abd-bvd.be/cah/papers/2012-2_Briys-Nock.pdf)>.

▪ Toute bibliothèque est un tribut à la géométrie. Une bibliothèque municipale est ordonnée selon les principes de la géométrie euclidienne. Les étagères y sont parallèles. Les livres y sont rangés par les bibliothécaires en fonction de leurs métadonnées thématiques. À angle droit avec les étagères sur lesquelles ils reposent, ils attendent patiemment d'être lus. Cette géométrie soignée est fort pratique. Elle facilite nos déambulations et nos découvertes. Mais qu'en est-il d'une bibliothèque numérique ? À quelle géométrie obéit-elle ? En quoi la découverte de cette géométrie intime permet-elle d'envisager de nouvelles pratiques bibliothécaires, de nouveaux modes de découvertes des livres ? Nous montrons dans cet article qu'une bibliothèque numérique est gouvernée par la géométrie non-euclidienne. Cette nouvelle géométrie est à la base même de nouvelles techniques d'apprentissage automatique et de nouvelles interfaces de recherche dont nous montrons les nombreuses applications possibles.

▪ Iedere bibliotheek is een ode aan de meetkunde. Een gemeentelijke bibliotheek wordt gerangschikt volgens de principes van euclidische meetkunde. De rekken staan parallel. De boeken worden geordend door de bibliothecarissen volgens hun thematische metagegevens. In een rechthoek met de rekken waar ze op staan wachten ze geduldig om te worden gelezen. Deze verzorgde meetkunde is erg handig. Het gemakkelijk onze wandelingen en ontdekkingen. Maar wat met een digitale bibliotheek? Aan welke meetkunde voldoet deze? In welk opzicht helpt deze intieme meetkunde om nieuwe bibliotheekkundige praktijken en nieuwe manieren om boeken te ontdekken te overwegen? We bewijzen in dit artikel dat een digitale bibliotheek bestuurd wordt door niet-euclidische meetkunde. Deze nieuwe meetkunde is de basis van nieuwe automatische leerprocestechnieken en van nieuwe opzoekingsinterfaces waarvan wij de talrijke toepassingen tonen.

## Les arpenteurs de la Terre plate

Au début de notre ère, les représentations de la Terre différaient notablement de celles que l'on connaît aujourd'hui. Les hommes d'alors ne disposaient pas des capacités d'observation actuelles, et les représentations cosmographiques exploitaient les connaissances locales du monde. Hormis les quelques informations globales qu'on savait utiliser à l'époque, comme la position des étoiles, on n'observait vraiment que ce qui se dessinait aux pieds des arpenteurs et dans leur voisinage. Contraint par ces visions locales, le monde était donc assez naturellement plat. Vinrent ensuite les intuitions, puis les premières observations, puis les premières méthodes techniques qui permirent de concevoir et de caractériser un monde sphérique. Certaines des méthodes utilisées arrivèrent, à partir d'un calcul local, à estimer avec une précision remarquable le périmètre de la Terre. L'une de ces méthodes est la géodésique d'Ératosthène, qui permet d'estimer

avec une marge d'erreur minimale le périmètre de la Terre à partir de l'observation des angles des rayons du soleil en deux villes distantes de moins de 1000 kilomètres : Syène et Alexandrie (voir fig. 1).

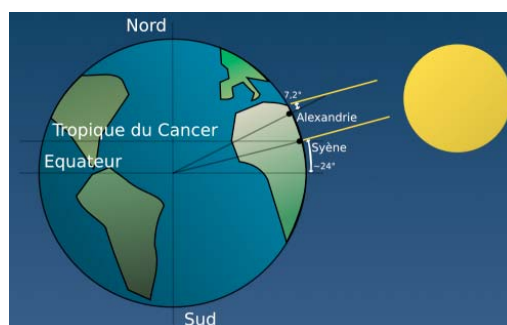


Fig. 1 : En remarquant la différence d'angle le jour du solstice d'été entre les rayons du soleil et la verticale, à Syène et à Alexandrie, Ératosthène en déduit facilement une très bonne estimation du périmètre de la Terre.

En somme, Ératosthène et ses illustres collègues, Pythagore, Euclide, Thalès de Milet, Archimède, Apollonius pour ne citer qu'eux assoient, la géométrie sur des bases solides et, ce faisant, nous dotent d'instruments qui, pour reprendre l'expression de l'Académicien des sciences Jean-Pierre Kahane, nous apprennent à appréhender l'espace.

## De la Terre plate à la bibliothèque

Quid de cet espace lorsqu'il s'agit de celui défini par une bibliothèque ? La question n'est pas si saugrenue que cela puisque déambuler le long des rayonnages d'une bibliothèque municipale, c'est aussi faire de la géométrie. Nos yeux balayaient ces lignes droites et parallèles que sont les étagères. Classer les livres sur des étagères, c'est encore faire de la géométrie. Cela implique de placer dans un ordre pré-établi des livres à angle droit avec les planches destinées à les accueillir. Finalement, une bibliothèque est une bâtisse (à la géométrie) familière, un endroit que nous connaissons généralement depuis notre tendre enfance. Adultes, il arrive que nous assemblions à la maison notre propre bibliothèque. Ce qui rend toute bibliothèque avenante à nos sens, en particulier à nos yeux, c'est le fait qu'elle soit un lieu organisé, optimisé dirait le mathématicien. La bibliothèque est taillée "géométriquement" à nos mesures. Elle réunit en un même lieu deux espaces, deux géométries radicalement différentes. Le premier espace est l'espace ambiant, ce que nos yeux voient : couloirs, rayonnages, étagères, livres, etc. C'est souvent celui que nous trouvons chaleureux et dans lequel nous éprouvons du plaisir. C'est un lieu rassurant, organisé, méthodique, feutré. Le deuxième est plus mystérieux. C'est l'espace d'information que recèle le contenu, les phrases, les mots de tous les livres stockés dans la bibliothèque. C'est celui qui effraie tant le Général Stumm dans *L'homme sans qualités* de Musil lorsqu'il visite la bibliothèque impériale :

*"Nous avons parcouru les rangs de ce colossal magasin et, je puis te le dire, ça ne m'a pas autrement frappé : ces rangées de livres ne sont pas plus impressionnantes qu'une parade de garnison... Mais que penses-tu que me réponde le bibliothécaire quand je vois que notre promenade s'éternise et lui demande combien de volumes contenait exactement cette absurde bibliothèque ? Trois millions et demi, me répondit-il ! de la manière que j'avais envisagée, il m'aurait fallu dix mille ans pour venir à bout de mon projet !"*

