



## Évolutions des moteurs de recherche sur internet



### Principes et état des lieux

rapport  
méthodes de recherche  
problèmes des entreprises



### Diversification des contenus

recherche personnalisée  
méthodes sociales  
méthodes innovantes  
méthodes hybrides



### Pertinence des réponses

recherche orientée  
méthodes de réponses  
méthodes personnalisées  
méthodes innovantes



### Visualisation des résultats

recherche hybrides  
méthodes graphiques  
méthodes à distance  
méthodes innovantes  
méthodes personnalisées



A. Singhal : « *Our work on interpreting user intent is aimed at returning results people really want, not just what they said in their query* » (2008, <http://googleblog.blogspot.fr/2008/07/technologies-behind-google-ranking.html>), cf. également E. Schmidt (2010, <http://www.wsj.com/articles/SB10001424052748704901104575423294099527212>)

- comprendre le contenu des pages indexées : association de concepts ; ex. : site officiel de la [Sprovieri Gallery] même si Londres n'est pas mentionné dans la requête ;
- comprendre les requêtes : suggestions orthographiques ; synonymes ; recherches similaires ;
- comprendre les utilisateurs : géolocalisation ; personnalisation (données utilisateurs) ; traduction

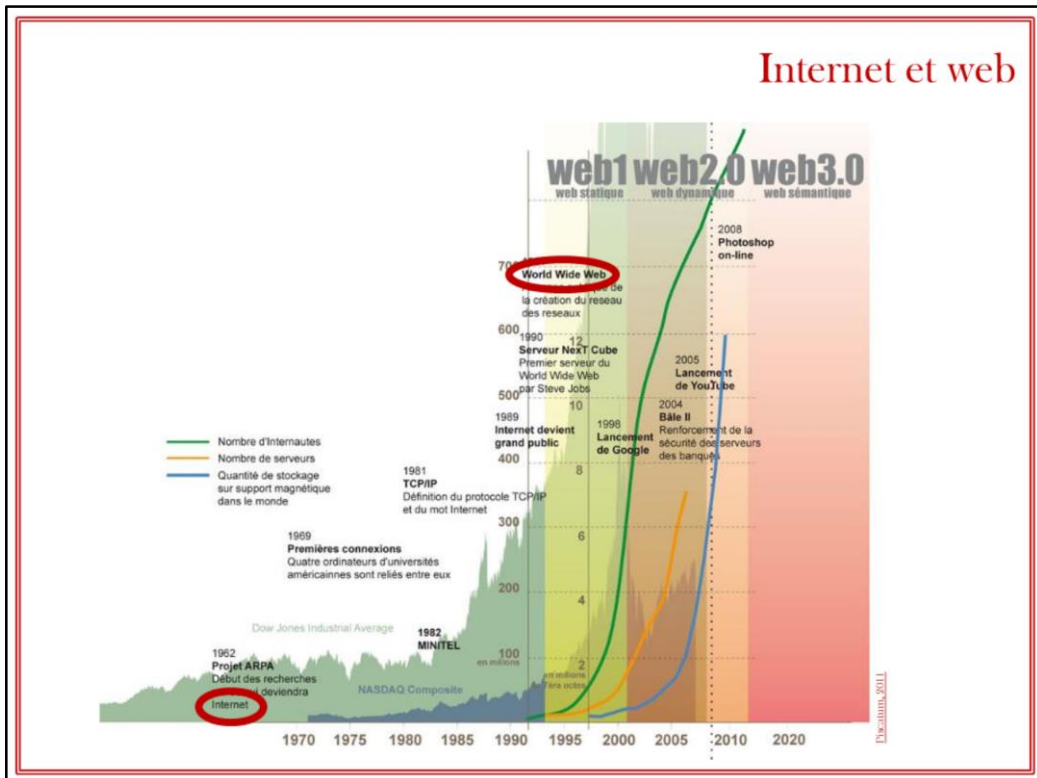
O. Ertzscheid : « Les moteurs de recherche ne fonctionnent plus sur un modèle « Donne-moi ce que je tape » (travail sur l'occurrence des mots-clés choisis) mais « Donne-moi ce que je veux » (travail sur l'adéquation des résultats de recherche au profil de l'utilisateur, ou au profil d'un macro-ensemble de requêtes semblables) » (2008, [http://archivesic.ccsd.cnrs.fr/index.php?halsid=ehi2g4f97ccsor77v7d0c61851&view\\_this\\_doc=sic\\_00325690&version=1](http://archivesic.ccsd.cnrs.fr/index.php?halsid=ehi2g4f97ccsor77v7d0c61851&view_this_doc=sic_00325690&version=1))



## Principes et état des lieux

rappels  
moteurs de recherche  
pratiques des internautes

## Internet et web



### Internet

- réseau de réseaux
- fin des années 1960
- applications et services divers : courrier électronique (mail), messagerie instantanée (IM), forums de discussion, transfert de fichiers (FTP), pair à pair (P2P), web (www)...

### World Wide Web (www)

- début des années 1990
- protocole HTTP
- ensemble de pages HTML (textes, images, liens...) avec une adresse URL

## 1 min sur le web



### Taille

- utilisateurs : 3 MM. d'internautes dans le monde (2014, <http://internetworldstats.com/stats.htm>)
- 858 M. de sites (Netcraft, 05/2015, <http://news.netcraft.com/>),  $30 \times 10^{12}$  pages individuelles (Google, 03/2013, <http://www.google.com/insidesearch/howsearchworks/thestory/>)

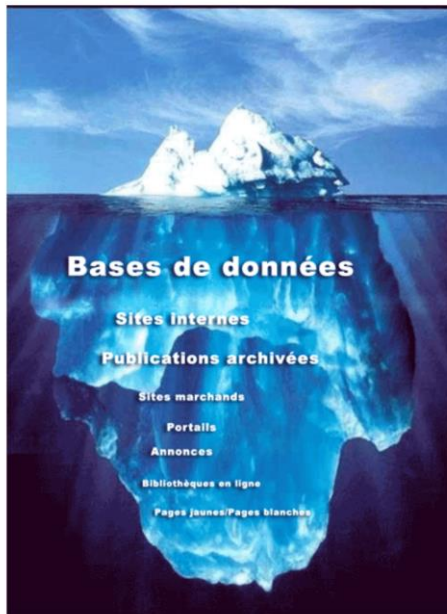
### Contenu

- langues : 55,5 % du contenu en anglais, 3,9 % en français (W3Techs, 05/2015, [http://w3techs.com/technologies/overview/content\\_language/all](http://w3techs.com/technologies/overview/content_language/all))

### Quelques types

- Facebook : 1,4 MM. de comptes
- blogs : 200 M. ?

## Web visible et web invisible



### web visible

indexé par les moteurs  
de recherche

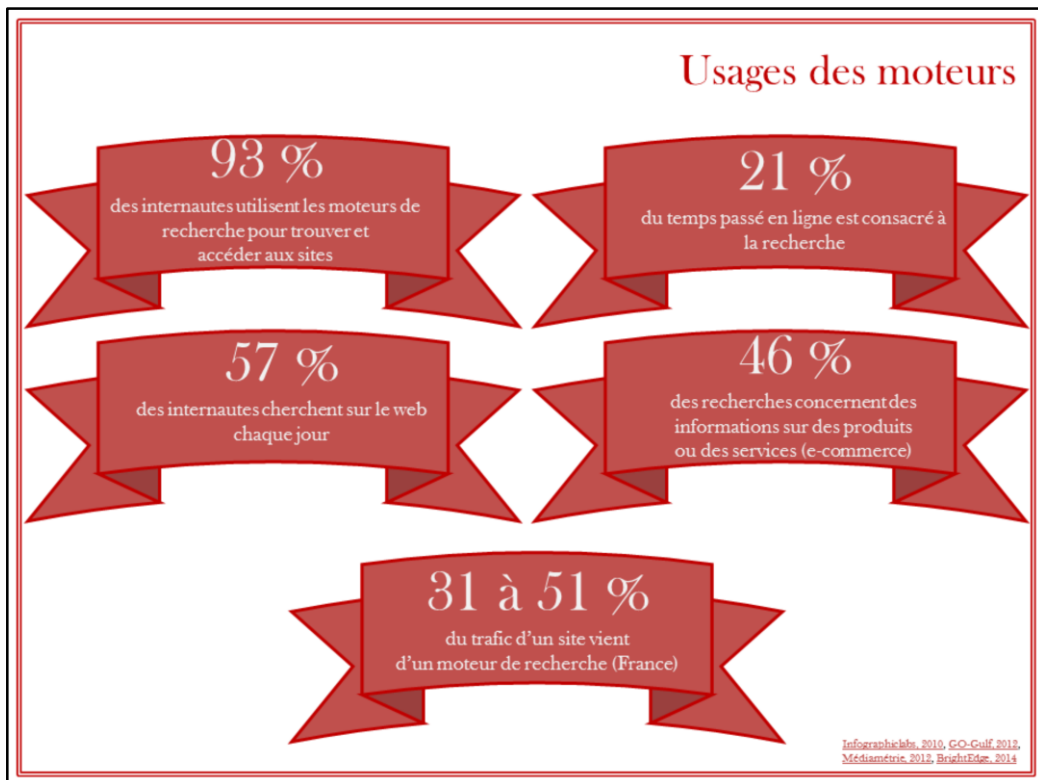
### web invisible / profond

Christophe Asselin, 2001.2

**Différents niveaux de visibilité** : sites des médias et des institutions ; blogs et sites personnels ; médias sociaux (forums, réseaux sociaux, microblogging)

**Web profond** : étude BrightPlanet (2001, <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104>) : le web profond serait 500 fois plus vaste que le web indexé

- 2007 : 12 % visible
- 2008 : 25-30 % visible
- actuellement : 40-60 % ?, dont les moteurs n'indexent que 5 % ? : au moins 2 milliards de pages (Worldwidewebsize, 05/2014, <http://www.worldwidewebsize.com/>) : impossibilité de tout indexer



### Recherche :

**3 besoins identifiés** : « *the need behind the query* » : 3 types de Andrei Broder (2002, « A taxonomy of web search », <http://www.sigir.org/forum/F2002/broder.pdf>) :

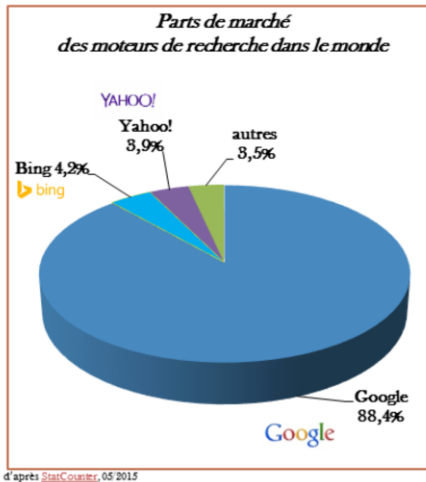
- navigation : pour trouver un site, une URL particulier que l'on a souvent déjà en tête (on le connaît ou on pense qu'il existe) : ex. : site d'une institution ou d'une marque, site déjà vu... ; moteur de recherche comme favori internet ou portail ;
- information : pour trouver des informations précises et détaillées sur un sujet donné, mais sans avoir de site en tête ; utilisation de plusieurs outils, consultation de plusieurs pages ; pages généralement statiques : ex. : descriptions, statistiques, instructions... - recherches les plus fréquentes ;
- transaction : pour faire quelque chose en ligne, peut concerner des pages dynamiques : ex. : horaires de transports, achats en ligne, téléchargement de logiciels...

3 types de **mission** de Jacquelyn Krones (2011, Microsoft, <http://searchengineland.com/exploring-the-shift-in-search-behaviors-with-microsofts-jacquelyn-krones-85750>) :

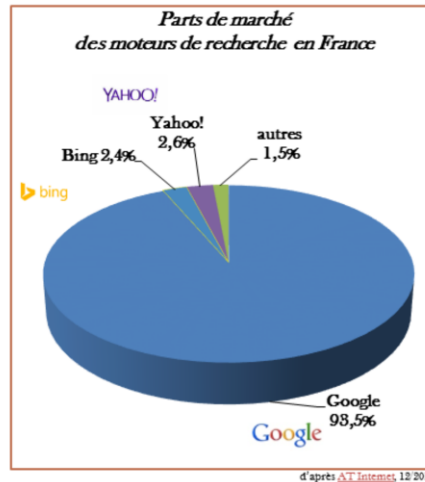
- « *mission mode* » : un but bien défini → mener la recherche de façon efficace : ex. : trouver une information ponctuelle ;
- « *excavation* » : le but n'est pas totalement circonscrit, il peut y avoir plusieurs réponses → trouver la meilleure réponse : ex. : trouver le meilleur service en comparant différentes réponses ;
- « *explorations* » : en fonction de ses centres d'intérêt pour étudier un sujet ou se divertir → trouver des réponses nouvelles et originales.

## Part de marché des moteurs

### Monde

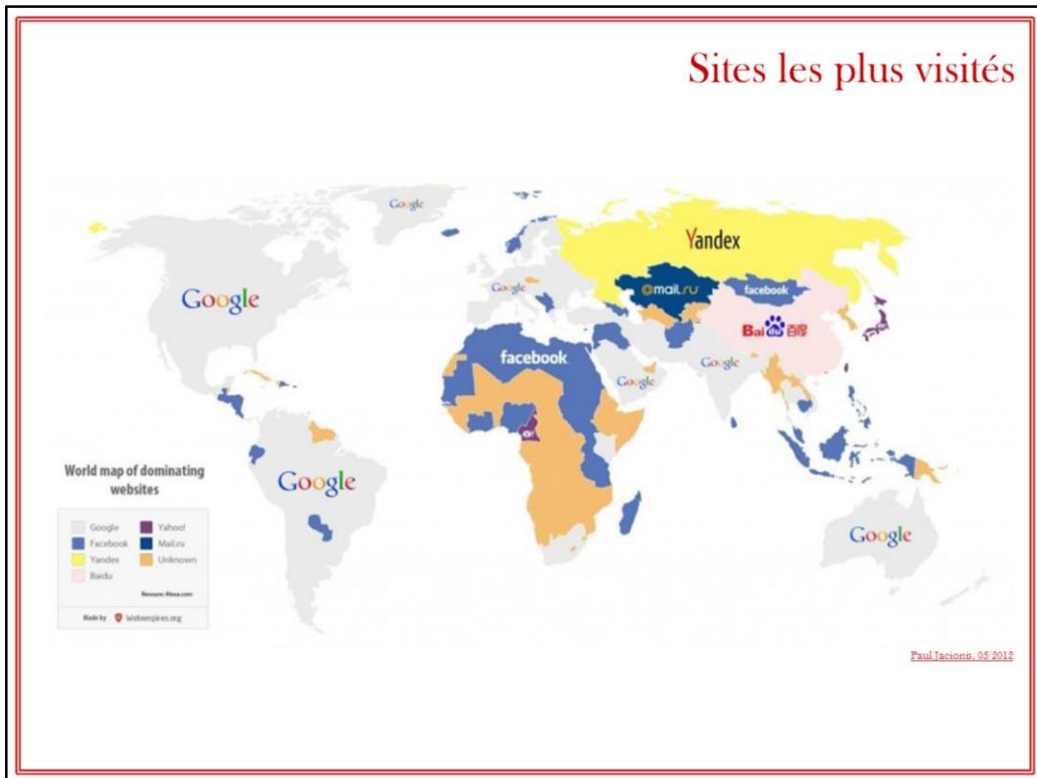


### France



**Google** [cf. O. Duffez, 2013, <http://www.webrankinfo.com/dossiers/google/chiffres-cles>] :

- connaît 30 000 MM. de documents sur le web
- crawle 20 MM. de pages par jour
- chaque minute : 2 M. de recherches
- 98,3 requêtes mensuelles par visiteur unique
- en 2012 : 1 200 MM. de requêtes (67 % du total mondial)



Sur **30 premiers sites mondiaux** (Alexa, <http://www.alexa.com/topsites>, au 1<sup>er</sup>/05/2015):

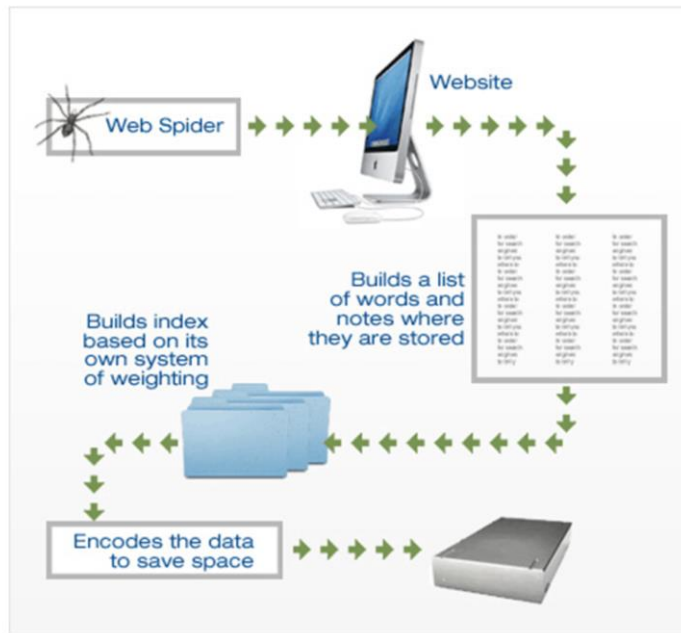
- Google.com : 1er site mondial (+ Google India : 13<sup>e</sup>, Japan : 18<sup>e</sup>, Google.de : 23<sup>e</sup>)
- Yahoo! : 4<sup>e</sup> [portail] (+ Yahoo Japon : 15<sup>e</sup>)
- Baidu : 5<sup>e</sup>
- Yandex : 20<sup>e</sup>
- Bing : 26<sup>e</sup>

## Enjeux technologiques et économiques

- l'industrie du *search*
  - des services qui ouvrent, fusionnent et ferment  
ex. : Qwant : ouverture en version expérimentale (bêta)  
Yahoo : technologie Bing (07/2009, renouvelé en 04/2015)  
Blekko : racheté par IBM (03/2015)
  - des services en évolution permanente  
ex. : Google : 500 changements effectués chaque année
- le marché du *search*
  - rôle central de la publicité comme modèle économique  
Google : 90 % de son chiffre d'affaire  
→ captation et monétisation de pans entiers de nos activités numériques, notamment avec le développement de services conjoints (mails, alertes...) (O. Ertzscheid)
  - développement du référencement et du SEO (*Search engine optimization*)
  - un marché brassant des milliards d'euros...  
2009 : « *How finding stuff on the web became a \$20 billion business* » (source)  
société Google : chiffre d'affaire de 48 MM. \$ et profit de 10 MM \$ (2014)
  - ... mais dont le modèle économique n'est pas assuré (ex. de Google)

**Recherche d'un modèle économique** : ex. : Doona † « premier moteur de recherche à but humanitaire, généreux et solidaire » (association loi 1901) et Backle † ou Ecofree « *energy saving search* » (écran noir consommant moins d'énergie que les écrans blancs), cf. également Charigle : <http://www.charigle.com/fr/index.php> ; Goodsearch : <http://www.goodsearch.com/>  
autres exemples de moteurs solidaires : [http://mashable.com/2012/12/18/search-social-good/#\\_](http://mashable.com/2012/12/18/search-social-good/#_) et liste <http://www.usearch.be/category/humanitaire/>

## Fonctionnement d'un moteur de recherche



**Fonctionnement** (cf. M.-L. Malingre et A. Serres, 2008, <http://fr.slideshare.net/jdeyaref/moteurs-de-recherche-sortir-de-google-presentation>, p. 6-14)

- **collecte des données :**

- *crawler* : robot qui parcourt le web pour aspirer les pages de sites ;
- étendue et profondeur de la collecte : liens, taille des fichiers ;
- paramétrages : nature des fichiers collectés (type de documents, délai du crawl et de mise à jour de l'index) : variable

en fonction du type d'information et du site

→ recul du web profond et accélération du rythme pour certains types de documents

- **indexation des données :**

- méthodes d'analyse : linguistique (reconnaissance des mots) / statistique (fréquence des occurrences = mots-clés)
- niveaux d'analyse : morphologique (reconnaissance du mot), lexical (lemmatisation), syntaxique (grammaire), sémantique (reconnaissance des concepts)

- **gestion des requêtes et des résultats :**

- gestion des requêtes (*information retrieval*) : recherche simple/avancée, filtres
- gestion des résultats : critères de classement (pertinence [« classement des résultats fondés sur la fréquence d'apparition et la localisation des termes de la requête dans une page web »], popularité [« calcul de la notoriété d'un site en fonction du nombre de liens pointant vers le site »], catégorisation [« répartition des documents dans des catégories »])
- gestion de la présentation des résultats : représentation graphique (listes, cartographie...)

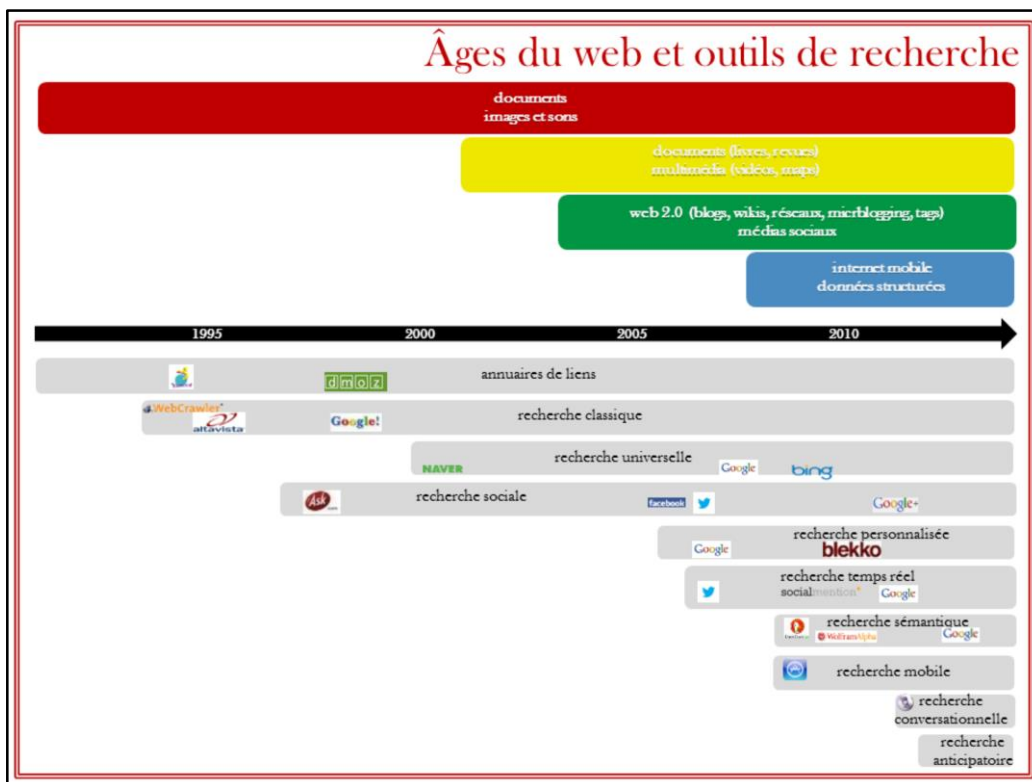
! chaque moteur a sa propre couverture du web, ses propres index, algorithmes et fonctionnalités

**Critères de typologie possibles** (cf. M.-L. Malingre et A. Serres, 2008, <http://fr.slideshare.net/UrfistRennes/stage-outils-recherchehorsgoogleoctobre2010-5504650> et O. Andrieu, 2011, <http://weburfist.univ-bordeaux.fr/wp-content/uploads/2013/02/andrieu2011.pdf>)

- offre de ressources : moteurs généralistes/spécialisés (type de document, de contenu...);
- implication des internautes : moteurs collaboratifs, personnalisables ...;
- mode d'indexation des données : moteurs morpho-syntaxiques, sémantiques, indexation sociale...;
- saisie de la requête : aide à la requête, requête textuelle ou non...;
- recherche de réponses pertinentes : avis, personnalisation, temps réel, géolocalisation...;
- présentation des résultats : texte, représentation graphique...

## Limites possibles des moteurs de recherche

- collecte des données
  - indexation du temps réel
  - prise en compte des contenus diversifiés : multimédia (vidéos), web 2.0, avis et opinions...
  - problème du *spamdexing* (référencement abusif) et des fermes de contenu
- compréhension des requêtes
  - interrogation avec des caractères spéciaux et signes de ponctuation  
développement avec les réseaux sociaux (@, #, \_) notamment, cf. [Google](#)  
→ SymbolHound (<http://symbolhound.com>), pour la programmation essentiellement
  - contextualisation des termes, ex. : Java
- présentation des résultats
  - véritables découvertes (sérendipité) alors que classement par popularité ?  
→ BananaSlug (<http://www.bananaslug.com>)  
propose des termes aléatoires pour chercher dans la « longue traîne » des résultats
  - place des nouveautés ou pages mal référencées alors que classement par popularité ?  
→ Million Short (<http://millionshort.com>)  
exclut les sites les plus populaires (100 à 1 M. de sites) (données de Bing)



Sources : notamment O. Ertzscheid, 2005, [http://affordance.typepad.com/mon\\_weblog/2005/11/moebius\\_le\\_web\\_.html](http://affordance.typepad.com/mon_weblog/2005/11/moebius_le_web_.html) et A. Serres, <http://www.sites.univ-rennes2.fr/urfist/ressources/histoire-dinternet-et-du-web-darpanet-au-web-semantic?destination=ressources>

### 1989-1994 : émergence du web : documents, images et sons – navigation par *browsing*

- premiers outils : annuaires de liens : Archie (1990, simple listing d'annuaires de fichiers FTP), Yahoo (1994, annuaire d'URL commentées)
- premiers moteurs (indexation du texte intégral de chaque page) : WebCrawler (1994)

### 1994-2004 : « web 1.0 » : web documentaire, plutôt statique – navigation par *searching* (recherche d'informations par des algorithmes de recherche)

#### 1994-2000 : essor de trois familles d'outils (moteurs, annuaires, métamoteurs) ;

- \* moteurs : indexation morphologique, indice de pertinence (occurrences des termes); domination d'Altavista (1995, *search tips*, langage naturel, balises *keywords* prises en compte, recherche multimédia) ; Google (1998, *one box*, syntaxe ET par défaut)

- \* grands annuaires généralistes : Yahoo, Open Directory / Dmoz (1998, collaboratif)

#### 2000-2004 : explosion de la « bulle internet » et émergence du web 2.0 (Wikipédia, blogs, réseaux sociaux)

- développement de Google ; indice de popularité (*Page rank*)
- diversification des médias
- nouveaux types d'interface (moteurs de recherche visuels...)

### 2004-2008 : « web 2.0 » / web social (réseaux et médias sociaux) : web plus dynamique, web de conversations – navigation par *subscribing*

- domination de Google ; Bing (2009)
- recherche universelle (tous les types de documents via une interface unique : Google (2007 : textes, images, vidéos, cartes, livres, actualités, blogs...))
- personnalisation : Google (2005) : meilleur classement des sites souvent visités, prise en compte de données locales, etc.

### 2009- : internet « SOLOMO » (social, local, mobile) et sémantique (métadonnées et données structurées)

- recherche temps réel : SocialMention (2008), intégration de Twitter et des données du web 2.0 : Google (2011-2012)
- recherche sociale : recherche et recommandations liées aux membres de son réseau et à la géolocalisation : Facebook Graph
- indexation sémantique : DuckDuckGo (2008), WolframAlpha (2009), *Knowledge Graph* de Google (2012)
- développement de l'utilisation mobile : Appstore (2008) - développement des apps et diminution de la recherche via des moteurs
- 2012-2013- : recherche conversationnelle : app. Apple-Siri (2011), Google Voice Search (2012) - compréhension du langage naturel, y compris de manière orale ;
- agents intelligents et recherche anticipatoire (Google Now...)

## *The Future of Search*

- « La recherche va devenir encore plus **personnelle** et Google sera capable de savoir ce que vous cherchez en fonction de ce que vous êtes.
- Les moteurs de recherche ne trouveront plus seulement mais **interpréteront** ce qu'ils trouveront en générant leurs propres algorithmes.
- Des avancées en matière d'intelligence artificielle et de compréhension du langage naturel amèneront à **des descriptions et à une compréhension approfondies** des pages web.
  - Les **réponses**, et non les liens, deviendront plus courantes.
  - La recherche **fera des choses**, plutôt que simplement les suggérer.
- Nos vies digitales seront combinées en **une seule plateforme interrogeable.** »

Morgan Rose Dickey, 01/2013

Question du *Future of search* ancienne et liée aux évolutions des moteurs de recherche

**2008** : Google, dominant, vient de lancer la recherche personnalisée et la recherche universelle

→ billet « The future of search » de Marissa Mayer sur le blog officiel de Google, (2008, <http://googleblog.blogspot.fr/2008/09/future-of-search.html>) où souligne les éléments à développer :

- modes d'interrogation (non textuels) → développement de la recherche inversée (image, son) ;
- médias (trouver tous les médias, même les plus riches comme les vidéos) → développement de l'*universal search* ;
- personnalisation (fournir des réponses en fonction de l'internaute qui les pose) : localisation, socialisation → prise en compte données personnelles et réseaux sociaux, aide à la requête ;
- langue → outil de traduction automatique

« *Search is a 90-10 problem. Today, we have a 90% solution. [...] Search is a science that will develop and advance over hundreds of years. Think of it like biology and physics in the 1500s or 1600s: it's a new science where we make big and exciting breakthroughs all the time. However, it could be a hundred years or more before we have microscopes and an understanding of the proverbial molecules and atoms of search. Just like biology and physics several hundred years ago, the biggest advances are yet to come.* »

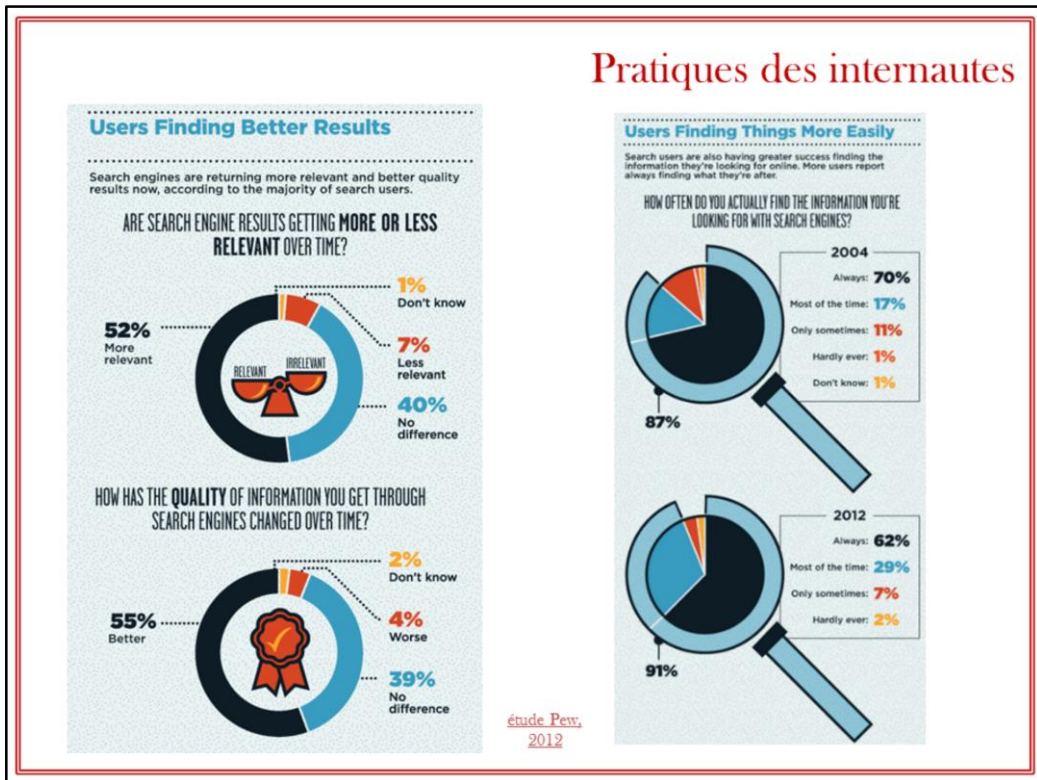
**2012** : Google, toujours dominant, vient de lancer son projet sémantique *Knowledge Graph* et annonce des projets pour le mobile, alors que les réseaux sociaux sont devenus incontournables :

- « *Search will become even more personal, and Google will be able to know what you're looking for based on who you are. Search engines will not only find, but interpret what they find by generating their own algorithms.*
- *Advances in artificial intelligence and natural language understanding will result in deeper descriptions and understanding of web pages.*
- *Answers, not links, will become more prevalent.*
- *Search will do things, rather than simply suggest things.*
- *Our digital lives will be combined into one searchable platform* » (M. R. Dickey, 2013, <http://www.businessinsider.com/this-is-the-future-of-search-2013-1>)

comprendre le comportement et les intentions de l'internaute

→ question corollaire : « Have we reached peak Google ? » au sujet du modèle économique de Google principalement basé sur la publicité sur ordinateur (*Forbes*, 2012, <http://www.forbes.com/sites/timworstall/2012/10/20/have-we-reached-peak-google/>)


## Pratiques des internautes



**Fiabilité** : en 2014, les moteurs de recherche sont considérés par les internautes comme plus fiables que les médias traditionnels. Ils sont même la première source d'information pour 31 % des sondés (O. Andrieu, 2015, <http://www.abondance.com/actualites/20150127-14638-les-moteurs-de-recherche-plus-fiables-que-les-medias-traditionnels-etude.html>).

Sur les critères de **satisfaction**, cf. G. Sterling, 2014, <http://www.referenceur.be/etude-de-satisfaction-moteurs-de-recherche-google-en-tete-13240.html>

## Pratiques des internautes



66 % des **requêtes** : 1-3 mots  
80 % des requêtes : + d'1 mot ([Hitwise, 2009](#))  
½ avec des mots, ½ avec des questions ([Nielsen Research, 2015](#))  
augmentation du nombre de recherches avec 4 mots ou plus  
→ les moteurs doivent être performants pour les requêtes longues, mots-clés rares... - effet longue traîne

déclin de la **recherche avancée** (cf. Google)  
développement des filtres *a posteriori* ([G.R. Notes, 2012](#))  
→ personnalisation des résultats après coup

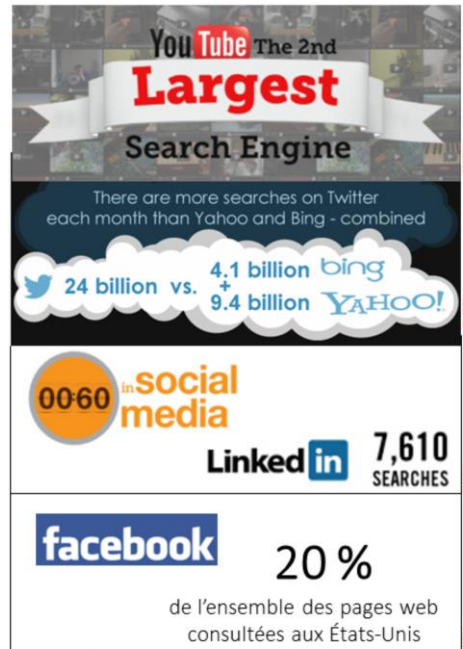
3 premiers **résultats** : 60-80 % des clics  
5 sites récupèrent 20 % du trafic des moteurs ([Experian, 2013](#))  
→ les moteurs sont utilisés comme portail et favoris internet (ex. : recherche de marques, actualités)

90 % des utilisateurs vont régulièrement sur la **2<sup>e</sup> page** de résultats  
([Aquire et SurveyMonkey, 2013](#))

Sites concernés : Facebook, YouTube, Yahoo, Wikipédia et Amazon ; 500 sites = ½ du trafic

## Évolution des pratiques

- Diversification des outils de recherche



La recherche d'informations est bien plus large que la simple question des moteurs de recherche : en 2012, 54 % des internautes ont recours aux moteurs de recherche contre 32 % aux **réseaux sociaux** pour trouver de nouveaux contenus en ligne, contre 18 % en 2010 (étude Forrester Research, 2013, <http://www.zdnet.fr/actualites/les-reseaux-sociaux-2e-de-la-recherche-web-39791699.htm>). Tendance confirmée en 2014, le poids des moteurs de recherche baisse dans le trafic au profit des réseaux sociaux (étude Shareaholic, 2014, <http://business.lesechos.fr/entrepreneurs/web/4424760-nombre-de-visiteurs-les-moteurs-de-recherche-generent-moins-de-traffic-101133.php#> et Médiamétrie pour la France, <http://www.emarketinglicious.fr/webmarketing/reseau-sociaux-futurs-moteurs-recherche-mots-cles>).

