

Status quo dans le net ou comment s'arranger avec l'incontournable

Dr. Rüdiger Zimmermann et Jacques Paparo

Les preuves d'admiration et de critique suscitées par la montée fulminante d'une entreprise telle Google Inc. sont tout à fait hors commun. Depuis 1998, elle a rapidement gagné grâce à sa technologie et ses machines de recherche une position dominante dans ce domaine. De plus, Google Inc. a pénétré avec une dynamique incroyable dans des secteurs du marché qui n'ont rien en commun avec les objectifs de l'entreprise initiaux. Le domaine de la santé en fait partie par exemple.

C'était le stockage massif de données et la manière de leur présentation dans le réseau, ainsi que les activités multiples sur différents champs d'activité allant de la numérisation de masse de livres jusqu'au propre service d'email, qui provoquèrent une critique violente et continue.

Cette critique se concentre essentiellement sur trois points:

- Filtrage des résultats des recherches pour des raisons politiques (voir la Chine)
- Pratiques douteuses de protection des données individuelles (Google Analytics)
- Position monopolistique de l'entreprise¹

Et tandis que tous ces trois points nourrissent la critique en provenance des camps politiques libéraux, gauches et écologistes, la position monopolistique de Google pose un problème aussi au niveau national.

Objet de critique sont surtout les projets de numérisation et leur prédominance dans le monde anglo-saxon.

Le projet national français, la « Galica » était dès le début le défi ouvert d'une grande nation culturelle lancé à Google. Moins spectaculaire, la mise en place du portail d'information scientifique allemand « vascoda » poursuit le même but et sa perception interne est celle d'un projet *anti-Google* sensé garantir l'indépendance vis-à-vis d'une entreprise étrangère. On peut aller aussi plus loin pour expliquer la genèse des projets et des contre-projets et nous le savons tous : la construction et le lancement de l'Europeana n'auraient jamais été réalisés sans le défi nommé Google.

Surtout dans les cercles professionnels des bibliothécaires, le partenariat stratégique (Google Book Search) entre le monopoliste et les grandes bibliothèques aux USA et en Europe a suscité de nombreuses discussions échauffées. Chez nous, c'est la Bibliothèque Nationale Bavaroise qui entra en partenariat avec Google. Le manque de transparence de la politique d'entreprise et la mauvaise qualité des objets numérisés livrèrent ici les arguments de la critique des opposants à développement.

Au cours des dernières années, la direction de la bibliothèque de la Fondation Friedrich Ebert faisait elle-même plutôt partie du chœur des voix oppositionnelles et a toujours dénoncé la politique unilatérale et obscurantiste de Google. Néanmoins, ces derniers jours ont apporté beaucoup de changements. La Bibliothèque Nationale Française semble aviser dans un proche avenir une coopération étroite avec Google, la Bibliothèque Nationale Bavaroise et la Biblioteca de Catalunya² le font déjà et la Commission Européenne³ semble-t-elle aussi tenir compte de l'avance technologique de Google, estimé par les experts à cinq ans au moins. Notre comparativement beaucoup plus petite bibliothèque s'est également décidée de suivre ces impératifs et d'accepter des offres de services proposées par Google, en pleine conscience des risques qu'une telle coopération pourrait cacher.