L'espace ambiant de la bibliothèque physique se décompose en deux ensembles disjoints, et l'on

peut écrire :

### bibliothèque physique = livres + lecteurs

La distinction est naturelle, d'autant que les deux ensembles sont diamétralement opposés. Il y a l'ensemble des livres, physiquement inanimé et silencieux, mais toujours bien rangé par les bons soins d'Euclide aidé parfois de classifications pertinentes qui font que le sympathique reporter Tintin côtoie rarement *La Bête humaine*. Il y a aussi l'ensemble des lecteurs, vivants, remuants, choisissant leurs livres en des termes qui échappent souvent à ces classifications, et laissent Euclide perplexe sur l'utilisation de son équerre et de ses étagères pour percer le secret des déambulations du Général Stumm. Peu importe, Euclide nous répondrait peut-être que sa bibliothèque à lui n'est qu'une interface physique entre une nourriture de l'esprit (le livre) et l'esprit lui-même (le lecteur), et qu'en l'absence de meilleure "connectique", le Général Stumm n'a qu'à s'aider de son cheval pour réduire le temps de sa déambulation.

Mais sans plus d'aide, et avec ou sans cheval, Stumm se perdrait bien davantage dans les méandres de l'espace d'information de la bibliothèque physique. Ce deuxième espace est plus angoissant. Il l'est tellement qu'il oblige l'écrivain Alberto Manguel (dont la bibliothèque ne compte "que" trente mille livres) à se relever la nuit pour ne plus voir les étagères, mais imaginer, à l'instar du bibliothécaire de Musil, les liens inopinés entre les livres, entrevoir les autres possibilités que le premier espace n'autorise pas. Chaque nuit, il cherche en vain à faire coïncider les deux espaces.

Notre guide Euclide est très fier du premier espace. Il est intarissable sur le sujet. Il faut dire qu'il l'a en quelque sorte inventé. Souvenez-vous : entre deux livres passe une étagère et une seule ; le plus court chemin d'un livre à un autre, c'est l'étagère sur laquelle ils se trouvent ; sous deux livres passe une étagère rectiligne qu'on peut prolonger à l'infini ; l'angle droit que fait le livre X avec son étagère est égal à l'angle droit que fait le livre Y avec l'étagère parallèle du dessous ou du dessus etc. C'est vrai, Euclide n'a pas tout à fait dit ça. Mais, il vous dira que si vous appelez un point un livre et une étagère une droite, vous retrouverez les postulats qui l'ont rendu célèbre et, qu'après tout, il en découle un ordre : celui de la mieux rangée des bibliothèques. Si vous n'êtes pas convaincu, Euclide a toujours une équerre avec un angle parfaitement droit dans la poche. Vous pourrez vérifier. Mais, faites attention, Euclide est un malin et il se gardera bien de vous parler de l'espace qui provoque les insomnies de Manguel et les angoisses du Géné-

ral Stumm. Si vous le taquinez un peu, il vous dira que de toute façon, si les bibliothèques sont toujours là, c'est qu'il doit quand même avoir un peu raison. Et, dans le fond, il n'a pas tout à fait tort. Depuis la bibliothèque des Ptolémées d'Alexandrie jusqu'à la bibliothèque municipale de Grand'Rivière en Martinique, la géométrie d'Euclide a fourni la recette qui a permis à notre corps et à notre esprit de progresser, de s'enrichir. Nous pouvons baguenauder dans la bibliothèque, saisir les livres, les feuilleter et les lire. Notre guide, Euclide, a donc du mérite et on doit lui savoir gré de sa prouesse géométrique. Pour un peu, on pourrait dire (Euclide vivait à Alexandrie) que c'est en se promenant dans la bibliothèque qu'il a eu la révélation comme Newton et sa pomme, ou encore Archimède et sa baignoire. Newton et Archimède regardent et tous deux voient une loi de la physique. Euclide regarde et il voit en quelque sorte une loi ou encore une géométrie de l'information dont l'incarnation est précisément la bibliothèque. C'est cette même loi, cette même géométrie à laquelle nous recourons quotidiennement lorsque nous fréquentons les bibliothèques. Mais, attention, de la même façon qu'un livre peut en cacher un autre, une géométrie peut en cacher une autre. Pour paraphraser l'économiste français du 19<sup>e</sup> siècle, Frédéric Bastiat, il y a la géométrie que l'on voit et celle que l'on ne voit pas. Le titre du livre de Manguel, *La bibliothèque, la nuit*, prend ici toute sa signification. Dans la nuit, nos yeux nous sont de peu d'utilité. Nos yeux, ces capteurs euclidiens, sont hors jeu et c'est le moment où Manguel repense sa bibliothèque. Le plus court chemin d'un livre à un autre n'est plus l'étagère, car dans l'esprit nocturne de Manguel les livres s'émanent de leurs étagères et s'unissent au gré des associations électives de leur(s) propriétaire(s). Entre deux livres peuvent passer plus d'une étagère, car les étagères elles-mêmes se libèrent de leur carcan euclidien pour enfin permettre à un seul et même livre de se poser simultanément sur plusieurs d'entre elles. Euclide n'y peut rien, il dort ! Tout au plus fait-il des cauchemars à la simple évocation de ce sabbat nocturne où sa géométrie est mise à mal. Euclide nourrit toujours l'espoir qu'une fois la lumière revenue tout ceci n'aura été qu'un mauvais rêve. Il n'en est rien.

Imaginons un seul instant que nous supprimons la partie physique de la bibliothèque (en quelque sorte le premier espace qu'Euclide apprécie tant) et que nous ayons un

moyen de naviguer à travers l'avalanche de mots que constituent les livres de la bibliothèque. Tous nos repères (euclidiens) habituels s'effacent : nous sommes partout à la fois, au début de la bibliothèque, au fond, au milieu, sur les côtés etc. Les livres se mélangent, chacun d'entre eux peut être partout à la fois et nulle part. Les étagères se croisent et se recroisent à en donner la nausée à ce cher Euclide. Lecteurs et livres se mélangent allègrement, violant tous les postulats de la géométrie euclidienne.

Bienvenue dans le web, bienvenue au sein de la bibliothèque numérique ! Toute "l'affordance" (pour employer un mot à la mode qui signifie la capacité d'un objet à suggérer sa propre utilisation) dont nous avons dotés Euclide s'évanouit. Tout est information, pour ne pas dire déluge d'informations. L'espace feutré et borné de la bibliothèque physique (la Terre plate) laisse place à un monde aux dimensions intimidantes.