## Évolution des pratiques

- accroissement des **usages mobiles**
  - 2015 : accès à internet : uniquement par mobile > uniquement par ordinateur ([étude ComScore I.S. 2015](#))
  - 2013 : temps passé sur mobile > ordinateur ([étude StatCounter et comScore](#))
  - 2015: recherches mobiles augmentent plus vite que recherches sur ordinateur ([étude ComScore, 2015](#))
  - 2015 : recherches Google : smartphone (moteur + Google Play) > ordinateur + tablette ([étude Google, 2015](#))
  - des usages différents selon les outils ([étude Microsoft research, 2013](#)) - 50 % des recherches sur mobiles sont locales ([G. Sterling, 2013](#))
- développement de la **recherche par app**
  - 2014 : 86 % du temps sur smartphone : app. (web : 14 %) ([étude Flurry, 2014](#))
  - 2015 : 40 % de la recherche par app ? ([Mashable, 2015](#))
  - 2015-2016 : recherche sur internet : mobile (web + app) > ordinateur ? ([étude BIA/Kelsey, 2013](#))

→ évolutions des moteurs de recherche

**! Utilisation sur mobile** (i.e. smartphones et tablettes) ne signifie pas nécessairement en situation de mobilité : une part importante est faite à la maison, etc. (<http://searchengineland.com/highest-use-of-mobile-search-at-home-report-69557>)

**Utilisation du PC** pour internet pendant longtemps, faute de mieux : « *"The Great PC Exodus on the Internet is happening because the PC was never truly a consumer product,"* added Weide. *"Many consumers use them because there was no better alternative. Now, with the huge and growing installed base of more user-friendly tablets and smartphones, there are."*» (IDC, 2012, [http://www.idc.com/getdoc.jsp?containerId=prUS\\_23756512](http://www.idc.com/getdoc.jsp?containerId=prUS_23756512))

**Recherches de produits et de services** en majorité (G. Sterling, 2013,

<http://searchengineland.com/google-44-percent-of-searches-for-last-minute-holiday-gifts-will-be-mobile-91763>) – développement des apps permet de ne plus utiliser le moteur de recherche pour rechercher les sites les plus utilisés par un internaute

recherche avant tout de réponses pertinentes et peu nombreuses : « *These factors reveal a fundamental problem with the search today : it's not a dynamic and flexible process. Surely, in today's world we need more than a search page with a list of blue URL links to sort through when we're looking for recommendations, advice, diagnosis and other methods of finding and exploring information and products in the digital age* » (L. Hård, 2012, <http://venturebeat.com/2012/11/06/next-generation-search/>) : recherche de *tailored content*

Développement des « **mobile updates** » des moteurs de recherche pour favoriser les sites

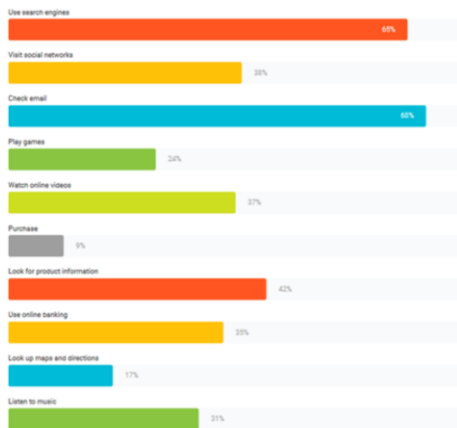
« *responsive* » (i.e. lisibles sur les terminaux mobiles), cf. Google (O. Andrieu, 2015,

<http://www.abondance.com/actualites/20150421-15051-la-mise-a-jour-google-sur-la-compatibilite-mobile-cest-aujourd'hui-etes-vous-prets.html>) et Bing (O. Andrieu, 2015,

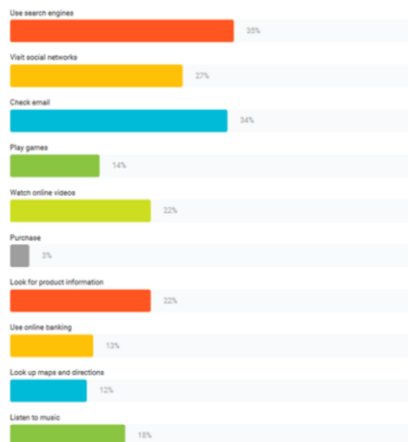
<http://www.abondance.com/actualites/20150515-15138-bing-va-aussi-lancer-son-mobile-update.html>)

# Évolution des pratiques

What online activities do people do on their **computers** at least weekly?



What online activities do people do on their **smartphones** at least weekly?

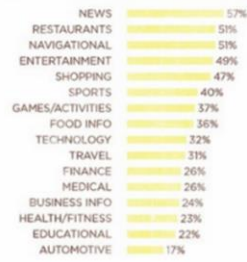


[Google Consumer Barometer 2014](#)  
(résultats France)

# Évolution des pratiques

## SMARTPHONE SEARCHES

PEOPLE WHO USE SMARTPHONES  
LOOK FOR A WIDE VARIETY OF  
INFORMATION.

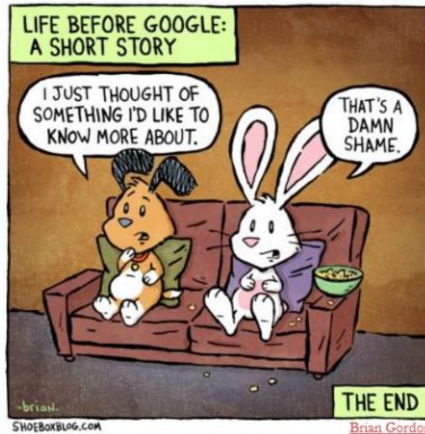


[Leveling d'opini Google 2011](#)

## Évolution des pratiques

« En quelques années seulement, nous sommes passés des moteurs de recherche [...] issus de la discipline académique un peu guindée de la recherche d'informations [*information retrieval*] à la recherche [*search*], tout simplement, qui est bien plus qu'un instrument et quelque chose se rapprochant d'une prothèse digitale »

T. Vanderbilt, 2013



« In just a few years we have gone from search engines [...] with their roots in the staid academic discipline of information retrieval, to, simply, "search", which is much more than an apparatus and something closer to a digital prosthesis. » (T. Vanderbilt, 2013, <http://www.wired.co.uk/magazine/archive/2013/01/features/the-future-of-search?page=all>)

**Recherche présente partout**, cf. Yelp, Foursquare, Facebook, Twitter : mode d'interactions avec des services

La recherche passerait de chercher ce que l'on connaît à chercher ce que l'on ne connaît pas, cf. principes d'auto-complétion de Google, corrections orthographiques

**Ne sont plus de simples outils de recherche d'informations**



## Diversification des contenus

recherche personnalisée  
recherche sociale  
recherche temps réel  
recherche inversée



La recherche personnalisée

## La recherche personnalisée

<b>Limite des moteurs traditionnels</b>	<p>« <i>One size fits all</i> » : mêmes résultats pour une même requête, quelque soit l'internaute et/ou le contexte de la requête          → ne convient plus forcément actuellement : un algorithme comme le PageRank initial de Google (principe de pertinence) est trop restreint. D'autant plus que les internautes savent ce qu'ils peuvent trouver (avis, opinions, comparaisons...) cf. <a href="#">interviews</a></p>
<b>Type de documents</b>	<p>tout type de document ou ressource adapté à l'internaute (ex. : centres d'intérêts, recherches fréquentes) et/ou au contexte de la requête (ex. : langue, lieu, historique de recherche)</p>
<b>Recherche</b>	<p>« <i>Search has one goal and one goal alone : to provide the most relevant information for the user in the fastest time possible</i> » (Ben Gomes)          évolution déjà ancienne (au moins les années 2000, cf. Amazon, eBay...)          deux axes : éléments explicites de la requête et implicites à l'internaute [cf. aussi <a href="#">journée DICEN 2012</a>]          - analyse statistique et sémantique des contenus : scores, classifications, similarités...          - étude des comportements des internautes (groupes et individuels) : traces</p>
<b>Outils</b>	<p>nombreux aspects :</p> <ul style="list-style-type: none"> <li>- moteur de recherche : requête (historique, profil, géolocalisation, choix des sources) ; résultats (classement)</li> <li>- extensions du navigateur (ex : Fast Search par SurfCanyon)</li> <li>- actions de l'internaute : saisie de la requête, actions sur les résultats, création de communautés et réseaux...</li> <li>- moteurs de recommandations / moteurs de découvertes</li> </ul>
<b>!</b>	<ul style="list-style-type: none"> <li>- rejoint des pratiques commerciales actuelles de personnalisation (cf. Nike, Coca-Cola...) et d'<a href="#">utilisation des données personnelles</a></li> <li>- assure une aide au choix et à la décision (critères particuliers, suggestions, réduction de l'abondance informationnelle) et fournit éventuellement des informations et des services supplémentaires (ex. : sérendipité)</li> <li>- permet de mieux connaître l'internaute et de le fidéliser (couplés avec d'autres services, ex. boîte mail... = économie de l'attention)</li> <li>- limite de la notion de popularité (sureprésentation des mieux notés), effet de « <i>filter bubble</i> » (cf. <a href="#">Don't bubble us</a>)</li> <li>- origine des données et algorithmes de traitement des résultats (manipulations et censure) ?</li> <li>? : risques d'utilisation des données personnelles et protection de la vie privée par rapport aux bénéfices des innovations technologiques ? cf. <a href="#">scandale PRISM (NSA) aux Etats-Unis</a> et <a href="#">interview d'O. Ertzscheid</a></li> <li>! : censures imposées de l'extérieur (censures politiques, questions juridiques comme droit à l'oubli...) ? cf. <a href="#">demandes de la CNIL et jugement canadien</a> (06/2015)</li> </ul>

D'après M. L. Malingre et A. Serres, 2009, <http://www.sites.univ-rennes2.fr/urfist/ressources/recherche-dinformation-sur-internet-recentes-evolutions-2009> et les différentes interventions de la journée d'étude du laboratoire DICEN sur les moteurs de recommandation (2012, <http://www.dicen-idf.org/evenement/journee-etude-moteurs-recommandation/>)

## La personnalisation de la recherche

### Une attente des internautes (2009)

#### PERSONALISATION DES RESULTATS

*Dans quelle mesure les résultats de  
recherche peuvent-ils être  
personnalisés ?*

Fournir des résultats de recherche personnalisés en fonction de **mon profil** (âge, sexe, centres  
d'intérêt..)

Fournir des résultats de recherche personnalisés en fonction de **ma localisation** actuelle  
(restaurant chinois, banque...)

Me permettre de sélectionner et pondérer les **critères de pertinences** utilisés par le moteur  
(nombre de visites, nombre de liens sortant....)

Ancien modèle : recherche des stations de métro à Paris ;

Nouveau modèle : recherche des stations de métro à Paris, mais prise en compte également de  
l'internaute (ex. : un utilisateur d'Iphone, situé à Paris)

## La personnalisation de la recherche

### Comment ?

- pratiques de l'internaute

« L'algorithme [de Google] utilise plus de 200 paramètres pour affiner une requête. Cela inclut le PageRank du site internet, la localisation géographique de l'internaute, les liens sur lesquels il clique généralement, la façon dont il modifie ses requêtes quand il n'est pas satisfait, ainsi que son historique de recherche »

[T. Vanderbilt, 2013,](#)  
voir également [infographie](#)

« *The algorithm uses over 200 signals to refine a search query. These include a website's PageRank, a searcher's geographic location, which links they usually click, how they modify their search queries when they are unsatisfied and their search history* » (T. Vanderbilt, 2013, <http://www.wired.co.uk/magazine/archive/2013/01/features/the-future-of-search?page=all>)

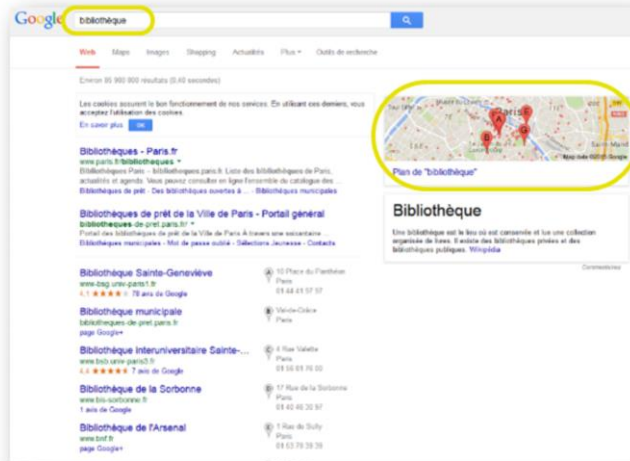
ex. : - pour un exemple sur les informations obtenues sur le type de navigateur avant même de saisir une requête, cf. <https://panopticlick.eff.org/>  
- en utilisant le navigateur TOR : message d'alerte : « Maximiser le navigateur Tor à l'écran peut permettre aux sites Web de déterminer votre taille de moniteur, laquelle peut être utilisée pour vous suivre à la trace. Nous recommandons que vous laissiez les fenêtres du navigateur Tor dans leur taille d'origine »

**Fonctionnalités en test** qui peuvent être conservées ou non

cf. Google : une dizaine de fonctionnalités en test pour chaque visite ; 500 changements effectués chaque année (B. Schwartz, 2012, <http://searchengineland.com/just-testing-google-searchers-may-see-up-to-a-dozen-experiments-141570>)

# La personnalisation de la recherche

- géolocalisation



développement avec l'internet social et mobile (SoLoMo)

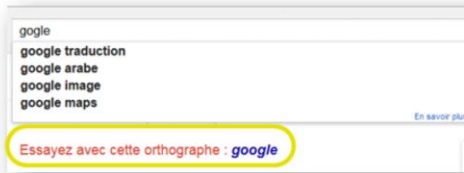
cf. [étude Neustar/15 miles/comScore 2013](#)

## Géolocalisation :

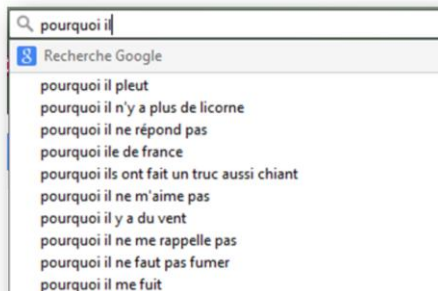
- « identifier l'emplacement géographique de l'internaute, afin de lui fournir des informations locales le concernant spécifiquement » (M.-L. Malingre et A. Serres, d'après *Netsources*, <http://www.sites.univ-rennes2.fr/urfist/ressources/recherche-dinformation-sur-internet-recentes-evolutions-2009>) : basé sur l'adresse IP
- recherche d'informations en incluant des informations locales, éventuellement couplé avec d'autres services (carte, itinéraire, liens sponsorisés...)
- développement avec les *mash up* et le développement des applications cartographiques ; lié notamment au marché de la publicité de proximité

## La personnalisation de la recherche


- aide à la requête



suggestions orthographiques



*Google instant search* : saisie semi-automatique et affichage des résultats en cours de frappe

[paramétrage :  ]

→ « la réponse avant la question »

(O. Ertzscheid)

*Google suggest* : suggestions de recherche

**Autocomplétion** : « *We no longer have to remember, as one Google engineer put it, we simply have to recognize* » (T. Vanderbilt, 2013, <http://www.wired.co.uk/magazine/archive/2013/01/features/the-future-of-search?page=all>)

- absence totale d'objectivité du moteur : résultats mêlant l'automatisation de la machine (classement des résultats) et les requêtes des humains « avec toute leur partialité et leur intelligence supérieure » (J. Battelle, 2005, p. 215 ; ex. <http://www.blogdumoderateur.com/meilleures-suggestions-google-suggest/>)

cf. aussi « Google, confessionnal du XXI<sup>e</sup> siècle », 2013, <http://geeko.lesoir.be/2013/11/16/google-confessionnal-du-xxie-siecle/>

- affaires portées en justice (racisme, diffamation...), mais décision de la Cour de cassation (19/06/2013) estime que le rapprochement des termes est dû à des traitements automatiques et aléatoires, sans volonté de la part de la société d'émettre un avis quelconque : simple **aide à la recherche**

cf. aussi les suggestions automatiques de services comme Facebook (A. Oury, 2013, <http://www.actualitte.com/societe/facebook-tu-lis-philip-kerr-essaye-hitler-ou-la-logique-de-l-algorithme-44619.htm>)

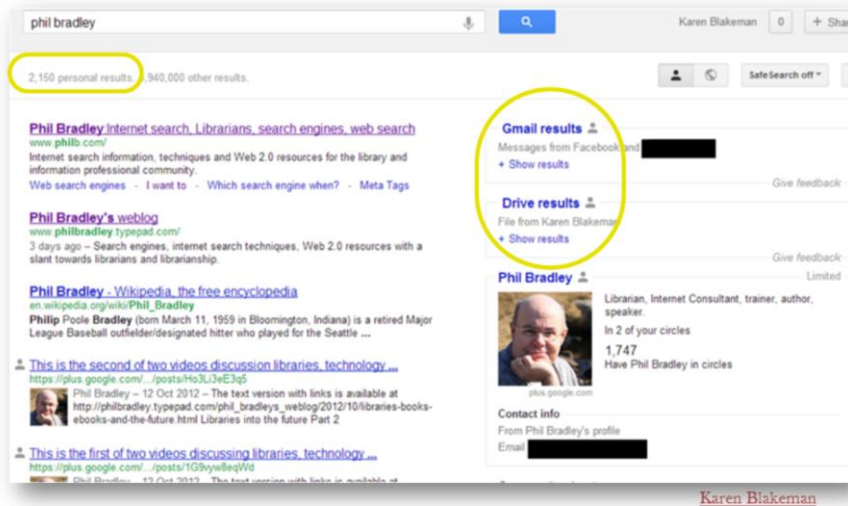
Autre exemple de personnalisation : **Google ne cherche pas toujours uniquement les mots indiqués**, et peut en laisser de côté

« *Because it's clear that people will often write long queries (with anywhere from 5 to 10 terms) for which there are no results. Google will then selectively remove the terms that are the lowest frequency to give you some results (rather than none). Bear in mind that 99% of searchers have no idea why they'd want to hard AND, and just get frustrated when they get no results. The soft AND is a way to reduce the overall frustration and give the searcher something to examine (and with luck, a chance to reformulate their query)* » (D. Russell, 2011, <http://www.rba.co.uk/wordpress/2011/11/08/dear-google-stop-messing-with-my-search/>)

--> développement d'une certaine **sérendipité** : « Sur Youtube, il y a toujours une réponse. La sérendipité est comme une pertinence seconde, qui vient se substituer à la réponse exacte » (A. Gunthert, cité in [http://affordance.typepad.com/mon\\_weblog/2010/02/ingenieries-de-la-serendipite.html](http://affordance.typepad.com/mon_weblog/2010/02/ingenieries-de-la-serendipite.html))

# La personnalisation de la recherche

- **identification** à un compte personnel



« Sometimes the best answer to your question isn't available on the public web — it may be contained somewhere else, such as in your email. We think you shouldn't have to be your own mini-search engine to find the most useful information — it should just work » (A. Singhal, 2012, <http://googleblog.blogspot.fr/2012/08/building-search-engine-of-future-one.html>)

# La personnalisation de la recherche



## Opportunités

Google, ne t'arrête pas de collecter mes données!  
[F. Maioo](#)

- ... selon Google (*How we use data*)
- apprentissage du moteur (ex. : [mots nouveaux](#))
- prédictions (ex.: *Google Flu Trends*)

## Enjeux

→ La réponse avant la question. Ou le complexe du scribe. [O. Ertzscheid](#)

- recherche moins diversifiée (limite la longue traîne des réponses)
- multiplication des stéréotypes (ex.: « les noirs... »)
- importantes [questions éthiques, économiques, géo-politiques...](#)

[Big Mother vs. Big Brother](#)

Question centrale des **données personnelles**

## La personnalisation de la recherche

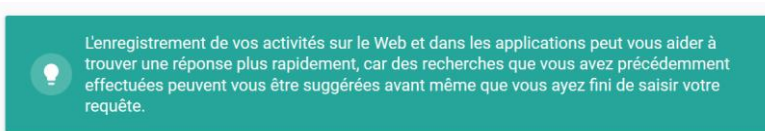
- une **préoccupation croissante** des internautes
  - 2/3 des internautes sont contre la personnalisation des résultats ([étude Pew, 2012](#))
    - problème d'une liste de résultats limitée
    - problèmes des données privées
  - programme PRISM et révélations d'E. Snowden sur les pratiques de la NSA (Nation Security Agency), 2013
  - 9 internautes US estiment importants de pouvoir contrôler l'accès à leurs informations personnelles ([étude Pew, 2015](#))
  - de plus en plus d'internautes US nettoient ou masquent leurs traces numériques ([étude Pew, 2015](#))
- une **réalité**
  - une même recherche faite par des internautes différents peut amener des résultats différents ([étude M.E. Bates, 2011](#)) - pour un exemple : *What happened to dinosaurs?*
  - personnalisation sur Google ([étude A. Hannak et al., 2013](#))
    - 12 % des résultats différents
    - basée surtout sur l'identification au service et sur l'adresse IP (localisation)

→ pour un exemple des informations que Google peut conserver si l'on fait ses recherches en étant connecté à un service Google (Gmail, Google Agenda, YouTube...) : <https://history.google.com>

→ pour un exemple de ce que l'on peut rapporter à Google, Facebook... : service <https://privacyfix.com>

« Surprisingly, we only find measurable personalization as a result of searching with a logged in account and the IP address of the searching user » (A. Hannak et al., 2013, <http://personalization.ccs.neu.edu/paper.pdf>)

Argument commercial cependant :



Ex. sur <https://history.google.com>

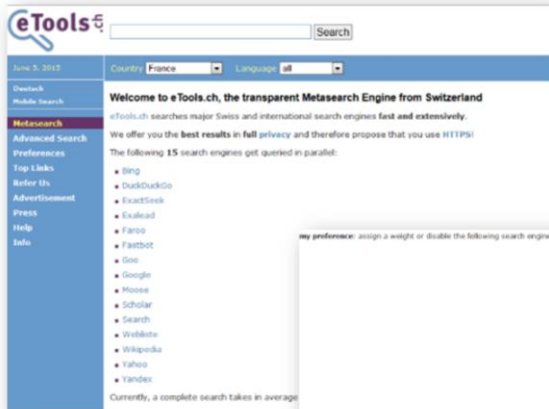
## La personnalisation de la recherche

### obtenir les résultats les plus neutres possible

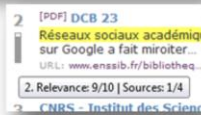
- paramétrer son **navigateur**
  - utiliser des plug-ins spécifiques (AdBlock Plus, Better Privacy, Disconnet, HTTPS Everywhere, Ghostery, Blur...)
  - nettoyer les « traces » : historique, cookies...
  - naviger anonymement (fenêtre privée de Firefox..., [navigateur TOR](#))
- maîtriser les **moteurs** de recherche
  - ne pas être identifié sur le service (O. Andrieu) Témoignages : [Karen Blakeman](#) et [Béatrice Fournie-Niou](#)
  - sur Google : gestion sur My Account : <https://myaccount.google.com> (informations personnelles et confidentialité, préférences, historique des recherches...)
  - utiliser [Disconnect Search](#), surcouche permettant d'interroger individuellement de façon anonyme les principaux moteurs de recherche (Google, Bing, Yahoo)
- utiliser des **moteurs** qui respectent la **confidentialité** des données personnelles
  - Ex. : Ixquick : <https://eu.ixquick.com/>
    - société hollandaise
    - label européen pour la protection des données personnelles (2008)
    - ✦ confidentialité des recherches (pas de conservations des adresses IP ni des cookies au-delà de 48h)
    - paramètres de recherche : <https://ixquick.com/fr/techniques-recherche-puissance.html>
    - recherche en 18 langues, mais pas de liste des moteurs interrogés
    - site associé [Startpage.com](#) permet d'interroger anonymement Google (l'option « proxy » permet ensuite d'aller sur les pages concernées anonymement)
  - DuckDuckGo : <https://duckduckgo.com/> (cf. *infra*)
  - Qwant : <https://www.qwant.com/> (cf. *infra*)
  - SwissCows : <https://swisscows.ch/?uiLanguage=de&region=fr-FR> (cf. *infra*)
  - Searx :
    - logiciel libre permettant de créer un métamoteur paramétrable
    - nombreuses instances dont <https://framabee.org/> et <https://searx.laquadrature.net/>
  - Mojeek : <http://www.mojeek.com/> (société anglaise)

## Un outil utile

Le métamoteur suisse eTools : <http://www.ertools.ch/>



préférences : poids des moteurs interrogés et nombre de résultats pour chacun, recherche par HTTPS...



classement des résultats par relevance et indication de la source

## Moteurs de recommandation

The screenshot shows the CANALSAT website's 'CAMPUS BAC' section. At the top, the 'CANALSAT' logo is on the left, and navigation links for 'S'ABONNER', 'ESPACE CLIENT', 'Appis', 'f', 'i', 'S'inscrire', and 'S'identifier' are on the right. Below this is a 'GUIDE TV' menu with options: 'ACTUALITES', 'LIVE TV', 'A LA DEMANDE', and 'CHAINES'. A secondary navigation bar includes 'Ce soir', 'Maintenant', 'Cette semaine', 'Campus', 'Eureka', 'Grille TV', 'Mes enregistrements', and 'Découvrez EUREKA'. The main heading is 'LA SELECTION CAMPUS' with a 'Découvrez Campus en vidéo' button. A highlighted text box explains: 'Le moteur de recommandation Education à la TV CAMPUS vous emmène plus loin dans toutes les matières de la 6ème à la terminale! Choisissez votre matière scolaire et découvrez les programmes CAMPUS sur les 7 jours à venir. à retrouver sur vos chaînes CANALSAT.' Below this is a carousel of three images: a man in a red tunic, a desert landscape, and a man in a suit. The 'EUREKA' program is highlighted with a text box stating: 'EUREKA est le premier moteur de recommandation personnalisée qui analyse les programmes que vous regardez pour vous en proposer d'autres qui vous plairont à coup sûr! Chaque soir, il sélectionne pour vous 6 programmes qui répondent à vos goûts et à vos habitudes télé. Pour cela, rien de plus simple, EUREKA est accessible dans le GUIDE de votre télé ou sur canalsat.fr, rubrique CE SOIR et LES SOIREES puis la sélection EUREKA.'

Ex. Campus (CanalSat - image 2012)

Une **tendance déjà ancienne**, cf. principe des annuaires, système du PageRank, suggestions Amazon dès début 2000 ? ; grand développement avec le web 2.0 (sites de vente, sites de réseautage...) et « moteurs de découverte personnalisée »

## Moteurs de recommandation

### Suggestions algorithmiques

- moteurs de **recommandation automatique**  
recommandations automatiques en fonction des profils et actions
  - biais : visées marketing et/ou mise en valeur de ce qui est déjà populaire
- Ex. : sites de commerce en ligne de type Amazon, cf. [Book Seer](#)
  - LibraryThing : <http://www.librarything.com>  
plateforme communautaire de catalogage de livres en ligne
  - Goodreads : <http://www.goodreads.com/>  
plateforme communautaire de catalogage de livres en ligne (racheté par Amazon)

Pour les bibliothèques de lecture publique notamment voir ABF, 2013,  
<http://www.abf.asso.fr/3/135/393/ABF/madame-machine-pouvez-vous-me-conseiller-un-bon-livre-les-nouveaux-outils-web-de-recommandation-de-lectures?p=7&p2=0> et L. Wiart, 2014,  
<http://www.inaglobal.fr/edition/article/lecteurs-quels-sont-vos-reseaux>

## Moteurs de recommandation

- fonctionnalités d'outils de **recherche scientifiques**

- Ex. : Google Scholar : <https://scholar.google.fr/>  
moteur de recherche scientifique



The current state and **future of search** based software engineering  
M. Harman - 2007 Future of Software Engineering, 2007 - dl.acm.org  
Abstract This paper describes work on the application of optimization techniques in software engineering. These optimization techniques come from the operations research and metaheuristics research communities. The paper briefly reviews widely used ...  
Cité 453 fois [Autres articles](#) Les 14 versions Citer Enregistrer

- Ex. : Mendeley : <https://www.mendeley.com>  
logiciel de gestion de références bibliographiques  
base de données accessible à la recherche : <https://www.mendeley.com/research-papers/search/>



**Searching for the future of search**  
Oliver Ryan in Fortune (2005)  
he article discusses on the testing of the Google television (TV) software by Google Inc., which is set for public demonstration on May 20, 2011. It says that Google TV aims to allow access to everything available on regular TV channels and internet...  
[Save reference to library](#) [Related research](#)

### Suggestions algorithmiques et humaines

- les « *Genome Project* »  
recommandations en fonction de l'« ADN » des ressources, établi (« *curated* ») par des professionnels et des traitements informatiques (cotes pour chaque « gène »)  
- modèle économique ?
- Pandora : <http://www.pandora.com>  
Music Genome Project (1999) : genre, tempo, type d'instrument, chanteur...  
- accessible uniquement aux États-Unis, Australie et Nouvelle-Zélande
- Booklamp † : <http://booklamp.org/index.php>  
Book Genome Project (2003) : structure, style d'écriture, termes...; partenariat avec des éditeurs  
(racheté par Apple pour contrer Amazon)
- Art.sy : <http://artsy.net/theartgenomeproject> [article]  
Art Genome Project (2012) : techniques, mouvement, thèmes...; partenariat avec des musées, galeries...
- Jinni : <http://www.jinni.com/>  
Movie Genome Project (2012) : genre, période, lieu...

## Moteurs de recommandation

- **moteurs émotionnels / d'envies**
  - CultureWok: <http://www.culturewok.com/>  
moteur « sensible » de contenus culturels : cinéma, jeux, livres et musique
  - Moteur d'envies du Livre de Poche : <http://www.livredepoche.com/moteur-d-envies>
    - peu de contenu
- **moteurs d'inspirations** (exemples)
  - Niice : <http://www.niice.co/>  
moteur de recherche visuelle pour les « créatifs » - recherche par mots-clés
  - moteur d'inspiration d'Havas voyages : <http://idees.havas-voyages.fr/>  
recherche par curseurs

## L'omniprésence de la personnalisation

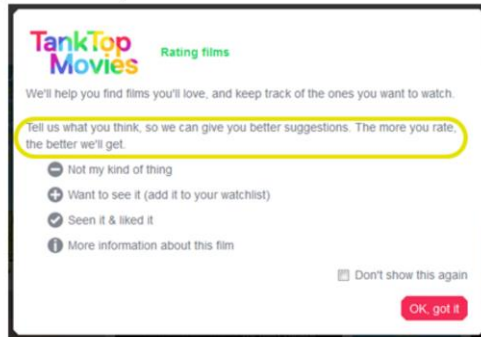
- développement de la **combinaison de différents types de services**

ex. : services associant les algorithmes et l'humain ([article](#))

ex. : moteurs de recommandations personnalisées

- nécessite de donner l'accès à des données personnelles (connexion, app. de smartphone)

- ex. : Tank Top Movies (UK) : <http://www.tanktop.tv>  
moteur de suggestions et comparateur de services de VOD



- ex. : app. Random ([article](#))  
moteur prédictif de découverte



## La recherche sociale

« Pour un bon tuyau, demandez  
à vos amis plutôt qu'à Google » [\(A.S. Moreau\)](#)

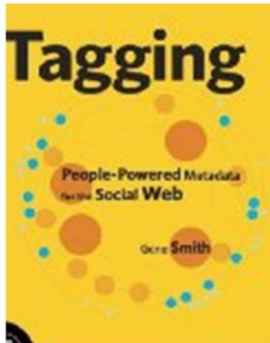
## La recherche sociale

<b>Limite des moteurs traditionnels</b>	<ul style="list-style-type: none"> <li>- moteurs surtout factuels et de généralités (recherches du quotidien)</li> <li>- peu prise en compte des réseaux sociaux (web invisible) et avis/opinions</li> <li>- certaines plateformes 2.0 non indexées par les moteurs de recherche</li> <li>- peu de visibilité des signaux faibles (pas assez importants pour être visibles dans les premiers résultats)</li> </ul> <p>→ marché potentiel important (un moteur de recherche Facebook pourrait capter <b>25 % de parts de marché</b>)</p>
<b>Type de documents</b>	<ul style="list-style-type: none"> <li>- tout type de document ou données</li> <li>- recommandations, avis, opinions</li> <li>- experts et groupes</li> </ul>
<b>Recherche</b>	<ul style="list-style-type: none"> <li>- recherche sociale : « mode de recherche qui utilise les réseaux sociaux et réseaux d'experts et est mené dans des espaces de travail partagés, ou met en jeu des techniques de <i>social data mining</i>, ou met en jeu des processus d'intelligence collective pour améliorer la recherche » (C. Deschamps, d'après E.h. Chi et B.M. Evans, 2009)</li> <li>- concept cependant un peu flou (V. Mesguich) : utilisation de moteurs de recherche classiques sur des contenus issus des médias sociaux / mode de recherche basé sur des usages sociaux (recommandation...)</li> <li>- n'a pas attendu l'essor du web 2.0 !, ex. : Ask, forums, listes de discussion...</li> <li>- peut être liée à la présence de tags (folksonomie) ou de hashtags (mots-clé marqués par #) - ex. : la recherche « Tour Eiffel » dans Google et avec un mot-clé (O. Ertzscheid)</li> </ul>
<b>Outils</b>	<p>trois types :</p> <ul style="list-style-type: none"> <li>- recherche sociale collective : moteurs de recherche proposant des résultats issus des médias sociaux (moteurs généralistes, moteurs spécifiques aux outils du web 2.0)</li> <li>- recherche sociale « <i>friend-filtered</i> » : intégration dans les résultats généralistes de résultats venant de données de membres de son réseau (partage de liens, abonnement à des services comme Twitter...)</li> <li>- recherche sociale collaborative : ex. service de questions/réponses</li> </ul>
<b>!</b>	<ul style="list-style-type: none"> <li>- qualité de l'information obtenue ? : basé avant tout sur la confiance</li> <li>- critères de classement : temps, popularité, pertinence ? ; poids des algorithmes créant une <b>fausse sérendipité</b></li> <li>- la recherche via moteur spécifique ne fonctionne pas toujours sur l'ensemble des services annoncés → importance de se reporter au service concerné directement</li> <li>- sur les réseaux sociaux : résultats de la recherche pouvant varier selon la taille du réseau et les abonnements souscrits</li> </ul>

32 % des internautes utilisent les réseaux sociaux pour trouver de nouveaux contenus en ligne contre 18% en 2010 (étude Forrester Research, 2013, <http://www.zdnet.fr/actualites/les-reseaux-sociaux-2e-de-la-recherche-web-39791699.htm>)

Sur les limites algorithmes / curation humaine, cf. C. Madsbjerg et M. Ramsey-Elliott, 2013, <http://www.fastcodesign.com/3018862/how-to-fix-streaming-music-services>

## La recherche sociale



[Library of Congress](#)



[Metadata.](#)  
[Internet --Social aspects.](#)

[Library Thing](#)



classification **folksonomies** **folksonomy** information information architecture information  
retrieval information technology internet IT libraries library 2.0 librarything **metadata** non-fiction read  
social **social bookmarking** **social media** social networking social software **social tagging** **social**  
**web** tag **tagging** **tags** taxonomy to read web **web 2.0**

exemple de contenu social : les tags (folksonomie)  
→ d'autres possibilités de recherche (cf. [O. Ertzscheid](#))