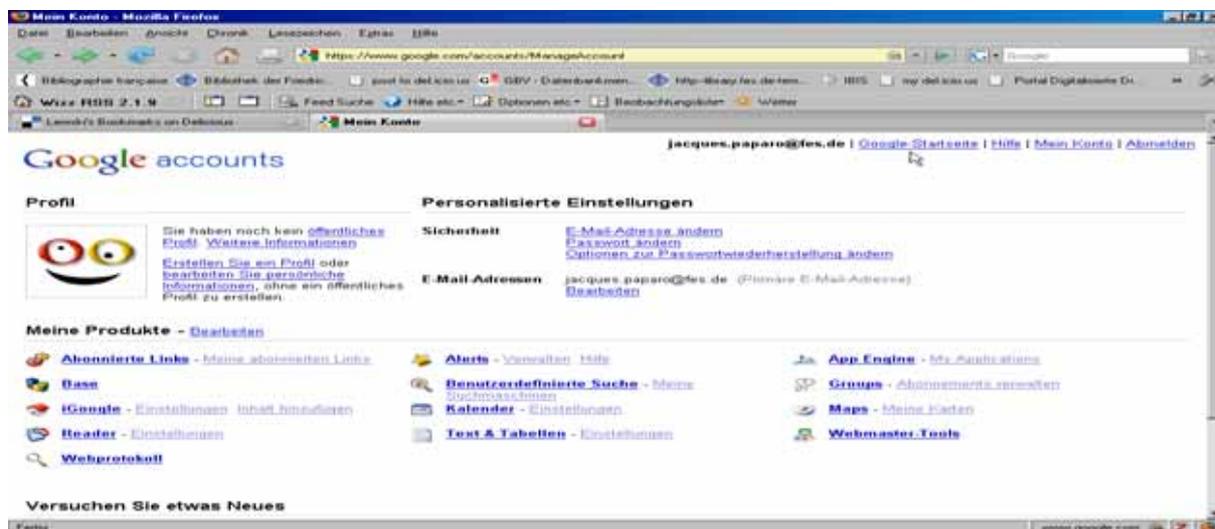
¹ La ministre de la justice, Brigitte Zypries (SPD), rapproche à Google un comportement illicite et met en garde contre sa position monopolistique concernant la numérisation des livres. Actuellement, l'utilisation est libre, mais cela peut changer rapidement : "L'héritage mondial de livres ne doit pas tomber dans les mains d'un monopoliste dicte les prix et décide qui a droit à un accès aux livres" (Interview donné à Rainer Wehaus et apparu le 07.09.2009 dans Stuttgarter Nachrichten online <http://www.stuttgarter-nachrichten.de/stn/page/detail.php/2188529>)

² <http://www.esquerra.cat/actualitat/35.000-llibres-del-fons-de-la-biblioteca-de-catalunya-ja-estan-disponibles-a-internet-a-tr>

³ «L'Europe doit ouvrir un nouveau chapitre dans le domaine des livres numériques et des droits d'auteur»: déclaration commune de Mme Reding et de M. McCreevy à l'occasion des rencontres Google Books. Reference: MEMO/09/376 Date: 07/09/2009. Texte de la déclaration à lire sur le site : <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/09/376&format=PDF&aged=0&language=FR&guiLanguage=en>

Et maintenant les projets dans leur application concrète dans le travail de la bibliothèque de la FES.

Un article paru le 24.10.2006 dans Spiegel Online⁴ marque le début d'une synergie que est passée jusque présent par trois étapes. Vu la vélocité de l'expansion de Google, des étapes suivantes ne peuvent pas être exclues. Pour commencer à utiliser les services offerts, il est nécessaire de créer un compte personnel.

