



Partager de la connaissance sur Internet

Jean-François Abramatic¹



Robert Cailliau, Jean-François Abramatic et Tim Berners-Lee fêtent les 10 ans du W3C à Sophia-Antipolis.

Bonsoir, voici donc la dernière présentation de cette soirée. L'exposé va porter sur le Web. J'ai la chance d'avoir travaillé sur une application qui tourne au-dessus de l'internet, mais qui pourrait tourner au-dessus de RINA, ou au-dessus d'un réseau plus sécurisé, etc. J'ai la chance de parler d'un sujet qui est indépendant du réseau grâce au design original de l'internet et de la forme en *hourglass*², d'une couche de standards qui sépare les applications de l'infrastructure.

Je vais parler des connaissances et de la manière dont on partage les connaissances sur

l'internet. Comme nous sommes réunis aujourd'hui pour parler d'histoire et que je suis de la génération du Web design, j'ai essayé de suivre des liens pour savoir à quel moment j'allais commencer le récit. J'aurais pu parler de la communication orale, de la communication écrite, puis de l'internet. Mais cela nous aurait amené trop loin. Il est tard. J'ai choisi de partir de 1849. La première raison est que c'est cent ans avant ma naissance, c'est un compte rond, j'aime bien ça. La seconde raison est qu'une des leçons de la ruée vers l'or est que les marchands d'outils ont fait plus d'argent que ceux qui ont cherché de l'or. En d'autres termes, pour gagner sa vie, il valait mieux vendre des pelles et des pioches que d'essayer de faire fortune avec l'or que l'on avait trouvé.

1. Informaticien, pionnier du Web.

2. En sablier, comme on dit en anglais.

« Gov Stanford » : communications for the 1900's

1824-1893



First Transcontinental Railroad
completed in 1869: « Last Spike »



Leland Stanford Junior University
founded in 1885: « Main Quad »

Une personne a très bien fait cela. Vous en avez déjà entendu parler dans l'exposé de Gérard Le Lann : c'est Leland Stanford. Stanford a fait, une première fois, fortune en vendant des pelles et des pioches. Il a ensuite investi l'argent gagné dans une compagnie ferroviaire avec trois autres collègues. L'image du haut ci-dessus est l'image où il plante le dernier clou de la liaison transcontinentale entre l'*Union Pacific* qui vient de l'est et la *Central Pacific* qui est celle dont il est l'actionnaire et le président-directeur général. Il a gagné beaucoup d'argent en faisant ça, c'est la chance de sa vie. La partie la plus triste est qu'il n'a eu qu'un seul enfant, un fils, qui est mort jeune. Stanford et son épouse ont décidé d'investir leur fortune pour créer une université, qui est devenue l'université de Stanford.

En essayant de choisir une date à partir de laquelle j'allais partir, j'ai trouvé que finalement Stanford avait contribué, à la fin du XIX^e siècle, à construire les éléments de communication du XX^e siècle : tout d'abord le train, puisqu'à l'époque il n'y avait pas de réseau de communication et qu'il fallait donc transporter les publications, les livres pour partager les connaissances, et puis évidemment une université pour enrichir et transmettre la connaissance. Cette université est devenue une des meilleures au monde. J'ai trouvé que c'était une bonne introduction pour le reste de l'exposé.

Un demi-siècle et deux guerres mondiales, voici le papier que le monde du Web considère comme étant vraiment le document « initial ». Pour beaucoup des acteurs du Web, si jamais il y a un papier original, c'est « *As we may think* », écrit par Vannevar Bush au lendemain de la Seconde Guerre mondiale. En 1945, Vannevar Bush est, pour faire simple, ministre de la recherche aux États-Unis. Il l'a été pendant la guerre. Il a animé les équipes d'ingénieurs qui ont développé les choses qui étaient nécessaires pour faire la guerre. Et puis là, on est en 1945 et il se pose la question de savoir ce qu'ils vont devoir faire maintenant. Son message est de dire que les hommes de science doivent désormais se tourner vers la tâche énorme de rendre la