## Moteurs de recherche « traditionnels »

**pas/peu d'interrogation** des médias sociaux par ces moteurs

- Google  
[renouveau](#) du partenariat avec Twitter après un premier partenariat (2009-2011)
  - Bing social : ! en version US seulement, [exemple](#) (avec authentification au service)  
intégration de résultats de Twitter, Facebook, Quora...  
mais peu satisfaisant
  - Qwant : <http://www.qwant.com/>  
recherche « Social » : Twitter
- + extension Wajam : <http://www.wajam.com/>  
permet d'ajouter les résultats de ses propres réseaux sociaux à ceux de Google, Bing et Yahoo

## Moteurs de recherche sociale

- contenu
  - ! **nombreux dysfonctionnements** sur la plupart de ces moteurs
  - Icerocket : <http://www.icerocket.com/>  
blogs, Twitter
    - dysfonctionnements pour Facebook ?
  - Social Searcher : <http://www.social-searcher.com>  
fonctionne uniquement pour Twitter et Google +, dysfonctionnement pour les autres ?  
sous-site spécifique pour Google : <http://www.social-searcher.com/google-social-search/>
- partage d'un lien
  - LinkTally : <http://linktally.com/>  
partage sur Google +, Facebook, LinkedIn et Twitter
  - SharedCount : <http://www.sharedcount.com>  
partage sur Facebook, Twitter, Google +, Pinterest, LinkedIn et StumbleUpon
- tags (mots-clés)
  - Tagboard : <http://tagboard.com>  
moteur de recherche de tags (*hashtags*) en temps réel  
6 services : Twitter, Facebook, Instagram, Google+, Vine,  
+ possibilité de créer un tagboard personnalisé sur un tag
  - Cur.to : <http://www.cur.to>

## Moteurs de sites spécialisés

### • Twitter

- moteurs du site Twitter
  - recherche simple : <https://twitter.com/search-home>
  - recherche avancée : <https://twitter.com/search-advanced> (mots, personnes, lieux, autre)
    - intérêt des # (*hashtags*)
    - + possibilité de croiser les champs (ex. mots-clés et compte...), notamment via [opérateurs booléens](#)
    - + intervalles de dates
    - ! liste des résultats rangés par défaut par « Top » ; privilégier « Direct » pour résultats les plus récents
    - n'est pas exhaustif et parfois des dysfonctionnements
- autres moteurs spécifiques à Twitter
  - Topsy : <http://topsy.com/>
    - archives de Twitter depuis 2006, y compris les tweets supprimés si indexés
    - recherche avancée : <http://topsv.com/advanced-search> et filtres
    - + recherche de liens partagés
    - + classement des tweets en fonction des RT (*retweets*), mais possibilité de les classer par ordre antéchronologique et chronologique
    - + alertes mail
  - Hashtagify : <http://hashtagify.me/>
    - recherche de *hashtags* (#)

## Moteurs de sites spécialisés

- **références bibliographiques (niveau académique)**
  - Mendeley : <http://www.mendeley.com/>  
propriété d'Elsevier  
3 millions de membres, plutôt STM, 520 millions de références, 264 000 groupes  
[formulaire de recherche avancée](#)  
suggestions automatiques
  - Zotero : <https://www.zotero.org/>  
[moteur de recherche](#) mais peu satisfaisant  
utile surtout pour trouver des personnes et non des références bibliographiques
  - Citeulike : <http://www.citeulike.org/>  
8 M. de références  
[critères de recherche avancée](#)  
suggestions automatiques

## Moteurs de sites spécialisés

- *social bookmarking*

- Delicious : <http://www.delicious.com/>  
plateforme de favoris internet  
moteur accessible via [Help](#)
- Diigo : <https://www.diigo.com/>  
plateforme de favoris internet  
[formulaire de recherche](#)  
[formulaire de recherche de personne](#) : champs, opérateurs  
! pour des termes associés, construire sa requête ainsi eau\_potable ou « eau potable »  
ou encore **eaupotable**
- StumbleUpon : <http://www.stumbleupon.com/>  
moteur de découvertes, recommandant des pages en fonction des intérêts de  
l'internaute
  - nécessite une authentification

complémentarité avec les moteurs de recherche  
ex. : la recherche « veille » dans Google et Diigo (T. Faragasso)

## Moteurs de sites spécialisés

- *social news*

sites communautaires permettant de voter pour une page proposée par un utilisateur

- Digg : <http://digg.com>
- Reddit : <http://www.reddit.com/>

- plateformes de contenus

plateformes permettant aux internautes de mettre du contenu en ligne

- textes : Scribd (<http://www.scribd.com>), Calaméo (<http://www.calameo.com>)
  - multimédia : Youtube (<https://www.youtube.com/>), Flickr (<http://www.flickr.com>), rech. avancée : <http://www.flickr.com/search/advanced/>
  - supports d'interventions : Slideshare (<http://www.slideshare.net>)
  - curation de contenus : Scoop.it (<http://www.scoop.it>), Storify (<http://storify.com>), Pinterest (<https://www.pinterest.com/>), Paper.li (<http://paper.li>)
  - documents scientifiques : Figshare (<http://figshare.com/>), myExperiment (<http://www.myexperiment.org/>)
- + complémentarité avec les moteurs de recherche - ex. : [P. Cuenot](#)  
- moteurs de recherche souvent peu satisfaisants (non accessibles sur la page d'accueil, sans fonctionnalités avancées ou filtres...)  
→ passer par les moteurs généralistes (Google...) avec des recherches de type [XXXX site:YYY]

## Moteurs de sites spécialisés

- réseaux sociaux

réseaux sociaux généralistes

- Facebook : <http://www.facebook.com/>
- Google + : <https://plus.google.com/>

réseaux sociaux professionnels

- LinkedIn : <https://fr.linkedin.com>
- Viadeo : <http://fr.viadeo.com>

réseaux sociaux académiques

- Academia : <http://academia.edu/>
- ResearchGate : <http://www.researchgate.net/>

permettent de trouver de l'information (références, documents, forums, etc.) et des experts...

mais :

- moteurs de recherche parfois inaccessibles sur la page d'accueil
- résultats parfois non exhaustifs et pouvant varier selon les abonnements souscrits et les comptes suivis

→ passer par les moteurs généralistes (Google...) avec des recherches de type [XXXX site:YYY]  
ou [XXX inurl:YYY], [exemple](#)

## Sites spécialisés

- **questions-réponses**

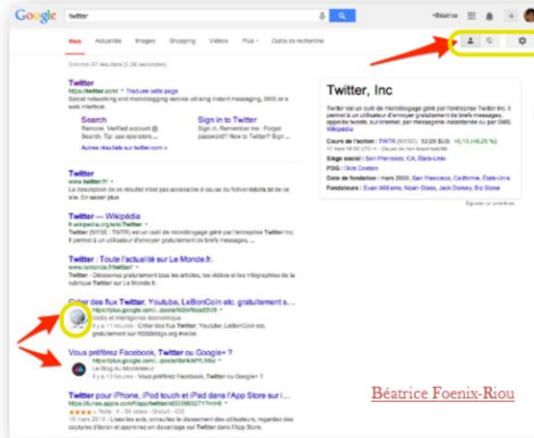
services avec une histoire déjà ancienne, ex. : [Ask.com](http://ask.com) (cf. C. Tisserand-Barthole, *Netsources*, 108)

+ possibilité d'interroger en langage naturel

- Quora : <https://www.quora.com/>  
membres évalués (système de « crédits ») – insiste sur la qualité des réponses  
possibilité de s'abonner à des comptes  
anglais seulement
  - nécessite une authentification, mais résultats indexés par Google
- StackExchange : <http://stackexchange.com/sites#>  
but : proposer des pages de références  
nombreuses thématiques différentes
- ChaCha : <http://www.chacha.com/>  
[étude de 2013](#) : meilleur service de questions-réponses face Google, Bing et Siri  
réponse assurée par une communauté de « Guides », sélectionnés et payés  
catégories  
anglais seulement
- Yahoo! Answers : <http://answers.yahoo.com/> [version fr. : <https://fr.answers.yahoo.com/>]  
librement interrogeable  
12 langues  
catégories  
membres évalués (points et niveau)  
sur des sujets professionnels, utiliser l'interface internationale
- Wikianswers : <http://answers.wikia.com/wiki/Wikianswers>

## La recherche « *friend filtered* »

Google Search plus Your World : Google +, Facebook, Twitter...



dès identification sur un service Google  
possibilité de faire apparaître les résultats personnels ou non



Masquer les résultats privés

### Résultats « personnels » :

- *personal results* : messages postés sur Google + par les personnes suivies ;
- pages signalées (via bouton +1 ou postées sur Google +) ;
- profils Google +

# La recherche « friend filtered »

Bing: Facebook, Twitter, LinkedIn...

The image shows a Bing search results page for the query "costa rica". The search bar at the top left contains the text "costa rica" and shows "378,000,000 RESULTS". The main content area displays several search results, including "Costa Rica All-Inclusive" from vacationscostarica.com, "costa rica" from cheapcaribbean.com, "Costa Rica Travel Guide" from bing.com/travel, and a Wikipedia entry for "Costa Rica". On the right side of the page, there is a social media sidebar. At the top of the sidebar, it says "Danny" with a profile picture and "25" friends. Below this, there is a search bar for friends and a list of "FRIENDS WHO MIGHT KNOW" with profile pictures and names. At the bottom of the sidebar, it says "PEOPLE WHO KNOW" and "ACTIVITY". A red arrow points from the search results area to the social media sidebar.

WEB IMAGES VIDEOS MAPS PLACES NEWS MORE

bing Preview

costa rica

378,000,000 RESULTS

**Costa Rica All-Inclusive**  
www.vacationscostarica.com · Worry-Free, Luxury Vacations to tropical Costa Rica We're Experts!

**costa rica**  
www.CheapCaribbean.com/CostaRica · Unbelievable Deals to Costa Rica! Book Now.

**Costa Rica Travel Guide**  
bing.com/travel  
Map · Flights · Attractions  
Explore destinations, attractions, and photos. Find great deals on flights and hotels.

**Costa Rica - Wikipedia, the free encyclopedia**  
en.wikipedia.org/wiki/Costa\_Rica ·  
History · Politics · Geography · Climate · Economy · Foreign relations  
Costa Rica is a country in Central America, bordered by Nicaragua to the north, Panama to the southeast, the Pacific Ocean to the west, and the Caribbean Sea to the east.

**Costa Rica - Travel, Real Estate, Relocation & Business**  
costarica.com ·  
Costa Rica! Up-to-date information on travel, relocation, business, living, real estate and investment in Costa Rica.

**Costa Rica All-Inclusive**  
CostaRicaResorts.com  
Save up to 50% on all-inclusive Costa Rica vacations.

**Free Costa Rica Fact Sheet**  
www.InternationalLiving.com  
For people considering Living, Visiting, or Retiring in Costa Rica

**Own a Home in Costa Rica**  
www.costaricainvest.ie  
Find the Hottest Property Listings & properties for You

**Costa Rica Home Listings**  
costa rica real estate  
largest property database with map daily posted by owners and realtors

**Costa Rica Villas**  
costa-rica.caribbeanway.com  
Exclusive Vacation Rentals View Our Extensive Catalogue Online  
See your message here

RELATED SEARCHES

Danny 25

Ask friends...

FRIENDS WHO MIGHT KNOW

Patti Bailey likes Health Check Costa Rica: Worry Free Medical Travel to Costa Rica

Frank Graber has photos of costa rica

Joseph Hunkins likes Visit Costa Rica - The Costa Rica Tourist Board

Jothan Frakes likes Visit Costa Rica - The Costa Rica Tourist Board

Lee Mills likes Costa Rica!

PEOPLE WHO KNOW

Lisbeth Arelys Carmona Rivas knows about costa rica

ACTIVITY

Danny Sullivan

## La recherche collaborative



« Quand on cherche quelque chose de compliqué, c'est à dire quelque chose que l'on sait que l'on ne trouvera pas directement sur Google, ou alors au prix d'une reformulation complexe en différents mots-clés séparés entre eux par des opérateurs variés autant que booléens, bref quand on a une question longue en langage naturel et qu'on a la flemme ... on va poser sa question sur Twitter. Et on obtient en général presque immédiatement la réponse. [...] Twitter est le nouveau Google. [...] Mais Twitter n'est pas que le nouveau Google. »

(O. Ertzscheid)

## La recherche sociale

	Potentiels	Limites
<b>formulation de la requête</b>	possibilité de faire des requêtes en langage naturel nécessité de réfléchir à la formulation	nécessité d'interroger les bons réseaux ou personnes pas toujours possibilité d'affiner la question au fur et à mesure
<b>type d'informations</b>	découverte de nouveaux domaines recherche d'avis sérendipité (= découverte et non plus recherche) possibilité de choix personnel de sources fiables	information « poussée » vers l'utilisateur
<b>type de résultats</b>	meilleure connaissance du contexte de la requête (tacite) individu qui donne sa valeur à l'information	recherche nécessitant parfois du temps résultats variant suivant le réseau ou les abonnements souscrits
<b>qualité de l'information</b>	informations fraîches (cf. comparaison avec bases de données professionnelles) et non pas seulement populaires, traitées par un algorithme	critères de classement des résultats ? ! de popularité / autorité parfois réponses courtes ou hors-sujet (entrer en contact) résultats en temps réel, non validés et produisant du bruit (conversation)
<b>réseau</b>	possibilité de développer un réseau	effet de « filter bubble » (E. Pariser) : on a plus confiance dans les membres de son réseau et informations peuvent être biaisées peur de passer pour un incompetent

d'après C. Deschamps, A. Serres et C. Tisserand-Barthole

TABEAU 2 – POTENTIELS D'USAGE ET LIMITES DES STRATÉGIES-TYPES DE RECHERCHE SOCIALE D'APRÈS EVANS ET EVANS [25] ET EVANS, KAIRAM ET PIROLLI [28]

	Collective social search	Friend-filtered social search	Collaborative social search
Potentiels d'usage	- Recherches préliminaires ou exploratoires - Découverte de nouveaux domaines - Recherche d'une solution - Recherche sans objectif précis -> on ne sait pas à quels types de résultats s'attendre	- Recherche de recommandations, de conseils, d'avis -> on accorde aux membres de son réseau un niveau de confiance supérieur à ceux qui ne le sont pas	- Processus itératif qui va permettre d'améliorer ou reformuler sa question - Réponse à des questions où le tactile est important : comment faire ceci -> on engage une conversation - Un réseau social proche (famille, amis) connaît mieux votre contexte et peut répondre en conséquence
Limites	- Résultats fournis par l'agrégation de données provenant de personnes que l'on ne connaît pas -> confiance limitée - Résultats en temps-réel qui peuvent entraîner du bruit (résultats sportifs, buzz, etc.) sur des questions de fond	- Les membres de notre réseau peuvent ne pas s'intéresser au domaine auquel notre besoin d'information est lié. On ne trouvera donc rien d'intéressant dans leurs contenus	- Une session de questions-réponses demande du temps - Risque d'en demander trop à la communauté -> crainte de devoir « rembourser » en temps passé - Crainte du demandeur que la question posée le fasse passer pour incompetent

« Le bilan semble donc mitigé pour ce qui est de la recherche sociale. D'une part (1) les utilisateurs y trouvent moins de résultats pertinents que dans les moteurs classiques, d'autre part (2) leur niveau de connaissance sur la question posée est moins élevé à l'issue de la recherche – ce qui semble logique par rapport au point (1) » (C. Deschamps, <http://fr.scribd.com/doc/40837799/Veille-et-recherche-d-information-a-l-heure-des-reseaux-sociaux>).

«En matière de recherche et de veille dans le domaine scientifique, il faudra donc penser à utiliser :

- les serveurs professionnels qui continuent d'offrir, moyennant finance, un contenu riche et varié, validé et fiable, structuré, avec des fonctionnalités de recherche extrêmement perfectionnées ;

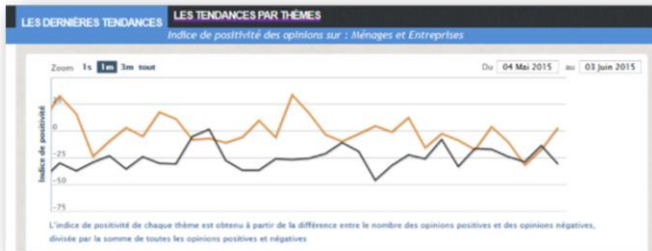
mais aussi les outils du web 2.0, majoritairement gratuits, qui fournissent des contenus très récents (quelquefois en temps réel), permettent de tirer parti de l'intelligence collective, facilitent l'accès aux documents en texte intégral et surtout permettent d'identifier relativement aisément des spécialistes et experts, malgré des possibilités de recherche à des années lumières de celles proposées par les serveurs » (C. Tisserand-Barthole, *Bases*, n°206).

## Les moteurs collaboratifs

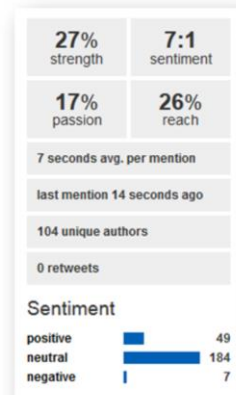
- Blippex : <https://www.blippex.org/>  
moteur de recherche *crowdsourced*: prend en compte l'activité des internautes (temps passé sur une page, etc.) pour classer les résultats de recherche  
nécessite d'installer l'extension de navigateur spécifique  
ne conserve pas les données personnelles  
encore en développement (quelques millions de pages référencées seulement)  
algorithme public afin de permettre une amélioration collaborative
  - ✦ filtre de date pour informations récentes (*seen in the last 30 days*)
  - ✦ filtre en fonction du temps passé sur les pages (*dwell factor*)
- Faroo : <http://www.faroo.com/>  
moteur de recherche en *peer-to-peer*: indexation et distribution sur le réseau des sites web visités par les internautes  
résultats classés en fonction de l'activité des internautes (temps passé sur une page, pages imprimées, etc.)
  - pas de fonctionnalités avancées ou de filtres
- SearchTeam : <http://searchteam.com/>  
moteur de recherche collaboratif - [présentation](#)  
possibilité de faire des recherches à plusieurs sur le web : recherche, partage et enrichissement
  - © 2011

## L'analyse de sentiments

Vers un « *Mood Graph* » ([E. Salinger](#)) ?



[Ecorama \(BFM Business\)](#) : suivi du moral des chefs d'entreprises et des particuliers français, à partir de messages sur les médias sociaux



Ex. d'une requête sur [Socialmention.com](#)

étude des réseaux sociaux pour y trouver des tendances, notamment Twitter ([A. Pak, 2011](#))  
→ fiabilité et efficacité de ce genre de services ?, ex. termes choisis dans l'[Hedonometer](#)

« *Mood Graph* » :

- part de plus en plus importante de l'expression d'émotions et d'humeurs
  - développement de services de traitement automatique des sentiments
- ex. [Rappler](#) : plateforme de news indiquant le ressenti de l'internaute après la lecture d'un article
- sentiments qui permettent de classer les résultats
  - plus large que le simple Like de Facebook ou le +1 de Google

[Ecorama](#) : établit un « indice de positivité » à partir de la mesure économique sur les réseaux sociaux



La recherche temps réel

## La recherche temps réel

<b>Limite des moteurs traditionnels</b>	<ul style="list-style-type: none"> <li>- pas d'indexation des contenus en temps réel : l'utilisateur interroge un index dont la mise à jour par des robots est variable selon les contenus (cf. erreur 404)</li> <li>- développement du contenu en temps réel sur internet (informations, réseaux sociaux, syndication de contenu), d'où recherche d'instantanéité             <ul style="list-style-type: none"> <li>→ développement d'outils facilitant l'indexation par les moteurs en temps réel (<a href="#">PubSubHubbub et RSS Cloud</a>)</li> <li>→ développement de moteurs de recherche « temps réel » ou moteurs de recherche synchrones : récent : une 20ne en 2009, mais recherche de modèles économiques viables et fonctionnalités</li> </ul> </li> </ul>
<b>Type de documents</b>	<ul style="list-style-type: none"> <li>- tout contenu récent (parfois quelques secondes ou quelques minutes), notamment réseaux sociaux et flux [informations (<i>breaking news</i>), conversations, mises à jour de blogs ou de statuts de réseaux sociaux, recommandations]</li> <li>- tendances</li> <li>- personnes</li> </ul>
<b>Recherche</b>	<p>« La recherche temps réel [<i>real time search</i>] peut être définie comme le fait d'accéder à des mises à jour, des informations, des articles de journaux, des billets de blogs, et toutes sortes de contenu au moment où il est publié sur le web, instantanément. Il s'agit d'accéder à l'information au moment où elle est rendue accessible » (<a href="#">W. Boswell</a>)</p>
<b>Outils</b>	<ul style="list-style-type: none"> <li>moteurs de recherche traditionnels</li> <li>moteurs de sites particuliers (ex. réseaux sociaux, Twitter)</li> <li>moteurs spécifiques temps réel, généralement sociaux, éventuellement spécialisés (images...)</li> <li>outils de veille et de suivi de l'e-reputation</li> <li>+ critères supplémentaires (<a href="#">C. Deschamps</a>) : représentation graphique, création de flux RSS, défilement de l'information en temps réel...</li> </ul>
<b>!</b>	<ul style="list-style-type: none"> <li>vrai temps réel ou temps différé : actualisation ?</li> <li>fiabilité de l'information (! effet de buzz, cf. piratage du compte Twitter de l'Associated Press, 04/2013)</li> <li>classement des résultats : pertinence, popularité, date ?</li> <li>sources d'information et types de médias ?</li> <li>exhaustivité ? antériorité des résultats ?</li> </ul>

« *Real Time Search can be defined as accessing updates, information, news articles, blog posts, and all sorts of content as it is published on the Web, instantaneously. It is all about access to information as it is made available* » (W. Boswell, <http://websearch.about.com/od/socialnetworking/p/What-Is-Real-Time-Search.htm>)

## Moteurs de recherche « traditionnels »

**développement de l'indexation** du temps réel par ces moteurs (web, actualités et médias sociaux)

- **Google**
  - web : + moteur de recherche Google : filtre « moins d'une heure »
  - actualités : Google Actualités, cf. *infra*
  - social : renouveau du partenariat avec Twitter après un premier partenariat (2009-2011)
  - tendances de recherche : <http://www.google.com/trends/>
- **Bing**
  - pas de filtre de date sur les résultats web
  - rubrique « Actualités » : possibilité de classer par « les plus récents »
  - mauvaise indexation des réseaux sociaux
- **Qwant** : <http://www.qwant.com/>
  - recherche « Social » : Twitter

## Moteurs de recherche sociale

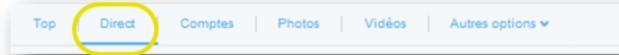
[reprise de *supra*]

- contenu
  - ! **nombreux dysfonctionnements** sur la plupart de ces moteurs
  - Icerocket : <http://www.icerocket.com/>  
blogs, Twitter
    - dysfonctionnements pour Facebook ?
  - Social Searcher : <http://www.social-searcher.com>  
fonctionne uniquement pour Twitter et Google +, dysfonctionnement pour les autres
- partage d'un lien
  - LinkTally : <http://linktally.com/>  
partage sur Google +, Facebook, LinkedIn et Twitter
  - SharedCount : <http://www.sharedcount.com>  
partage sur Facebook, Twitter, Google +, Pinterest, LinkedIn et StumbleUpon
- tags (mots-clés)
  - Tagboard : <http://tagboard.com>  
moteur de recherche de tags (*hashtags*) en temps réel  
6 services : Twitter, Facebook, Instagram, Google +, Vine,
    - + possibilité de créer un tagboard personnalisé sur un tag
  - Cur.to : <http://www.cur.to>

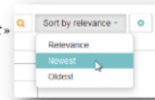
# Moteurs de sites spécialisés

## Twitter

- moteurs du site Twitter
  - recherche simple : <https://twitter.com/search-home> et recherche avancée : <https://twitter.com/search-advanced>
    - ! intérêt des # (*hashtags*) et des LT (*livenweets*, suivi d'événements en direct), ex. [#CNESweetup](#)
    - ! liste des résultats rangés par défaut par « Top » ; privilégier « Direct » pour résultats les plus récents
    - n'est pas exhaustif et parfois des dysfonctionnements



- autres moteurs spécifiques à Twitter
  - Topsy : <http://topsy.com/>
    - ! classement des tweets en fonction de leur *relevance*, mais possibilité de les classer par « Newest »
  - Trendsmap : <http://trendsmap.com/>
    - carte de tendances temps réel géolocalisées ; voir également <http://trends24.in/>
  - Hashtagify : <http://hashtagify.me/>
    - recherche de *hashtags* (#)
- autres outils spécifiques (*dashboards*)
  - Tweetdeck : <https://tweetdeck.twitter.com/>



## Type de documents


- actualités (1)

- ! **moteurs de recherche temps réel ≠ moteurs de recherche d'actualités (agrégation) !**

- actualités internationales et locales

Google actualités : <http://news.google.com>

« Google Actualités est un site d'actualités généré par ordinateur qui recueille les grands titres provenant de plus de 500 sources d'actualités en langue française dans le monde entier, regroupe les articles par thème et les affiche en fonction des intérêts de chaque utilisateur »

- + possibilité de personnaliser les résultats 
- + nombreuses versions locales (+ 72 éditions, 30 langues, 55 000 sources)
- + alertes et flux RSS

NewsLookup : <http://www.newslookup.com/>

agrégateur de titres du monde entier  
très orienté Etats-Unis, mais possibilité d'éditions régionales par grandes aires géographiques

Hubii : <http://www.hubii.com/>

moteur de recherche d'actualités géolocalisées

## Type de documents

- actualités (2)

- actualités européennes

Eufeds : <http://www.eufeds.eu/fr>

agrégateur de plus de 1 000 journaux (nationaux et locaux), mis à jour toutes les 20 minutes

NewsBrief : <http://emm.newsbrief.eu/NewsBrief/clusteredition/fr/latest.html>

projet du *Joint research centre* européen

+ 4 000 sites

dépêches d'actualités, mises à jour toutes les 10 min et classées automatiquement

✦ nombreux filtres (langues, localisation, thèmes...)

✦ graphiques de tendances, clusterisation...

à compléter par NewsExplorer : <http://emm.newsexplorer.eu/NewsExplorer/home/fr/latest.html>

- actualités francophones

WebPlanete news : <http://news.webplanete.net/>

867 sources d'informations francophones

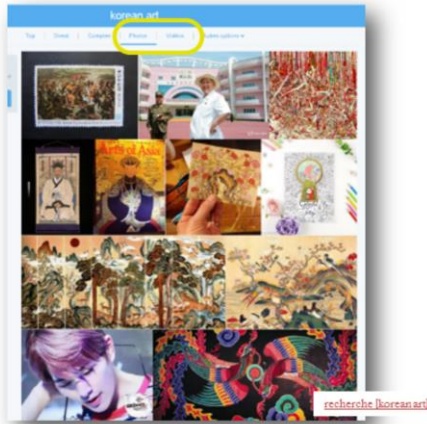
grandes thématiques

## Type de documents

- images et vidéos

essentiellement via Twitter

- Twitter : recherche (simple : <https://twitter.com/search-home> ou avancée : <https://twitter.com/search-advanced>), puis filtre



- Topsy : photos : <http://topsy.com/photos> et vidéos : <http://topsy.com/videos> (filtres *latest results* et *past 1 hour*)



## La recherche inversée

« Avec les années je me suis habitué  
à faire des recherches à partir  
d'un texte » [\(Ballinck\)](#)

## La recherche inversée

<b>Limite des moteurs traditionnels</b>	- recherche textuelle - prennent peu en compte l'information non écrite, alors que développement des pratiques audio et vidéo ( <a href="#">51 % du trafic internet US en 2010</a> )		
<b>Type de documents</b>	documents qui ne comportent pas de chaînes de caractères pour en caractériser le contenu (hors URL, nom de fichiers et métadonnées, tags) : images, multimédia		
<b>Recherche</b>	<p>recherche « inversée » : à partir de son (<a href="#">voice search</a> et autres), d'image inanimée (<a href="#">visual search</a>) ou animée et non à partir de texte cf. attentes des internautes (<a href="#">2009</a>)</p> <table border="1"> <tr> <td><b>TYPES DE DONNÉES SOUMISES</b> Quel genre de fichiers est transmis au moteur de recherche ?</td> <td>                 Soumettre une photo / image (portrait, objet, endroit...) pour que le moteur trouve l'original ou des informations liées                  Soumettre un fichier audio (chanson, histoire...) pour que le moteur trouve l'original ou des informations liées                  Soumettre un fichier vidéo (film...) pour que le moteur trouve l'original ou des informations liées             </td> </tr> </table> <p>ex. d'utilisations :</p> <ul style="list-style-type: none"> <li>- trouver des informations sur une image (origine, lieu), une musique (compositeur...), un film (réalisateur, acteurs...)</li> <li>- identifier les personnes sur une photo</li> <li>- identifier les différentes utilisations et réutilisations d'une image, par exemple dans le cadre de recherches pour des questions de droit</li> <li>- intérêt pour les langues difficiles à écrire (idéogrammes...)</li> <li>- vérifier des informations (<a href="#">images retouchées...</a>)</li> </ul> <p>est parfois associé à des recherches en temps réel ou textuelles</p>	<b>TYPES DE DONNÉES SOUMISES</b> Quel genre de fichiers est transmis au moteur de recherche ?	Soumettre une photo / image (portrait, objet, endroit...) pour que le moteur trouve l'original ou des informations liées Soumettre un fichier audio (chanson, histoire...) pour que le moteur trouve l'original ou des informations liées Soumettre un fichier vidéo (film...) pour que le moteur trouve l'original ou des informations liées
<b>TYPES DE DONNÉES SOUMISES</b> Quel genre de fichiers est transmis au moteur de recherche ?	Soumettre une photo / image (portrait, objet, endroit...) pour que le moteur trouve l'original ou des informations liées Soumettre un fichier audio (chanson, histoire...) pour que le moteur trouve l'original ou des informations liées Soumettre un fichier vidéo (film...) pour que le moteur trouve l'original ou des informations liées		
<b>Outils</b>	moteurs généralistes (Google) moteurs spécifiques		
<b>!</b>	<p>domaine encore en développement :</p> <ul style="list-style-type: none"> <li>- techniques de reconnaissance d'image actuelles : <a href="#">25% de succès seulement</a> (homme : 69%)</li> <li>- nécessité d'indexer le contenu de ces documents automatiquement (ex. pour le son : TAL : traitement automatique de la langue ; pour l'image : <a href="#">recherches de Google</a>) ou humainement (documentalistes : projet <a href="#">Sync Notes par l'INA</a> pour les <a href="#">JT</a> ; internautes <a href="#">pavés ou non</a>)</li> </ul> <p>enjeux commerciaux (<a href="#">livres, magazines, vêtements...</a>) et <a href="#">sociétaux</a></p>		

quel est le point commun entre

**Déterminer l'âge d'une personne sur une photo, How-Old**

Déterminer l'âge d'une personne qui est sur une photo ne va pas toujours de soi, surtout quand on n'est pas très physionomiste. Pour aider ceux et celles ayant ce genre d'interrogations existentielles, Microsoft vient de mettre à disposition un service en ligne appelé **How-Old** censé déterminer automatiquement l'âge et le sexe d'une ou plusieurs personnes apparaissant sur une photo. L'outil, présenté comme une démonstration technologique du savoir-faire des ingénieurs de Microsoft, s'appuie sur un algorithme de reconnaissance faciale pour faire ses calculs et afficher ses estimations en seulement quelques secondes.

**How-Old** s'utilise depuis un navigateur Web quelconque à partir d'un des exemples des photos des personnes saines ou qui sont en groupe qui vont servir de cobayes à la démonstration. Il est ainsi possible de rechercher une photo d'une personne depuis la zone de saisie pour effectuer une requête de recherche sur Bing. Il ne reste plus qu'à choisir une photo d'une personne parmi les résultats trouvés en cliquant sur le bouton **Use This Photo**. Si le logiciel de reconnaissance identifie un visage humain, la réponse est quasi instantanée.



*#Microsoft en 2015*



**Google veut compter les calories sur vos photos de nourriture**



Un sandwich bœuf et fromage dans un pain blanc avec des légumes. Saucis, accompagnement.

M.C. **Google** a pu compter les calories sur vos photos de nourriture. **Google** a pu compter les calories sur vos photos de nourriture.

L'entreprise regarde votre smartphone et dit ce que vous mangez. Grâce aux images, elle peut identifier les aliments et les calories. Google a pu compter les calories sur vos photos de nourriture.

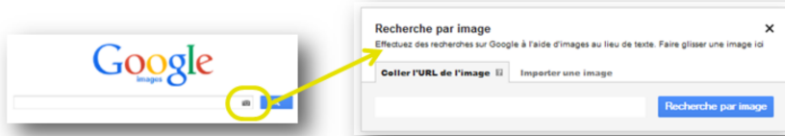
Pour en savoir plus sur **Google** et ses produits, visitez [www.google.com](http://www.google.com).

*#Google en 2015*

Reconnaissance d'images couplée à des **bases de données**, éventuellement (dans le cas de How-old) associée à des systèmes autoapprenants (*deep machine learning*) ex. avec <https://www.imageidentify.com/> de Wolfram (J. Lowensohn, 2015, <http://www.theverge.com/2015/5/13/8603531/wolfram-image-identification-site-trained-by-chewbacca>)

- la recherche par image

- Google images : <http://www.google.fr/imghp?hl=fr&tab=wi>



- + possibilité de glisser-déposer l'image
- + hypothèse d'identification, proposition d'images similaires et identiques
- + possibilité de croiser la recherche par image et la recherche textuelle


- TinEye : <http://www.tineye.com>

11,4 MM. d'images (issues du web, mais aussi d'autres collections : GettyImages, stocks...)

- + peut retrouver l'image même retouchée ou recadrée

- + complémentaire de Google images (possibilité d'utiliser l'extension du navigateur Chrome [Fast Image Research](#) pour faire une recherche combinée sur les deux sites)

- Bing images : <http://www.bing.com/images?FORM-Z9LH>

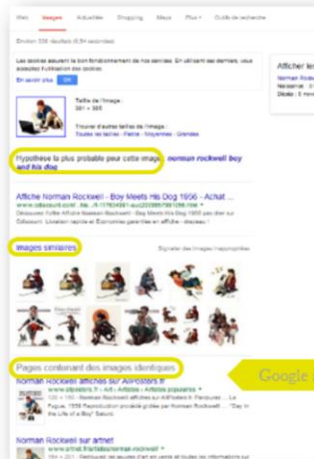
« correspondance d'image »  Correspondance d'image

- moins de résultats et de fonctionnalités qu'avec les deux précédents

[Encore ?](#)

Enjeu important : Flickr : 3 100 photos téléchargées / min.