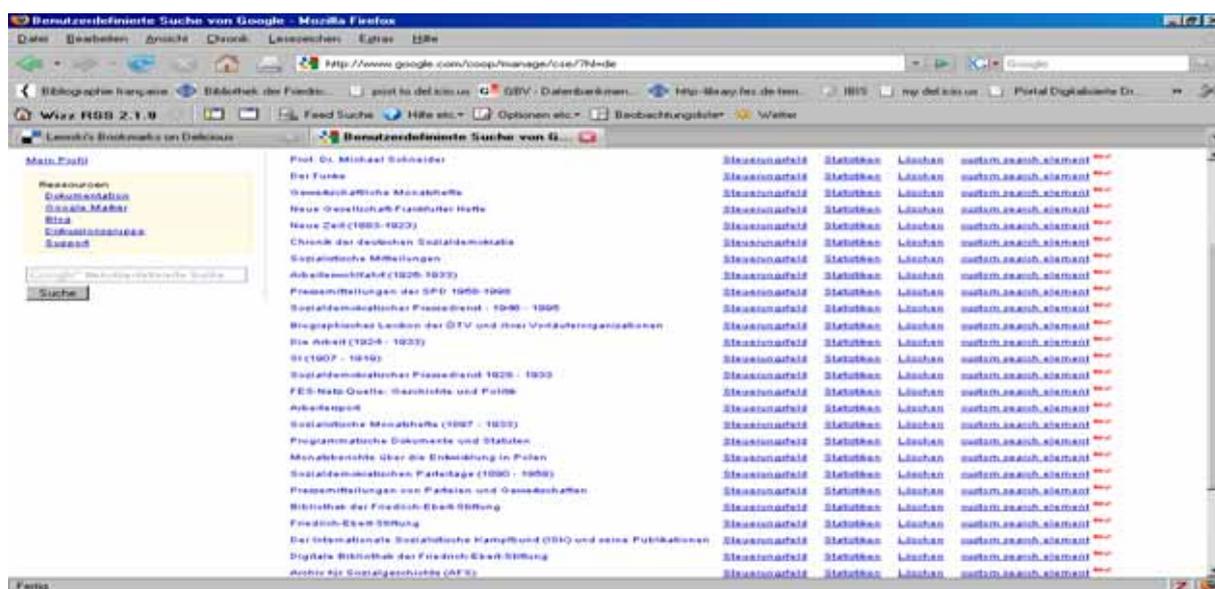


Une fois en possession d'un profile/compte, rien ne nous empêche de commencer à « créer » nos propres moteurs de recherche. Google nous propose⁵ le service suivant :

« Avec la recherche personnalisée Google, vous pouvez exploiter la puissance de Google et créer une expérience de recherche personnalisée pour votre propre site Web.

- Incluez un ou plusieurs sites, ou encore des pages Web spécifiques.
- Placez un champ de recherche et des résultats de recherche sur votre site Web.
- Personnalisez l'aspect des résultats pour qu'ils s'intègrent parfaitement à votre site. »

A l'heure actuelle, nous avons défini quelque vingt-six moteurs de recherche personnalisée, qui sont intégrés à notre site web.

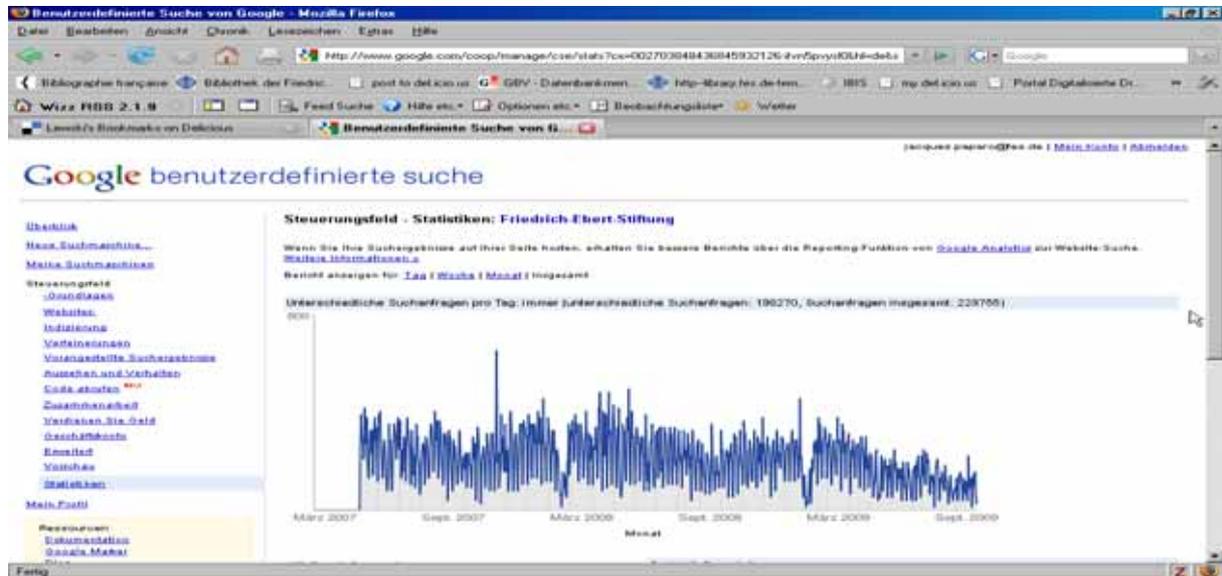


⁴ <http://www.spiegel.de/netzwelt/tech/0,1518,444304,00.html>

⁵ <http://www.google.com/coop/cse/?hl=fr>

Nous avons commencé l'expérience avec la recherche de trois sites :