connaissance facilement accessible les uns aux autres. C'est une chose de faire des marteaux, des microscopes, des objets, des équipements, on en a fait depuis toujours. Mais il va falloir aussi développer des instruments pour partager les connaissances, pour représenter l'héritage qui nous a été transmis par ceux qui sont venus avant nous. Ce document est considéré comme inaugural car, si vous lisez cet article, vous verrez qu'il propose de développer une machine (la machine MEMEX) qui présente des concepts semblables à ceux des hyperliens que l'on va trouver dans le Web cinquante ans plus tard, tout cela alors que le monde numérique dans lequel nous vivons aujourd'hui n'existe pas. L'invention du transistor n'a pas encore eu lieu. L'invention du transistor date de 1947. Le support de l'imagination de Vannevar Bush est donc complément analogique. Il reste que les concepts sont là et vont être repris par d'autres ensuite. Évidemment la machine MEMEX ne verra jamais le jour. Mais le papier, lui, restera une inspiration pour ceux qui viendront ensuite.

J'ai essayé de faire un exposé qui n'est pas seulement technique, mais qui essaie aussi de comprendre la dynamique et les relations entre les personnes qui nous ont amenés là où nous en sommes. Frederick Terman est bien connu des gens de la *Silicon Valley*. Il était un étudiant de Vannevar Bush. Donc là le lien, c'est que Frederick Terman est un étudiant de Vanne-



var Bush, étudiant en électronique et en *radio engineering*. Il aura plus tard quelques étudiants célèbres. J'imagine que vous connaissez au moins Hewlett et Packard. Et peut-être un peu moins Russell & Sigurd Varian. Hewlett et Packard créeront la première start-up issue de Stanford, dans un garage à Palo Alto. Décidément, on aura souvent parlé de Palo Alto dans cette soirée. Et David Packard jouera aussi un rôle après la guerre, dans l'évolution de Stanford, il sera au *board* de l'université auprès de Frederick Terman et ils participeront à la création du *Stanford Research Institute*, du *Stanford Research Park*. On dit souvent de Frederick Terman qu'il est le père de la *Silicon Valley*. De plus, c'est quelqu'un qui, certes, était un scientifique, un ingénieur, mais il a aussi été un professeur (il a eu des étudiants célèbres) et puis il a été administrateur de Stanford et du parc industriel autour de Stanford. Il est avec quelqu'un dont Valérie Schafer vous a déjà parlé, Doug Engelbart.

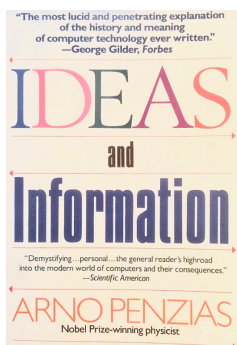


En 1968, Doug Engelbart est un des responsables du *Stanford Research Institute* que Terman a contribué à créer. Doug Engelbart fait la *Mother Of All Demos*, dont Valérie vous a parlé tout à l'heure. Cette démo a lieu lors d'une conférence ACM-IEEE à San Francisco, comme il y en a eu beaucoup depuis et il y en aura beaucoup encore dans l'avenir. Le caractère exceptionnel tient dans le fait que la démo dure une

heure et demie, ce qui est exceptionnel à ce moment-là (et qui serait aussi incroyable aujourd'hui de toute façon), mais la deuxième chose est que la démo est complètement enregistrée, alors que nous sommes en 1968. Aujourd'hui, vous pouvez aller sur YouTube et chercher « *Mother Of All Demos* », vous obtenez le lien vers l'ensemble de la démo. Elle est aussi coupée en petits morceaux. Si vous n'avez pas envie de consacrer une heure et demie dans la même soirée, vous pouvez vous la passer sous forme d'une série, vingt minutes par vingt minutes... Cette démo a été développée au *Stanford Research Institute* et elle démontre le système que Doug a appelé NLS pour *oNLine System*. Comme Valérie l'a dit, elle présente l'ensemble des concepts de l'informatique individuelle que l'on utilise aujourd'hui, y compris l'infrastructure logicielle comprenant les documents structurés, les liens hypertextes, et pas seulement la souris qui a rendu célèbre Engelbart qui en est l'inventeur. C'est la première démo en public qui utilise une souris, à l'époque faite en bois. Il y aussi une démonstration de vidéo-conférence : alors qu'il est à San Francisco, Doug engage un travail coopératif avec un collègue qui est à Menlo Park et l'on voit apparaître le curseur de la souris de l'autre personne, qui est donc à quelques dizaines de kilomètres, au moment où c'est lui qui prend le contrôle de l'application. Je vous encourage vraiment à regarder cette démo... morceau par morceau si c'est trop long pour vous. C'est surtout le fait de se souvenir qu'elle a été faite en 1968 qui à mon avis fait la différence. J'ajoute, s'agissant de Doug, que j'ai eu la chance de le rencontrer quand j'étais président du W3C (il était à la fin de sa carrière et il est venu nous voir à Boston). Comme je dis souvent à mes enfants et petits-enfants « *les meilleurs sont souvent les plus humbles* ». Je me souviens de ce déjeuner avec Tim Berners-Lee et Doug Engelbart dans une brasserie à côté du MIT. Doug parlait d'une voix douce, et il se montrait admiratif des travaux que l'on était en train de faire alors qu'il en était à l'origine. Vraiment, on ne peut pas oublier ce genre de moment.