Car arpenter une bibliothèque municipale, c'est un peu arpenter la Terre plate. Qu'importe si tout ceci n'est qu'une approximation locale, un arpentage de voisinage. C'est largement suffisant (avec la complexité des bibliothécaires) : nous restons toujours à quelques pas des livres et des étagères. En revanche, arpenter une bibliothèque numérique, c'est une autre paire de manches.

## L'arpenteur local de la librairie numérique plate

En effet, l'information dans laquelle nous baignons quotidiennement héberge des mondes dont la géométrie n'est pas la même que celle du monde euclidien dans lequel nous vivons. Les bibliothèques et les librairies numériques sont deux de ces mondes. À l'instar des géomètres grecs aux prises avec les dimensions de la Terre, l'entreprise américaine Amazon fait partie des premiers arpenteurs des librairies virtuelles de très grande taille. Dans ces mondes un peu par-



Fig. 2 : Recommandations fournies par Amazon.

ticuliers, la recherche d'informations cosmographiques a pour objectif principal d'aider les lecteurs à repérer leurs livres en s'appuyant sur le comportement des autres lecteurs.

C'est ainsi qu'*Amazon* fournit des recommandations sur la base d'un arpentage local de la librairie virtuelle : les lecteurs qui ont acheté mon livre ont aussi acheté tel ou tel livre. On ne cherche donc pas à voir plus loin que le "nez" des listes de lecture.

La première utilité des informations consiste à faciliter la recherche d'ouvrages, la base de données étant justement de très grande taille. Comment construire un outil de recommandation simple et efficace permettant de trouver les quelques livres susceptibles d'intéresser un lecteur, au coeur d'une montagne de livres accessibles ? La solution exploitée par *Amazon* consiste à ne surtout pas s'éloigner du lecteur, et à arpenter **localement** la bibliothèque, en utilisant comme base du déplacement les comportements de lecture des lecteurs. Pour dépasser la simple proximité géographique (thématique) - mais statique - de l'étagère physique (reproductible à l'identique dans la bibliothèque numérique en utilisant les classifications ad hoc, Dewey par exemple), *Amazon* a en effet utilisé la proximité virtuelle des livres dans les habitudes de lecture. On conseille à un lecteur les livres qui sont associés, dans les habitudes de lecture des autres lecteurs, aux livres qu'il a déjà lus. Le gros avantage de cette technique est qu'elle donne un côté dynamique à la recommandation : que les habitudes de lecture changent, et les recommandations changent dans la foulée ; qu'un best-seller soit introduit dans la bibliothèque, et les recommandations le captent quasi immédiatement à hauteur de son statut de best-seller.

Ce mode de calcul, tel qu'il est pratiqué, fait toutefois penser à un arpentage très limité de la librairie numérique, puisqu'on ne sélectionne que des livres immédiatement voisins des livres issus des habitudes de lecture (d'achat devrait-on plu-



Fig. 3 : Focus sur un ensemble de livres qui omet les livres hors cadre à l'instar de l'algorithme de recommandation d'*Amazon*.

tôt préciser). On ne tient donc pas compte des livres se trouvant à peine un pas plus loin, peut-être très proches des livres déjà lus - voire plus proches encore que certains livres recommandés. D'une certaine manière, en faisant l'impasse sur les structures géométriques globales de la librairie virtuelle, cette façon locale de procéder s'apparente à celle des arpenteurs de la Terre plate, et considère donc que la librairie est plate. Pour prendre une autre image, la librairie numérique telle qu'imaginée par *Amazon* n'est pas sensiblement différente de la photo d'un tas de livre, "oubliant" les ouvrages qui sortent du cadre.

On peut bien sûr objecter que les recommandations "à la *Amazon*" ne nécessitent finalement qu'un arpentage local de ce monde, et pas la capacité d'aller voir plus loin que le bout du nez des listes de lecture. Néanmoins, à l'échelle imposée par le numérique, la vision de la librairie qu'en donne *Amazon* reste étriquée, et ses recommandations approximatives. L'arpentage local ne rend pas "correctement" compte de la librairie physique ou de la bibliothèque municipale. Face à la globalité numérique, l'arpenteur doit devenir global.

## L'arpenteur global de la bibliothèque numérique non plate

Chez Cyberlibris, nous avons la passion des bibliothèques numériques et des riches univers qu'elles engendrent. Cette richesse va d'ailleurs au-delà de celle des librairies numériques car, contrairement à *Amazon* qui n'est pas en mesure de savoir si les livres achetés sont lus (de suite) ou pas, nous sommes en mesure pour chaque bibliothèque numérique bâtie par Cyberlibris de saisir les lectures en ligne, à la volée. C'est un avantage immense, comme on va le voir dans les lignes qui suivent.

Nous avons donc repris l'idée d'Ératosthène. Les géométries des mondes numériques sont beaucoup plus riches que celle du monde réel qui nous entoure. Pour comprendre la vraie géométrie de chaque bibliothèque numérique (académique, professionnelle ou grand public), telle qu'induite par les livres et les lecteurs, nous avons en quelque sorte multiplié à l'envi les mesures locales pour en capturer les particularités géométriques jusque dans leurs moindres détails. Force a été de constater que cette bibliothèque numérique n'est ni plate, ni ronde. Elle peut prendre des formes très diverses en fonction des livres, des lecteurs, des habitudes de lecture et de leurs variations. Qui peut le plus peut le moins : la modélisation complète de la géométrie de la bibliothèque numérique nous

permet évidemment de saisir le vrai voisinage d'un livre, et donc de réaliser des recommandations plus pertinentes que celles pratiquées par *Amazon*.

Mais, là où *Amazon* est obligé de s'arrêter aux recommandations locales dans la valorisation de sa librairie, du fait de l'ingénuité de sa méthode d'arpentage, nous pouvons nous permettre de monter considérablement en exigence dans l'utilisation de la description globale de notre bibliothèque. Notre vision d'ensemble de la bibliothèque nous permet par exemple de trouver automatiquement l'analogie de la classification "à la Dewey" qui serait la mieux adaptée pour "calculer" les étagères de livres et les présenter à l'ensemble des lecteurs. Évidemment, cette classification est dynamique : elle change avec les habitudes de lecture et présente une architecture globale de la bibliothèque toujours en phase avec son lectorat.

Notre vision d'ensemble de la bibliothèque nous permet également de proposer au lecteur de regarder bien plus loin que le bout de son livre, en quelque sorte de passer du "focus" au "contexte" et vice versa. En lui proposant des parcours de lecture partant de ce livre, nous sommes à même de reproduire à l'identique dans la bibliothèque virtuelle les promenades qu'un visiteur réaliserait pour partir à la découverte d'un lieu géographique. Partant d'un livre sur la cuisine épicée, la machine est à même de construire une véritable promenade initiatique amenant le lecteur à la découverte des pays d'origine des épices en question, voire même de contrées insoupçonnées, réelles ou fictives, ayant pourtant un lien indirect avec ces épices, car associées via les parcours et pratiques de lecture des autres lecteurs. L'intelligence communautaire devient alors une excellente source de sérendipité<sup>2</sup>.