- notre base de données abritant nos propres publications « Digitale Bibliothek » créée le 25 octobre 2006, donc un jour après l'article de Spiegel Online
- le serveur de la bibliothèque tout seul depuis février 2007
- les serveurs de la Fondation Friedrich Ebert et de la bibliothèque de la Fondation en ligne depuis mars 2007



Après avoir évalué cette première étape en tant que succès, nous avons défini des moteurs de recherche sur des pages Web spécifiques correspondant à des projets de notre bibliothèque – ce sont surtout des projets de numérisation, soit de périodiques soit de sources de l'histoire du mouvement ouvrier. Spécifique pour ces projets est le fait, qu'ils possèdent des méta données très professionnelles et sont sur notre serveur en mode texte ou html. Avec un minimum de ressources, nous avons réussi de mettre en valeur des collections qui ne font plus partie du deep web et qui peuvent facilement et rapidement être recherchées. La qualité et la pertinence des résultats de la recherche sont impressionnantes.

Le 30 octobre 2008, le Blog⁶ de Google annonça une nouvelle mini-révolution sous le titre suivant: : A picture of a thousand words? Il est question de reconnaissance de texte grâce à la technologie OCR (Optical Character Recognition) par Google. Le principe consiste à pouvoir convertir une image de 1000 mots en 1000 mots qui peuvent être indexés comme n'importe quelle page web. Des tests concernant nos propres objets numériques en mode image nous apportèrent la preuve que les robots de Google ont déjà trouvé, reconnu à force de OCR et indexé en plein texte de nombreux documents de notre serveur. La conséquence logique était de créer des moteurs de recherche pour ces projets aussi et de les intégrer, comme les précédents déjà, à notre site. A l'heure actuelle, les caractères de l'écriture Fraktur (all. Frakturschrift), parfois appelée écriture gothique, ne sont pas traités avec OCR et par conséquent restent hors du projet. Il est très probable que, suite à la coopération avec la Bibliothèque Nationale Bavaroise, le répertoire de caractères reconnus automatiquement grâce à OCR s'élargisse de ces derniers.

Cette coopération est pour nous majoritairement avantageuse. Bien sûr, les règles à suivre sont instaurées par des inconnus, le tout est très anonyme et impersonnel. Nous avons suivi les conseils donnés et préparé avec l'aide d'un logiciel offert sur la page des administrateurs des plans de nos sites (sitemaps) en xml et nous les actualisons dès qu'il y a des changements. Les robots nous donnent des informations supplémentaires très utiles quant à la structure et accessibilité de nos sites. Un problème se pose tout de même. Des projets qui sont de petite taille, Google les indexe très vite et complètement. Autrement quand il est question d'indexer par exemple 300.000 objets, comme notre projet⁷ de collection de communiqués de presse de quelque deux-cents partis politiques et syndicats. Là, la quantité des pages indexées grandit de jour en jour, mais est loin d'être complet. Malheureusement, nous n'avons aucune possibilité outre le sitemaps d'influencer le partenaire Google. C'est le prix de la gratuité.

⁶ <http://googleblog.blogspot.com/2008/10/picture-of-thousand-words.html>

⁷ Pressemitteilungen von Parteien und Gewerkschaften: http://library.fes.de/inhalt/digital/pd_search.html

Est-ce la fin de l'histoire? Bien sûr que non. Google est en train d'indexer des audio et des vidéos (Google Audio Indexing technology - GAUDI⁸). Et encore plus récent, aux Etats-Unis, les utilisateurs jouissent d'encore plus de précision lors des recherches de photos. Google Images propose une nouvelle fonctionnalité permettant d'affiner les options de recherche tels dimension, type et couleur et obtient ainsi des résultats plus pertinents.⁹ Notre bibliothèque ne possède point de collections qui soient tanguées par ces nouvelles méthodes – c'est aux collègues archivistes et mediathécaires de déployer leur fantaisie et de mieux mettre en valeur leurs collections.

⁸ <http://labs.google.com/gaudi/static/faq.html>

⁹ <http://googleblog.blogspot.com/2009/07/search-options-now-on-google-images.html>