Enfin, il faut ajouter la contribution de Doug à la naissance de l'internet. Nous avons vu, dans l'exposé de Gérard Le Lann, que SRI faisait partie des quatre nœuds de la démo de 1969. Un des *Network Information Centers* était dans les locaux du SRI sous la responsabilité de Doug. Enfin, pour la petite histoire, quelques années plus tard, sur le parc industriel de Stanford, Xerox installera le PARC (*Palo Alto Research Center*), qui est devenu pour ceux qui ont des cheveux blancs, un endroit mythique, où la machine Alto, l'ancêtre des ordinateurs individuels, sera développée en 1972. La démo d'Engelbart précède ainsi de trois ans la disponibilité de la machine Alto et la suite de machines qui seront développées à Xerox PARC.

L'avantage d'être parti en 1849, c'est que je peux faire des grands sauts dans le temps en passant d'un transparent à un autre. Là, c'est un petit peu personnel. En 1989, je travaillais à développer des terminaux graphiques, qui, à l'époque, étaient connectés à un réseau local Ethernet. J'étais au cœur de l'impact de l'informatique individuelle ; en l'occurrence, dans mon cas, le travail des ingénieurs.

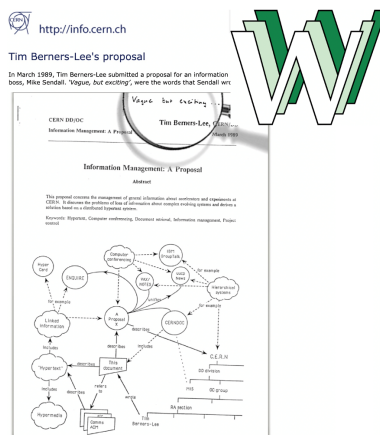


En 1989, est paru ce livre d'Arno Penzias : *Ideas and Information*. Arno est un prix Nobel de physique qui a été vice-président de *Bell Labs*. Il décrit dans ce livre comment l'informatique a changé, pour lui, la manière de faire son travail de recherche en radioastronomie. Il explique, par exemple, que pour mettre en œuvre la première manip' qu'il a faite pour faire sa thèse, il fallait mettre en correspondance des signaux. Il représentait ces signaux sur des bouts de papier et il les mettait en correspondance les uns avec les autres, pour essayer d'identifier des corrélations entre signaux émis depuis l'espace. Il explique que la même manip', au moment où il écrit le livre en 1989, deviendrait tri-

viale puisque les signaux seraient numériques et la mise en correspondance se ferait par un programme lui-même trivial. Donc, Penzias explique à quel point, en 1989, l'informatique a déjà changé la manière de faire la science. Et en fait, cela a attiré mon attention car vous vous souvenez sans doute qu'en 1989, la même année, comme Valérie l'a rappelé, le Web a été inventé. Nous avons fêté les trente ans du Web au printemps dernier au CERN à Genève.