Les approches locales telles que pratiquées par *Amazon* ont un inconvénient que notre vision d'ensemble de la bibliothèque numérique n'a pas : en dehors des ouvrages les plus consultés, une part importante des livres moins lus court le risque de se retrouver orpheline de recommandations, puisque celles-ci se concentrent uniquement sur les livres les plus lus. Ce problème n'existe pas dans notre approche, puisque la construction virtuelle de la bibliothèque utilise absolument toute l'information sur les livres (y compris leurs métadonnées) et les pratiques de lecture. En somme, et pour simplifier :

### **Bibliothèque numérique = Information**

Cet arpentage global suggère qu'ensemble nous pouvons être "plus forts" qu'Euclide. Alors prêtons-nous à rêver, prenons de l'altitude et jouons

aux apprentis géomètres. En déambulant dans la bibliothèque numérique, allons chercher les quelques principes qui la gouvernent. Empruntons les pas qu'Euclide a un jour empruntés pour définir sa géométrie et tâchons de répondre à une question ambitieuse : quelle est donc la géométrie de la bibliothèque numérique si elle n'est pas euclidienne ? En d'autres termes, compte tenu des contenus de la bibliothèque et des pratiques de lecture des usagers, quelle est la meilleure organisation de ladite bibliothèque, c'est-à-dire celle qui offre la meilleure "affordance numérique" ?

Si nous parvenons à formuler la réponse, alors, à l'aide en particulier des technologies modernes d'apprentissage automatique, nous pourrions organiser en un clin d'œil (ou à peu près) ce qui a pris des siècles aux bibliothèques euclidiennes et qui, quoi qu'en dise notre ami Euclide, est loin d'être optimal (souvenez-vous, le jour, la nuit !). La réponse nous permet de résoudre quatre problèmes cruciaux :

- Identifier les communautés d'affinité, c'est-à-dire les sous-ensembles de membres de la bibliothèque qui sont plus ou moins proches les uns des autres (proches veut dire géométriquement proches) ;
- Identifier la bibliothèque idéale de chacun et recommander à chacun les livres que d'autres qui appartiennent à sa communauté d'affinité ont lus, mais qu'il n'a pas encore découverts ;
- Identifier la structure idéale de la bibliothèque (un rêve à la Borges) dans son ensemble idéal pour chacun d'entre nous, et en tirer des promenades parmi les livres qui permettent à partir d'un livre de dériver "intelligemment" vers d'autres livres, c'est-à-dire de déambuler astucieusement, de naviguer "au long cours" parmi les livres ;
- Enfin, last but not least, bâtir une interface utilisateur qui capture à l'écran en un seul objet la bibliothèque idéale, qui soit telle que, contrairement aux résultats d'une recherche *Google*, tous les objets intelligemment organisés soient visibles d'un coup d'oeil, un seul, et qui permette d'afficher de façon intelligible **tous** les résultats de n'importe quelle recherche en un seul écran.

Le troisième objectif est bien connu des bibliothécaires euclidiens qui investissent beaucoup de leur temps dans la classification et le rangement corollaire des ouvrages. Malheureusement, le résultat de ce travail est statique. Il correspond à cet espace dans lequel Euclide est notre diligent guide. Notre ambition est de faire mieux,

c'est-à-dire d'une pierre deux coups. Si nous découvrons la géométrie virtuelle de la bibliothèque, celle qui remplace la géométrie d'Euclide, alors nous sommes à même, munis de notre nouvelle trousse de géomètre, de calculer la structure idéale de la bibliothèque.

Il en ressort une bibliothèque recomposée sur des bases autres que celles d'Euclide. Dans cette bibliothèque, le plus court chemin d'un livre à un autre n'est plus une droite, une étagère. Pour employer un terme savant, la "géodésique" unissant deux livres, le point de trajet le plus court entre eux, n'est pas droite. Ce n'est plus l'étagère chère à Euclide. Difficile à imaginer ! Le globe terrestre permet une analogie pédagogique utile. Si je place un livre en un point du globe et un autre livre en un autre point, le plus court chemin entre eux ne sera pas celui du vol d'oiseau mais celui de la géodésique passant entre ces deux livres. La géodésique correspond à l'arc de grand cercle, cercle qui épouse la sphère et a le même diamètre qu'elle. Les navigateurs et les aviateurs parlent d'"orthodromie". N'avez-vous jamais remarqué que les avions qui desservent par exemple Los Angeles depuis Paris frôlent quasiment le cercle arctique ? Ils montent très haut vers le Nord. Ils suivent une géodésique, c'est-à-dire un lieu de plus court chemin qui n'est pas une droite. Euclide en aurait le mal de mer (ou de l'air) : les droites sont devenues courbes !

Qui plus est, la bibliothèque devient dynamique. Elle change au gré des visites et des consultations des abonnés. Euclide s'y perd, car ce n'est jamais la même. Elle est en distorsion permanente. Ces distorsions sont provoquées par la communauté d'utilisateurs elle-même. En quelque sorte, et pour reprendre Euclide, les livres sont des points et les distances qui les séparent sont "distordues" par les comportements de la communauté. La géométrie, c'est aussi la communauté. Par un livre passe plus d'une étagère. Un livre peut en effet appartenir à plusieurs étagères simultanément. Imaginez notre brave Euclide courant d'une étagère à l'autre, livres en mains, chaque fois qu'un nouvel usager pénètre dans la bibliothèque de façon à reconfigurer la bibliothèque en fonction des nouvelles informations apportées par le nouvel arrivant. Son équerre ne lui sera pas d'un grand secours. Si Euclide ne veut pas périr d'essoufflement, il doit admettre que sa géométrie n'est plus adaptée. Il doit adopter une nouvelle géométrie.