- la recherche par image - exemples d'utilisations



Google images



Mike Licht, *Tiro* blogger after Norman Rockwell



TinEye

- la recherche par image - exemples d'utilisations



« Repérer les retombées presse d'une exposition grâce aux images »  
[A. Vahonne](#)



« A la recherche d'un singe photographe »  
[P. Pecatte](#)

Voir également  
« Du bon usage des photos de stock », [P. Pecatte](#)



**Exemple 1** : retrouver les retombées presse et web, au niveau mondial, de l'exposition *La petite veste noire*, ayant eu lieu à Tokyo (2012) et qui est accompagné d'un livre de photos portant le même titre  
problème central : question de la langue

→ intérêt de la recherche par l'image (celle de l'affiche et celle de la couverture du livre) : permet d'interroger sans maîtriser la langue

**Exemple 2** : documenter une image brute : retrouver son origine, le contexte d'apparition et toute autre information (date, photographe, publication...), voire sa réutilisation  
problème central : photo abondamment reprise, voire modifiée

→ intérêt de la recherche par l'image, couplée à une recherche textuelle classique

- la recherche d'images par similarité et couleurs
  - CultureCam : [http://culturecam.eu/?shared=1427196094594\\_a2d5d1448150f49807caa3036bdd7bde](http://culturecam.eu/?shared=1427196094594_a2d5d1448150f49807caa3036bdd7bde)  
images Europeana
  - PicsLikeThat : <http://www.picslikethat.com/>  
images Fotolia, recherche par similarité
  - MulticolorEngine (TinEye Labs) : <http://labs.tineye.com/multicolor>  
10 M. d'images Flickr en CC, recherche par couleur
    - ✦ possibilité d'indiquer jusqu'à 5 couleurs
    - ✦ lien vers Flickr pour obtenir plus d'informations
  - Chromatik (Exalead) : <http://chromatik.labs.exalead.com/>
    - ✦ fonctionnalités de recherche avancée (lieu, licences, tags...)
    - ✦ possibilité d'indiquer plusieurs couleurs (possibilité également de croiser plusieurs couleurs à la fois directement à partir du moteur de recherche images : <http://www.exalead.com/search/image/>, recherche avancée)
    - référence des images ? (Flickr ?)
  - Krazydad : <http://krazydad.com/colrpickr/>  
d'après une sélection d'images sur Flickr
- la recherche d'images par dessin
  - Search Google by drawing : <http://search-by-drawing.fullstackoptimization.com/>
  - Retrievr : <http://labs.systemone.at/retrievr>  
recherche dans une sélection d'images Flickr

- développement de la recherche visuelle ([M. Boland, 2010](#))
  - axes :
    - recherche par codes
      -  code barres
      -  QR code
    - traitement des images (« redocumentarisation »)  
ex. Google Goggles ([démonstration](#))
    - réalité augmentée  
[exemples](#)
  - moyens
    - accessibles par lecteurs de codes barres, téléphones mobiles, *smartphones* et webcam
    - développement avec les *smartphones*
  - limites
    - exemples surtout de la vie quotidienne (transports, commerce)
    - question des éléments techniques et des données disponibles, mais intérêt aussi pour la recherche, cf. [app. Pl@antNet](#) pour les plantes

**Recherche visuelle** (M. Boland, <http://blog.kelseygroup.com/index.php/2010/01/07/when-and-how-will-visual-search-arrive/>, 2010) :

- lecteurs de QR code
- lecteurs de code barres [UPC : *Universal Product Code*]
- traitement des images : ex. : Google Goggles (Google Mobile App, <http://www.google.com/mobile/goggles/#text>, Android et iPhone) : photos prises sont comparées à des bases de données d'images pour identifier des objets (paysages, livres, vin, œuvres d'art, produits), ex. de démonstration : <http://www.youtube.com/watch?v=ezc108DTaug> ; avec éventuellement géolocalisation et intelligence artificielle
- réalité augmentée : présentation du monde physique via l'écran d'un appareil (smartphone, tablette...) auquel un système informatique superpose en temps réel des informations 2D ou 3D (informations, éléments commerciaux...), cf. M. Boland, 2010, <http://searchenginewatch.com/article/2064238/Mobile-Search-Time-for-a-Hearing-and-Vision-Checkup>; ex. *plane finder AR* qui combine réalité augmentée, géo-localisation et étude des vols par avion

mais articles *Search Engine Watch*, 2010 :

- développement avec développement des *smartphones* et applications
- surtout des exemples commerciaux ou de réalité augmentée
- question centrale des données disponibles

- la recherche par sons

- Midomi : <http://www.midomi.com/>  
recherche de musique  
+ recherche textuelle et recherche vocale

d'autres possibilités également du côté des app.  
ex. [Shazam](#), [SoundHound](#), [Sound search](#), [encore ?](#)  
et qui intéressent de grands groupes comme [Facebook](#) ou [Free](#)

Base : fréquence + amplitude + tempo

- la recherche vocale

- intérêt

- faire des recherches par la parole
    - obtenir des réponses
    - obtenir des recommandations

- outils

- développement avec les *smartphones* - ce sont des téléphones après tout
  - ex. app. Siri (Apple)
  - app. Google Voice search

- développement sur les navigateurs

- ex. Google Chrome - *démonstration*



- limites : traitement automatique de la langue

- quelle compréhension du langage naturel ? (= *searchese* = de L. Gannex), problème de l'accent

- ex. d'un *comparatif Siri / Google search (en anglais)*

- apprentissage du service pour s'améliorer (reconnaissance vocale)

- poursuite de la recherche dans ce domaine

**Recherche vocale** : plus simple que recherche visuelle (cf. Marissa Mayer, 2009, <http://www.guardian.co.uk/technology/2009/jul/08/google-search-marissa-mayer>)

Déjà existence de moteurs parlants : ex. : Speegle qui reprenait les résultats de Google avec une synthèse vocale

- la recherche vidéo

enjeu : indexer le contenu multimédia (texte, son, séquences...)

recherche « *speech to text* »

- Blinkx : <http://www.blinkx.com/>  
association des techniques de reconnaissance vocale et visuelle, établissement d'idées et de thèmes - affirme disposer du « *most advanced video engine* »  
index de plus de 35 M. d'heures interrogeables et plus de 800 partenaires médias
  - assez commercial et très orienté États-Unis
  - manque d'informations sur les informations permettant le classement - pas de recherche avancée
- TV News search and borrow : <http://archive.org/details/tv>  
initialement prévu pour « aider les citoyens à mieux comprendre les questions et les candidats aux élections américaines de 2012 »  
435 000 *broadcasts* (San Francisco et Washington DC) depuis 2009, mise à jour journalière
  - + reconnaissance des entités nommées
- Voxlead : <http://voxleadnews.labs.exalead.com> (arrêté en 01/2014, mais toujours accessible)
  - + 9 langues disponibles
  - + pour la France : radios et télévisions (informations)
  - + reconnaissance des entités nommées
  - + mise en contexte du terme recherché
  - transcription automatique laissant parfois à désirer
  - critère de classement des résultats ?
  - mise à jour de l'index ?

Enjeu important : YouTube : 100 h. de vidéos téléchargées / min.

recherche de contenu par indexation des vidéos



- reconnaissance faciale
- reconnaissance de chansons dans un clip
- reconnaissance d'objets vidéo

} brevets Google  
2011-2012

- reconnaissance de séquences, de personnes, d'objets, de lieux (cf. images)
- quelles conséquences pour la vie privée ? (A. Chéron)

**Pas nouveau** : déjà des moteurs précédents : Viewdle † : moteur de reconnaissance vidéo : analysait les vidéos image par image et ajoutait des informations contextuelles en convertissant la voix en texte  
**Google**, *Automatic large scale video object recognition* : projet d'un référentiel de noms de 50 000 objets

cf. déjà travaux sur des images de chats (cf. L. Shin, 2012, <http://www.smartplanet.com/blog/science-scope/google-brain-simulator-teaches-itself-to-recognize-cats/12969>)

repérage de formes dans la vidéo avec variables (formes, couleurs, luminosité, vitesse de déplacement...)

→ permettre d'avoir des informations sur les contenus d'une image ou d'une vidéo (qui sont les acteurs, le titre d'une musique, l'adresse d'un lieu...)

+ développement des travaux sur l'intelligence artificielle

développement prévisible avec les TV connectées (K. Freeman, 2013, <http://mashable.com/2012/01/31/video-search-future/>)

## Vers une convergence des techniques



Ex. *Google glasses* avec reconnaissance vocale  
et reconnaissance visuelle

Ex. *Blippar*

*Google glasses* et ce genre de services nécessitent de **développer les différents types de recherche** non textuelles : son, image (notamment reconnaissance faciale) mais également un gros potentiel commercial (cf. J. O'Brien, 2013, <http://mashable.com/2013/03/12/visual-search/>)



## Pertinence des réponses

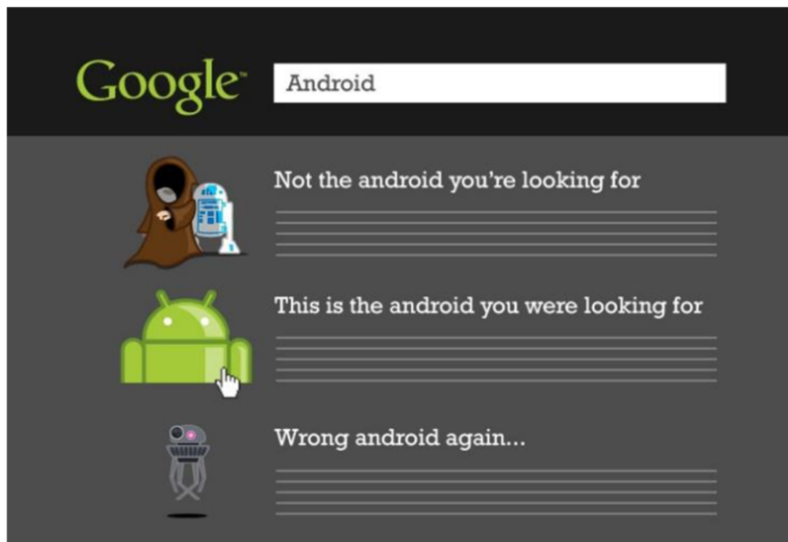
recherche sémantique  
moteurs de réponses  
recherche anticipatoire  
recherche conversationnelle



## La recherche sémantique

se faire comprendre de la machine  
comme par un humain

## La recherche sémantique



## La recherche sémantique

<b>Limite des moteurs traditionnels</b>	<ul style="list-style-type: none"> <li>- « dictature du mot-clé » (O. Andrieu) : « il est clair que si la recherche basée sur les mots-clés est incroyablement puissante, elle est également incroyablement limitative » (M. Mayer) : limites morphologiques et syntaxiques : problèmes des homonymes (ex. : Saint-Louis), synonymes (chat/félin), phrases « atomiques » (York New York)...</li> <li>- pas/peu de compréhension du langage naturel (ex. : une chanson qui fait danser, un endroit où aller pour un 1<sup>er</sup> rendez-vous cf. <a href="#">étude 2009</a>)</li> <li>- pas/peu de compréhension du contexte de la recherche (ex. : les éléments biographiques d'un personnage célèbre ne comportent pas forcément le mot « biographie » comme mot-clé) : usager qui doit trier les informations</li> <li>- classement des résultats par popularité et quantité (indexation du texte intégral) et non pertinence → pas de compréhension du contexte de la recherche et du contenu des documents</li> </ul>
<b>Type de documents</b>	<ul style="list-style-type: none"> <li>- tous documents, mais surtout des connaissances (informations qualifiées) et non des informations brutes</li> <li>- données</li> </ul>
<b>Recherche</b>	<ul style="list-style-type: none"> <li>- pratiques du <i>clustering</i> (regroupement et visualisation en fonction de catégories) : pratique de <i>data mining</i>, permettant une classification automatique des termes, notamment avec reconnaissance d'entités nommées</li> <li>- recherche sémantique : « La recherche sémantique désigne l'habileté d'un moteur de recherche de déterminer ce que vous voulez dire quand vous cherchez quelque chose et de vous fournir des résultats qui ne correspondent pas nécessairement aux mots que vous utilisez dans votre requête » (M. Karch) : prend en compte l'intention de l'internaute et le contexte de la recherche (localisation, désambiguïsation des termes, synonymes, relations entre les termes) et pas seulement les mots de la requête ; identifie des associations et des concepts (champ lexical et sémantique)</li> </ul>
<b>Outils</b>	<ul style="list-style-type: none"> <li>- moteurs généralistes (Google, Bing)</li> <li>- moteurs spécifiques</li> <li>- extensions de navigateurs (ex. : ClearForest Gnosis)</li> <li>+ pratique fréquente du « <i>mashup</i> » (mélange d'informations et/ou de services)</li> </ul>
<b>!</b>	<ul style="list-style-type: none"> <li>- fonctionne mieux avec des requêtes longues</li> <li>- domaine encore en développement : indexation des contenus, traitement sémantique... (cf. N. Aussenc-Gilles)</li> <li>- outils devant permettre de limiter les réponses inappropriées (comme les spams) et favorisant la <i>contextualisation</i>, mais : <ul style="list-style-type: none"> <li>- désambiguïsation des termes et place pour l'information non factuelle (cf. Abraham Lincoln)</li> <li>- validité et fraîcheur des informations ? (cf. <i>Henn VIII wives</i> ou <i>What happened to dinosaurs ?</i> sur le <i>Knowledge graph</i>)</li> <li>- rôle éventuel des internautes dans la production des données et dans l'amélioration des résultats du moteur (<i>crowdsourcing</i>), mais ressources humaines limitées → développement de traitements algorithmiques (<i>Knowledge vault</i> de Google)</li> <li>- visée potentiellement commerciale (ex. : conserver l'internaute plus longtemps sur une page, vendre des services..., cf. O. Erzscheid) → contrôle de l'information par les moteurs et liberté de l'internaute ?</li> </ul> </li> <li>- interfaces pas toujours intuitives et faciles d'utilisation</li> </ul>

rappel : **indexation des données** (M.-L. Malingre et A. Serres)

méthodes d'analyse : linguistique (reconnaissance des mots) / statistique (fréquence des occurrences = mots-clés)

niveaux d'analyse : morphologique (reconnaissance du mot), lexical (lemmatisation), syntaxique (grammaire), sémantique (reconnaissance des concepts)

« *It's clear that while keyword-based searching is incredibly powerful, it's also incredibly limiting* » (M. Mayer, 2008, <http://googleblog.blogspot.fr/2008/09/future-of-search.html>)

« *The two types of analysis that are commonly attempted to categorize search engine results are clustering and semantic search, both of which are hot research topics* » (Greg, 2013, <http://blog.blekko.com/2013/01/14/cracking-the-search-category-problem/>)

clustering : « techniques statistiques (méthode des mots associés avec matrices de cooccurrences) qui permettent d'extraire automatiquement des termes dans les pages de résultats trouvés par un moteur ou métamoteur, afin d'affiner ou de réorienter la recherche » : lever des ambiguïtés, préciser une recherche (V. Mesguich et A. Thomas) – idées généralement déduites des résultats de la recherche initiale, par exemple « art coréen » peut proposer « peinture », « céramique »

question de l'IA (intelligence artificielle) : associer les mots qui apparaissent sur le web à des entités qui ont un sens et qui ont des relations avec d'autres – ce que le cerveau fait naturellement,

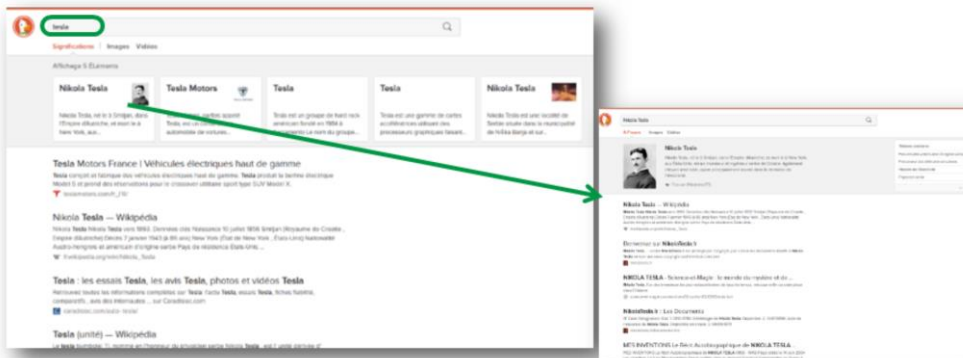
ex. : distinction des termes dans « il livre le lit » / « il lit le livre » ; association de termes comme « scrutin », « vote », quand recherche du terme « élection »

ex. : Barack Obama (D. Sullivan, 2013, <http://moz.com/blog/the-evolution-of-search>, partie sur la « search 5.0 »)

visées également économiques : par exemple : intérêt pour la cuisine indienne associé aux différents plats possibles (curry, tikka...), cf. E. Swallow, 2011, <http://mashable.com/2011/03/04/semantic-search-media/>

# La recherche sémantique

- DuckDuckGo : <https://duckduckgo.com/> [référence au jeu du mouchoir, « *duck duck goose* » en anglais] moteur de recherche interrogeant une 50ne de sites (dont Bing, ou son propre moteur, et des sites *crowd-sourced* comme Wikipédia)  
web, images et vidéos  
limite les sites commerciaux et les recherches locales  
[aide à la syntaxe, paramètres](#) (langue...)
  - + assure confidentialité des recherches : pas de stockage des adresses IP, peu d'utilisation des cookies... et rend possible recherche anonyme → pas de personnalisation des résultats (auto-corrrection, géolocalisation...)
  - + permet de désambiguïser les termes d'une recherche en cas d'homonymes



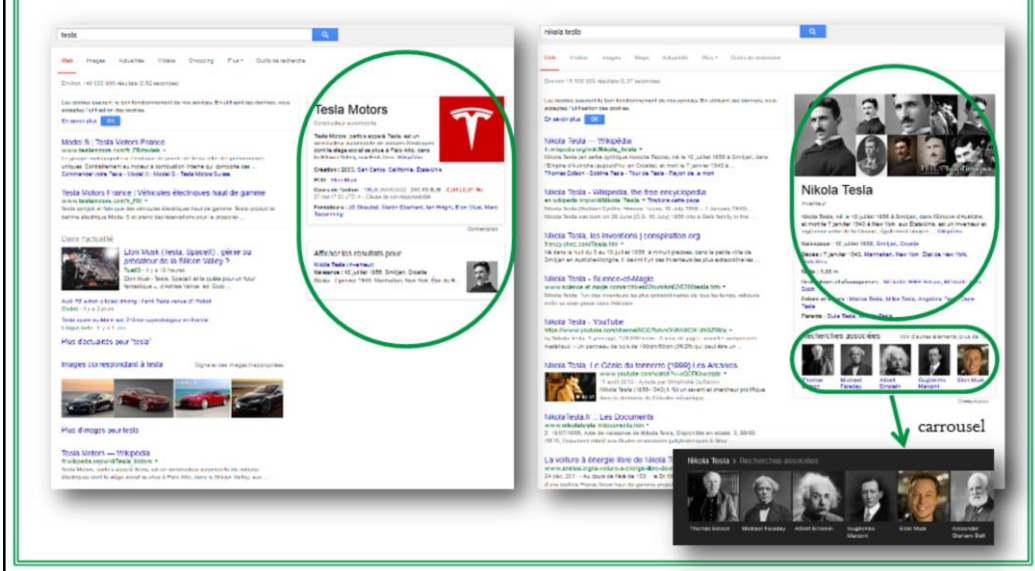
- + possibilité de faire des recherches via des catégories et mots-clés (**!bang**), ex. : **!images**, **!Google**

# La recherche sémantique

- Google Knowledge Graph : <http://www.google.fr/>

identification d'entités nommées (personnes, lieux...), permettant de désambigüiser le terme - « things, not strings »

entité, permettant d'élargir la recherche (recherches associées...)



Google est déjà un « moteur de réponses », cf. pour formules mathématiques, météo, bourse...

**Knowledge Graph** (05/2012 aux États-Unis, 12/2012 en France) : A. Singhal, 2012, <http://googleblog.blogspot.fr/2012/05/introducing-knowledge-graph-things-not.html> et <http://googlefrance.blogspot.fr/2012/12/lancement-en-france-du-knowledge-graph.html>

modèle intelligent (« graph ») qui puisse comprendre les entités nommées et leurs relations entre elles – cf. travail antérieur de Google sur la « roue magique »



Image Eduscol, <http://eduscol.education.fr/numerique/dossier/archives/visite-virtuelle-chapelle-sixtine/recherches-sur-internet/roue-magique>

- **sources** : résultats issus de Wikipédia, Freebase, CIA World Factbook, mais aussi de données extraites du contenu publié sur le web grâce à des « algorithmes d'extraction » (mise à profit de l'intelligence collective) et les données antérieures de Google (données des livres et cartes) : données factuelles, non sujettes à caution ou spéculation – possibilité de signaler des erreurs

- **taille** : + 1 MM. d'entités (2015) et 18 MM. de faits et relations

- **fonctionnement** : établir des données factuelles, mais aussi des relations entre des objets et des concepts – fonctionne notamment à partir de l'étude des requêtes des internautes (requêtes les plus posées, requête venant après telle autre...) ex. pour un scientifique : distinctions ; pour un auteur : livres ; pour un acteur : films, etc. ; les données indiquées sur Tom Cruise répondent ainsi à 37% des questions suivantes

- **3 objectifs** :
  - trouvez ce que vous cherchez, ex. Taj Mahal
  - obtenez le meilleur résumé
  - allez plus loin (trouvez des informations que vous ne saviez pas que vous cherchiez) : faire ressortir des connexions que la recherche traditionnelle ne verra pas : ex. Einstein et Gandhi : tous les deux pacifistes à la fin de leur vie
- **interface** : nouvelle présentation des recherches associées grâce à un carrousel
- **limites** :
  - utilisation de sources alimentées par la communauté (cf. Dbpedia)
  - fiabilité des données : juin 2012 : 20 % d'erreurs (B. Schwartz, <http://searchengineland.com/googles-knowledge-graph-errors-126098>)
  - mise à jour des données ?, plus lente que la page de résultats ?
  - aspects commerciaux : fournir une réponse plutôt qu'un lien permet de garder l'internaute sur Google (et ses publicités) plus longtemps

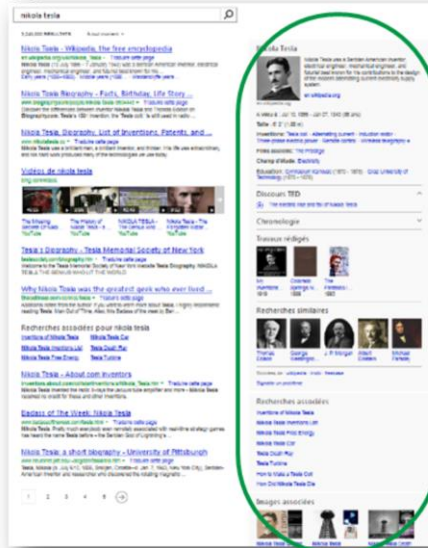
# La recherche sémantique

- Bing Satori : <http://www.bing.com/?scope=web&mkt=en-US&FORM=HDRSCI>

dès le formulaire de recherche



sur la page de résultats



**Sources** : résultats issus de Wikipédia, IMDB et Freebase, notamment

Contrairement au Knowledge Graph de Google, plus orienté sur les entités et les faits, Satori est plus orienté sur les actions/transactions : acheter de la musique, trouver des paroles ou des tickets... (S. Gallagher, 2012, <http://arstechnica.com/information-technology/2012/06/inside-the-architecture-of-googles-knowledge-graph-and-microsofts-satori/>) : orientation plus commerciale (production de catalogues et de sites de commerce en ligne)

# La recherche sémantique

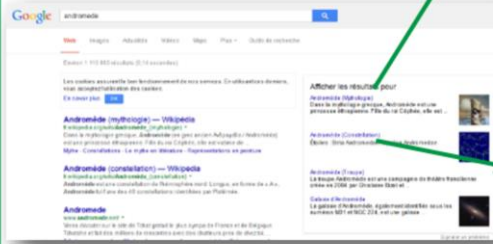
intérêts : - aide à la sélection

- éliminations de corrélations inintéressantes (raccourcis de recherche) -> liste de résultats plus courte, mais avec plus de contenu (parfait pour mobiles)
- suggestions d'idées et de pistes nouvelles

limites : - corrélations automatiques, donc potentiellement

restrictives ou trompeuses

- basé avant tout sur le crowdsourcing donc limité en nombre -> projet de Knowledge vault



## La recherche sémantique



The image displays two screenshots of Google search results. The left screenshot shows the search for "barack obama âge" (Barack Obama's age). The search bar is highlighted with a green circle. The results page features a knowledge panel for Barack Obama, stating he is 53 years old (born August 4, 1961). It includes a photo of Barack Obama and a list of related figures: Hillary Clinton (67 years), Vladimir Putin (62 years), and Michelle Robinson-Obama (51 years). Below the panel are search results from Staragora.com and Wikipedia.

The right screenshot shows the search for "barack obama femme" (Barack Obama's wife). The search bar is also highlighted with a green circle. The results page features a knowledge panel for Michelle Robinson-Obama, stating she is the wife of Barack Obama and was born in 1992. It includes a photo of Michelle Obama and a link to her Wikipedia page. Below the panel are image results for "barack obama femme" and a link to her Wikipedia page.

permet de pointer directement sur une réponse  
et non un document  
→ moteur de recherche d'entités

Autres exemples : [tour Eiffel taille], [Paris population], [Léonard de Vinci œuvres]  
Moteur de recherche capable de comprendre les relations

## La recherche sémantique

- surtout des usages commerciaux
  - ex. : Yummly : <http://www.yummly.com>  
moteur de recettes de cuisine
  - Monster : <http://www.monster.com/>  
site de recrutement ([source](#))
- exemples de moteurs de recherche sémantique spécialisés
  - HealthMash : <http://healthmash.com>  
médecine  
© 2011
  - Symbolab : <http://symbolab.com/>  
équations mathématiques : « *The stated goal of the site is to provide the most relevant search results that are theoretically and semantically similar, rather than visually* »  
indexe texte intégral et équations d'encyclopédies, dictionnaires, publications scientifiques, conférences... (ex. : MIT, MathWorld...)  
recherche avec symboles mathématiques, chiffres, notations scientifiques, texte  
+ éditeur d'équations

What is Semantic Search? (<http://company.hakia.com/whatis.html>, site †)

1- Handling morphological variations : temps, singulier/pluriel...

2- Handling synonyms with correct senses : ex. : cure, heal, treat,.. etc. avec la requête [Is there a cure for ALS?]

3- Handling generalizations : ex. : disease = GERD, ALS, AIDS, etc.) avec la requête [Which disease has the symptom of coughing?]

4- Handling concept matching : ex. : migraine avec la requête [What treats headache ?]

5- Handling knowledge matching : ex. : swine flu = H1N1, flu=influenza

6- Handling natural language queries and questions : ex. : what, where, how, why, etc.

7- Ability to point to uninterrupted paragraph and the most relevant sentence

8- Ability to Customize and Organic Progress : recherche spécifique à un contexte précis

9- Ability to operate without relying on statistics, user behavior, and other artificial means : pas de classement par popularité , mais en fonction du degré de pertinence par rapport à la requête ; intérêt surtout pour le contenu temps réel

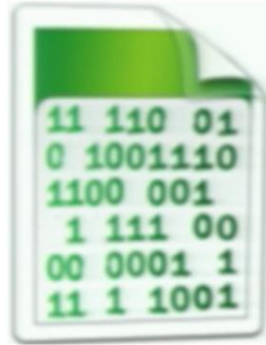
10- Ability to detect its own performance

## Le web sémantique

Que voyez-vous ?



Que voit la machine ?

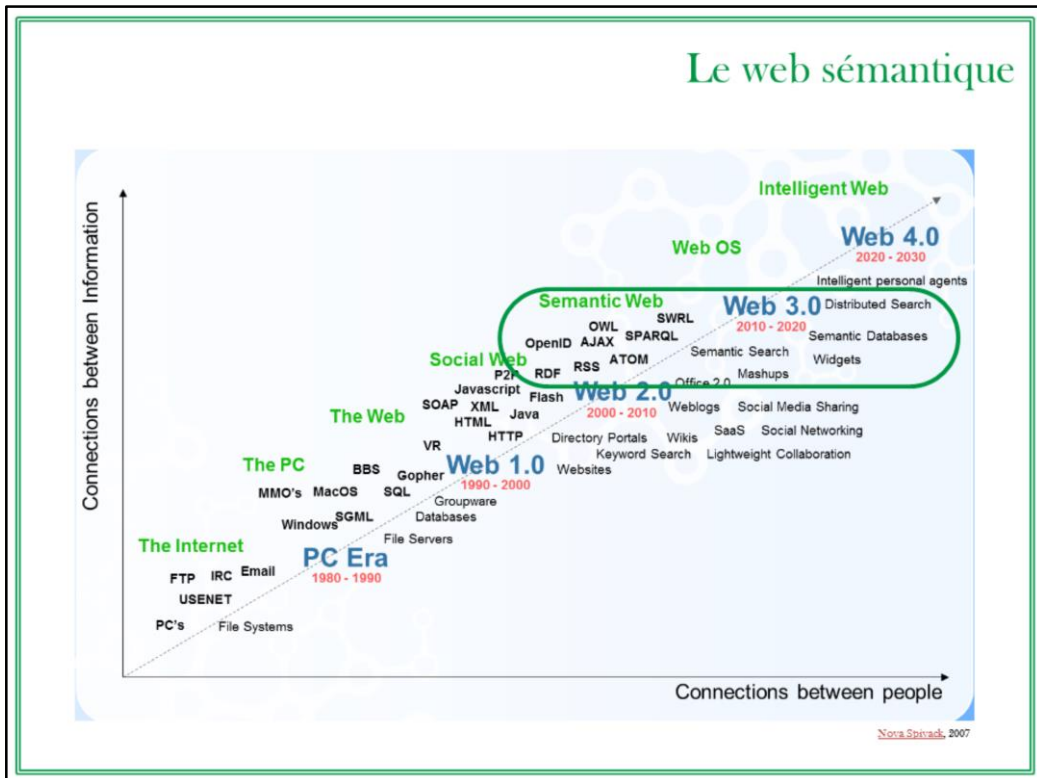


Emmanuel Bernié, 2014

**But** : permettre aux machines de comprendre « intelligemment » des données structurées par ailleurs  
Le web sémantique n'est ni de l'intelligence artificielle (machines qui comprennent de façon intelligente) ni du traitement automatique des langues, mais une structuration des données par des humains en leur appliquant un formalisme, une syntaxe permettant un lien entre les machines

Rôle de plus en plus important des données structurées pour les moteurs de recherche (ex. : B. Starr, 2015, <http://searchengineland.com/structured-data-serps-googles-patents-tell-us-ranking-universal-search-219205>)

## Le web sémantique



**Web** : contenus et données, mais pas de structuration de l'information fine, ni de sémantisation du contenu : la plupart des données structurées se trouvent dans des bases de données (web invisible)

**Web sémantique** (1998) : ne concerne pas les documents en langage naturel : « *The Semantic Web will enable machines to COMPREHEND semantic documents and data, not human speech and writings* » (T. Berners-Lee, J. Hendler et O. Lassila,

<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>)

« *The semantic web is a web of data, in some ways like a global database* » (T. Berners-Lee, 1998, <http://www.w3.org/DesignIssues/Semantic.html>) : formel et structuré

→ un environnement technique qui doit permettre de résoudre le problème de l'interopérabilité des bases de données (approche normative)

## Le web sémantique

**RDF** est un modèle de triplets *i.e.*  
toute connaissance est décomposée en  
( **sujet** , **predicat** , **objet** )



**doc.html** a pour auteur **Fabien**  
**doc.html** a pour thème **Musique**

F. Gandon

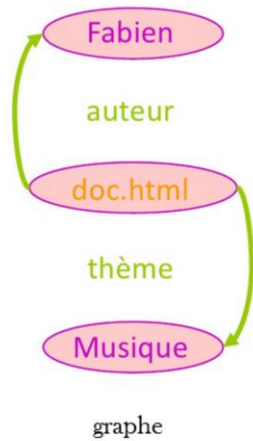
« Qu'est-ce donc que le web sémantique ? Fondamentalement, c'est juste un moyen de décrire les choses d'une façon « compréhensible » par un ordinateur. Bien sûr, il ne s'agit pas de compréhension, mais de logique, comme on l'apprend dans les lycées » (P. Ford, 2009, [http://www.ftrain.com/google\\_takes\\_all.html](http://www.ftrain.com/google_takes_all.html))

**Web de données** : ensemble de technologies et de normes  
spécification de **données/entités** et de leurs **relations**  
ne s'arrête pas aux données textuelles mais à des morceaux de données (nombres, dates, entités, données factuelles...)

### 3 éléments de base :

- identifiants web : URI
- norme : RDF (*resource description framework*, syntaxe XML pour permettre l'échange des graphes) : système de triplet (sujet, prédicat, objet)
- modèle : graphe typé qui permet de modéliser les triplets

## Le web sémantique



```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-
  rdf-syntax-ns#"
  xmlns:inria="http://inria.fr/schema#" >

  <rdf:Description
    rdf:about="http://inria.fr/xr/doc.html">
    <inria:author rdf:resource=
      "http://inria.fr/~fabien#me" />
    <inria:theme>Music</inria:theme>
  </rdf:Description>

</rdf:RDF>
```

F. Gandon

RDF

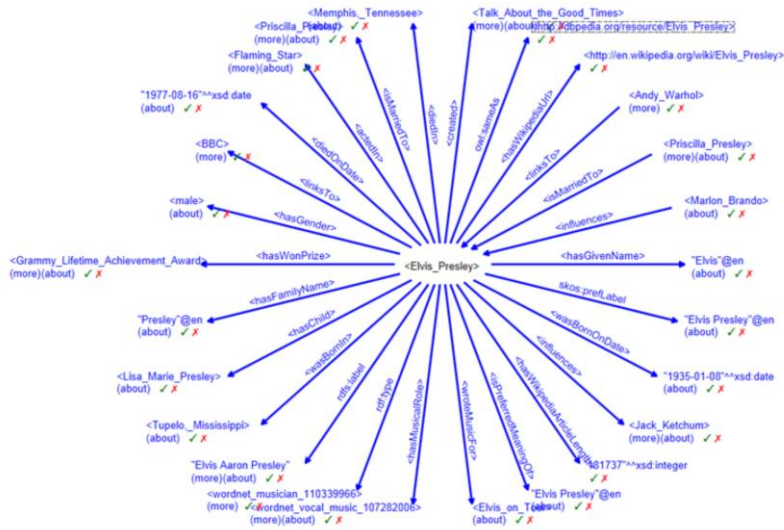
Exemple 1 : question « *What is the birth date of George W. Bush ?* » : *True Knowledge would reason from the facts « George W. Bush is a president », « George W. Bush is a human being », « A president is a subclass of human being », « Date of creation is a more general form for birth date », and « the 6th of July is the date of creation for George W. Bush », to produce the simple answer, « the 6th of July »* (présentation Evi, sur Wikipédia, [http://en.wikipedia.org/wiki/Evi\\_%28software%29](http://en.wikipedia.org/wiki/Evi_%28software%29)) ;  
Exemple 2 : le moteur voit une page indiquant que « X est la fille de Y », il peut déduire que « Y est la mère de X » ; capable de comprendre que si la femme de mon oncle est ma tante, la femme d'oncle Bob est ma tante

Basé sur des **ontologies** (indexation des ressources avec des outils communs et normalisés, OWL)

Ex. : relier des concepts similaires utilisés dans des vocabulaires différents et/ou des langues différentes (ex. : Révolution française = France – 1789-1799 (Révolution))

Utilisation de balises sémantiques permettant de donner une signification aux éléments

# Le web sémantique



From <http://yago-knowledge.org>  
 Give us feedback:  
 ✓ = fact is correct  
 x = fact is false

ex. Elvis Presley sur [Yago](#)

## Le web sémantique

- bases de données sémantiques : deux exemples
  - Freebase : <http://www.freebase.com/>  
acquis par Google en 2010  
base communautaire de données structurées et reliées entre elles (personnalités, lieux et choses)  
45,6 M. de sujets et 2,9 MM. de faits  
- validité et fraîcheur des informations ?  
ex. : [Nikola Tesla]
  - DBpedia : <http://dbpedia.org/About>  
projet universitaire et communautaire  
base de données extraites de Wikipédia (125 langues, 3 MM. d'informations) et structurées au format du web sémantique (triplets RDF)  
pour la France : dans le projet Sémanticpédia <http://www.semanticpedia.org/>  
doit être un des noyaux du web de données  
- moteur pour informaticiens (interrogation en SPARQL)  
- validité et fraîcheur des informations ?  
ex. : [Nikola Tesla]  
[Give me all Italian musicians from the 18th century]  
[Which skyscrapers in China have more than 50 floors and have been constructed before the year 2000. ?]  
  