En 1989 donc, alors que Penzias est radioastronome, Tim Berners-Lee travaille au CERN (Centre de recherche en physique des hautes énergies). Ni l'un ni l'autre n'est au *Computer Science Department* de Stanford ou à Inria ou dans un laboratoire d'informatique. Ils ont une vision de l'informatique, ça c'est clair, mais ils sont dans un univers complètement applicatif. Penzias mentionne le rôle des réseaux mais il n'a pas la vision que Tim va avoir en faisant la proposition ci-contre : « *Information Management : a Proposal* ». Il s'agit de la proposition faite à la direction du CERN. Le commentaire de son chef est : « *vague but exciting* ». À l'époque, Tim a une

vision, et il va développer le World Wide Web. Ce logo vert et blanc est le premier logo que Tim va utiliser pour symboliser son travail. En 1989, il fait cette proposition. En 1991, le papier qu'il présente à la conférence Hypertexte est refusé comme étant trop trivial, et donc il a juste droit à un poster – pour ceux qui font de la recherche vous connaissez la différence entre avoir une publication et un poster. C'est ainsi qu'il fait la première démo publique du Web dans les couloirs de la conférence, il n'a pas le droit à être dans l'amphi. La question importante que l'on doit



se poser est la suivante : pourquoi est-ce que le comité de programme de la conférence – je ne suis pas allé leur demander, donc ce n'est que mon interprétation que je partage avec vous – pourquoi est-ce que le comité de programme n'accepte pas le papier ? Eh bien c'est assez simple : quand on veut utiliser des concepts qui ont été développés pour fonctionner sur un ordinateur individuel et que l'on a la vision (ou l'arrogance ou l'ambition) de les faire fonctionner sur un réseau mondial tel que l'internet, on choisit la version la plus simple. HTTP est un protocole simple, donc les gens des protocoles disent « bof ». S'agissant d'HTML, pour une conférence hypertexte, là c'est « *brain damaged* », c'est un jouet d'enfant. Le langage lui-même est d'une simplicité telle que l'on ne va quand même pas accepter un papier de recherche pour quelque chose qui est d'un tel niveau de simplicité. Quoi qu'il en soit, la suite montrera que Tim avait quand même raison. À l'adresse <https://news.netcraft.com/archives/category/web-server-survey/>, on peut voir l'évolution de l'usage du Web. En octobre, c'est ce mois-ci, il y a eu un milliard trois cent millions etc. de sites identifiés. Et surtout, encore aujourd'hui, une croissance de neuf millions par rapport au mois précédent, par rapport au mois de septembre. De toute façon, vous vous en servez, vous le savez, le Web se développe tous les jours. Mais sachez que l'on est encore dans des ordres de grandeurs de croissance toujours très importants.

Le Web est la seule chose sur laquelle j'ai décidé d'avoir deux transparents... C'est le sujet où j'en sais le plus. Sur le transparent qui vient, après 1989, la proposition, 1991 la démo à San Antonio, nous sommes en 1994 pour la création de W3C, le *World Wide Web Consortium*. Là, il faut s'arrêter un petit moment. Ce n'est pas de la technique, mais à l'occasion d'un anniversaire je pense qu'il faut prendre un peu de recul et élargir la réflexion. Tim aurait pu créer une start-up. D'ailleurs, en ce moment, il en a créé une, à 60 ans, je pourrais dire un mot dessus si vous voulez, après. Mais à ce moment-là, en 1994, il décide de faire autre chose, c'est-à-dire de créer un consortium. Début 1994, je pense que l'on a eu un peu de chance, que j'ai eu un peu de chance : le président du CERN est Hubert Curien, et le président de Inria est Alain Bensoussan. Évidemment, les deux présidents se connaissent, Hubert Curien a été ministre de la recherche. Le navigateur Mosaic, qui a permis le déploiement du Web, est disponible depuis 1993. Dans mon cas par exemple, j'étais très copain avec Gilles Kahn et un soir au téléphone, depuis Sophia-Antipolis, il me dit « *Tu devrais faire FTP de Mosaic pour voir ce que c'est* ». Je fais le *download* dans les semaines et mois qui suivent, on se met à essayer de comprendre ce que l'on pourrait faire sur le Web. Donc l'affaire prend son essor et Alain Bensoussan décide d'inviter Curien, Berners-Lee et Cailliau à venir à Rocquencourt au printemps 1994. Là aussi, c'est un souvenir que l'on ne peut pas oublier. On se retrouve autour de la table, Gilles, Alain, Vincent Quint qui avait développé un éditeur de documents structurés qui pouvait être utilisé pour enrichir le Web ainsi qu'Hubert, Tim et Robert. Comme