S'il l'accepte, alors le Général Stumm est sauvé. En pénétrant dans la bibliothèque numérique, même sans son cheval, il se retrouvera rapidement aiguillé par la totalité des visiteurs ayant laissé une trace géométriquement proche de la sienne. Il ira vers la promenade qui lui permettra

d'appréhender globalement la quantité d'informations de la bibliothèque en quelques heures en lieu et place des milliers d'années qu'il prévoyait initialement d'y passer physiquement. L'agencement idéal de la bibliothèque est donc indissociable de l'identification des communautés d'affinités, de leurs traces géométriques. Sans ces traces et malgré son sens de l'orientation militaire, le Général Stumm avouerait sa défaite. Nous sommes donc ramenés au premier problème évoqué plus haut. Quelles sont ces communautés d'affinités ? Comment les "calculer" ? Calculer, le verbe semble bien aride, voire aux antipodes de ce qui fait la relation entre un livre et son lecteur, une bibliothèque et sa/son bibliothécaire. Mais, après tout, lorsque nous étions en compagnie d'Euclide, nous faisons déjà du calcul, un calcul familier, basé sur les postulats élémentaires de sa géométrie, mais un calcul quand même.

La différence, Euclide nous ayant faussé compagnie, réside dans le calcul nouveau que nous devons utiliser. Ce n'est plus celui d'Euclide. Cela n'est pas bien grave car un grand principe demeure que personne ne songera à renier : celui de la passion pour les lecteurs, celui de l'envie de mieux les connaître pour mieux les servir. Il n'est pas rare qu'une bibliothécaire recommande un livre à une lectrice qu'elle a appris à connaître. C'est d'ailleurs en général à ce souci de partage bibliographique que l'on reconnaît les bibliothécaires passionnés. Imaginons maintenant que la bibliothécaire nous indique non seulement une nouvelle lecture, mais aussi (par une alchimie mystérieuse qu'elle seule utilise) un ou plusieurs lecteurs, des gens que nous ne connaissons pas forcément. Ces gens sont (d'après notre bibliothécaire et son alchimie) nos voisins d'affinités, mais nous l'ignorons. Car c'est le travers de toute bibliothèque physique chère à Euclide de ne pas pouvoir réaliser cette alchimie : nous y sommes réduits au silence. Parler dérange les autres et il est fréquent qu'Euclide nous dise "*chut !*". Nous y perdons tous, à titre individuel et à titre collectif. Euclide n'y peut rien et notre bibliothécaire non plus. Non pas qu'elle soit incompétente ! Elle est simplement submergée par l'ampleur de la tâche et la masse d'informations à collecter et à traiter. Tâche qui viendrait s'ajouter à toutes celles qu'une bibliothèque euclidienne requiert quotidiennement. Son alchimie a ses limites !

L'équation **bibliothèque numérique = information** prend une nouvelle fois tout son sens. Nous gérons individuellement dans la bibliothèque numérique des informations qui peuvent être décortiquées dans leur ensemble par ces outils que le brave Euclide n'a osé imaginer. Tout se passe comme si notre bibliothécaire pouvait re-



garder par dessus l'épaule de chacun des lecteurs de sa bibliothèque (c'est en fait cela le début de la vraie alchimie), enregistrer l'information, la retraiter et, enfin, nous inviter à changer de chaise pour nous rapprocher de nos voisins ainsi calculés. On imagine aisément le ballet de chaises, de livres et de lecteurs que cela représenterait. À l'issue de ce ballet, le deuxième problème évoqué plus haut trouverait sa résolution naturelle : chaque lecteur serait à même de découvrir l'étagère d'ouvrages trait d'union de la micro-communauté à laquelle il appartient.

Ce que la bibliothèque euclidienne ne peut facilement se permettre, la bibliothèque numérique le prend en charge : elle organise ce ballet numérique porteur de toutes les lectures, tous les bavardages, de toutes les étagères personnelles et de tous les conseils et commentaires. Cette observation est lourde d'implications techniques, mais riche de conséquences. Lourde parce qu'elle exige des développements mathématiques conséquents et des temps de calculs non négligeables. Certes, les ordinateurs modernes facilitent la tâche. Il n'en reste pas moins que nous sommes dans la position de l'architecte auquel on a confié la mission d'édifier une bibliothèque, avec de surcroît la possibilité de façonner ses propres outils numériques pour la construire.



Fig. 4 : La préhension "Gutenberg".

Riche parce que nous avons de bonnes chances d'en sortir plus savants. Riche surtout parce que la bibliothèque numérique mélange toute l'information statique de ses livres et toute l'information dynamique de ses communautés en un seul espace ambiant numérique, dans lequel s'auto-définissent sa propre géométrie, ses propres interactions, ses propres règles d'évolution et de mélanges communautaires.

Reste pourtant un point sur lequel Euclide pourrait bien prendre sa revanche : l'interface livres-lecteurs. Car, après tout, qu'avons-nous à proposer en lieu et place de ces merveilleux endroits que sont les bibliothèques municipales, universitaires etc. ? Bonne question qui est sans aucun doute la question cruciale qui, faute de réponse

convaincante, mettrait à bas tout notre effort géométrique. C'est le quatrième problème que nous avons identifié ci-dessus : celui de l'interface. Et, c'est précisément cette interface que nous avons baptisée : *The (smart) DICE (Digital Content Exploration)*.

## **THE (smart) DICE : comment arpenter globalement et intelligemment toute bibliothèque numérique**

L'avènement des tablettes tactiles de type *iPad* n'est pas anodin. Avec ces tablettes, nous récupérerons l'usage de nos mains et de nos doigts. La bibliothèque numérique que l'on consulte à partir de ce type de tablettes devient réellement bibliothèque "digitale". On retrouve la préhension si caractéristique de l'usage du livre papier : nos pouces opposés aux autres doigts permettent de saisir confortablement la tablette. Cette préhension nous assure une mobilité inégalée et c'est sans doute ce qui explique, d'après les études récentes, le fort usage domestique des tablettes. L'utilisateur "fait corps" avec la tablette et peut ainsi l'utiliser où bon lui semble<sup>3</sup>. Certes, il en va de même du livre papier que nous pouvons lire en tout lieu. Cette lecture est usuellement une lecture "d'immersion". Le lecteur est concentré sur son livre : pour employer un terme de linguistique cognitive, le livre est le focus, l'objet de l'attention. Lire *Harry Potter* ou *L'Homme sans qualités* requiert indubitablement une concentration, un temps durant lequel le lecteur est seul, livre en mains, absorbé par le texte. Pour reprendre la description de Nicholas Carr :

*"Le sentiment de complétude est ce qui fait de la lecture d'un livre papier une expérience si satisfaisante. La nécessité de cette complétude pousse l'écrivain dans ses derniers retranchements. Cette complétude doit se sentir de la première de couverture à la quatrième de couverture"*<sup>4</sup>.

C'est précisément ces dimensions d'autonomie et de complétude qui font de l'écriture et de la lecture du livre Gutenberg des expériences uniques. Les frontières du livre que nous saisissons entre nos mains définissent le principe même d'une expérience riche et pourtant singulièrement fermée.