→ ne peuvent généralement pas s'interroger directement  
(et ne sont pas faits pour cela !)  
→ interrogation du web sémantique ≠ recherche sémantique

### Autres exemples :

#### données institutionnelles

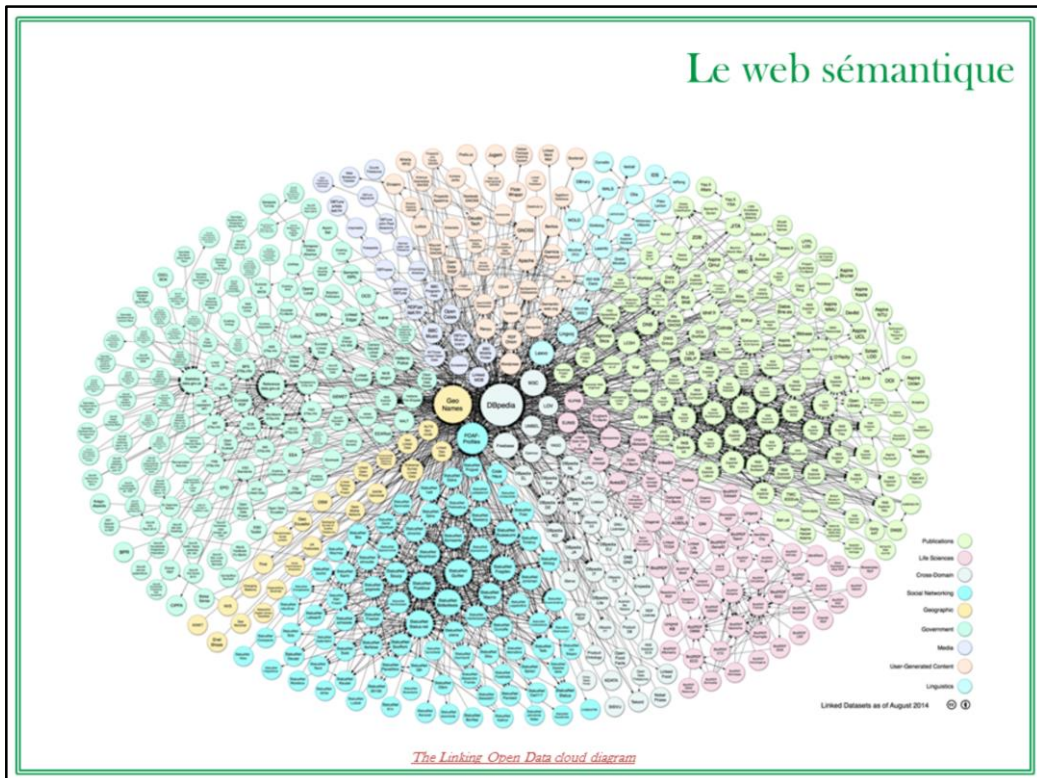
- Data.gov : <http://www.data.gov/>
- Data.gov.uk : <http://data.gov.uk/>
- CIA World factbook : <https://www.cia.gov/library/publications/the-world-factbook/>

#### données géographiques

- GeoNames : <http://www.geonames.org>
- LinkedGeoData: <http://linkedgeodata.org/About>

#### données culturelles

- MusicBrainz : <http://musicbrainz.org/>
- Linked Movie Database : <http://linkedmdb.org/>



**Web de données** (*linked data*, 2007) : application des technologies du web sémantique pour exposer, échanger et traiter des données structurées sur le web : extension des principes du web de documents aux données (HTML → RDF, URL → URI, etc.)

? possibilité de trouver de nouvelles relations ?

« *linking open data* » : exposer des données publiques

Quelles places pour les moteurs de recherche ?

- **technologies sémantiques ≠ recherche sémantique** [interrogation RDF : langage de requête SPARQL (*SPARQL protocol and RDF query language*)]
- « *One thing to always remember is that the Web of the future will have BOTH documents and data. The Semantic Web will not supersede the current Web. They will coexist. The techniques for searching and surfing the different aspects will be different but will connect. Text search engines don't have to go out of fashion* » (T. Berners-Lee, 2008, <http://dig.csail.mit.edu/breadcrumbs/blog/4>) ; question importante à cause du développement de la taille du web et sa complexité
- mettre en forme le « **chaos informationnel** » et permettre aux moteurs de recherche de trouver des informations plus pertinentes car catégorisées

# Le web sémantique

VIAF  
Fichier d'autorité international virtuel

**Rechercher**

Critère de recherche: mot(s) nom(s) | Fichier source: Tout VIAF | Recherche: Nikola Tesla, 1856-1943 | **Rechercher**

18 Entrées trouvées pour votre recherche sur **Nikola Tesla, 1856-1943**

Formes autorisées	Type de nom	Exemple de titre associé
1 <a href="#">Tesla, Nikola, 1856-1943</a>  <a href="#">Tesla, Nikola</a>  <a href="#">Nikola Tesla</a>  <a href="#">Tesla, Nikola (American engineer, inventor, 1856-1943)</a>  <a href="#">1856-1943 نيكولا تسلا</a> 	Personne	Articles Članci Dnevnik istraživanja iz Kolorado Springsa, 1899-1900

[notice Nelson Mandela](#)

VIAF (*Virtual International Authority File*) : <http://viaf.org>  
notices d'autorité de 34 bibliothèques dans le monde (dont nationales), OCLC  
créé initialement par et pour les bibliothèques, mais devenu depuis un élément  
important du [Linked data](#)

## Le web sémantique

- exemples de réalisations : moteurs de recherche français
  - Thèses.fr : <http://www.theses.fr/>  
base de données des thèses françaises
  - Isidore : <http://www.rechercheisidore.fr/>  
moteur de données francophones en SHS (normalisation, catégorisations automatiques...)
  - Data.bnf.fr : <http://data.bnf.fr/>  
fiches de référence sur les auteurs, les œuvres et les thèmes, permettant d'interroger les différents catalogues et bases de données de la BnF
  - HdA Lab : <http://hdalab.iri-research.org/hdalab/>  
à partir du portail Histoire des arts
  - Centre Pompidou (virtuel) : <http://www.centrepompidou.fr/>  
site du Centre Pompidou, permettant d'interroger des « silos » variés d'informations

**But : « briser les silos »** (Emmanuelle Bermès) : faire dialoguer des bases de données bâties sur des modèles différents ; transparent pour l'utilisateur

→ indexation non plus seulement des documents mais de leur contenu (qui peut être relié et enrichi) (cf. S. Pouyllau, 2011)

→ permet de mieux comprendre les requêtes des internautes

→ permet d'afficher des pages de résultats plus pertinentes (qualification et validation des informations à partir de leurs relations, catégorisation...)

## Le web sémantique

- exemples de réalisations : *mash-ups* (1)
  - à partir de données librement accessible sur le web
    - ex. : Dbpedia
      - Ex. : BBC Wildlife finder : <http://www.bbc.co.uk/nature/>
    - ex. : MusicBrainz
      - Ex. : BBC music : <http://www.bbc.co.uk/music> - Ex. : [Queen](#)

Tim Berners-Lee (<http://dig.csail.mit.edu/breadcrumbs/node/215>) : favoriser les connections entre les données et quelques soient les applications

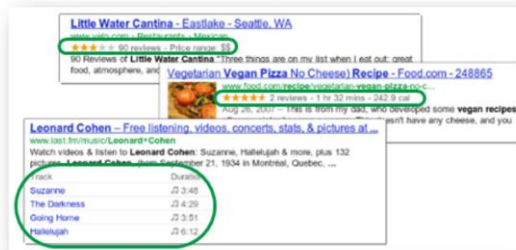
Ex. d'un vol aérien : possibilité d'avoir toutes les informations intégrées (site de voyage ou de la compagnie aérienne)] et de réutiliser les données, de façon parfois inattendue par les producteurs [ex. du mouvement *Linked Open data*]

Possibilité de *mash-ups* : applications composites permettant de produire un nouveau service en mixant du contenu et /ou des services d'autres applications

ex. : par exemple, durant une émission sur les œuvres d'art sur une chaîne de France Télévisions, on pourrait synchroniser l'émission à une application sur sa tablette PC avec les données de la Bibliothèque Nationale de France (BnF), avec les peintures du Louvre et du musée du Quai Branly (« SPARQL », *Wikipédia*, <http://fr.wikipedia.org/wiki/SPARQL>, consulté le 11/06/2013)

## Le web sémantique

- exemples de réalisations : *mash-ups* (2)
  - à partir de données fournies par les webmestres (microformats)  
[schema.org](http://schema.org) [Google, Bing, Yahoo, Yandex]  
Ex. *rich snippets* de Google : amélioration des résultats (événements, recettes, produits, critiques...) par des informations supplémentaires (textes, classements, photos...)



- à partir de données librement accessibles et de données fournies par les webmestres  
Ex. *Knowledge Graph* de Google et Bing

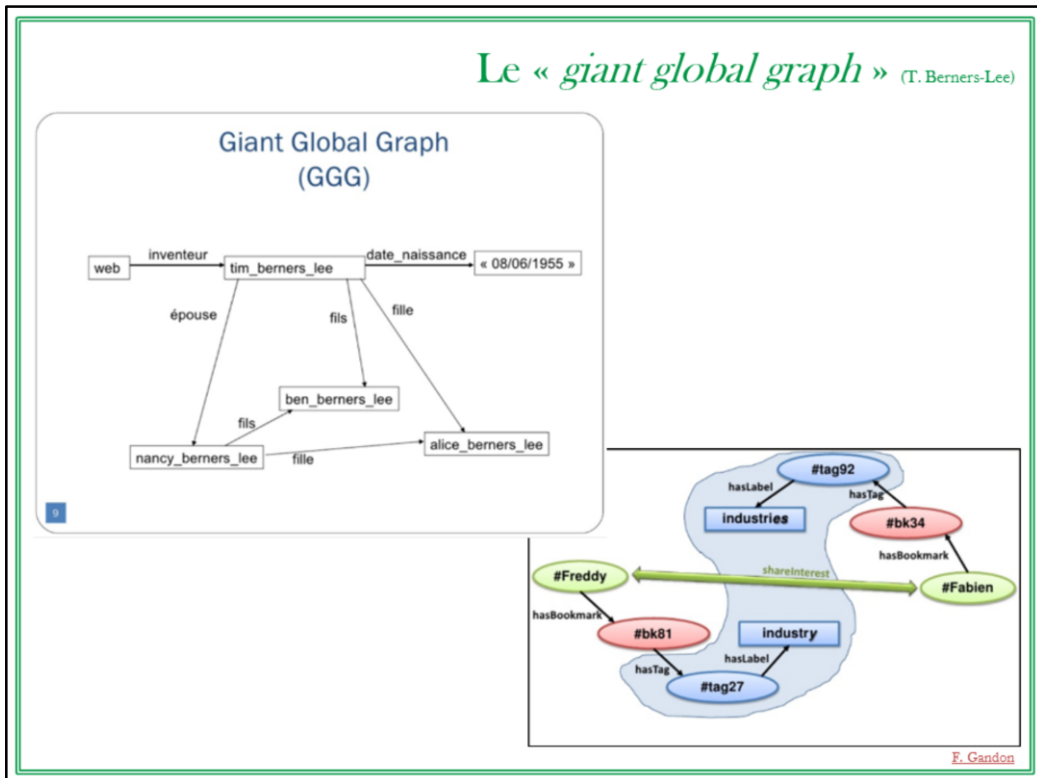
**Données structurées sur le web** : pas seulement du web sémantique (données reliées), mais aussi structuration de contenus (RDFa, microformats comme [schema.org](http://schema.org)...) (T. Francart, 2014, <http://blog.sparna.fr/2014/09/04/linked-open-data-cloud-nouvelle-version/>)

### Moteurs de recherche :

- outils comme Knowledge graph ne sont pas encore du web sémantique (Mondeca, 2013, <https://mondeca.wordpress.com/2013/02/20/ceci-n-est-pas-le-web-semantique/>)
- fonctionnent sur des microformats (métadonnées HTML) que les webmestres peuvent utiliser pour identifier leur contenu d'une manière reconnaissable par les principaux moteurs de recherche et qui leur permettent d'améliorer la présentation des résultats (la balise « meta-description » peut apparaître sous forme de snippets (résumé extrait de la page) dans la page de résultats
- ces microformats ne sont pas visibles par les internautes mais par les navigateurs et les moteurs : la sémantisation du web doit faciliter le travail des moteurs de recherche

Peut se faire manuellement cf. projet [schema.org](http://schema.org) (Bing, Google, Yahoo et Yandex, 06/2011, <http://schema.org/> : collection de schémas), ou automatiquement (cf. Facebook : le remplissage de formulaires d'inscription permet de taguer les informations avec « nom de personne », « date de naissance », « lieu », « centre d'intérêt »)...

## Le « giant global graph » (T. Berners-Lee)



Knowledge Graph de Google : surtout orienté sur les informations et les faits (Wikipédia, sources vérifiées ou validées...) plus que sur les centres d'intérêts de l'internaute et les interactions sociales

Or, développement du web sémantique dans le domaine des données sociales (réseaux sociaux) – encore plus vrai à l'ère du *big data*

→ évolution du WWW au GGG (*Giant Global Graph*) (T. Berners-Lee, 2008, <http://dig.csail.mit.edu/breadcrumbs/node/215>)

Trois niveaux :

- *net* (infrastructures : relie des ordinateurs),
- *web* (contenus : relie des documents)
- *social graph* (relations et connexions : relie des éléments, hommes et ressources) : intérêt d'exprimer les relations au niveau supérieur aux documents, via la technologie du web sémantique

Réseaux typés + opérateurs = meilleure analyse de l'information

« *Graph Search on consumed/verified/validated information, which is a core component of the Semantic Web, is now considered key for the future of search **in both search and social engines*** » (B. Starr, 2013, <http://searchengineland.com/semantic-graph-based-search-the-future-face-of-search-156461>)

## La Graph search de Facebook



Ex. : Facebook : propose un *open graph* : succède au *social graph*, pour créer des relations entre les personnes et les contenus, cf. C. Jourdain, 2011, <http://www.camillejourdain.fr/10-verites-open-graph-de-facebook/> : possibilité de rajouter à des pages externes des contenus venant de Facebook ; la page/le site « liké » devient un objet du *social graph*

modèle repris ensuite par d'autres, comme LinkedIn

→ emporter le graphe hors de Facebook et faciliter le partage de contenus de Facebook (ex. : bouton Like sur différents sites)

**Moteur de recherche *Graph search*** : moteur de recherche interne, mais avec des suggestions sociales, en fonction des avis d'amis ; pour l'instant limité : *people* (amis), *photos* (photos d'amis ou likées), *places* (restaurants, hôtels) et *interests* (jeux, émissions de télé, films, groupes de musique)

ex. de requête : « Je cherche une femme de 18 à 25 ans qui est allée à Cuba et qui a lu les essais politiques de Che Guevara » (*Archimag*, n°267)

### Points d'attention :

\* extension en-dehors de Facebook, à l'*open graph* ? – éléments de Facebook déjà présents sur Bing notamment  
 \* intérêt réel pour la recherche ? (F. Cavazza, 2013, <http://www.fredcavazza.net/2013/01/16/facebook-lance-son-moteur-de-recherche-mais-manque-dinspiration/>)

\* limites :

- nécessite de construire la requête (pas simple succession de mots-clés) ;
- n'interroge que les données publiques, mais déjà des informations très précises, qui bien que publiques pouvaient être enfouies : meilleure visibilité, cf. <http://actualfacebookgraphsearches.tumblr.com/> ;
- dépend du réseau et de l'activité sur Facebook, notamment de ses amis ;
- fonctionne surtout pour des recherches d'avis et de recommandations et des recherches personnelles, moins pour une recherche internet classique ;
- renforce la *filter bubble*

\* important pour les bibliothèques : P. Bradley, 2013,

[http://philbradley.typepad.com/phil\\_bradleys\\_weblog/2013/01/why-the-new-facebook-graph-search-is-important-for-librarians.html](http://philbradley.typepad.com/phil_bradleys_weblog/2013/01/why-the-new-facebook-graph-search-is-important-for-librarians.html)



## Les moteurs de réponses

« un lien n'est pas du tout une réponse » [\[O. Etzioni\]](#)

« *Answer engines* » / « *Knowledge engines* »


« *A link is not at all an answer* » (O. Etzioni, 2013,  
<http://www.wired.co.uk/magazine/archive/2013/01/features/the-future-of-search?page=all>)

## Les moteurs de réponse

<b>Limite des moteurs traditionnels</b>	<ul style="list-style-type: none"> <li>- comprennent mal le langage naturel</li> <li>- fournissent des liens plutôt qu'une réponse</li> <li>- mais déjà direction prise pour les entités ou les actions (cf. <i>Knowledge graph</i> de Google et Satori de Bing)</li> </ul>
<b>Type de documents</b>	informations factuelles essentiellement
<b>Recherche</b>	recherche par mots-clés ou en langage naturel
<b>Outils</b>	<p>outils spécifiques, reposant sur les avancées actuelles (recherche sémantique, web sémantique...) : doivent comprendre la question, les expressions et le contexte et proposer une réponse collectant les données les plus pertinentes et objectives - proposer des sortes d'« encyclopédie[s] interactive[s] » (<a href="#">M. Maver</a>)</p> <p>→ plus seulement des outils de recherche mais des sources d'information : éditorialisation de contenu</p>
<b>!</b>	<ul style="list-style-type: none"> <li>- domaine encore en développement : indexation des contenus, traitement sémantique, compréhension du langage naturel, développement de l'intelligence artificielle, synthèse automatique...</li> <li>- outils majoritairement anglo-saxons (US)</li> <li>- intérêt pour certains types de recherche uniquement (informationnelles, ex. : [quel est le temps prévu pour aujourd'hui ?], [combien mesure la tour Eiffel ?]), moins pour d'autres (navigationnelles et transactionnelles) qui ont besoin d'une succession de liens (informations complémentaires, contradictoires...) ; pas de synthèse de plusieurs pages par ex. <ul style="list-style-type: none"> <li>→ quel moyen de comparaison si une seule réponse ?</li> </ul> </li> <li>- établissement de la base de connaissance : <ul style="list-style-type: none"> <li>- sources d'informations ?</li> <li>- validité et fraîcheur des informations ? pré-déterminée ou mise à jour en continu ?</li> <li>- quelle présentation des éléments pour laquelle la question n'a jamais été posée et/ou dont la réponse ne se trouve pas dans la base de connaissance (cf. <a href="#">Google</a> : 15 % des requêtes quotidiennes sont nouvelles) ? cul-de-sac ?</li> </ul> </li> <li>- visée potentiellement commerciale (réponses sponsorisées encore plus fortement car diminution du nombre de résultats affichés)</li> </ul>

15 % des questions que les gens posent à Google sont totalement nouvelles et n'ont jamais été posées  
 Reconstruction de documents à partir des requêtes des usagers (J.-M. Salaün, 2012, <http://blogues.ebsi.umontreal.ca/jms/index.php/post/2012/06/06/Web-de-donn%C3%A9es%2C-Google%2C-Wikip%C3%A9dia%2C-les-liaisons-dangereuses>)

## Les moteurs de réponse



The screenshot shows the WolframAlpha interface for the query 'puma'. It displays various pieces of information:

- Input interpretation:** mountain lion (animal)
- Scientific name:** *Puma concolor*
- Alternate scientific name:** *Felis concolor*
- Alternate common names:** puma | cougar | mountain lion | mountain cat | catamount | panther
- Taxonomy:**

kingdom	Animalia (animals)
phylum	Chordata (chordates)
class	Mammalia (mammals)
order	Carnivora (carnivores)
family	Felidae (cats)
genus	Puma
species	<i>Puma concolor</i> (mountain lion)
- Biological properties:**
  - maximum recorded lifespan: 23.8 years (in captivity)
  - Physical dimensions: head and body length (1.1 to 1.8) meters

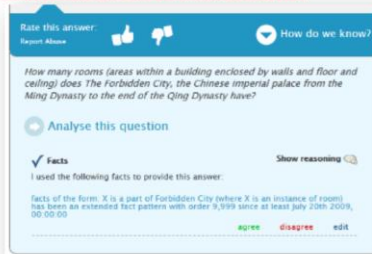
- WolframAlpha : <https://www.wolframalpha.com/>
  - « *computational knowledge engine* » : moteur de connaissances, à partir d'une analyse linguistique et de données *curated* (équipe de rédacteurs)
  - n'indexe pas le web : interroge sa propre base de données, à partir de données issues notamment du web invisible
  - alimente d'autres moteurs (Bing, DuckDuckGo, Siri...)
  - espace personnel : préférences, favoris, téléchargements...
  - + offre payante permettant de télécharger ses propres données
  - + nombreux tableaux et graphiques
  - contenu essentiellement textuel
  - seulement pour recherches factuelles (peu d'actualités, pas de pages de liens...), statistiques ou computationnelles (ex. [[cars in California](#) / [California population](#)])
  - résultats très inégaux, plus orienté STM que SHS (ex. [[tetrahedron](#)] / [[French revolution](#)])
  - pas de listes des sources utilisées [*external sources* : à titre informatif seulement]
- ex. [[zodiac](#)]
- [[NASDAQ vs. New York stock exchange](#)]
- [[father's mother's sister's son](#)]
- [liste d'exemples](#)

« Wolfram/Alpha introduces a fundamentally new way to get knowledge and answers— not by searching the web, but by doing dynamic computations based on a vast collection of built-in data, algorithms, and methods » → n'est plus un moteur de recherche web à proprement parler

## Les moteurs de réponse

développement surtout sur des applications mobiles

- voir également applications smartphones Siri et Google Voice search
- Evi : <http://www.evi.com>  
groupe Amazon  
est avant tout une application mobile, mais accessible aussi sur ordinateur  
interroge des bases extérieures (Freebase, Dbpedia, Quora...) et permet la contribution des internautes (possibilité de signaler des erreurs)  
ex. : [*How many rooms does the Forbidden City have?*]



désambiguïsation des termes  
pour trouver le sens le plus  
probable de la question

- Kngine: <http://www.kngine.com> [pour Knowledge Engine]  
est avant tout une application mobile, mais accessible aussi sur ordinateur  
base de données continuellement mise à jour sur le web : ! lenteurs de la requête  
app. la plus performante ([source](#)) ?  
+ moteur de réponse, mais propose également une recherche web en complément

Rôle du développement des **appareils mobiles** : recherches souvent ponctuelles (une adresse, simple référence...) : souhait d'avoir la réponse et non une liste de liens (peu visibles de toute façon)

En 11/2012 : tests sur les applications les plus performantes : Kngine répond à 54 %, contre 26 % pour Siri et 25 % pour Evi (A. Ha, 2012, <http://techcrunch.com/2012/11/30/kngine/>)



La recherche anticipatoire

## La recherche anticipatoire

<b>Limite des moteurs traditionnels</b>	portent sur de l'existant (documents, données) et n'ont pas de fonctionnalités prédictives, alors que : - le développement du web permet de mieux connaître le comportement des internautes et leurs données (statuts et réseaux) et d'établir « la base de données de nos intentions » ( <a href="#">J. Battelle</a> , 2005, i.e. les comportements les plus prévisibles) ; ex. : Zeitgeist de Google, suggestions automatiques de services, choix des informations présentées dans le <i>Knowledge graph</i> ; - le développement des outils intelligents (smartphones, smartwatches...) et des objets connectés, l'usage du <i>cloud computing</i> et de la synchronisation des données permettent de mieux connaître le comportement d'un internaute déterminé et ses données (comportements prévisibles, centres d'intérêt...)
<b>Type de documents</b>	données surtout factuelles
<b>Recherche</b>	recherche anticipatoire ( <i>anticipatory search</i> ) / prédictive / proactive ( <i>proactive search</i> ) : « le stimulus [de la recherche] n'est plus un mot-clé, mais plutôt un changement d'état » ( <a href="#">S. Weisz</a> ) dans le cas des agents intelligents des smartphones : recherche contextuelle ( <i>contextual search</i> ) : « le logiciel créé pour apporter l'information utilise toujours des moteurs de recherche pour obtenir l'information, mais l'utilisateur n'a pas à interagir avec le moteur de recherche directement » ( <a href="#">C. Mims</a> ) : les recherches s'effectuent automatiquement en fonction des informations disponibles sur le comportement de l'internaute (union de l'expérience utilisateur et de l'intelligence artificielle) → vers des « <i>decision engines</i> » ( <a href="#">A. Friedman</a> ), outils d'aide à la décision : « [les gens] veulent que Google leur dise ce qu'ils devraient faire ensuite » ( <a href="#">E. Schmidt</a> )
<b>Outils</b>	agents intelligents comme Google Now, Apple Siri et Microsoft Cortana outils de <i>monitoring</i>
<b>!</b>	développement surtout pour des usages personnels ou stratégiques (bourse, tremblements de terre <a href="#">exemple</a> ), défense <a href="#">exemple</a> ...) analyses des données (web en général, monitoring des réseaux sociaux, analyse des données personnelles de l'internaute [mail, agenda...]), ex. : <a href="#">Stephen Wolfram</a> , et donc, importants enjeux économiques (cf. <a href="#">ZDNet</a> ) problèmes : archivage des données et questions d'identité numérique données personnelles pouvant être transmises cf. demandes de gouvernement (PRISM) voir également <a href="#">O. Ernscheid</a> , <a href="#">Le web : entre stratégies prédictives et technologies injonctives</a>

Ex. du Zeitgeist de Google créé en 2001 (J. Battelle, 2005) : classement des recherches les plus populaires : visibilité des centres d'intérêts des internautes, modes, intentions (« *database of intentions* »)...

comprend les requêtes, les résultats, les chemins empruntés par l'utilisateur (clics), en temps réel  
→ outil important pour étudier les internautes et leur proposer des produits et services liés, cf. développement des liens commerciaux sponsorisés (système AdWords – publicité par mots-clés) et des publicités ciblées, mais aussi outil pouvant être exploité à toute autre fin que commerciale (surveillance...)

« *Anticipatory search* » (O. Thomas, 2013, <http://readwrite.com/2013/05/15/google-search-anticipatory-system-io13#awesm=~o91GOpQyv9yO4t>) - cf. E. Schmidt, 2010 (<http://searchengineland.com/schmidt-great-stage-search-is-autonomous-personal-50014>) ou « recherche contextuelle » (cf. C. Mims, 2014, <http://qz.com/205689/context-this-is-what-comes-after-search/>) : « *In the near future of search, you won't be looking for Internet content anymore – it will be looking for you* » (B. Proffitt, 2013, <http://readwrite.com/2013/04/25/forget-searching-for-content-soon-content-will-be-searching-for-you#awesm=~o91yedU762HiO7>).

## La recherche anticipatoire



[Google Now](#)

rôle du contexte (localisation, objets connectés...) pour comprendre et prédire  
→ « recherche contextuelle » (au-delà du simple moteur de recherche)

### Contexte :

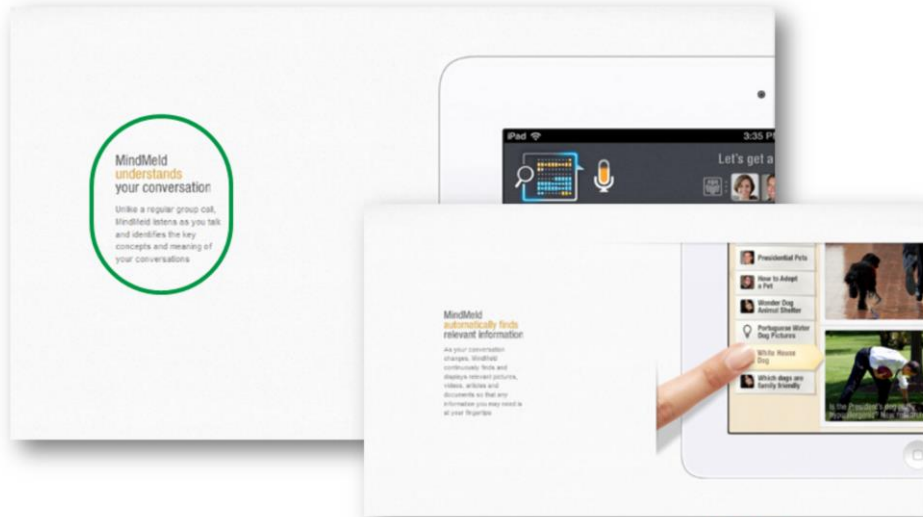
- recherche contextuelle (recherche personnalisée, géolocalisation...) cf. C. Mims, 2014, <http://qz.com/205689/context-this-is-what-comes-after-search/>
- développement d'agents intelligents permettant de chercher de l'information et de gérer sa vie, cf. Siri d'Apple (moteur de recherche, et outil permettant d'interagir, ex. ouvrir une application, etc.) - développement avec les mobiles
- phénomène du *big data* (internet + personnels), cf. « *quantified self* » : possibilité d'automatiser certaines tâches et prédire des comportements (cf. O. Ertzscheid, 2013, [http://affordance.typepad.com/mon\\_weblog/2013/04/prediction-psychosociologie-data-centrique.html](http://affordance.typepad.com/mon_weblog/2013/04/prediction-psychosociologie-data-centrique.html)) et internet des objets (dialogue entre les objets, ex. : téléphones, montres, ordinateurs, caméras de sécurité, détecteurs...) → recherche des données fournies par ces objets, ex. : Shodan (cf. B. Proffitt, 2013, <http://readwrite.com/2013/04/26/how-the-internet-of-things-will-revolutionize-search#awesm=~o91JOFn1SnsGuG>)

*Google Now* (2012) et son développement *Now On Tap* (2015) : association du contexte, des réponses et de l'action (<https://www.youtube.com/watch?v=7V-flGMDsmE&feature=youtu.be>, 49'40) : application qui doit apporter tous les éléments nécessaires, sans même les demander : pense-bêtes, météo, unes de journaux, résultats sportifs, informations sur un trajet... : “*you get just what you need to know, right when you need it*” (<http://www.youtube.com/watch?v=pPqliPzHYyc>)

- répondre à n'importe quelle question, même personnelle comme heure d'une réservation au restaurant ou temps nécessaire pour aller travailler : “search” étendue à tous les domaines
- résultats toujours plus personnalisés : « *personal search* » (E. Schmidt)
- ! commerciaux : pour que le service soit le plus performant, il convient que l'internaute utilise le plus de services Google et donne donc accès à toujours plus de données personnelles
- évolutions techniques : ce ne sera plus le web qui fournira des données à interroger, mais également les objets

# La recherche anticipatoire

Encore plus loin ?



MindMeid [app.] - image 2014

identification de concepts dans des conversations et présentation de documents en relation



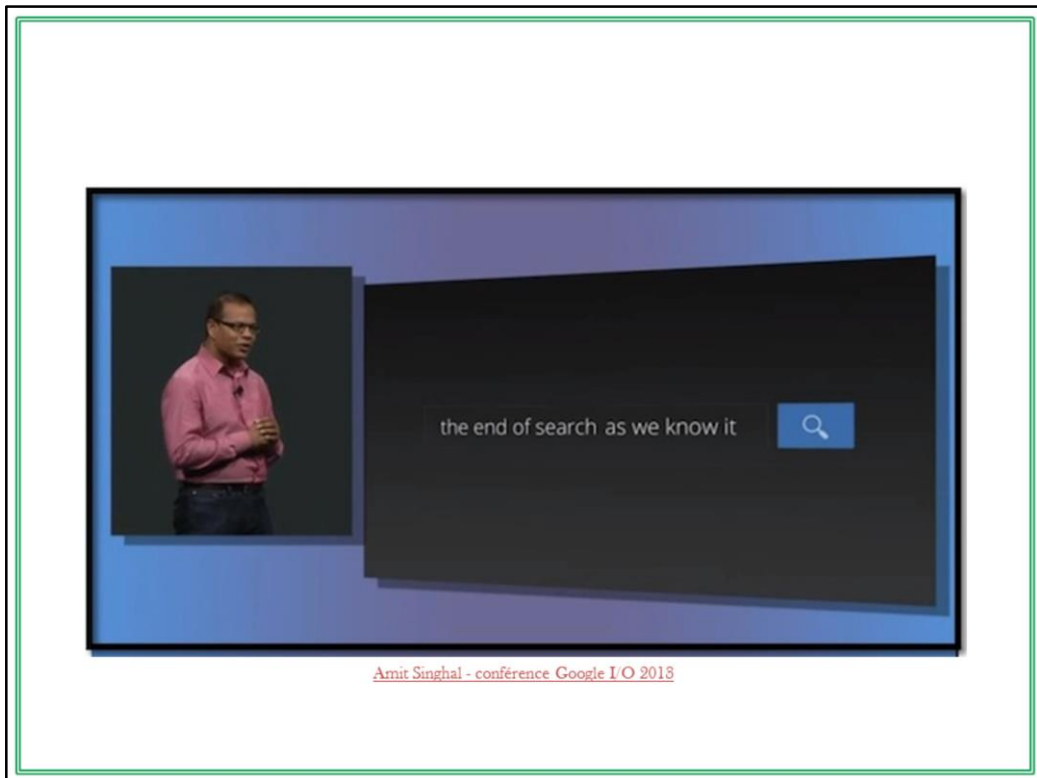
La recherche conversationnelle

## La recherche conversationnelle

<b>Limite des moteurs traditionnels</b>	comprennent mal le langage naturel et les questions orales
<b>Type de documents</b>	informations factuelles essentiellement
<b>Recherche</b>	- recherche par mots-clés ou en langage naturel ( <i>conversational</i> = [langage] courant) cf. algorithme <a href="#">Hummingbird</a> de Google (09/2013) - l'utilisateur peut diriger le moteur vers l'information qu'il recherche
<b>Outils</b>	- outils spécifiques, reposant sur les avancées actuelles (recherche sémantique, traitement automatique du langage naturel...), capables de comprendre une question orale, voire d'y répondre oralement également ou de demander des clarifications - réponses doivent être immédiates (réponse à un besoin et une action, cf. <a href="#">C. Warren</a> )
<b>!</b>	- domaine encore en développement - outils anglo-saxons (US) : problème de la compréhension d'accent non américain - intérêt surtout pour les mobiles, ou si loin d'un clavier (handicapés, mains occupées...) - ne fonctionne pas encore en saisissant la requête par écrit - mêmes problèmes que les moteurs de réponse : base de connaissance ?, validité et fraîcheur des informations ?, problème d'une réponse unique, marchandisation des réponses... - dérives possibles ? : présentation orale d'une réponse unique, et non d'une liste de liens textuels → « nous éviter de lire, c'est nous éviter de comparer » ( <a href="#">O. Ertzscheid</a> )

Se faire comprendre la machine : **Graal des informaticiens** – cf. ordinateurs de Star Trek (F. Manjoo, 2013, [http://www.slate.com/articles/technology/technology/2013/04/google\\_has\\_a\\_single\\_towering\\_obsession\\_it\\_wants\\_to\\_build\\_the\\_star\\_trek\\_computer.html](http://www.slate.com/articles/technology/technology/2013/04/google_has_a_single_towering_obsession_it_wants_to_build_the_star_trek_computer.html))

« Nous gardions jusqu'ici deux avantages considérables sur les moteurs de « recherches » : celui de leur poser les questions que nous voulions, et celui d'être capable d'anticiper sinon sur les réponses, du moins sur la typologie des sites de réponses qu'ils nous fourniraient (un article de presse mais lequel ? une vidéo mais laquelle ? une page Wikipédia mais laquelle ?). Le passage au stade oral enterre notre dernière capacité d'anticipation. Nous n'attendons plus rien de ces « recherches » parce que pour l'essentiel d'entre elles nous ne les « projetons » plus, parce que leur coût cognitif est nul et que sa satisfaction immédiate épuise l'idée même d'une projection pourtant toujours possible dans un processus de requête élaboré ou non déjà prescrit (*Google Suggest*). Et qu'il existe une corrélation claire entre le coût cognitif des requêtes et le coût transactionnel des réponses : plus le premier est faible, plus le second est élevé. Telle est la martingale du capitalisme cognitif des liens sponsorisés » (O. Ertzscheid, 2013, [http://affordance.typepad.com/mon\\_weblog/2013/05/il-ne-lui-manquait-plus-que-la-parole.html](http://affordance.typepad.com/mon_weblog/2013/05/il-ne-lui-manquait-plus-que-la-parole.html))