j'avais participé auparavant au consortium X Window, je savais comment fonctionnait un consortium. Et Tim m'explique – on est en mars 1994 – qu'il a l'intention de créer un consortium, qu'il a rencontré Michael Dertouzos, le directeur du laboratoire d'informatique du MIT, et qu'il pense s'inspirer de ce que Bob Scheifler avait fait quelques années avant sur le standard X Window. La conversation s'engage immédiatement même si c'était la première fois que je rencontrais Tim. Il ajoute : *« Il y a une différence : je souhaite que le consortium soit mondial. Certes, il sera au MIT car le MIT a l'expérience d'avoir fait le premier, mais je veux qu'il ait un pied en Europe et un pied en Asie. »* À ce moment-là, c'est le CERN qui est le candidat évident, pour devenir l'hôte européen du consortium. On se quitte comme ça. Une semaine plus tard, Tim m'envoie un mail pour demander à me voir de nouveau. Il revient seul cette fois-ci, en me disant *« Je ne suis pas sûr que le CERN va le faire donc je veux avoir un plan B »*. Inria devient donc un plan B. Je pense que l'on est convaincant comme la suite de l'histoire le montre.

Fin 1994, le CERN reçoit de ses États membres la mission de faire le *Large Hadron Collider* (LHC qu'ils ont fait depuis), qui a permis de découvrir le boson de Higgs et d'engendrer quelques prix Nobel supplémentaires. Évidemment, le CERN se voit attribuer un budget inférieur à celui qui a été demandé. Le directeur du CERN décide alors qu'il est hors de question de faire autre chose que le LHC. Toutes les activités périphériques sont arrêtées. Quelques jours après Noël 1994, Alain Bensoussan reçoit un coup de fil du CERN ou de la commission européenne qui lui dit qu'il faut un plan B, et comme on avait préparé le plan B, Inria a été choisi. C'est ainsi qu'en janvier 1995, Inria est devenu l'hôte européen du W3C. Quelques années plus tard, Bernard Larrousturou décidera de transférer cette responsabilité à ERCIM qui est, lui-même, un consortium de centres de recherches européens dont Inria est membre bien entendu. S'agissant de l'université de Keio, c'est moi (je serai à ce moment-là président de W3C), qui irai négocier l'accord avec Keio pour qu'un hôte du W3C soit ouvert au Japon, et plusieurs années plus tard, il y aura également un hôte en Chine.

Un consortium qu'est-ce que c'est ? C'est, en premier lieu, une collection d'acteurs, à peu près 450 aujourd'hui. Qu'est-ce que ça produit ? Des recommandations, pas 8 000 comme pour l'internet comme on l'a dit tout à l'heure, seulement 300 pour le Web. Ça s'intéresse à des questions génériques comme l'accessibilité en créant la *Web Accessibility Initiative*, évidemment ça s'intéresse à l'aspect international du Web. Plus récemment, le W3C s'est orienté vers des secteurs verticaux – on continue à monter dans les couches – tels que le paiement, l'automobile, la publication numérique, les aspects de décentralisation, de net-neutralité, de design par le bas, d'universalité – le fait que le Web puisse tourner partout –, et le fait que le consortium prend ses décisions sur la base du consensus. Sur ce dernier point, cela a été l'essentiel de ma responsabilité. J'étais président, donc quand il y avait des tensions

entre les membres, des désaccords, il fallait comprendre les pour et les contre. Il nous arrivait souvent que Tim parle à une entreprise et moi à une autre parce qu'elles étaient en tension. On jouait le rôle d'intermédiaires pour faire que, finalement, on obtienne le consensus sur les standards du Web. Le W3C développe également des recommandations pour les composants du Web : URL, URI, HTTP, HTML, CSS, XML, RDF, OWL, WCAG, etc. URL et HTTP ont continué à être gérés à l'IETF avec la participation de W3C. Par contre, HTML, XML, etc. et bien d'autres sont issus de ce travail coopératif du consortium, où la dernière fois que l'on a évalué, on pense qu'il y a environ 10 000 ingénieurs, ou contributeurs qui ont participé, à un moment ou à un autre, à la conception de l'ensemble de ces standards.



Computer Networks and ISDN Systems 30 (1998) 107–117



The anatomy of a large-scale hypertextual Web search engine¹

Sergey Brin², Lawrence Page²

Computer Science Department, Stanford University, Stanford, CA 94305, USA

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype, with a full-text and hyperlink database of at least 24 million pages is available at <http://google.stanford.edu/>.