Cette fermeture est un paradoxe à l'heure d'un web ultra-connecté, hyper-relié qui semble faire de chacun de nous non plus des fourmis du contenu, mais des cigales de l'hyperlien. L'idée d'un espace fermé est antinomique du web comme le rappelle Nicholas Carr :

*"L'idée de bords, de limites est antithétique du web qui dissout les frontières. En effet, la force et*

*l'utilité du web en tant que technologie provient de la façon dont toute forme autonome est détruite pour faire partie d'un tout plus grand, en permanente mutation, sans forme précise. Le web n'est pas un assemblage de choses, c'est un assemblage d'extraits, de bribes et de morceaux. Un livre électronique est donc un oxymore. Mettre le texte d'un livre en ligne revient à orchestrer une collision entre deux forces technologiques et esthétiques contradictoires. L'une des deux doit l'emporter. Ou bien le web acquiert des limites, ou bien le livre perd les siennes*"<sup>5</sup>.

Le web s'inscrit dans cette tension permanente et brouillonne entre le focus et le contexte au terme de laquelle l'important n'est apparemment plus de "savoir" mais de "savoir où cela se trouve". C'est cette tension qu'*Amazon* tente aujourd'hui d'exploiter avec son *Kindle Fire* et son système *X-Ray*. Pour faire simple, chaque livre électronique acheté sur *Amazon* "embarque" un ensemble d'informations (par exemple *Wikipedia*) liées au contenu du livre ou plus exactement aux phrases qu'*Amazon* juge intéressantes<sup>6</sup>. Le lecteur peut alors s'évader du texte pour approfondir dans *Wikipedia* ou ailleurs les phrases ou les mots intéressants du texte. Mais, là encore, il s'agit d'un arpentage local. Le texte est arpenté au moyen des ressources jugées adéquates par *Amazon*. Quel que soit le jugement que l'on porte sur ces assauts portés au livre et à son auteur, il nous semble une fois de plus que cet arpentage local est étrié et procède d'une mauvaise interprétation de la tension focus-contexte.

Cette tension est paradoxalement bien mieux maîtrisée dans une bibliothèque municipale ou dans une librairie. Le livre, objet du focus, est en contexte. Il a des voisins proches et lointains, visibles à l'œil nu, organisés selon les préceptes d'Euclide. La bibliothèque et la librairie sont les lieux du contexte et, dans la mesure où nous pouvons y lire des livres, elles sont aussi le lieu du focus. Mais, il s'agit d'un contexte bien particulier, ordonné selon les règles de l'art du bibliothécaire et du libraire. Comme nous l'avons vu, la bibliothèque numérique procède d'un autre art (ou plutôt d'une autre science), celui de la géométrie de l'information. En capturant l'information (sur les livres et surtout sur les relations des lecteurs aux livres), la bibliothèque numérique définit son architecture et la dynamique de celle-ci. La tension focus-contexte se résout alors naturellement dès lors qu'elle exprime le passage "sans heurt" du livre (focus) vers la bibliothèque numérique (son contexte) et vice-versa.

La résolution de cette tension s'exprime par une interface que nous avons baptisée *The Dice* (*Digital Content Exploration*). *The Dice* correspond à

ce souci de restituer à chaque lecteur à tout moment en toute continuité une image **locale** et une image **globale** du livre qu'il est en train de lire. L'idée est de pouvoir passer progressivement du livre lu à sa position dans la totalité de la bibliothèque numérique. C'est un peu l'expérience qu'offre un planétarium lorsque partant d'une planète, il montre comment celle-ci se situe dans l'univers céleste. Que l'on soit enfant ou adulte, l'expérience du planétarium est émouvante et unique : focus et contexte y font bon ménage. Les frontières de chaque planète se dissolvent pour laisser place à un espace qui coupe le souffle et invite à la méditation, au recul. Mais, ce n'est pas pour autant que les planètes disparaissent ou qu'elles en deviennent moins intéressantes. Bien au contraire, chacune dans sa singularité prend un relief particulier. Aucune n'est diluée, chacune fait l'objet d'interrogations particulières.

Il en va des livres et des bibliothèques numériques comme des planètes et du planétarium. Faut-il d'ailleurs rappeler combien astronomie et géométrie s'accordent ! C'est cet émerveillement intelligent que *The Dice* veut relayer. Certes, il ne s'agit plus de la géométrie de corps célestes mais d'une géométrie (de l'information) sociale. L'ambition de *The Dice*, c'est-à-dire de l'interface qu'il représente, est "simple" :

- Être intuitif, facile d'utilisation et ludique ;
- Situer tout livre dans la bibliothèque numérique préalablement calculée (agencée) en tirant parti des informations données par les livres et les lecteurs ;
- Visualiser **toute** la bibliothèque, c'est-à-dire tous les livres sans exception en un seul espace ;

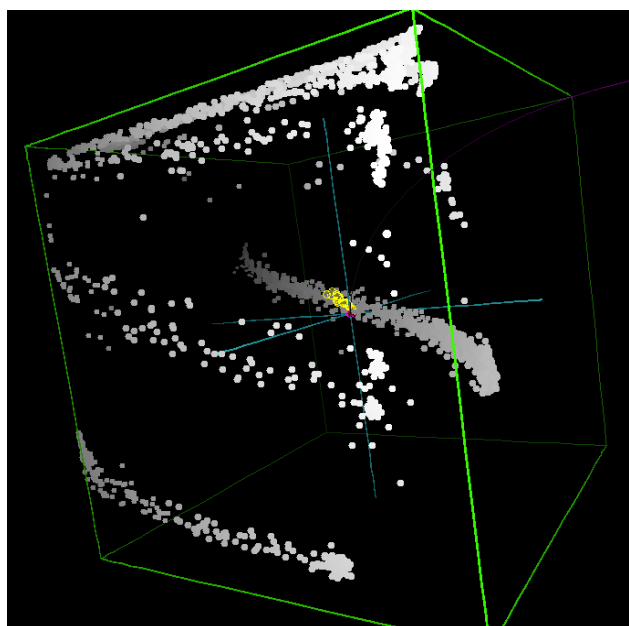


Fig. 5 : Toute la bibliothèque numérique est contenue dans le Dice.

Chaque livre occupe une position issue de l'analyse des contenus et des habitudes de lecture.