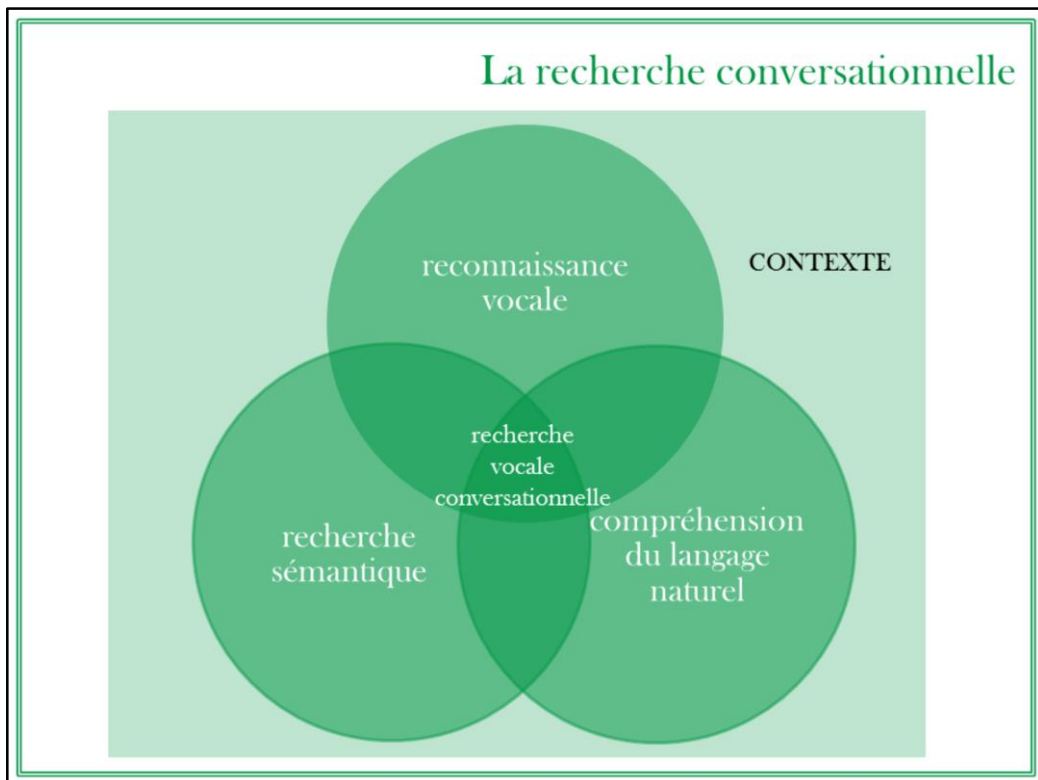
[Amit Singhal - conférence Google I/O 2013](#)

« At Google I/O today, we gave an update on where we are in building the search engine of the future - a search engine that can answer your questions, have a conversation with you, and even give you useful things without you ever having to ask » (2013, <http://insidesearch.blogspot.fr/2013/05/a-multi-screen-and-conversational.html>) – présentation d'Amit Singhal :

<http://www.youtube.com/watch?v=wjsA4J1BOrY> :

- répondre : *Knowledge Graph*
- converser : phrase « OK, Google » pour que le moteur réponde vocalement, mais service encore à développer (D. Sullivan, 2013, <http://searchengineland.com/googles-impressive-conversational-search-goes-live-on-chrome-160445>) + connexion avec services Google (Gmail, Google calendar)...
- anticiper : *Google Now* : cf. *supra*

Progrès considérable de la reconnaissance automatique : 23 % d'erreurs en 2013 contre 8 % en 2015 (Google I/O 2015, <https://www.youtube.com/watch?v=7V-flGMDsmE&feature=youtu.be>, 46'30)



« *The future of search will be context based and not query based, says the [Google] search team* » (D. Olanoff, 2012, <http://thenextweb.com/google/2012/08/08/google-opens-up-about-the-future-of-search-serves-100-billion-searches-every-month/>)

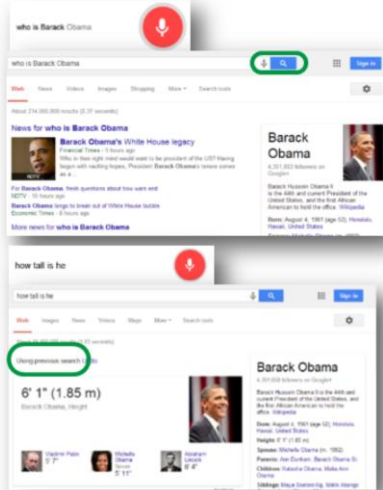
« *conversational search* » (Amit Singhal) : « *Conversational search is search that tries to understand context, that makes educated guesses, that takes voice input, that parses homonyms and adapts to mobile environments, and that understands the same user across multiple devices* » (L. Gannes, 2013, <http://allthingsd.com/20130314/how-search-is-evolving-finally-beyond-caveman-queries/>)

ex. :

Ex. : recherches sur Barack Obama : « qui est Barack Obama ?, qui est sa femme ?, combien mesure-t-il ? » : nécessité de développer l'historique de la recherche pour que les trois questions soient reliées et comprendre que « sa » et « il » renvoient à « Barack Obama »

## La recherche conversationnelle

Ex. : Google sur Chrome

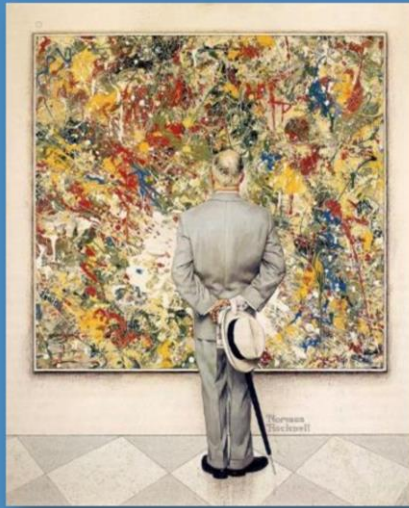


Présentation Google I/O keynote 2013 (2:00:00)  
implanté également sur la version française  
Exemples d'utilisations

Ex. : Evi



Existe depuis déjà quelques années sur les ordinateurs avec des logiciels spécialisés en reconnaissance vocale (et payants) comme Dragon NaturallySpeaking  
Compréhension du **langage naturel**, de manière suivie : pronoms, références à des éléments précédents, etc. : « mémoire immédiate « répondante » » (O. Ertzscheid, 2013, [http://affordance.typepad.com/mon\\_weblog/2013/05/il-ne-lui-manquait-plus-que-la-parole.html](http://affordance.typepad.com/mon_weblog/2013/05/il-ne-lui-manquait-plus-que-la-parole.html)), rôle de l'historique de recherche  
Inférences d'éléments de réponses à partir d'éléments de **contexte**, ex. : « *Will it rain tomorrow ?* »



## Visualisation des résultats

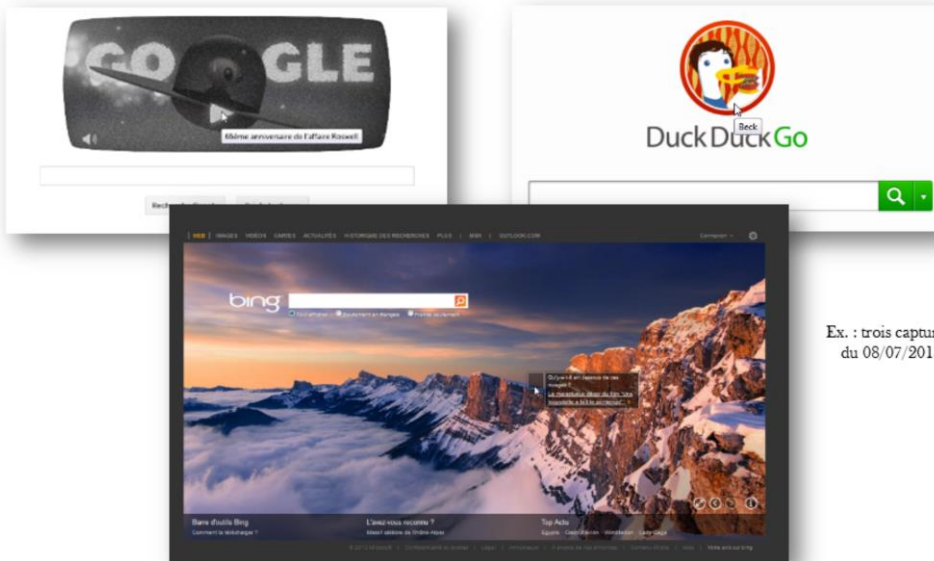
moteurs linéaires  
moteurs graphiques  
moteurs à *clusters*  
moteurs visuels  
moteurs multimédia

## Visualisation des résultats

<b>Limite des moteurs traditionnels</b>	<ul style="list-style-type: none"> <li>- accent mis sur la compréhension de la requête (contextualisation et sémantique)</li> <li>- présentation linéaire (liste de résultats), répondant à des logiques matérielles (taille de l'écran et lisibilité)</li> <li>- études d'<i>eye-tracking</i>, montrent que les premiers résultats sont plus visibles que ceux du bas de la page</li> <li>- peu d'évolutions visuelles, en dehors des <i>cards</i> et <i>snippets</i> sémantiques ou de services plus gadgets (ex. : <a href="#">Google tendances de recherche</a>)</li> </ul>
<b>Type de documents</b>	tous, mais surtout images
<b>Recherche</b>	<p>recherche textuelle</p> <p>« c'est peut-être l'étape qui propose le plus de possibilités, mais aussi la plus complexe à changer (la force des habitudes...) » (<a href="#">O. Andrieu</a>)</p>
<b>Outils</b>	<ul style="list-style-type: none"> <li>- moteurs linéaires</li> <li>- moteurs (carto)graphiques</li> <li>- moteurs à clusters</li> <li>- moteurs visuels</li> <li>- moteurs multimédia</li> </ul>
<b>!</b>	<ul style="list-style-type: none"> <li>- intérêt de la visualisation des résultats : compréhension plus rapide que le texte, rapprochements potentiels d'idées voire désambiguïsation et analyse (lutte contre infobésité)...</li> <li>- outils généralement en version anglaise uniquement</li> <li>- souvent simple habillage de résultats Google, avec peu/pas de recherche avancée</li> <li>- outils souvent lents et parfois difficiles à utiliser</li> <li>- peu d'interaction</li> <li>- utilité pas toujours prouvée (au mieux, moteurs d'inspiration ; au pire, gadgets)</li> <li>- peu de services sonores et/ou parlants</li> <li>- nombreux services morts (<a href="#">Kartoo</a>, <a href="#">Grokker</a>, <a href="#">SearchCube</a>...)</li> <li>→ impact marginal sur le secteur pour l'instant (accent mis surtout sur l'amélioration de la page de résultats, plus que sur visualisation)</li> </ul>

## Visualisation des résultats

Des interfaces suscitant le désir d'information (B. Simonnot)



« L'objectif [des moteurs] n'est plus, selon nous, de répondre à des besoins mais de susciter des désirs d'information » (B. Simonnot, 2012, p. 136)

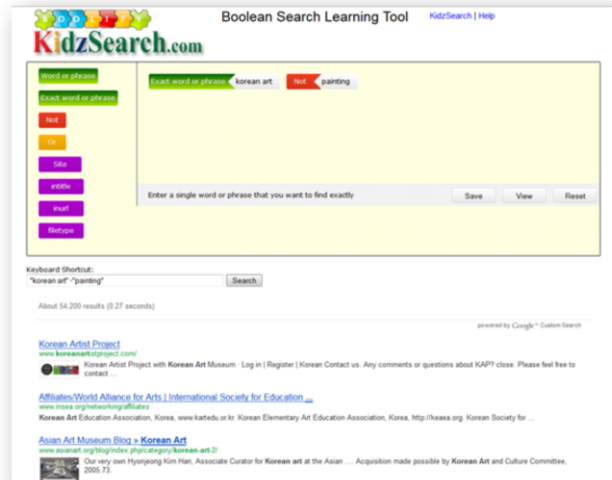
Ex. : doodles de Google ou DuckDuckGo, images cliquables de Bing : l'internaute peut s'informer sur des sujets auxquels il n'aurait pas pensé *a priori* ; sorte de prescription du moteur

mais pourquoi ces choix-là et pas des autres ? les trois captures d'écran ont ainsi été faites le même jour à quelques minutes d'intervalle : Google présente le 66<sup>e</sup> anniversaire de l'affaire Roswell, DuckDuckGo présente Beck et Bing montre le massif du Vercors

→ les moteurs de recherche proposent désormais également des contenus éditorialisés

# Visualisation des résultats

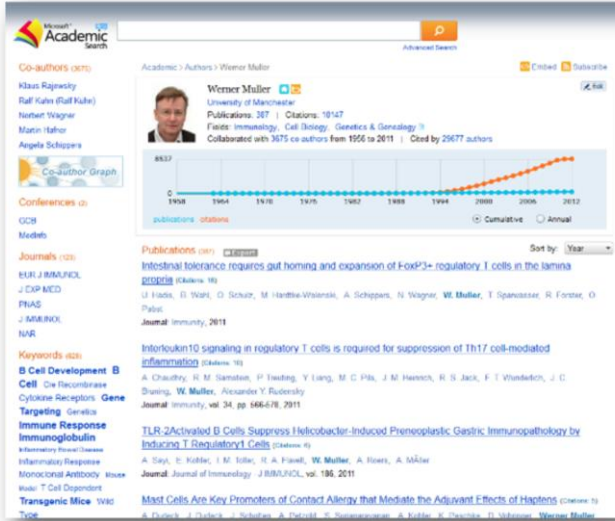
## Des interfaces ludiques et pédagogiques



Ex. : Boolify de KidzSearch : utilisation pédagogique de Google recherche [« korean art » -painting]

# Visualisation des résultats

Des résultats toujours plus graphiques



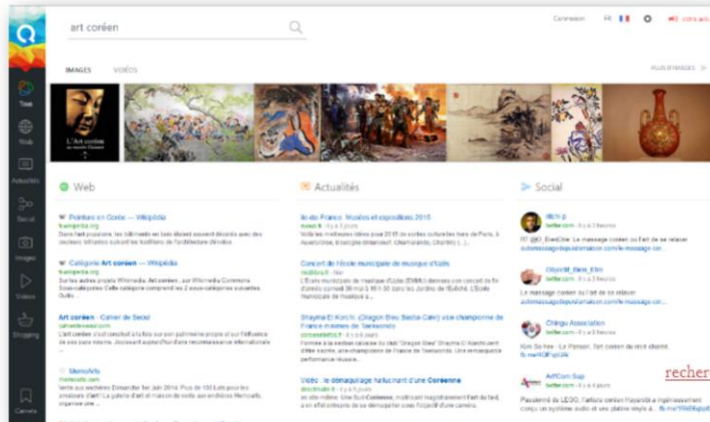
moteur scientifique : Microsoft Academic search (fin annoncée, captures 2014)



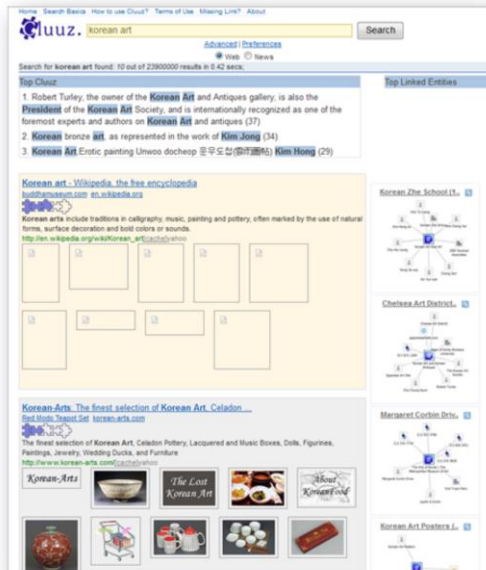
réseaux de co-auteurs, de citations...

# Moteurs linéaires

- **Qwant** : <http://www.qwant.com>  
moteur de recherche français, ouvert officiellement en 07/2013 (éléments techniques) - nouvelle version en 4/2015  
couverture de 15 langues et 25 pays  
présentation en catégories (web, actualités, social, images, vidéos, shopping, « Knowledge graph »), mais interface web standard possible  
respecte les recommandations de la CNIL sur les données personnelles (privacy)
  - pas de recherche avancée, tris...une bonne alternative à Google ?

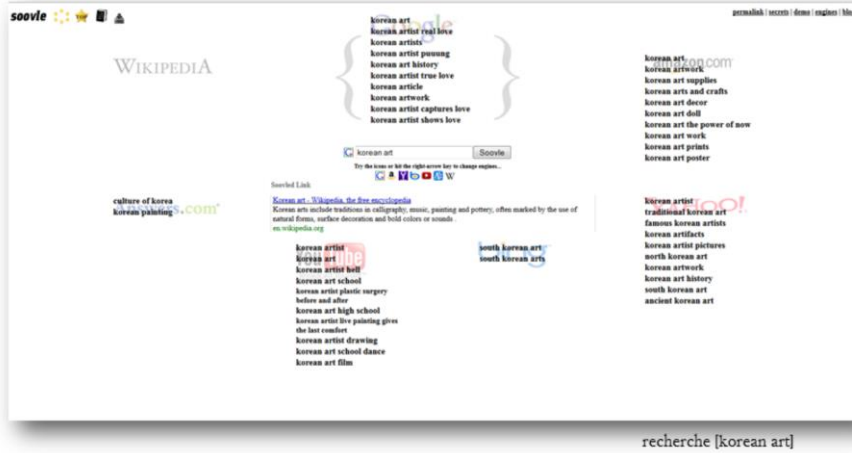


- Cluuz : <http://cluuz.com>  
technologie Yahoo  
présentation étoffée de chaque résultat  
(mots-clés, images...) et arborescence  
(graphe sémantique)  
mise en valeur des entités  
- © 2008 , pas de résultats récents  
! quelques lenteurs



# Moteurs linéaires

- Soovle : <http://www.soovle.com/>  
métamoteur interrogeant plusieurs moteurs et bases de données (dont Google, Bing,....,jusqu'à 15 simultanément) et présentant les recherches associées  
- uniquement sur les sites anglais



- SearchPoint : <http://www.searchpoint.com>
  - ! © 2012
  - ! quelques lenteurs

The screenshot shows a search engine interface with the search term 'korean art' entered. The results are listed as follows:

- (1) [Korean art - Wikipedia, the free encyclopedia](#)  
Korean arts include traditions in calligraphy, music, painting and pottery, often marked by the use of natural forms, surface decoration and bold colors or sounds.  
[http://en.wikipedia.org/wiki/Korean\\_art](http://en.wikipedia.org/wiki/Korean_art)
- (2) [Korean Arts, The finest selection of Korean Art, Celadon](#)  
The finest selection of Korean Art, Celadon Pottery, Lacquered and Music Boxes, Dolls, Figures, Paintings, Jewelry, Wedding Ducks, and Furniture  
<http://www.korean-arts.com/>
- (3) [Korean art | Britannica.com](#)  
Korean art, the painting, calligraphy, pottery, sculpture, lacquerware, and other fine or decorative visual arts produced by the peoples of Korea over the centuries.  
<http://www.britannica.com/EBchecked/topic/719289/Korean-art>
- (4) [Korean Art | LACMA](#)  
Korean wrapping cloths (known as boggi in Korean) are celebrated for both their form and function. As wrapping cloths, boggi were used ubiquitously in premodern ...  
<http://www.lacma.org/art/collection/korean-art>
- (5) [Korean Art, Korean Paintings & Framed Artwork by Korean](#)  
24 hr shipping on most orders. Completely customizable Korean art from Zazzle.com - Select your favorite Korean designs for art prints and posters of all sizes.  
<http://www.zazzle.com/korean-art>
- (6) [Korean Art Posters at AllPosters.com](#)  
Korean Art Posters at AllPosters.com. Choose from over 500,000 Posters and Art Prints. Value Framing. Fast Delivery. 100% Satisfaction Guarantee.  
[http://www.allposters.com/-di/Korean-Art-Posters\\_c130353\\_item](http://www.allposters.com/-di/Korean-Art-Posters_c130353_item)
- (7) [Korean Art - Asia - Art](#)  
Korean art is fascinating because although it has been deeply influenced by Chinese art, the most productive periods in terms of art often do not coincide between the ...  
[http://www.asia-art.net/korean\\_art.html](http://www.asia-art.net/korean_art.html)
- (8) [Korean art](#)  
Introduction. This site is to introduce and promote Korean art and artists that haven't been well known to global art viewers so far. We may be able to find many ...  
<http://www.artstudio.com/>
- (9) [About The Art and Culture of Korea - Korean Arts](#)  
Offering the finest selection of Korean Celadon Pottery, Lacquered and Music Boxes, Dolls, Figures and Paintings. 韓國의 예술과 문화에 대해 ...  
[http://www.korean-arts.com/about/art\\_of\\_korea.html](http://www.korean-arts.com/about/art_of_korea.html)
- (10) [Korean Art | Thematic Essays | Heilbrunn Timeline of Art](#)  
Art of the Korean Renaissance, 1400-1600. ... List of Rulers of Korea, Mountain and Water Korean Landscape Paintings, 1800-1870. Paintings From the East Asian Art

recherche [korean art]

The search results are accompanied by a circular navigation menu with the following items: IMAGE, SPECIAL, SOCIETY, NORTH, ARTISTIC, HISTORY, ANTIQUITIES, POSTERS, CONTEMPORARY, and MUSEUM.

# Moteurs graphiques

- Touch Graph : <http://www.touchgraph.com/seo>
  - organisation des résultats en fonction des liens entre les sites internet (données Google)
  - nombre de résultats volontairement limité
  - disponible également pour [Amazon](#) (livres, musique, films)
  - ! nécessite Java
  - ! quelques lentesurs

The screenshot displays the Touch Graph web application interface. On the left, a search results pane for 'art coréen' lists eight items, including Wikipedia articles, Wikimedia Commons pages, and various art-related websites. On the right, a network graph visualizes the relationships between these sites. The central node is 'ART COREEN'. It is connected to several other nodes, including 'w', 'runningart', 'cahierdesou', 'nupartcenter', 'universita.fr', 'boulquedesmusees.fr', 'artsdelacoreehypotheses', 'gumet.fr', 'lacma', 'visuarts-cork', 'metuseum', 'koreanartantiques', 'asianart', 'korean-arts', 'asiast', and 'britannica'. The interface also includes a search bar at the top, a 'Search' button, and options for 'Remove', 'Export Data', and 'Export Graph'.

[recherche \[korean art\]](#)  
[suivie de recherche \[art coréen\]](#)

## Moteurs graphiques

- Instagrok : <http://www.instagrok.com>  
relations entre des concepts  
malgré possibilité de filtres de différents niveaux, de niveau scolaire plutôt primaire ou secondaire  
(informations basiques)





## Moteurs à clusters

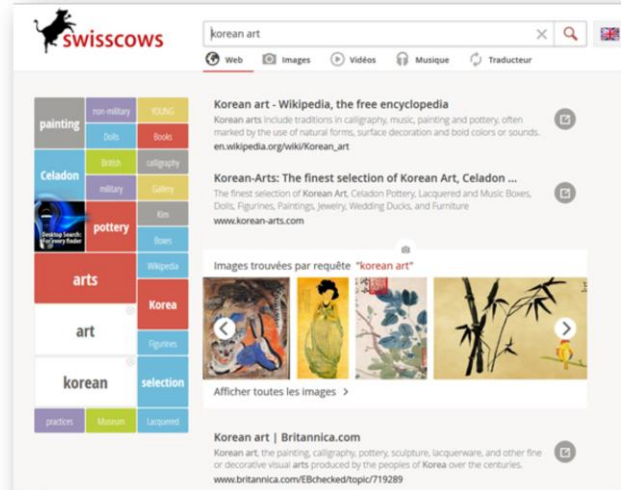
- Yippy : <http://yippy.com/>  
anciennement Clusty  
moteur de recherche classant les 200 à 500 premiers résultats des moteurs dans des dossiers

The screenshot shows the Yippy search engine interface. The search bar at the top contains the text "korean art". Below the search bar, there are tabs for "clouds", "sources", "sites", and "time". The main content area is divided into two columns. The left column lists various categories with their respective document counts, such as "All Results (425)", "Martial Arts (125)", "Culture (98)", "Photos (42)", "Film (15)", "Museum (28)", "Pottery (8)", "Collections, School & Teachers (6)", "International, Seoul (6)", "Reviewed, Book (2)", "LACMA (2)", "Following Information Was Released By Radio Free (2)", "Italian, Virginia Beach (2)", "Military, Historical (2)", "Other Topics (1)", "Performing (17)", "Free (18)", "Sale (18)", "Antiques, Japanese (8)", "Magazine (14)", "Resources (11)", "Jewelry (18)", "State Of The Art (11)", "Media (8)", "Market, Blog (8)", "Norwegian, Japanese (18)", "Automotive (8)", and "School District (8)". The right column displays search results for "Cluster Museum contains 29 documents". The first result is "LACMA strikes 'unprecedented' museum deal w/...", followed by "National Jazz Museum in Harlem", "Durham, NC", "Welcome to the International Dunhuang Project", "Bona Museum & Gallery", "The Metropolitan Museum of Art: Arts of Korea", "One Line", and "Pots of it Korean art (coming out of the shadows)".

recherche [korean art]

## Moteurs à clusters

- Swisscows : <https://swisscows.ch>  
moteur de recherche suisse  
ne collecte pas les données personnelles

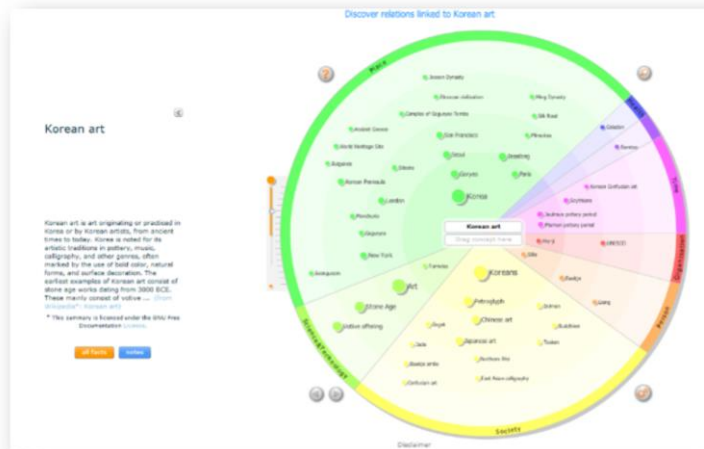


[recherche \[korean art\]](#)



## Moteurs à clusters

- Vionto : <http://en.vionto.com/show/>  
représentation visuelle de concepts (jusqu'à 500) et de leurs relations
  - + possibilité de pondérer les catégories
  - résultats seulement issus de Wikipédia ?

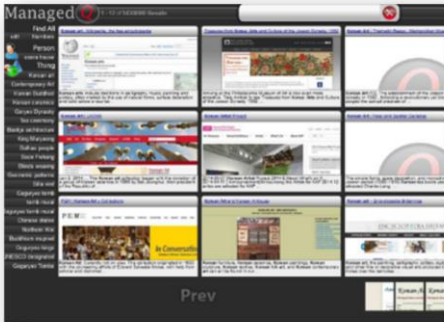




## Moteurs visuels

### captures d'écrans (*Thumbshots*)

- **ManagedQ** : <http://managedq.com>  
fonctionnalité permettant de trouver des termes de recherches associées dans les pages de résultats
- **RedZ** : <http://redz.com>  
recherche web, images, vidéos  
! capture d'écran ne correspond pas toujours à la réalité  
liste de vignettes comme ManagedQ ou carroussel



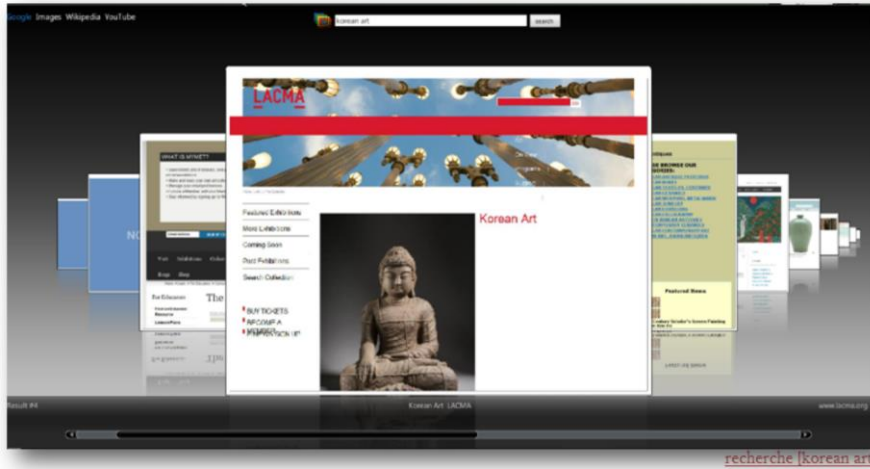
[recherche \[korean art\]](#)



[recherche \[korean art\]](#)

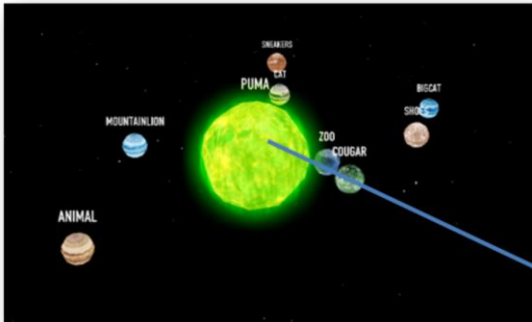
## carrousels

- Spacetime 3D : <http://app.spacetime3d.com>  
recherche Google et Wikipédia (images et YouTube ne fonctionnent plus)



## Moteurs visuels

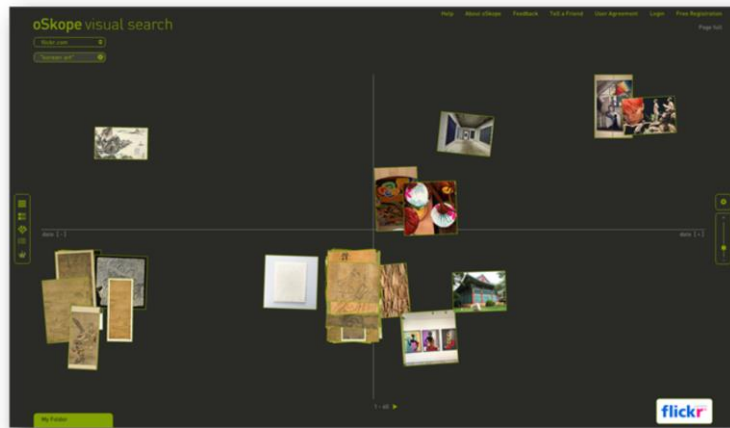
- Tag Galaxy : <http://taggalaxy.de/>  
recherche de photos (Flickr) par tags  
présentation 3D  
! pas de désambiguïisation des termes



recherche puma

### autres types de présentation

- oSkope : <http://www.oskope.com/>  
recherche Flickr, Fotolia...  
différents types de présentation (grille, pile, liste...)



## Moteurs multimédia

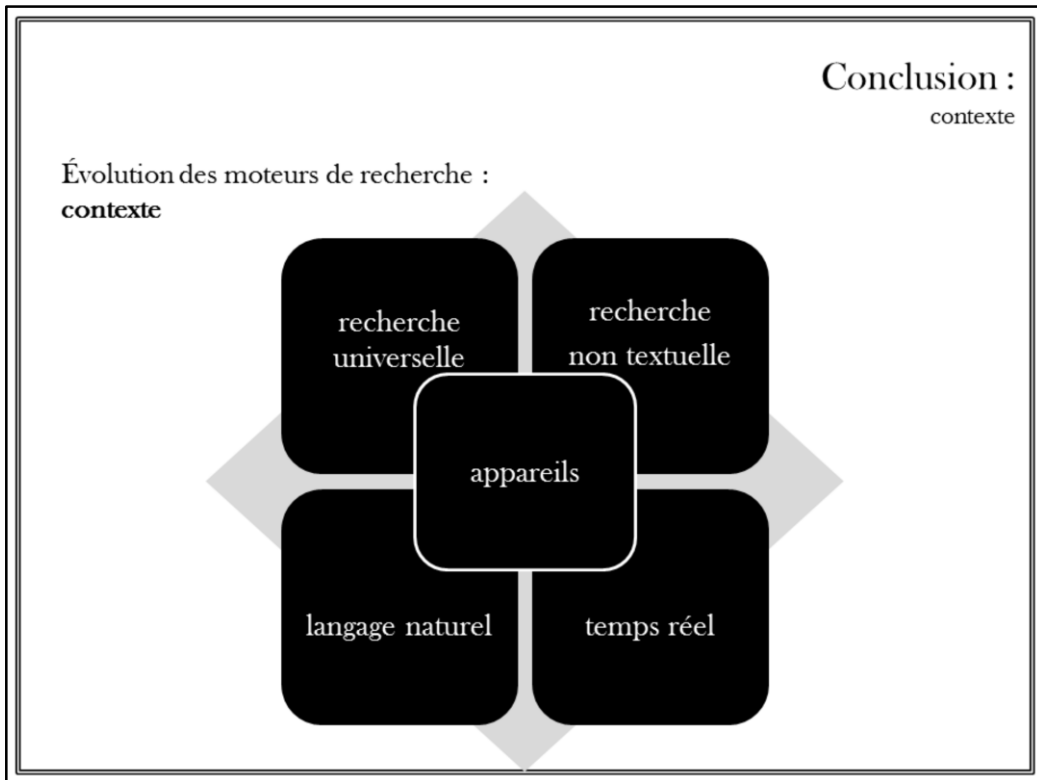
- Oamos : <http://www.oamos.com>  
plusieurs filtres, dont mode (*objective* > *entertaining*)  
présence d'images, de sons (*meta-inspiration engine*)  
divers types de présentation (vignettes, bulles, textes, tags...) via la colonne de gauche, avec remix possible
  - peu ergonomique
  - résultats pas toujours très pertinents par rapport à la requêtesources ?