To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of Web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the Web, very little academic research has been done on them. Furthermore, due to rapid advances in technology and Web proliferation, creating a Web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale Web search engine – the first such detailed public description we know of to date.

Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want. © 1998 Published by Elsevier Science B.V. All rights reserved.

Keywords: World Wide Web; Search engines; Information retrieval; PageRank; Google

1. Introduction

The Web creates new challenges for information retrieval. The amount of information on the Web is growing rapidly, as well as the number of new users inexperienced in the art of Web research. People are

likely to surf the Web using its link graph, often starting with high quality human maintained indices such as Yahoo³, or with search engines. Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and cannot cover all exotic topics. Automated search engines that rely on keyword matching usually return too many low quality matches. To make matters worse, some advertisers attempt to gain people's attention by taking measures meant to mislead

¹<http://www.yahoo.com/>

² Corresponding author.

³ There are two versions of this paper – a longer full version and a shorter preprint version. The full version is available on the Web at the conference CD-ROM.

E-mail: sergey.p@cs.stanford.edu

0166-3557/98/\$10.00 © 1998 Published by Elsevier Science B.V. All rights reserved.
PII S0166-3557(98)00101-X

Après, une fois que l'on a défini la plateforme et qu'elle rassemble le consensus, on laisse des milliers de fleurs fleurir. Et donc, je vais parler de deux fleurs. Une est née en 1998, lorsque Sergey Brin et Larry Page publient le papier ci-contre lors d'une des conférences World Wide Web. Je reconnais le logo puisqu'il s'agit de la conférence qui a eu lieu à Brisbane deux ans après celle que j'avais organisée à Paris. Le titre du papier est *The Anatomy of a large-scale hypertextual Web search engine*. C'est la première fois que Page et Brin expliquent comment ils s'y sont pris pour faire un outil de recherche – je me souviens d'un de leurs traits d'humour, ils disaient « *pour tester si un moteur de recherche est correct, il faut lui demander d'aller se chercher lui-même* ». Quand on allait sur Google et que l'on tapait « Google », on obtenait Google. Alors que

quand on tapait « Altavista » sur Altavista (le concurrent du moment), on n'obtenait pas Altavista. Donc, ils se sont dit « *On va gagner parce que notre algorithme est plus solide et plus résilient* ». Quelques chiffres : en allant sur *internet live stats* cet après-midi avant d'arriver, il y avait 4 336 569 351 requêtes qui ont été faites dans la journée d'aujourd'hui. C'est intéressant aussi de savoir ce qui s'est passé les premiers jours. En décembre 1998, alors que le site est encore en bêta, il est à 10 000 requêtes par jour. En février 1999, il en est à 500 000 et en août 1999 il en est à trois millions. Il faut avoir une capacité de « scalabilité » qui est assez impressionnante, quand même. Et je pense que cette capacité-là, surtout pour les informaticiens, je pense que ce challenge, le fait que le nombre de sites augmente – là il y a trente

milliards de documents indexés, etc. – cela veut dire que l'espace sur lequel on travaille explose, et je ne sais pas si ça vous est déjà arrivé, mais quand je fais une requête Google, la réponse est immédiate. Donc le fait d'être immédiat en 1998 ou immédiat maintenant, pour un informaticien, c'est un peu de boulot quand même, hein. C'est un peu de boulot sur l'algorithme, c'est un peu de boulot sur l'infrastructure technique qui est mise en jeu, etc. D'ailleurs, il m'est arrivé, en allant dans la *Silicon Valley*, de discuter avec des gens qui étaient chez Google (Louis Monier par exemple) : les gens qui s'occupent du moteur de recherche sont « protégés » chez Google. C'est la partie la plus confidentielle de l'activité de l'entreprise. En tout cas, c'est la première « fleur » dont je voulais parler. Pour une raison qui n'a pas à voir avec le succès de l'entreprise mais qui a à voir avec le rôle que les moteurs de recherche jouent dans l'utilisation du Web. Le Web ne serait pas ce qu'il est aujourd'hui s'il n'y avait pas d'outils de recherche qui permettent de se servir de manière efficace des informations qui sont apportées tous les jours.