- Suggérer des parcours complets et intelligents (au sens de la géométrie de l'information) de lecture à partir de n'importe quel livre ;
- Permettre de lancer une recherche dans la bibliothèque (par exemple en texte intégral) dont, premièrement, tous les résultats (livres) sont visibles à l'œil nu, en une seule entité (contrairement à *Google*) et, deuxièmement, tous les résultats (livres) sont situés dans leur voisinage aux autres livres qui, bien que ne contenant pas les mots-clés recherchés, leur sont voisins au sens des habitudes de lecture.

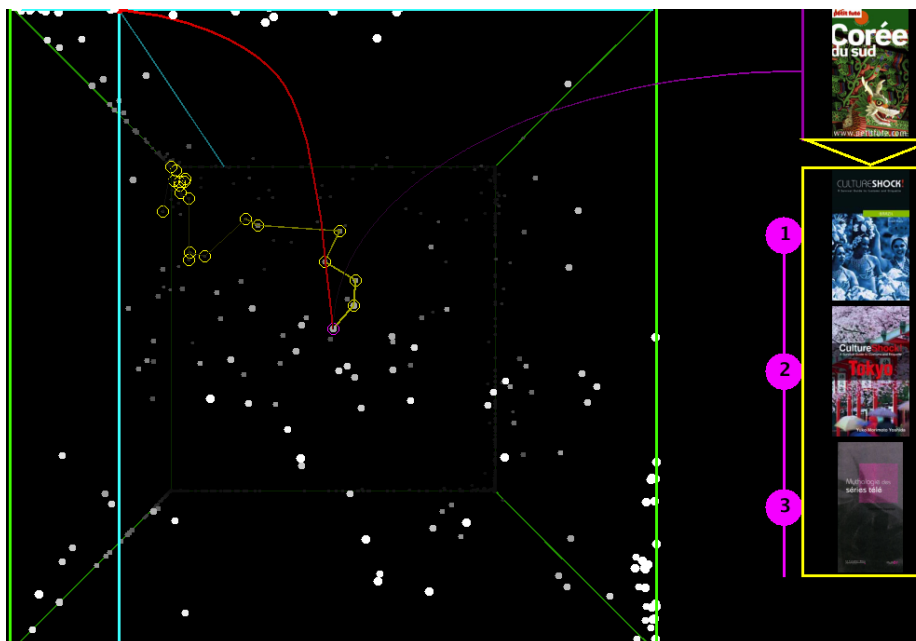


Fig. 6 : Le plus court chemin d'un livre à un autre n'est plus la droite.

Afin de réussir ce quadruple pari, nous avons défini une interface originale. Dit plus simplement, toute la bibliothèque numérique issue des livres et des lecteurs est contenue dans un cube. Elle s'affiche donc en trois dimensions comme le montre la figure 5.

Chaque point à l'intérieur du cube est un livre de la bibliothèque. Le point central du cube à l'intersection des deux droites bleues est le livre que l'on est en train de lire. Les points jaunes figurent un parcours intelligent de lecture possible à partir de ce livre. Au voisinage du livre central (focus), la bibliothèque se comporte comme si elle était euclidienne. Au fur et à mesure que l'on s'éloigne du livre central et que l'on se déplace vers les bords du cube, l'approximation euclidienne n'est plus valable. Si on la maintenant, la bibliothèque ne pourrait être contenue intégralement dans le cube. On retrouve ici l'idée chère à Gutenberg et à Nicholas Carr d'un espace "self-contained" qui, à l'instar du livre papier, offre au lecteur une expérience dense, homogène et cohérente : le livre est dans la continuité de la bibliothèque, la bibliothèque est dans

la continuité du livre. Ce caractère compact est rendu possible par la mise en correspondance mathématique de l'espace initial avec le nouvel espace défini par le cube. L'utilisateur a ainsi l'impression de se promener de façon "euclidienne" dans le cube alors même que le plus court chemin entre deux livres dans le cube n'est plus une droite mais une courbe (dite géodésique).

Dans la figure 6, la courbe rouge illustre ce parcours le plus rapide vers le guide sur la Corée du Sud depuis le livre de départ.

On perçoit également tout l'intérêt du *Dice* dans la visualisation de résultats de recherche. On sait combien les recherches à l'aide de *Google* peuvent être frustrantes. Des centaines de pages s'affichent et la plupart du temps nous n'en consultons que les trois premières, avec le risque de manquer les résultats qui nous intéressaient. *The Dice* donne des dimensions complètement nouvelles à la recherche d'informations numériques comme en témoigne la figure 7. Premièrement, tous les résultats sont visibles d'un bloc : il s'agit du parcours jaune

que l'on peut visualiser à son aise et avec un gain de temps et donc de productivité remarquable. Deuxièmement, cette recherche (notre focus) est mise en contexte. Au voisinage des points jaunes se trouvent des points blancs qui peuvent s'avérer tout aussi intéressants que lesdits points jaunes. Certes, les points blancs ne correspondent pas à la recherche stricto sensu (texte intégral, mots-clés), mais leur voisinage des points jaunes signifie qu'ils sont à même d'apporter un enrichissement au lecteur, enrichissement que seule l'analyse des comportements de lecture a permis de mettre à jour. Enfin, la recherche revêt un caractère ludique et on imagine aisément combien *The Dice* est adapté à la géométrie et à l'ergonomie d'une ardoise tactile<sup>7</sup>. Jamais le passage du livre à son contexte bibliothécaire n'a été aussi simple et aussi riche. Enfin, avec *The Dice*, la "long tail"<sup>8</sup> chère à Chris Anderson devient aussi visible que la "short tail" ! Certes, un livre peut ne jamais être consulté, mais les métadonnées qui lui sont associées l'apparient à d'autres livres qui eux sont consultés. Ainsi, la sagesse collective rejoint-elle la sagesse d'autorité (des métadonnées).

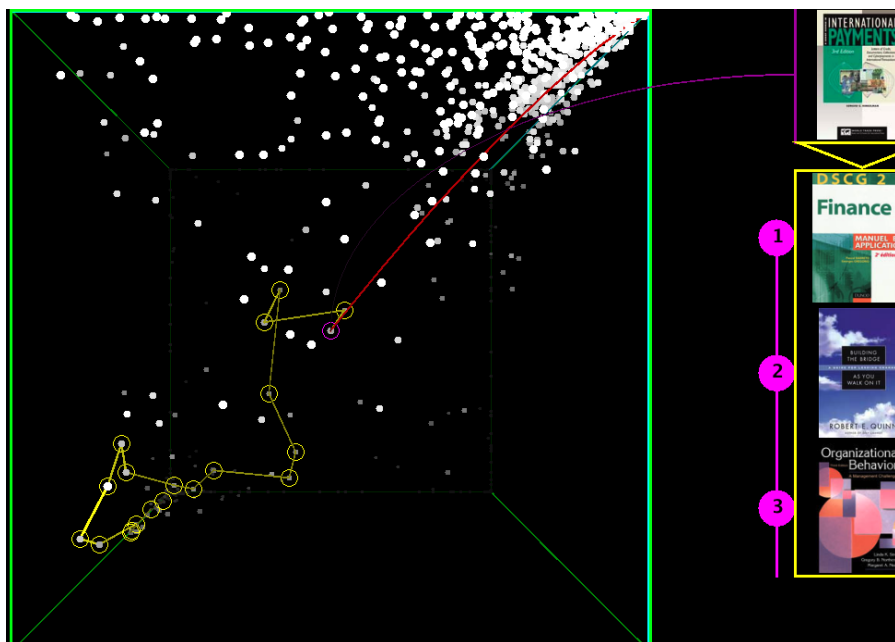


Fig. 7 : En jaune, l'affichage de **tous** les résultats d'une recherche au sein de la bibliothèque.  
Chaque résultat est mis en contexte.