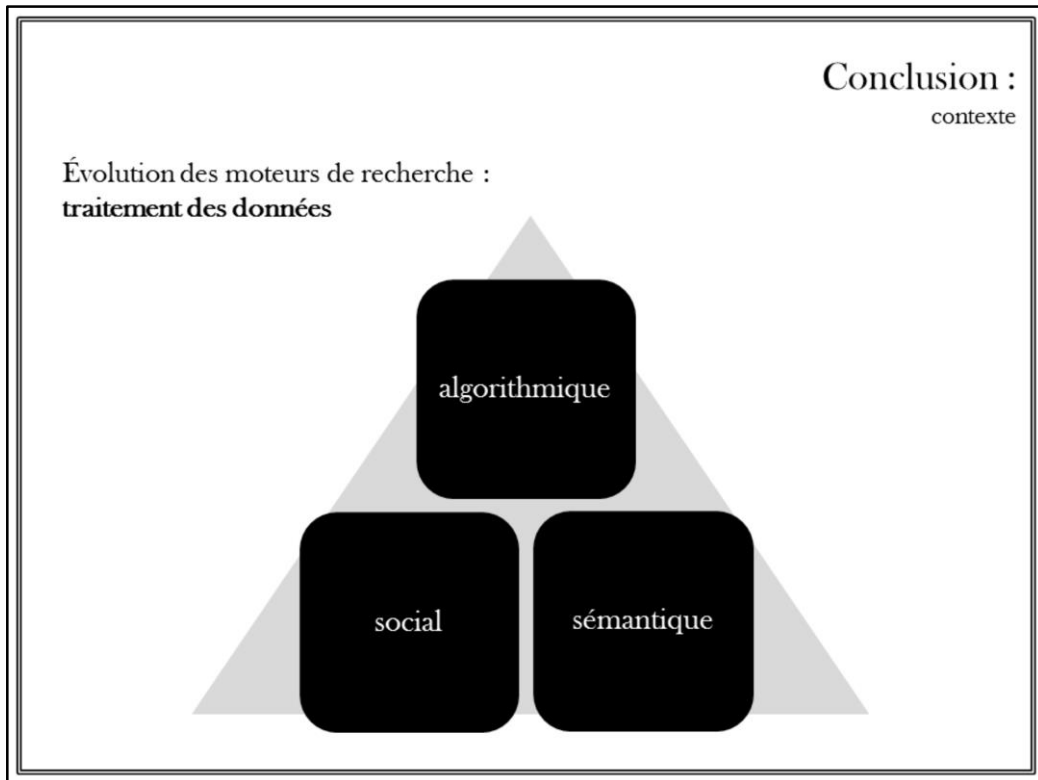
recherche [- korean art]



## Conclusion



« "In order to win the new battle of search, you're going to have to win all of the elements of the unified experience," Winarsky says. "You better be a dominant player in social networks ... you better be a great player in artificial intelligence and speech recognition ... you need to be able to understand the content of videos and images, and you need a far better interaction experience that enables you to better understand the human interaction with the tablet." » (in J. Van Grove, 2011, <http://mashable.com/2011/09/13/tablets-changing-search/>)



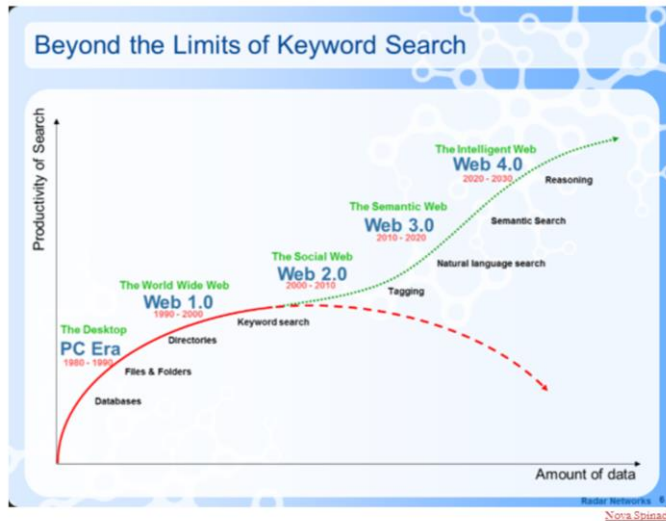
**Problème** : beaucoup d'outils restent dans des logiques qui ont déjà fait leurs preuves (cf. manque de diversité pour la présentation des résultats)  
quelle force d'innovation et quels moyens (humains, techniques et financiers) face à Google ?  
de nouveaux modes d'accès l'information, moins basés sur les questions techniques que sociales (V. Mesguich, 2011)

Ex. : *Five types of music discovery* (S. Dredge, 2014,  
<http://www.theguardian.com/technology/2014/mar/19/music-discovery-spotify-apps-facebook>) :

- *friends*,
- *the crowd*
- *curators*
- *algorithms*
- *serendipity*

Conclusion :  
contexte

Évolution des moteurs de recherche :  
diminution de la recherche par mot-clé



Doc Sheldon : vraie recherche sémantique : à l'horizon 2025 : les données et les techniques sont déjà là  
(<http://searchenginewatch.com/article/2222577/Semantic-Search-in-2025>)

#### Question cruciale des données :

- *qualité et fiabilité* : pour l'instant, web sémantique reste encore basé sur des initiatives communautaires (Freebase, Wikipédia) et peu sur des initiatives institutionnelles, permettant d'avoir des informations validées
- *croisement des données* et, surtout, leur *synthèse* :
  - recherches complexes : « *Does Bill Cosby come from a large family ?* », « *How Bill Cosby's wife's next door neighbor voted in the 1948 presidential election ?* » (D. Sheldon), « Dans quelles circonstances le troisième président des Etats-Unis abjura-t-il de sa conception de l'esclavage ? » (J. Battelle), quelles sont les universités de la côté ouest des Etats-Unis qui ont des droits d'inscriptions inférieurs à 30 000 \$ ?
  - synthétiques : ex. : « *What does Bill Slawski say about patents ?* » (D. Sheldon) : présentation de rapports et résumés sur les différentes positions et tendances...
  - possibilité d'obtenir des réponses personnalisées : ex. : recherche en vue d'un achat dont les réponses prendraient en compte les besoins, les attentes et les moyens de l'internaute (goûts, prix, localisation, avis sur internet et recommandations d'amis...)
  - saurait ce qu'on a déjà vu
  - capable de distinguer les documents des personnes : ex. : suggestion que l'on pourrait obtenir la meilleure réponse en consultant telle personne, mais qu'on perdrait son temps en consultant tel document (J. Battelle)

#### Question cruciale des techniques

cf. achat d'un ordinateur quantique Google/NASA (05/2013), dont l'un des buts est de mieux comprendre le langage naturel, et le fonctionnement selon un modèle neuronal pour développer l'intelligence artificielle

« En somme, le moteur de recherche de l'avenir ne ressemblerait plus vraiment à ceux que nous connaissons aujourd'hui. Il s'agirait plutôt d'un agent intelligent, ou comme Larry Page me l'a suggéré, d'un bibliothécaire de référence maîtrisant toutes les connaissances humaines » (J. Battelle, 2005, p. 228)  
cf. déjà les services de questions/réponses en ligne de type Guichet du savoir, Rue des facts... - métamoteur dans les différents services : <http://www.sqrpro.fr/moteur-de-recherche-dans-la-base-de-connaissance/>

Conclusion :  
Google ?

*The Best Search Engines of 2015*  
([Paul Gil](#))

DuckDuckGo  
Google Scholar  
Ask  
Bing  
Mahalo  
Dogpile  
Yippy  
Webopedia  
Yahoo  
The Internet Archive  
**Google**

*The Best Search Engines of 2014*  
([Wendy Boswell](#))

1. Bing  
2. Indeed  
**3. Google**  
4. The Internet Archive  
5. Pinterest  
6. DuckDuckGo  
7. WolframAlpha  
8. IMDB  
9. USA.gov  
10. Twitter

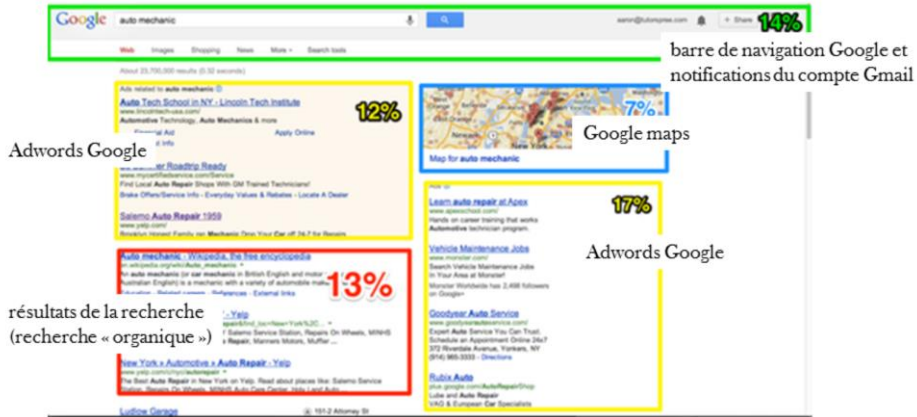
→ place toujours centrale de Google

Google : toujours un temps d'avance

« Malheureusement, beaucoup d'outils récents restent dans ces logiques où des acteurs ont déjà fait leurs preuves et sont donc difficilement remplaçables » (F. Jeanne-Beylot, <http://fjb.blogs.com/weblog/2013/03/a-qwant-un-moteur-de-recherche-nouvelle-g%C3%A9n%C3%A9ration-.html>)

Conclusion :  
Google ?

*How Google is Killing Organic Search ? (A. Harris)*



Google est-il en train de tuer la recherche « organique »  
au profit de ses seuls produits ?

cf. problèmes avec la Commission européenne pour abus de position dominante (ZDNet, 04/2013)

diminution des résultats de la recherche organique au profit de contenus produits par Google et générant des revenus

les moteurs de recherche se font de plus en plus pourvoyeurs d'informations (médiation d'informations et de services, éditorialisation des contenus) et non plus simples outils de recherche : abandon de toute objectivité ; le service dirige l'internaute vers ses propres services, qui seraient à même de répondre à ses intentions (J. Battelle, 2005, p. 215)

mais déjà très présent chez Yahoo depuis début des années 2000 : « Chez Yahoo, les résultats organiques ne forment qu'une possibilité (extrêmement importante) parmi toute une variété de services proposés par la société. Quand on effectue une requête, les résultats naturels sont toujours présents, mais sont accompagnés d'autres produits développés par l'entreprise pour répondre à l'intention sous-jacente qu'elle devine dans les termes introduits » (J. Battelle, 2005, p. 219)

Conclusion :  
quelle fiabilité des résultats ?

- une attente des internautes (2009) :

PRESENTATION / AFFICHAGE DES  
RESULTATS  
Comment pouvons-nous améliorer  
l'affichage et la visualisation des  
résultats ?

M'aider à identifier parmi tous les résultats ceux qui sont les plus fiables (questions de santé, scientifique...)

- quelles avancées ?

ex. : Google

- échec du projet d'AuthorRank pour meilleure prise en compte de la notion d'*authorship*
- vers une analyse de la véracité des faits pour classer les résultats ?

→ vers des *Knowledge vaults*

- auto-apprenants (quelle place pour la fiabilité ?)
- permettant de développer le croisement d'informations et la réalité augmentée (quelles limites ?)
- disposant de toujours plus de données personnelles (quelle protection ?)
- gérés par des grandes entreprises (quelle ouverture des données ?)

*AuthorRank* : via la balise rel=author, introduite en 2011

- permet aux webmasters d'améliorer le classement de leurs pages en mettant en avant la notion d'*authorship* (qualité et traçabilité de l'auteur), notamment grâce au profil Google + de la personne
- plus seulement popularité de la page, mais aussi de son auteur

## Conclusion : les « jardins fermés »

La recherche d'informations au prisme des « [jardins fermés](#) » (T. Berners-Lee)

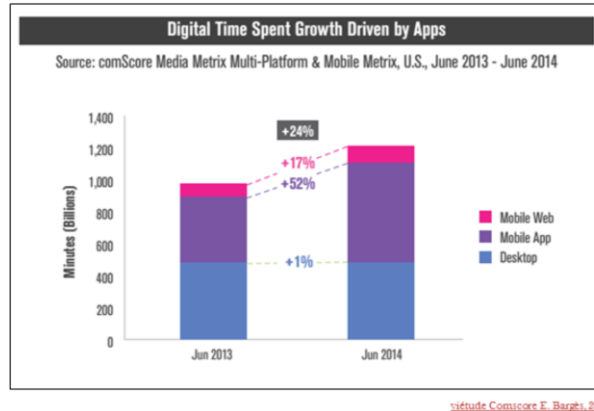
Ex. 1 : Google : mise en valeur d'écosystèmes propres



interface 2013 : YouTube = vidéos

## Conclusion : les « jardins fermés »

### Ex. 2 : développement des app. au détriment du web mobile



→ un web de moins en moins universellement « *searchable* » (T. Battelle)

? quelle innovation (T. Berners-Lee)

? vers une convergence search/app sur mobile ? cf. annonce de Now On Tap à Google I/O 2015 (48')

développement prévisible d'internet et de la recherche sur toute sorte d'appareils : ordinateurs, *smartphones*, et autres objets (TV, automobile...) – rappel : internet est bien plus large que le seul web, et comprend aussi les applications pour *smartphones* et tablettes numériques

« *The largest issue with search is that we learned about it when the web was young, when the universe was “complete” – the entire web was searchable! Now our digital lives are utterly fractured – in apps, in walled gardens like Facebook, across clunky interfaces like those in automobiles or Comcast cable boxes. Re-uniting our digital lives into one platform that is “searchable” is to me the largest problem we face today* [...] *So we all search now, all the time, across all manner of artifacts, large and small. But our searches are not federated – we can't search across these repositories, as we could across that wonderful, vast, loosely joined early world of the web. We've lost the connection* » (J. Battelle, <http://battellemedia.com/archives/2012/08/what-is-search-now-disjoined.php>)

applications mobiles ressemblent à des bases de données fermées, qui ne sont pas indexables par les moteurs de recherche et donc interrogeables (cf. O. Ezratty,

<http://www.oezratty.net/wordpress/2011/enjeux-appstorisation-internet/> et T. Worstall,

<http://www.forbes.com/sites/timworstall/2012/10/20/have-we-reached-peak-google/>)

→ évolution des pratiques de recherche : moins « *searching* » que « *getting* » (C. Anderson et M. Wolff, [http://www.wired.com/magazine/2010/08/ff\\_webrip/](http://www.wired.com/magazine/2010/08/ff_webrip/))

Rapprochement du search et des apps : service *Now on tap* d'Android comme assistant permettant de lancer des recherches à partir du contenu de certaines applications : « *This means that you can get contextual search information around almost anything you're doing, provided there is text and data that Google can pull from the app itself* » (J. Topolsky, <http://www.bloomberg.com/news/articles/2015-05-28/what-google-just-announced-is-a-bombshell>).

## Conclusion

les « jardins fermés »

« *Is search changing ?* » (M. Cuban, via O. Ertzscheid)

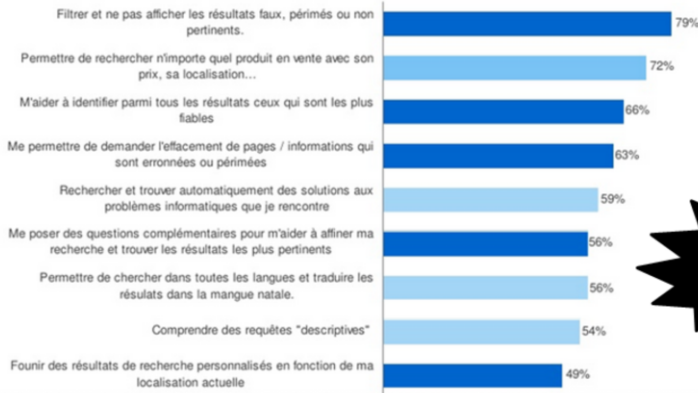
« Les moteurs de recherche traditionnels perdent de leur importance. Si je veux des images, je vais aller sur Pinterest ou Flickr plutôt que sur Google ou Bing. Si je suis sur Facebook, c'est là que je ferai une recherche pour trouver ce dont j'ai besoin et le trouver directement [...] ; si ce sont des informations, je les obtiendrai directement de Twitter et je suivrai un site indiqué en lien ; si j'ai besoin de savoir ce qui se passe dans le métier, j'obtiendrai les données sélectionnées pour moi par l'un de mes outils de veille spécial tablettes, et si je veux quasiment tout le reste, j'irai le chercher en utilisant une app. si je suis par monts et par vaux »

Phil Bradley

*« Traditional search engines are becoming less important. If I want images I'll go to Pinterest or Flickr rather than Google or Bing. If I'm on Facebook I'll do a search there to find what I need and go straight off to it [...], if it's news, I'll get that directly from Twitter and jump straight to a site that's linked, if I need to see what is happening in the profession I'll get that data pulled up for me by one of my tablet based news curation tools, and if I want most of anything else, I'll search for it using an app if I'm out and about » (P. Bradley)*

## Et des attentes seulement partiellement résolues

TOP des fonctionnalités les plus attractives



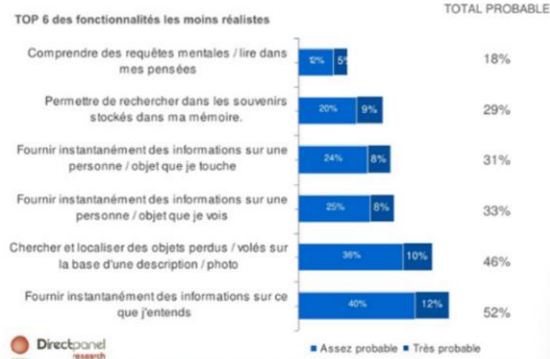
[étude Google](#)

Types de fonctionnalités les plus attractives



## Et encore d'autres pistes

La plupart des nouvelles fonctionnalités testées semblent réalistes 



[étude Google 2009](#)

[Google Nose \(1<sup>er</sup>/04/2013\)](#)

**Chien mouillé**

Une odeur qui a du chien. Notes de musc et de serviette mouillée. Google Aromathèque

[Sentir](#) [Partager](#) [Fonctionnement](#)

D'autres internautes ont également senti :

-  Beurre de cacahuète
-  Pastille de menthe
-  Piles mal installées
-  Taxi
-  Pamplém...

Eric Schmidt (ancien PDG de Google) : « Google vise à vous aider à trouver n'importe quoi » (J. Battelle, p. 223)

cf. développement des réponses issues des services associés (Gmail + Google+ +...)

→ découverte sur internet = documents et personnes

mais, dans un monde ultra-connecté, recherche beaucoup plus large que ça, cf. objets équipés de puces RFID – la recherche pourra porter sur tout, ex. : bagage perdu...



## Bibliographie

# Bibliographie

## Généralités sur la recherche d'informations

- Aline Bouchard. *Recherche d'informations sur internet (perfectionnement) : méthodologie et outils disponibles*. URFIST de Paris. Présentation. 135 p. 2015. [en ligne]. Disponible sur : <http://slidehare.net/URFISTParis/recherche-d-information-sur-internet>.
- Tony Paragiso. *Mode de recherche : recherche par mots-clés*. Présentation. 71 p. 2012. [en ligne]. Disponible sur : [http://dolara.fr/wp-content/uploads/2012/11/Diaporama\\_moteurs\\_recherche.pdf](http://dolara.fr/wp-content/uploads/2012/11/Diaporama_moteurs_recherche.pdf) [lien inactif au 15/04/2013].
- Béatrice Foenix-Niou. *Recherche évaluée sur Internet : mode d'emploi : outils et méthodes pour explorer le Web, Web visible, Web invisible, Web social, Web temps réel*. Paris : Lavoisier : Bases publication, DL 2011. 367 p.
- Marie-Laure Maligou et Alexandre Serres. *Approfondissement des moteurs de recherche : - sortir - de Google...*. URFIST de Rennes. Présentation. 38 p. 18/11/2008. [en ligne]. Disponible sur : <http://slidehare.net/idevaref/moteurs-de-recherche-sortir-de-google-presentation> et mise à jour *Outils de recherche du web hors Google...*. URFIST de Rennes. Présentation. 53 p. 21/10/2010. [en ligne]. Disponible sur : <http://www.sites.univ-renne2.fr/urfist/ressources/moteur-de-recherche-sortir-de-google>, avec documents complémentaires.
- ... *Quels outils pour quelles recherches ? Panorama des outils de recherche du web*. URFIST de Rennes. Présentation. 88 p. 13/03/2012. [en ligne]. Disponible sur : <http://www.slidehare.net/URFISTRennes/quels-outils-pour-queles-recherches-panorama-des-outils-du-web>.
- Véronique Meiguach et Armelle Thomas. *Net recherche 2013. Surveille le web et trouve l'information utile*. Préface d'Olivier Antineu. 5<sup>e</sup> éd. refondue et mise à jour. Paris : ADBS éd., DL 2013. 263 p. (- Information et stratégie).
- Université européenne de Bretagne. *Form@doct Formation à distance en information documentation pour les doctorants*. [en ligne]. Disponible sur : <http://guide-formadoc.ueb.eu/>.

# Bibliographie

## Internet et web

- Chris Anderson et Michale Wolff. - The web is dead. Long live the internet -. *Wired*, 17/08/2010. [en ligne]. Disponible sur : [http://www.wired.com/magazine/2010/08/ff\\_webrip/](http://www.wired.com/magazine/2010/08/ff_webrip/).
- Michael K. Bergman. - White paper : the deep web : surfacing hidden value -. *JEP*, vol. 7, n°1, 08/2001. [en ligne]. Disponible sur : [http://quod.lib.umich.edu/cgi/text/text-ids?c=jep;view=text;main.idno=3336451\\_0007\\_104](http://quod.lib.umich.edu/cgi/text/text-ids?c=jep;view=text;main.idno=3336451_0007_104). [étude BrightPlanet].
- Sylvio Boudreau. - Web 1.0, 2.0 ou 3.0 - De quel type êtes-vous ? -. *ConceptArt multimédia* 2/02/2010. [en ligne]. Disponible sur : <http://franco.ca/internet/index.cfm?cvo=home&id=10711&M=599&item=1500>.
- Olivier Ertzscheid. *Etat des lieux, histoire du web de la bibliothèque. Web 2.0, 3.0*. Présentation, 28 p. 2013. [en ligne]. Disponible sur : [http://fr.sldelivre.net/olivier/histoire\\_weboc-15923635](http://fr.sldelivre.net/olivier/histoire_weboc-15923635).
- . - Mœbius, le web 2.0 et Darwin. (Chapitre 1) -. *Affordance.info*, 27/11/2005. [en ligne]. Disponible sur : [http://affordance.rpnetad.com/mon\\_weblog/2005/11/moebius\\_le\\_web.html](http://affordance.rpnetad.com/mon_weblog/2005/11/moebius_le_web.html).
- Olivier Eratty. - Les enjeux de l'appropriation de l'Internet -. *Opinions libres*, 22-26/12/2011. [en ligne]. Disponible sur : <http://www.ocerintj.net/nordpress/2011/enjeux-appropriation-internet/>.
- Martin. - Les sites Internet les plus visités dans le monde -. *Webzooz*, 19/11/2012. [en ligne]. Disponible sur : <http://www.webzooz.fr/les-sites-internet-les-plus-visites-dans-le-monde/>.
- Fingdom. - Internet 2012 in numbers -. *Royal pingdom*, 16/01/2013. [en ligne]. Disponible sur : <http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/>.
- Alexandre Serres. *Histoire d'internet et du web : d'Arpanet au web sémantique*. Présentation, URFIST de Rennes, 29/03/2013. [en ligne]. Disponible sur : <http://www.ices.univ-rennes2.fr/urfiit/ressources/histoire-dinternet-et-du-web-darpanet-au-web-semantique?destination=ressources>.

# Bibliographie

## Recherche et moteurs de recherche

- Admin. - Top 5 alternative search engines that aren't Google -. *WallStreet.com*. 04/03/2013. [en ligne]. Disponible sur : <http://www.wallstreet.com/top-5-alternative-search-engines-that-arent-google/>.
- Tom Anthony. *The evolution of search*. Présentation, 26/05/2013, 96 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/TomAnthony/the-evolution-of-search>.
- Olivier Andrieu. *Evolution des moteurs de recherche sur le web*. Présentation, URFIST de Bordeaux, 48 p. 05/04/2011. [en ligne]. Disponible sur : <http://webufist.univ-bordeaux.fr/wp-content/uploads/2013/02/andrieu2011.pdf>.
- . - France : 1 visiteur sur 2 vient des moteurs de recherche! -. *Abondance*. 07/05/2012. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20120507-11423-france-1-visiteur-sur-2-vient-des-moteurs-de-recherche.html>.
- Charles Arthur. - Google's Manisa Mayer on the importance of real-time search -. *The Guardian*. 08/07/2009. [en ligne]. Disponible sur : <http://www.guardian.co.uk/technology/2009/jul/08/google-search-manisa-mayer>.
- John Battelle. *La révolution Google. Comment les moteurs de recherche ont réinventé notre économie et notre culture*. Paris : Eyrolles, 2006. VIII-279 p. Traduction de *The search : how Google and its rivals remove the rules of business and transformed our culture*.
- Michael Boehl. - Yandex Just Passed Bing to Become 4th Largest Global Search Engine -. *Search Engine Watch*. 07/02/2013. [en ligne]. Disponible sur : <http://searchenginewatch.com/article/2242374/Yandex-Just-Passed-Bing-to-Become-4th-Largest-Global-Search-Engine>.
- Wendy Boswell. - The best search engines of 2014 -. *Websearch*. 2014. [en ligne]. Disponible sur : <http://websearch.about.com/od/Top-10-Lists/tp/The-Best-Search-Engines-of-2013.htm>.
- Andrei Broder. - A taxonomy of web search -. *SIGIR Forum*. automne 2002, vol. 36, n°2, 8 p. [en ligne]. Disponible sur : <http://www.sigir.org/forum/F2002/broder.pdf>.
- Alex Chris. - Top 10 search engines in the world -. *Reliablesoft.net*. 02/01/2013. [en ligne]. Disponible sur : <http://www.reliablesoft.net/top-10-search-engines-in-the-world/>.
- Antoine Crochet-Damais. - Quant : les secrets techniques du moteur de recherche français -. *Journal du net*. 25/04/2014. [en ligne]. Disponible sur : <http://www.journaldu.net/solutions/cloud-computing/quant-les-couilles-techniques.html>.
- Olivier Ertzscheid. *Evolution des moteurs de recherche*. Présentation, URFIST de Bordeaux, 66 p. 05/04/2011. [en ligne]. Disponible sur : <http://webufist.univ-bordeaux.fr/wp-content/uploads/2013/02/moteur2011.pdf>.
- . - Méthodes, techniques et outils. Qu'y aura-t-il demain sous nos moteurs ? -. *Documentaliste, sciences de l'information* vol. 48, n°3, 10/2011. p. 10-11. Version auteur disponible en ligne sur : [http://archivis.eceid.cnrs.fr/sic\\_00634397.fr](http://archivis.eceid.cnrs.fr/sic_00634397.fr).
- . - Moteurs de recherche : des enjeux d'aujourd'hui aux moteurs de demain -. *Métadonnées : mutations et perspectives*. Séminaire INRIA, 29 septembre-3 octobre 2008, Dijon. Liette Caldern, Bernard Hidoine et Jacques Millet. (coord.) - Paris : ADBS éd., 2008. 322 p. p. 59-59. Version auteur disponible sur : [http://archivis.eceid.cnrs.fr/index.php/hakim?chuzel67cc0977d0c618318vtes\\_das\\_doc%3E00325690%3Eversion1](http://archivis.eceid.cnrs.fr/index.php/hakim?chuzel67cc0977d0c618318vtes_das_doc%3E00325690%3Eversion1).
- Future of search series -. *Mashable*. Dossier. 2011. [en ligne]. Disponible sur : <http://mashable.com/category/future-of-search-series/>.
- Liz Gannes. - How search is evolving—finally!—beyond catman queries -. *AllThingsD*. 14/03/2013. [en ligne]. Disponible sur : <http://allthingsd.com/20130314/how-search-is-evolving-finally-beyond-catman-queries/>.
- Paul Gil. - The best search engines of 2015! -. *Net for beginners*. 2015. [en ligne]. Disponible sur : <http://netforbeginners.about.com/od/internet101/tp/The-Best-Search-Engines-of-2015-12.htm#step-headings>.
- Lars Härd. - Can Google compete with the next generation of search engines ? -. *VentureBeat*. 12/11/2012. [en ligne]. Disponible sur : <http://venturebeat.com/2012/11/06/next-generation-search/>.
- The hidden value of longtail SEO. *Infographic*. 03/2013. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20130503-13376-infographie-la-longue-traine-seo.html>.



# Bibliographie

## Pratiques des internautes

- Olivier Andrieu - Les moteurs de recherche génèrent 51% du trafic (étude) -. *Abondance*. 1<sup>er</sup> 09 2014. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/2014/09/01-14229-les-moteurs-de-recherche-generent-51-du-traffic-quotidien.html>
- Les moteurs de recherche plus fiables que les médias traditionnels (étude) -. *Ibid*. 27/01/2015. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/2015/01/27-14633-les-moteurs-de-recherche-plus-fiables-que-les-medias-traditionnels-etude.html>
- Pierrick Aubert - Les réseaux sociaux : 2<sup>e</sup> de la recherche web -. *ZZNet*. 23/06/2013. [en ligne]. Disponible sur : <http://www.zynet.fr/actualites/les-reseaux-sociaux-2e-de-la-recherche-web-32791699.htm>
- Emmanuel Barthe - Discussion entre documentalistes-formateurs. Les facettes : un moteur de recherche révolutionnaire ? Quel est le meilleur mode de recherche ? Facettes, moteur simple (à la Google) ou recherche avancée ? -. *Précédent.org*. 19/03/2013. [en ligne]. Disponible sur : <http://precident.org/2013/03/19/les-facettes-un-moteur-de-recherche.html>
- Wendy Boswell - Fast facts : the need for speedy information online -. *About.com Web search*. 01/2013. [en ligne]. Disponible sur : <http://websearch.about.com/od/researchtipsandtools/a/Fast-Facts-The-Need-For-Speedy-Information-Online.htm>
- Business insider - Mobile search: how smartphones are disrupting the internet's biggest business -. *Business insider*. 20/04/2013. [en ligne]. Disponible sur : <http://www.businessinsider.com/smartphones-are-disrupting-mobile-search-2013-4>
- Paras Chopra - How location & small screen size impact search behavior on mobile devices -. *Search engine land*. 26/09/2012. [en ligne]. Disponible sur : <http://searchengineland.com/how-location-and-small-screen-size-impact-search-behavior-on-mobile-devices-133551>
- Alex Cocotas et Henry Blodgett - The future of mobile (slide deck) -. *Business insider*. 22/03/2012. [en ligne]. Disponible sur : <http://www.businessinsider.com/the-future-of-mobile-deck-2012-3?page=1>
- Frédéric Dempuré - Nombre de visiteurs : les moteurs de recherche génèrent moins de trafic -. *Les échos*. 27/06/2014. [en ligne]. Disponible sur : <http://business.lesechos.fr/54424760-nombre-de-visiteurs-les-moteurs-de-recherche-generent-moins-de-traffic-101133.php>
- Jerry Ditchler - Building for the next moment -. *Google Inside AdWords*. 05/05/2015. [en ligne]. Disponible sur : <http://adwords.blogspot.fr/2015/05/building-for-next-moment.html>
- GO-Gulf. *How people spend their time online*. Infographic. 02/2012. [en ligne]. Disponible sur : <http://www.go-gulf.com/blog/online-time/>
- Google. *Think mobile. The mobile movement. Understanding smartphone users*. Présentation, 04/2011. 39 p. [en ligne]. Disponible sur : <http://www.thinkwithgoogle.com/research/studies/the-mobile-movement.html>
- Google Consumer barometer. Etude. 2014. [en ligne]. Disponible sur : <http://www.consumerbarometer.com/en>
- Greg - Searcher behavior -. *Search Engine Showdown*. 02/04/2013. [en ligne]. Disponible sur : <http://searchengine showdown.com/2013/04/searcher-behavior/>
- Gord Hotchkiss - Exploring the shift in search behaviors with Microsoft's Jacquelyn Kroes -. *Search engine land*. 15/06/2011. [en ligne]. Disponible sur : <http://searchengineland.com/exploring-the-shift-in-search-behaviors-with-microsofts-jacquelyn-kroes-34750>, suivi de <http://searchengineland.com/exploring-a-new-search-landscape-with-microsofts-jacquelyn-kroes-39164>, 12/08/2011.
- IDC - Consumers increasingly using mobile devices as their default gateway to the internet, IDC says -. *IDC.com Press release*. 29/10/2012. [en ligne]. Disponible sur : <http://www.idc.com/getdoc.jsp?containerID=pr12082756519>
- Simon Khalaf - Apps solidify leadership six years into the mobile revolution -. *Flurry insights*. 1<sup>er</sup> 04/2014. [en ligne]. Disponible sur : <http://flurrymobile.nurkix.com/blog/115191864500-apps-solidify-leadership-six-years-into-the-mobile-VRHrQGYS10K>
- Jessica Lee - 46% of searchers now use mobile exclusively to research [study] -. *Search engine watch*. 01/05/2013. [en ligne]. Disponible sur : <http://searchenginewatch.com/article/2265547-46-of-Searchers-Now-Use-Mobile-Exclusively-to-Research-Study>
- Adam Lella - Number of mobile-only internet users now exceeds desktop-only in the U.S. -. *ComScore*. 28/04/2015. [en ligne]. Disponible sur : <http://www.comscore.com/Insights/Blog/Number-of-Mobile-Only-Internet-Users-Now-Exceeds-Desktop-Only-in-the-U.S.>

# Bibliographie

- Leveling. *Using search engines. Part II: How people search*. Infographic. 2013. [en ligne]. Disponible sur : <http://www.marketingprofi.com/chip/2013/11/692/how-people-search-online-infographic>
- MDGadvertising. *Is personalized search getting too personal?* Infographic. 04/2012. [en ligne]. Disponible sur : <http://www.mxdm.com/2012/04/04/is-personalized-search-getting-too-personal-infographic.php>, d'après l'étude Pew internet. *Search engine use 2012*. 09/03/2012. [http://pewinternet.org/~/media/Files/Reports/2012/FIT\\_Search\\_Engine\\_Use\\_2012.pdf](http://pewinternet.org/~/media/Files/Reports/2012/FIT_Search_Engine_Use_2012.pdf).
- Mushroom networks. *YouTube The 2nd largest search engine* 2013. [infographie]. [en ligne]. Disponible sur : <http://davidfroy.fr/2013/08/infographie-youtube>.
- Nathan Safran. *Psychology of the searcher*. Exute Nile Research. 28/04/2015. [en ligne]. Disponible sur : <http://bluelinersearch.com/psychology-searcher>.
- Fabian Ropan. - On passe plus de temps à surfer depuis un mobile que depuis un ordinateur fixe -. *Le blog du modérateur*. 21/08/2013. [en ligne]. Disponible sur : <http://www.blogdumoderateur.com/internet-mobile-augmentation-duree/>.
- Yang Song et al. - Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance -. *Inproceedings: Microsoft research*. 05/2013. 11 p. [en ligne]. Disponible sur : <http://research.microsoft.com/apps/usbui/default.aspx?id=133343>.
- Greg Sterling - Google: 50% of mobile search is local -. *Screentwerk*. 1<sup>er</sup> 10/2012. [en ligne]. Disponible sur : <http://screentwerk.com/2012/10/01/google-50-of-mobile-search-is-local>.
- Mobile search use stats: big at home, when watching TV, while running errands -. *Search engine land*. 23/08/2011. [en ligne]. Disponible sur : <http://searchengineland.com/higher-use-of-mobile-search-at-home-report-69337>.
- Report: mobile search queries 39 percent of total biz growth modest -. *Ibid*. 26/03/2015. [en ligne]. Disponible sur : <http://searchengineland.com/report-mobile-search-queries-39-percent-of-total-biz-growth-modest-217391>.
- Study suggests search engines not as popular on mobile devices -. *Ibid*. 08/08/2012. [en ligne]. Disponible sur : <http://searchengineland.com/study-suggests-search-engines-not-as-popular-on-mobile-devices-129903>.
- Benjamin Thiers. - Les réseaux sociaux, futurs moteurs de recherche en ligne sans mots-clés -. *Marketing101*. 21/10/2014. [en ligne]. Disponible sur : <http://www.marketing101.fr/webmarketing/reseaux-sociaux-futurs-moteurs-recherche-mots-clés>.
- Yufei Xue et al. - Unfocused search: web searchers behave without clear intent -. *Journal of computational information systems*. 2013, vol. 9 (6). p. 2471-2481. [en ligne]. Disponible sur : [http://www.ijcisi.com/publishedpapers/2013\\_9\\_6\\_2471\\_2481.pdf](http://www.ijcisi.com/publishedpapers/2013_9_6_2471_2481.pdf).