Et la deuxième fleur dont je veux parler c'est Wikipédia. Il m'est souvent arrivé de dire que si on avait fait tout le boulot qu'on a fait rien que pour qu'il y ait Wikipédia, ça valait déjà le coup. Wikipédia, comme vous le savez, c'est la manière ultime, comme il est dit ici, de rassembler la connaissance humaine de manière *bottom-up*. Et donc là aussi, quelques chiffres : dans la page que je suis allé chercher sur *Web Archive*, qui date de 2001, Jimmy Wales dit « *Aujourd'hui, Wikipedia a 3 000 pages et l'objectif est d'arriver à 100 000 donc, au boulot* ». Aujourd'hui, il y a quarante millions d'articles en 299 langages sur Wikipédia. Alors, bien sûr, si vous y allez, de temps en temps, ces jours-ci, on vous demande un peu d'argent, d'accord, mais d'abord, c'est épisodique, c'est de temps en temps, et par rapport aux services que ça rend, enfin, je ne sais pas comment vous êtes mais moi je n'hésite jamais à mettre un peu d'argent dans Wikipédia.



De plus, cela va me permettre de faire le lien avec ma conclusion, Wikipedia a un impact sur l'échange des connaissances qui mérite qu'on s'y arrête un peu. L'image ci-contre est un *screenshot* du site des réunions de Lindau. Vous n'êtes pas obligés de savoir, mais j'ai appris récemment ce

qu'étaient les réunions de Lindau. Il s'agit de réunions annuelles où des chercheurs ayant reçu des prix Nobel rencontrent des jeunes étudiants. Une fois par an, à Lindau, une ville au bord du lac de Constance, en Allemagne, les prix Nobel et les jeunes

brillants étudiants passent quelques jours ensemble, à échanger. L'an dernier, ils sont passés à l'étape d'après. Elizabeth Blackburn, qui est un prix Nobel en biologie, a pris l'initiative de dire – c'est pour ça qu'elle parle de *new ways in science* – qu'elle est convaincue qu'on ne fera plus jamais de la science de la même manière qu'on a fait de la science jusqu'à présent. Nous sommes en train de passer de la science de Gutenberg à la science de Berners-Lee. La science de Gutenberg c'est un chercheur qui fait des publications, qui sont lues par d'autres chercheurs, le canal d'échange est constitué par les publications. La science de Berners-Lee c'est les chercheurs qui échangent des publications, des données et du logiciel. Cela permet ainsi, par exemple, d'échanger des résultats intermédiaires. Le Web un peu « jeune » qui ne méritait pas de faire une publication, aurait pu être présenté sous forme de résultats intermédiaires et ainsi recueillir le concours de gens qui travaillent sur le même domaine. Un autre exemple de nouvel usage est de lancer des programmes multidisciplinaires. Si on met à disposition les données, par exemple, les gens du spatial mettent à disposition les données d'observation de la Terre, les agronomes vont pouvoir s'en servir pour faire des études d'agronomie. Bref, on ne va plus faire de la science de la même manière. Et donc, lors des meetings de Lindau, l'année dernière, Elizabeth Blackburn a fait une proposition de dix mesures – je vous invite à aller sur le site <http://lindau-declaration.org> si ça vous intéresse – parmi lesquelles plusieurs ont trait à la science ouverte, dix mesures à mettre en place de manière à vraiment instancier la manière de faire de la science à la mode Berners-Lee et donc vous pouvez y aller, commenter... Ils se sont donnés un an pour recevoir les commentaires et suggestions de n'importe qui sur le Web.

Je voulais terminer là-dessus parce que je pense qu'on n'a encore rien vu. « *We ain't seen anything yet* » comme disait Tim déjà en 1998. Je pense que l'on a construit, pour nous tous, des outils qui vont fournir des résultats que l'on n'imagine pas encore. Je préférerais vous laisser sur cette note optimiste. Je sais bien, je ne suis pas naïf – certains ne me connaissent pas mais je ne suis pas un Bisounours –, je sais quels sont les problèmes de cybersécurité, de protection de la vie privée, je suis au courant de tout ça. J'y ai travaillé, j'y ai dépensé beaucoup d'énergie. Mais je pense que les anniversaires nous font prendre un peu de recul, et quand je prends du recul, je n'ai pas de doute : on a fait quelque chose qui va servir à l'ensemble de l'humanité. Je vous remercie.