### Let's dice !

On notera que lire et relire sont les anagrammes de lier et relier. Dans une bibliothèque numérique, la lecture (connectée) des livres crée une valeur informationnelle en jetant des ponts entre livres, entre lecteur et entre livres et lecteurs. Les livres et les lecteurs sont (re)liés par le truchement de leurs contenus et des habitudes de lecture. L'enjeu, pour reprendre la belle expression de Frédéric Kaplan<sup>9</sup>, est "d'extraire, transformer et valoriser le minerai biographique". Il ajoute avec force pertinence que "trouver la structure sous-jacente à une série d'observations permet de la réécrire de manière plus compacte". *The Dice* est la partie émergée de ce (d)iceberg qui à l'instar du feuilletage d'un livre permet de feuilleter intelligemment (toute) la bibliothèque (celle que nos yeux euclidiens ne voient normalement pas).

*The Dice* rend l'invisible visible, intelligible et intelligent. Le focus est mis en contexte. Ce contexte est tissé de focus intelligemment disposés dont la totalité n'est autre que la bibliothèque

numérique elle-même. De cette intelligence naissent des parcours de lecture, comme autant de sentiers de grande randonnée dont l'emprunt est garant de belles et enrichissantes escapades. Tout cet édifice est rendu possible par un arpentage global de l'information, arpentage qui montre si besoin était combien les techniques d'arpentage local sont réductrices.

De fait, la philosophie du *Dice*, si bien incarnée par la bibliothèque numérique, embrasse un domaine qui dépasse largement celui des livres numériques. Musique, cinéma, annonces immobilières, destinations touristiques, etc., tout

contenu numérique, dès lors que l'on dispose des données appropriées, peut être "dicé". Et, il paraît évident qu'avec l'avènement des tablettes tactiles, on peut anticiper l'émergence d'une nouvelle génération de navigateurs web dont *The Dice* pourrait bien être la matrice.

**Éric Briys**

*Cyberlibris*

95, rue d'Amsterdam

75008 Paris

France

eric.b@cyberlibris.com

**Richard Nock**

*Ceregmia*

*Université des Antilles et de la Guyane*

Campus de Schoelcher

BP 7209

97275 Schoelcher Cedex

France

rnock@martinique.univ-ag.fr

*Novembre 2011*

### Références

Manguel, Alberto. *La bibliothèque, la nuit*. Éditions Actes Sud, 2006.

Saint-Exupéry, Antoine de. *Le Petit Prince*. Reynald et Hitchcock, 1943.

Anderson, Chris. *The Long Tail*. Hyperion, 2006.

Glissant, Édouard, *Traité du tout-monde. Poétique IV*. Gallimard, 1997.

Euclide; traduction de Henrion, Didier. *Les quinze livres des éléments géométriques*, 1632. disponible sur Gallica : <http://gallica.bnf.fr/ark:/12148/bpt6k68013g.image.r=Les+quinze+livres+des+%C3%A9l%C3%A9ments+g%C3%A9om%C3%A9triques+d'Euclide.f3.langFR> (consulté le 30 novembre 2011)

Bastiat, Frédéric. *Ce qu'on voit et ce qu'on ne voit pas*. Romillat, 1993.

Kaplan, Frédéric. *La métamorphose des objets*. FYP Éditions, 2009.

Borges, Jorge Luis. *Fictions*. Gallimard, 1991.

Carr, Nicholas. *The Shallows : What Internet is doing to our brains*. W.W. Norton, 2010.

Musil, Robert, *L'homme sans qualités*. Points Seuil, 1982.

## Notes

- 1 **DICE** : Digital Content Exploration. *The Dice* est une application propriétaire conjointement développée par le Professeur Richard Nock et Cyberlibris.
- 2 La sérendipité est la capacité, en réponse à une observation surprenante, de découvrir, d'inventer ou de créer ce qui n'était pas initialement recherché.
- 3 Bret Victor, ancien d'Apple, est plus réservé quant à ce retour des mains : Victor, Bret. *A brief rant on the future of interaction design* [en ligne]. <http://worrydream.com/ABriefRantOnTheFutureOfInteractionDesign/> (consulté le 30 novembre 2011).
- 4 Nick. The remains of the book. *Rough Type* [en ligne], 30 septembre 2011 (consulté le 30 novembre 2011). [http://www.roughtype.com/archives/2011/09/the\\_seethrough\\_1.php](http://www.roughtype.com/archives/2011/09/the_seethrough_1.php)  
"The sense of self-containment is what makes a good book so satisfying to its readers, and the requirement of self-containment is what spurs the writer to the highest levels of literary achievement. The book must feel complete between its edges".
- 5 Voir note 3.  
"The idea of edges, of separateness, is antithetical to the web, which as a hypermedium dissolves all boundaries, renders implicit connections explicit. Indeed, much of the power and usefulness of the web as a technology derives from the way it destroys all forms of containment and turns everything it subsumes into a part of a greater, ever shifting, amorphous whole. The web is an assembly not of things but of shards, of snippets, of bits and pieces. An electronic book is therefore a contradiction in terms. To move the words of a book onto the screen of a networked computer is to engineer a collision between two contradictory technological, and aesthetic, forces. Something's got to give. Either the web gains edges, or the book loses them".
- 6 Intéressantes à quel titre ? Celui du cybermarchand...
- 7 Nota bene : Les figures illustrent la version bêta du *Dice*. Nous travaillons à l'heure actuelle activement à son design final.
- 8 La fameuse "longue traîne". (<http://www.wired.com/wired/archive/12.10/tail.html>)
- 9 Kaplan, Frédéric. *La métamorphose des objets*. FYP, 2009.