# Bibliographie

## Google

- Tim Adams. - Google and the future of search : Amit Singhal and the Knowledge Graph -. *The Observer*, 19/01/2013. [en ligne]. Disponible sur : <http://www.guardian.co.uk/technology/2013/jan/19/google-search-knowledge-graph-singhal-interview>.
- Laurent Alexandre et David Angwin. *Google démocrate*. Paris : Naïve, 2011. 398 p. (roman)
- Michael Arrington. - Google CEO Eric Schmidt On The Future Of Search: 'Connect It Straight To Your Brain' -. *TechCrunch* 3/09/2009. [en ligne]. Disponible sur : <http://techcrunch.com/2009/09/03/google-ceo-eric-schmidt-on-the-future-of-search-connect-it-straight-to-your-brain/>.
- Olivier Andrieu. - Google arrête l'Authorship ! -. *Abondance*, 29/08/2014. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20140829-14219-google-arrete-authorship.html>.
- . - Google pourrait baser son algorithme sur l'exacitude des faits relatés dans une page web -. *Ibid.* 04/03/2015. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20150304-14796-google-pourrait-baser-son-algorithme-sur-l'exacitude-des-faits-relates-dans-une-page-web.html>.
- . - Matt Cutts confirme l'importance croissante de l'authorship -. *Ibid.* 07/06/2013. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20130607-12697-matt-cutts-confirme-l'importance-croissante-de-l-authorship.html>.
- . - Pour Matt Cutts, les liens vont perdre de l'importance dans les années à venir -. *Ibid.* 06/05/2014. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20140506-13833-matt-cutts-les-liens-vont-perdre-l'importance-les-annes-venir.html>.
- Atlantico. - Pourquoi le règne de Google sur les moteurs de recherche est révolu [Interview de Fabrice Epelboin] -. *Atlantico*, 2/11/2014. [en ligne]. Disponible sur <http://www.atlantico.fr/decryptage/pourquoi-le-regne-google-moteurs-recherche-est-revolu-fabrice-epelboin-1237543.html>.
- Alexandra Bensoussan-Sureau. - Top 5 des projets les plus fous de Google -. *Frenchweb*, 17/06/2013. [en ligne]. Disponible sur : <http://frenchweb.fr/top-5-des-projets-les-plus-fous-de-google-120677>.
- Karen Blakeman. - Google drops translated foreign pages -. *Karen Blakeman's blog*, 17-18/05/2013. [en ligne]. Disponible sur : <http://www.kba.co.uk/wordpress/2013/05/17/google-drops-translated-foreign-pages/>.
- Julie Bort. - Nearly 500 millions searches a day are for things Google has never heard of -. *Business insider*, 13/05/2013. [en ligne]. Disponible sur : <http://www.businessinsider.com/500m-things-google-has-never-heard-of-2013-5>.
- Bianca Bosker. - Google's future: Marissa Mayer explains what the search engine will look like in 20 years -. *The Huffington post*, 28/03/2012. [en ligne]. Disponible sur : [http://www.huffingtonpost.com/2012/03/28/marissa-mayer-google-future\\_n\\_1385549.html](http://www.huffingtonpost.com/2012/03/28/marissa-mayer-google-future_n_1385549.html).
- Flavien Chantrel. - 12 suggestions de Google suggest complètement décalées -. *Le blog du modérateur*, 22/07/2013. [en ligne]. Disponible sur : <http://www.blogdumoderateur.com/meilleures-suggestions-google-suggest/>.
- David Degrelle. - Etude Eye-Tracking : Comment la recherche universelle modifie la lecture sur une page de résultats Google -. *Iere position*, 22/01/2008. [en ligne]. Disponible sur : <http://www.1ere-position.fr/blog-eye-tracking-recherche-universelle-et-tem-de-lecture-sur-une-page-de-resultats-google>.
- Emmanuel Delid. - Le Web 1.2 : Ben Gomes présente le Knowledge Graph de Google -. *L'usine nouvelle*, 04/12/2012. [en ligne]. Disponible sur : <http://www.usinenouvelle.com/article/le-web-1-2-ben-gomes-presente-le-knowledge-graph-de-google-N127220>.
- Olivier Duffez. - Google en chiffres : l'entreprise, ses produits, ses salariés -. *WebRankInfo*, 19/04/2013. <http://www.webrankinfo.com/dossiers/google-chiffres-cles>.
- Béatrice Foenix-Niou. *Google : trucs et astuces pour les professionnels de l'info-doc*. Présentation, 15/01/2013. 35 p. [en ligne]. Disponible sur : <http://fr.kdelshare.net/bfoenix/bfrabd/2013-google-trucs-et-astuces-pour-les-professionnels-de-linfo-doc-1602859/>.
- Google. *The evolution of search*. Vidéo, 6' 21". 21/11/2011. [en ligne]. Disponible sur : <http://www.youtube.com/watch?v=mTBSK7wCnD4>.
- Google. *Google algorithm changes*. Infographie, 2013. [en ligne]. Disponible sur : [http://bushtech.net/article/1333816270\\_Google\\_search\\_engine\\_algorithm\\_change\\_history](http://bushtech.net/article/1333816270_Google_search_engine_algorithm_change_history).

# Bibliographie

- Google I/O 2015. *Keynote*. 28/05/2015. Vidéo. 2:08:27. [en ligne]. Disponible sur : <https://www.youtube.com/watch?v=7V-fGMDjmfE&feature=youtu.be>.
- Google's 200 ranking factors. *The complete list*. 2013. Infographie. [en ligne]. Disponible sur : <http://www.entrepreneur.com/article/226834>.
- Holman W. Jenkins Jr. - Google and the search of the future -. *The Wall Street Journal*. 14/08/2010. [en ligne]. Disponible sur : <http://www.wsj.com/articles/SB10001424052749704901104575433294099527212>.
- Marissa Mayer. - The future of search -. *Google official blog*. 10/09/2008. [en ligne]. Disponible sur : <http://googleblog.blogspot.fr/2008/09/future-of-search.html>.
- Drew Olanoff. - Google opens up about the future of search, serves 100 billion searches every month -. *The next web*. 08/08/2012. [en ligne]. Disponible sur : <http://thenextweb.com/google/2012/08/08/google-opens-up-about-the-future-of-search-serves-100-billion-searches-every-month/>.
- Rédaction de ZDNet. - Position dominante : la Commission européenne dévoile les concessions de Google -. *ZDNet*, 25/04/2013. [en ligne]. Disponible sur : <http://www.zdnet.fr/actualites/position-dominante-la-commission-europeenne-devoile-les-concessions-de-google-3973912.htm>.
- Chuck Rosenber. - Improving photo search: a step across the semantic gap -. *Google research blog*. 12/06/2013. [en ligne]. Disponible sur : <http://googleresearch.blogspot.ca/2013/06/improving-photo-search-step-across.html>.
- Daniel Russell. *Search onward search in the future. A look at where we are and the current vector*. Vidéo. 36'08". ELD <http://www.youtube.com/watch?v=QDBhPX1TFTI>.
- Alexandre Sagakian et Emile Moliner. *Etude - Search of the future - France. Google. Le moteur de recherche idéal vu par les internautes*. Etude Directpanel, 12/2009. 49 p. [en ligne]. Disponible sur : <http://fr.thedirectpanel.com/Parisee-directpanel-etude-google-search-of-the-future-12-2009>.
- Barry Schwartz. - Google's Knowledge Graph has an error 20% of the time -. *Search engine land*. 27/06/2012. [en ligne]. Disponible sur : <http://searchengineland.com/google-knowledge-graph-error-126093>.
- . - Just testing : Google users may see up to a dozen experiments -. *Search engine land*. 05/12/2012. [en ligne]. Disponible sur : <http://searchengineland.com/just-testing-google-searchers-may-see-up-to-a-dozen-experiments-141370>.
- Tom Simonite. - How Google plans to find the unGoogleable -. *MIT technology review*. 27/11/2012. [en ligne]. Disponible sur : <http://www.technologyreview.com/news/507451/how-google-plans-to-find-the-un-googleable>.
- Amit Singhal. - Building the search engine of the future, one baby step at a time -. *Inside search. Google official blog*. 08/08/2012. [en ligne]. Disponible sur : <http://googleblog.blogspot.fr/2012/08/building-search-engine-of-future-one.html>.
- . - Fifteen years on—and we're just getting started -. *Inside search. Google official blog*. 26/09/2013. [en ligne]. Disponible sur : <http://insidesearch.blogspot.fr/2013/09/fifteen-years-onand-were-just-getting.html>. Voir notamment le graphique *Google search timeline*.
- . - Introducing the Knowledge Graph: things, not strings -. *Inside search. Google official blog*. 16/05/2012. [en ligne]. Disponible sur : <http://googleblog.blogspot.fr/2012/05/introducing-knowledge-graph-things-not.html>. traduction/adaptation en français disponible sur : <http://googlefrance.blogspot.fr/2012/12/lancement-en-france-du-knowledge-graph.html>, 5/12/2012.
- . - A multi-screen and conversational search experience -. *Inside search. Google official blog*. 15/05/2013. [en ligne]. Disponible sur : <http://insidesearch.blogspot.fr/2013/05/a-multi-screen-and-conversational.html>.
- . - Technologies behind Google ranking -. *Inside search. Google official blog*. 16/07/2008. [en ligne]. Disponible sur : <http://googleblog.blogspot.fr/2008/07/technologies-behind-google-ranking.html>.
- Greg Sterling. - Google: 44 percent of searches for last-minute holiday gifts will be mobile -. *Search engine land*. 06/09/2011. [en ligne]. Disponible sur : <http://searchengineland.com/google-44-percent-of-searches-for-last-minute-holiday-gifts-will-be-mobile-91763>.
- Tom Wasserman. - How Google could lose its grip on mobile search -. *Mashable*. 22/10/2012. [en ligne]. Disponible sur : <http://mashable.com/2012/10/22/google-mobile-search-advertising/>.
- Tim Worstall. - Have we reached peak Google ? -. *Forbes*. 20/10/2012. [en ligne]. Disponible sur : <http://www.forbes.com/sites/timworstall/2012/10/20/have-we-cracked-peak-google/>.

# Bibliographie

## Recherche personnalisée et sociale

- Olivier Andrieu - Comment obtenir des résultats Google le plus neutre possible ? - *Abondance* 28/01/2012. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20120128-11251-comment-obtenir-des-resultats-google-le-plus-neutre-possible.html>.
- Un moteur de recherche Facebook pourrait capter un quart du marché du search (étude) - *Abondance* 19/09/2012. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20120919-11831-un-moteur-de-recherche-facebook-pourrait-capter-un-quart-du-marche-du-search-étude.html>.
- Carole Barthole - Blekko.com : va-t-il réussir là où tant d'autres ont échoué ? - *Netnouveau*, n°37, 07-08/2010, p. 4-7.
- Mary Ellen Bates - Is Google really filtering my news ? - *Librarian of fortune*, 02/09/2011. [en ligne]. Disponible sur : [http://www.librarianoffortune.com/librarian\\_of\\_fortune/2011/09/is-google-really-filtering-my-news.html](http://www.librarianoffortune.com/librarian_of_fortune/2011/09/is-google-really-filtering-my-news.html).
- Karen Blakeman - Oi, Google ! NO !! - *Karen Blakeman's blog*, 18/10/2012. [en ligne]. Disponible sur : <http://www.rba.co.uk/wordpress/2012/10/18/oi-google-no/>.
- *Personalisation and socialisation of search : embrace the change*. Présentation. *Online information* 2012, 21/11/2012, 28 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/KarenBlakeman/personalisation-and-socialisation-of-search-embrace-the-change>.
- Phil Bradley - Google results - down to 6 on a page and most of those are wrong! - *Phil Bradley's weblog*, 25/03/2013. [en ligne]. Disponible sur : [http://philbradley.typepad.com/phil\\_bradleys\\_weblog/2013/03/google-results-down-to-6-on-a-page-and-most-of-those-are-wrong.html](http://philbradley.typepad.com/phil_bradleys_weblog/2013/03/google-results-down-to-6-on-a-page-and-most-of-those-are-wrong.html).
- David Carle - Comment chercher sur Facebook comme un pro - *DavidCarle*, 11/03/2014. [en ligne]. Disponible sur : <http://www.davidcarle.com/2014/03/comment-chercher-sur-facebook-comme-un-pro/>.
- Erwan Cario - Le rapport risque-bénéfices est très flou pour les internautes. [Interview d'Olivier Ertzscheid] - *Ecrans*, 6/8/2013. [en ligne]. Disponible sur : <http://www.ecrans.fr/Le-rapport-risque-benefices-est-16770.html>.
- Jon Catoe - Why web personalization may be damaging our worldview - *Mashable*, 03/06/2011. [en ligne]. Disponible sur : <http://mashable.com/2011/06/03/filters-clickbait/>.
- Francine Clément - Art.3 : un nouveau venu dans le monde des arts visuels - *Theo Curvus*, 14/10/2012. [en ligne]. Disponible sur : <http://curvus.edu/dossier-articles/articles/18337-art-nouveaux-venus-dans-monde-des-arts-visuels>.
- Colin - Sérendipité : le hasard fait-il encore les choses ? - *L'œil au carré*, 16/03/2015. [en ligne]. Disponible sur : <http://www.oeil-au-carre.fr/le-blog/2015/03/16-serendipite-hasard>.
- Patrick Cuendet - Veille : peut-on se passer des moteurs de recherche ? - *Veille & internet, mode d'emploi*, 08/07/2013. [en ligne]. Disponible sur : <http://patrickcuendet.wordpress.com/2013/07/08/veille-peut-on-se-passer-des-moteurs-de-recherche/>.
- Christophe Deschamps - La recherche d'information et la veille sur le web à l'heure des réseaux sociaux - *L'usage numérique. Séminaire INRIA 27/09-10/10/2010*, Anglet, Paris : ADBS éd., 2010, 204 p. p. 117-144. manuscrit auteur disponible sur : <http://fr.scribd.com/doc/40837799/Veille-et-recherche-d-information-a-l-heure-des-reseaux-sociaux>.
- *La recherche sociale d'information*. Présentation, URFIST de Paris, 06/06/2012, 177 p. Version antérieure disponible sur : <http://fr.scribd.com/doc/53249308/La-recherche-sociale-d-information-Do-not-track>. Série TV 2015, 7 épisodes. [en ligne]. Disponible sur : <http://donottrack-doc.com/fr/>.
- Olivier Ertzscheid - Chercher des textes : ouvrir des mon. La recherche à la hashtag) - *Abondance info*, 20/06/2013. [en ligne]. Disponible sur : [http://abondance.typepad.com/mon\\_weblog/2013/06/chercher-des-textes-ouvrir-des-mon.html](http://abondance.typepad.com/mon_weblog/2013/06/chercher-des-textes-ouvrir-des-mon.html).
- De la conquête du *farweb* à celle du *nearweb* - *Ibid.*, 03/02/2012. [en ligne]. Disponible sur : [http://abondance.typepad.com/mon\\_weblog/2012/02/far-web-neure.html](http://abondance.typepad.com/mon_weblog/2012/02/far-web-neure.html).
- La réponse avant la question. Oule complexe du scribe - *Ibid.*, 19/09/2010. [en ligne]. Disponible sur : [http://abondance.typepad.com/mon\\_weblog/2010/09/la-reponse-avant-la-question.html](http://abondance.typepad.com/mon_weblog/2010/09/la-reponse-avant-la-question.html).
- L'indexation est-elle soluble dans les boutons ? - *Ibid.*, 21/09/2011. [en ligne]. Disponible sur : [http://abondance.typepad.com/mon\\_weblog/2011/09/indexation-est-elle-soluble-dans-les-boutons.html](http://abondance.typepad.com/mon_weblog/2011/09/indexation-est-elle-soluble-dans-les-boutons.html).
- Ingénieries de la sérendipité - *Ibid.*, 03/02/2010. [en ligne]. Disponible sur : [http://abondance.typepad.com/mon\\_weblog/2010/02/ingenieries-de-la-serendipite.html](http://abondance.typepad.com/mon_weblog/2010/02/ingenieries-de-la-serendipite.html).

# Bibliographie

- . (My) state of search 2012 - . *Ibid.* 27/11/2012. [en ligne]. Disponible sur : [http://affordance.nyepad.com/mon\\_weblog/2012/11/my-state-of-search-2012.html](http://affordance.nyepad.com/mon_weblog/2012/11/my-state-of-search-2012.html).
- Sophian Fazen. - Le web prend le pli durecommande - . *Ecrans*. 24/04/2013. [en ligne]. Disponible sur : <http://www.ecrans.fr/Le-Web-prend-le-pli-durecommande-16283.html>.
- Tony Faragasso. *De la gestion de signets au social bookmarking. Delicious, Diigo, Zotero et quelques autres*. Paris : ADBS, 2011. 62 p.
- Benoit Fechner. - Comment empêcher Google de vous pister en ligne ? - . *France TV Info*. 12/05/2013. [en ligne]. Disponible sur : <http://www.francetvinfo.fr/comment-empêcher-google-de-vous-pister-en-ligne-381305.html>.
- Filter bubble - . *Wikipédia*. [en ligne]. Disponible sur : [http://en.wikipedia.org/wiki/Filter\\_bubble](http://en.wikipedia.org/wiki/Filter_bubble).
- Béatrice Foenix-Nou. - Google : résultats personnalisés et doodle rien que pour vous... - . *Le blog de recherche-evenillee.com* 18/03/2014. [en ligne]. Disponible sur : <http://blog.recherche-evenillee.com/2014/03/google-resultats-personnalises-et.html>.
- Sarah Frier. - Twitter reaches deal to show tweets in Google search results - . *Bloomberg Business*. 05/02/2015. [en ligne]. Disponible sur : <http://www.bloomberg.com/news/articles/2015-02-05/twitter-reach-to-reach-deal-for-tweets-in-google-search-results>.
- Google, confessionnal du XXI<sup>e</sup> siècle - . *Geeko - Le Soir*. 16/11/2013. [en ligne]. Disponible sur : <http://geeko.lesoir.be/2013/11/16/google-confessionnal-du-xxi-siecle>.
- Magali Guibert. - Trucs et astuces pour optimiser sa recherche dans LinkedIn - . *Netnoues* n°101, 11-12/2012. p. 10-13.
- Hubert Guillaud. - Du graphe social au graphe de l'humeur - . *InternetActu* 18/09/2013. [en ligne]. Disponible sur : <http://www.internetactu.net/2013/09/18/du-graphe-social-au-graphe-de-l-humeur/>.
- Aniko Hannak et al. - Measuring personalization of web search - . *Proceedings of the 22nd International World Wide Web Conference (WWW'13)*, Rio de Janeiro, Brazil, May 2013. 11 p. [en ligne]. Disponible sur : <http://personalization.csi.cmu.edu/paper.pdf>.
- Jérôme G. - Alternative Google : DuckDuckGo flambe - . *Génération NT* 19/06/2013. [en ligne]. Disponible sur : <http://www.generation-nt.com/google-duckduckgo-moteur-recherche-prim-is-quick-image-actualite-1747742.html>.
- Arif Jdey. - Le nouveau web : un web d'humeurs et d'émotions - . *Demain la veille*. 26/08/2013. [en ligne]. Disponible sur : <http://www.demain-la-veille.fr/2013/08/26/le-nouveau-web-un-web-d-humeurs-et-d-emoions/>.
- Clémence Joist. - La carte des réseaux sociaux - . *Archimag* n°265, 06/2013. p. 20.
- Natasha Lomas. - Funaref's predictive discovery engine gets new interface, expands from iPad to iPhone - . *TechCrunch*. 05/08/2013. [en ligne]. Disponible sur : <http://techcrunch.com/2013/08/05/funaref-iphone-app/>.
- Madame Machine, pouvez-vous me conseiller un bon livre ? : les nouveaux outils Web de recommandation de lectures... - . *ABF* 27/06/2011. [en ligne]. Disponible sur : <http://www.abf.asso.fr/3135393/ABF-madame-machine-pouvez-vous-me-conseiller-un-bon-livre-les-nouveaux-outils-web-de-recommandation-de-lectures?pe=28p20>.
- Mary Madden et Lee Rainie. - Americans' attitudes about privacy, security and surveillance - . *Few Research Center*. 20/05/2015. [en ligne]. Disponible sur : <http://www.fewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance>.
- Christian Madjberg et Morgan Ramsey-Elliott. - How to fix streaming music services - . *FastCompany*. 1<sup>er</sup>10/2013. [en ligne]. Disponible sur : <http://www.fastcodesign.com/3018862/how-to-fix-streaming-music-services>.
- Farhad Marjoo. - Google, ne t'arrête pas de collecter mes données ! - . *Slate.fr*. 23/04/2011. [en ligne]. Disponible sur : <http://www.slate.fr/story/36851/vie-privee-google-unle>.
- Véronique Mesguich. - Vers une recherche sociale ? - . *Netnoues*, n°92, 05-06/2011. p. 1-5.
- Les moteurs de recommandation*. Journée d'étude organisée par le laboratoire DICEN, 11/06/2012. Supports d'interventions disponibles en ligne sur : <http://www.dicen-ijff.org/evenement/journee-etude-moteurs-recommandation>.
- Francesca Musiani. - Bienvenue sur votre Amazon - : les systèmes de recommandation d'ouvrages - . *Lab'iHadopi* 28/03/2012. [en ligne]. Disponible sur : <http://labi.hadopi.fr/actualites/bienvenue-sur-votre-amazon-les-systemes-de-recommandation-d-ouvrages>.
- Antoine Oury. - Facebook : Tu lis Philip Ken ? Essaye Hitler, ou la logique de l'algorithme - . *Actualité* 26/08/2013. [en ligne]. Disponible sur : <http://www.actualite.com/societe/facebook-tu-lis-philip-ken-essaye-hitler-ou-la-logique-de-l-algorithme-44619.htm>.

# Bibliographie

- Alexander Pak et Patrick Paroubek. - Twitter as a corpus for sentiment analysis and opinion mining -. In Nicoletta Calzolari et al. (éd.). *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta. European Language Resources Association 2010* [en ligne]. Disponible sur : [http://www.lrec-conf.org/proceedings/lrec2010/pdf/385\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf).
- Tommy Pouilly. - Au bout de la personnalisation du web : paradis ou enfer? (1/3) -. *RSLN*. 18/08/2014. [en ligne]. Disponible sur : <http://www.rslnmg.fr/post/2014/08/18/Au-bout-de-la-personnalisation-du-web-paradis-ou-enfer-1-3>.
- . - Au bout de la personnalisation du web : paradis ou enfer? (3/3) -. *Ibid*. 21/08/2014. [en ligne]. Disponible sur : <http://www.rslnmg.fr/post/2014/08/21/Au-bout-de-la-personnalisation-du-web-le-miroir-aux-aloettes-3-3>.
- Lee Rainie et al. - Anonymity, privacy, and security online -. *PenResearchCenter*. 05/09/2013. [en ligne]. Disponible sur : <http://www.pensrcmet.org/2013/09/05/anonymity-privacy-and-security-online>.
- Claire Richard. - Qui de moi ou de l'algorithme filtre les posts de mes amis Facebook? -. *Rue89*. 08/05/2015. [en ligne]. Disponible sur : <http://rue89.nouvelobs.com/2015/05/08/lalgorithme-filtre-les-posts-facebook-amis-259031>.
- Greg Sterling. - ChaCha defeat Google, Bing and Siri in - Anvenquity - study -. *Search engine land*. 19/02/2013. [en ligne]. Disponible sur : <http://searchengineland.com/chacha-defeat-google-bing-and-siri-in-anvenquity-study-149044>.
- David Sullivan. - How the myth of Google censorship was busted by the EU & Canada -. *Marketing land*. 12/06/2015. [en ligne]. Disponible sur : <http://marketingland.com/google-censorship-myth-busted-132106>.
- Carole Tisserant-Barthole. - Quelques moteurs de questions réponses à la loupe -. *Netnouvelles*, n°103, 03-04/2013, p. 1-4.
- . - Veille et recherche scientifique 2.0 : au-delà des ressources classiques -. *Bases*, n°206, 09/2012, p. 1-6.
- Where is your next customer? Stop playing hide and seek. *The value of accurate local search* Etude Neustar, 15 miles. 6th Annual comScore Local search usage study. Infographic. 2013. [en ligne]. Disponible sur : <http://www.infographiconly.com/2013-local-search-study>.
- Louis Wiart. - Lecteurs, quels sont vos réseaux? -. *INAGlobal*. 13/01/2014. [en ligne]. Disponible sur : <http://www.inaglobal.fr/edition/article/lecteurs-quels-sont-vos-reseaux>.

# Bibliographie

## Recherche temps réel

- Wendy Boswell - What is realtime search? -. *About.com* [en ligne]. Disponible sur : <http://websearch.about.com/od/socialnetworking/p/What-is-Real-Time-Search.htm>.
- Christophe Deschamps - Comparatif de 14 moteurs de recherche 'temps réel' -. *Outils frosts*, 23/06/2009. [en ligne]. Disponible sur : <http://www.outilsfrosts.net/news/comparatif-de-14-moteurs-de-recherche-temps-reel/>. 1 nombreux services morts.
- Béatrice Foenix-Niou - Méthodologie : de l'intérêt de Twitter pour des recherches en temps réel; retour sur l'exemple Diigo.com -. *Le blog de recherche-veillee.com*, 26/10/2012. [en ligne]. Disponible sur : <http://blog.recherche-veillee.com/2012/10/de-linteret-de-twitter-pour-des.html>.
- Alain Garnier - Pourquoi Twitter a gagné la bataille du - social temps réel - grâce au Search -. *Net sources*, n°91, 03-04/2011. Disponible sur : <http://www.bases-net-sources.com/2011/03/pourquoi-twitter-gagne-la-bataille-du.html>.
- François Magnan - Recherche et veille sur les réseaux sociaux. 1 / Icerocket, simple et pratique -. *François Magnan - formateur consultant*, 01/01/2013. [en ligne]. Disponible sur : <http://www.francoismagnan.info/?p=46>.
- . - Recherche et veille sur les réseaux sociaux 2 Socialmention sonde les réseaux sociaux -. *Ibid.*, 01/01/2013. [en ligne]. Disponible sur : <http://www.francoismagnan.info/?p=54>.
- . - Recherche et veille sur les réseaux sociaux 3 Topsy : trouver du contenu recommandé et des experts sur le réseau Twitter -. *Ibid.*, 01/01/2013. [en ligne]. Disponible sur : <http://www.francoismagnan.info/?p=61>.
- Marie-Laure Malinque et Alexandre Serres. *Outils de recherche et de veille sur l'actualité (news, médias sociaux...)*. URFIST de Rennes. Présentation, 66 p., 11/12/2012. [en ligne]. Disponible sur : <http://fr.slideshare.net/UrfistRennes/taag-outils-recherche-actualite/20121211>.
- Véronique Metzgaich. *Veille et temps réel*. 5 à 7 de l'ADBS. 12/10/2012. Vidéos et présentation. [en ligne]. Disponible sur : <http://www.adbs.fr/veille-et-temps-reel-1-de-time-is-now-92903.htm?RFP=1266334569418>.
- Jason Petkov - Yahoo news stream will now include tweets -. *W3Reports*, 16/05/2013. [en ligne]. Disponible sur : <http://www.w3reports.com/2013/05/16/yahoo-news-stream-will-now-include-tweets/>.

# Bibliographie

## Recherche inversée

- Ajinkye. - Reverse image search engines, apps and its uses -. *Beebon* 14/02/2015. [en ligne]. Disponible sur : <http://beebon.com/2015/02/reverse-image-search-engines-apps-and-its-uses/>.
- Tayyab Babar. - What songs is this ? 10 mobile and web apps to identify songs -. *Beebon* 24/01/2015. [en ligne]. Disponible sur : <http://beebon.com/2015/01/what-songs-is-this-apps/>.
- Ballajack. - Déterminer l'âge d'une personne sur une photo, How-Old -. *Les infos de Ballajack* 04/05/2015. [en ligne]. Disponible sur : <http://www.ballajack.com/determiner-age-personne-photo>.
- Michael Boland. - The future of visual search : UPCs, QR codes, and augmented reality -. *Search Engine Watch* 29/10/2010. [en ligne]. Disponible sur : <http://searchenginewatch.com/article/2049179/The-Future-of-Visual-Search-UPCs-QR-Codes-and-Augmented-Reality>.
- . - Mobile search : time for a hearing and vision checkup -. *Search Engine Watch* 14/05/2010. [en ligne]. Disponible sur : <http://searchenginewatch.com/article/2064288/Mobile-Search-Time-for-a-Hearing-and-Vision-Checkup>.
- . - When and how will visual search arrive ? -. *Keygroup blog* 07/01/2010. [en ligne]. Disponible sur : <http://blog.keygroup.com/index.php/2010/01/07/when-and-how-will-visual-search-arrive/>.
- M.C. - Google veut compter les calories sur vos photos de nourriture -. *20minutes* 04/06/2015. [en ligne]. Disponible sur : <http://www.20minutes.fr/high-tech/1623347-20150604-google-veut-compter-calories-photos-nourriture>.
- Thomas Chaintreuil. *Réalités augmentées en bibliothèque : passerelles entre espaces physiques et numériques*. Présentation. 07/03/2012. 89 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/Fakom/rkr-augmentees-en-bibliotheque>.
- Loïc Chauveau. - Pi@Net : l'application smartphone qui identifie les végétaux -. *Sciences et avenir* 07/03/2013. [en ligne]. Disponible sur : [http://www.sciencesetavenir.fr/science/evolution/20130305\\_0850735/pi-net-l-application-smartphone-qui-identifie-les-vegetaux.html](http://www.sciencesetavenir.fr/science/evolution/20130305_0850735/pi-net-l-application-smartphone-qui-identifie-les-vegetaux.html).
- Antoine Chéron. - Reconnaissance vidéo : Google pousse encore plus loin son intrusion dans la vie privée -. *Adfameo* 13/09/2012. [en ligne]. Disponible sur : <http://www.adfameo.fr/devoilage-reconnaissance-video-google-pousse-encore-plus-loin-intrusion-dans-vie-privee-antoine-cheron-480800.html>.
- Jérôme Deiss. *Comment vérifier l'information à l'ère des médias sociaux*. 05/2013. Présentation. 25 p. [en ligne]. Disponible sur : [http://fr.slideshare.net/Veille\\_Digitale/verifier-l-information-les-medias-sociaux](http://fr.slideshare.net/Veille_Digitale/verifier-l-information-les-medias-sociaux).
- Olivier Ertzscheid. - D'après - Elle -, de Spike Jonze -. *Affordance info* 27/03/2014. [en ligne]. Disponible sur : [http://affordance.nypepad.com/mon\\_weblog/2014/03/apres-elle-ber-spike-jonze.html](http://affordance.nypepad.com/mon_weblog/2014/03/apres-elle-ber-spike-jonze.html).
- . - Il ne lui manquait plus que la parole -. *Ibid.* 23/05/2013. [en ligne]. Disponible sur : [http://affordance.nypepad.com/mon\\_weblog/2013/05/il-ne-lui-manquait-plus-que-la-parole.html](http://affordance.nypepad.com/mon_weblog/2013/05/il-ne-lui-manquait-plus-que-la-parole.html).
- . - Reconnaissance faciale ... de doi -. *Ibid.* 11/05/2015. [en ligne]. Disponible sur : [http://affordance.nypepad.com/mon\\_weblog/2015/05/reconnaissance-faciale-de-doi.html](http://affordance.nypepad.com/mon_weblog/2015/05/reconnaissance-faciale-de-doi.html).
- Fig. - Comment Google traduit les images en mots -. *Technology Review* -. *A lire ailleurs*. 04/12/2014. [en ligne]. Disponible sur : <http://akre@leuri.tumblr.com/post/104317135157-comment-google-traduit-les-images-en-mots>.
- Kate Freeman. - The future of video search -. *Mashable* 31/01/2013. [en ligne]. Disponible sur : <http://mashable.com/2013/01/31/video-search-future/>.
- Jérôme G. - Freebox Révolution : mise à jour et reconnaissance musicale -. *GNT - Génération nouvelles technologies*. 15/01/2014. [en ligne]. Disponible sur <http://www.generation-it.com/freebox-revolution-free-music-et-reconnaissance-music-jour-firmer-d-une-actualite-1339442.html>.
- Hubert Guillaud. - Le prolétariat du web accède à la conscience de classe -. *InternetActu* 08/12/2014. [en ligne]. Disponible sur : <http://www.internetactu.net/2014/12/08/le-proletariat-du-web-accede-a-la-conscience-de-classe>.
- Lauren Hochenson. - With audiorecognition, Facebook can help automatically check in to your favorite music, movies and TV -. *Cignom* 21/05/2014. [en ligne]. Disponible sur : <http://cignom.com/2014/05/21/with-audio-recognition-facebook-can-automatically-check-in-to-your-favorite-music-movies-and-tv/>.
- Yves Le Corre. - Beacon, reconnaissance d'image et vocale, RFID... Les nouvelles techno 2014 qui vont compter en 2015 -. *LSA*. 30/12/2014. [en ligne]. Disponible sur : <http://www.lsa-conto.fr/beacon-reconnaissance-d-image-et-vocale-rid-les-nouvelles-technos-2014-qui-vont-compter-en-2015-196394>.
- Joel Lowenstein. - Wolfram has created a website that will identify any image you throw at it -. *The verge* 13/05/2015. [en ligne]. Disponible sur : <http://www.theverge.com/2015/5/13/2659351/wolfram-image-identification-site-creates-benchmarks>.
- Pierre Nobis. - Chercher des images par l'image -. *Documentation 13hp*. Rouen 11/2008-11/2012. [en ligne]. Disponible sur : [http://documentation.13hp.rouen.fr/spip.php?article2028&rubrique\\_2028\\_15](http://documentation.13hp.rouen.fr/spip.php?article2028&rubrique_2028_15) ! quelques liens morts.

## Bibliographie

- James O'Brien. - How AR [augmented reality] will change the way we search -. *Mashable* 13/03/2013. [en ligne]. Disponible sur : <http://mashable.com/2013/03/13/visual-search/>.
- Patrick Peccatte. - A la recherche d'un singe photographe -. [en ligne]. Disponible sur : <http://culturevisuelle.org/deauv795>. Egalement paru dans *Nouveaux*, n°93, 07-08 2011, p. 10-11.
- . - Du bon usage des photos de stock -. *Culture visuelle*, 26/06/2011. [en ligne]. Disponible sur : <http://culturevisuelle.org/deauv737>. Egalement paru dans *Nouveaux*, n°92, 05-06 2011, p. 11-13.
- Roberto Pieraccini. - The future of voice search -. *Mashable* 13/06/2011. [en ligne]. Disponible sur : <http://mashable.com/2011/06/13/the-future-of-voice-search/>.
- Samy. - Google rachète une start-up canadienne pour améliorer sa reconnaissance vocale -. *FrAndroid* 18/03/2013. [en ligne]. Disponible sur : <http://www.frandroid.com/actualites-generales/133596-google-rachete-start-up-canadienne-reconnaissance-vocale/>.
- Laura Shin. - Google brain simulator teaches itself to recognize cats -. *Smart planet*, 27/06/2012. [en ligne]. Disponible sur : <http://www.smartplanet.com/blog/science/scope/google-brain-simulator-teaches-itself-to-recognize-cats/12999>.
- Danny Sullivan. - iOU users to get enhanced Google search app ahead of most android users -. *Search Engine Land* 08/08/2012. [en ligne]. Disponible sur : <http://searchengineland.com/ios-users-to-get-enhanced-google-voice-search-129995>.
- Aurélien Vadhon. - Repérer les retombées presse d'une exposition grâce aux images -. *Nouveaux*, n°97, 03-04/2012, p. 4-5. Disponible en ligne sur <http://fr.calameo.com/read/001251350a96540ee3589?authid=4A6OadT1Vkw>.
- Oriol Vinayal et al. - A picture is worth a thousand (coherent) words: building a natural description of images -. *Google research blog*, 17/11/2014. [en ligne]. Disponible sur : <http://googleresearch.blogspot.fr/2014/11/a-picture-is-worth-thousand-coherent.html>.

# Bibliographie

## Recherche sémantique

- David Amerland. - Five steps to a fully semantic web -. *David Amerland* 08/2014. [en ligne]. Disponible sur : <http://davidamerland.com/2014/05/20/five-steps-to-a-fully-semantic-web.html>
- Olivier Andrieu. - Knowledge Vault, le futur de la recherche sur Google ? -. *Abondance* 02/09/2014. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/2014/09/02/knowledge-vault-futur-recherche-google.html>
- Hervé Basset et François Bouais. - The semantic search in life sciences -. In Hervé Basset, David Stuart et Denise Silber. *From Science 2.0 to Pharma 3.0. Semantic search and social media in the pharmaceutical industry and STM publishing*. Oxford : Chandos Publishing, 2012. 276 p. p. 151-168.
- Emmanuelle Bernès. *Les bibliothèques sur le web : Enjeux, évolutions techniques : Web sémantique et Web de données*. Journée *Les catalogues au défi du Web - Saison 2*, BNF/CNFTT, 26/11/2014. Vidéo, 42'27. [en ligne]. Disponible sur : <http://video.orfu.fr/conferences/1-les-catalogues-au-defi-du-web-les-bibliothèques-sur-le-web>
- Tim Berners-Lee. - Giant global graph -. *Timbl's blog* 21/11/2007. [en ligne]. Disponible sur : <http://dig.csail.mit.edu/bradfordmb/node/215>
- . - Semantic web in the news -. *Timbl's blog* 27/03/2008. [en ligne]. Disponible sur : <http://dig.csail.mit.edu/bradfordmb/blog/4>
- . James Hendler et Ora Lassila. - The semantic web -. *Scientific American* com 17/05/2001. [en ligne]. Disponible sur : <http://www.cs.sand.edu/~colbeck/LBSC690/SemanticWeb.html>
- Karen Blakeman. - Google adds in-depth articles to results -. *Karen Blakeman's blog* 15/08/2013. [en ligne]. Disponible sur : <http://www.dba.co.uk/nordnet/2013/08/15/google-adds-in-depth-articles-to-results>
- Phil Bradley. - Why the new Facebook Graph search is important for librarians -. *Phil Bradley's weblog* [en ligne]. Disponible sur : [http://philbradley.vpnet.com/phil\\_bradley\\_weblog/2013/01/why-the-new-facebook-graph-search-is-important-for-librarians.html](http://philbradley.vpnet.com/phil_bradley_weblog/2013/01/why-the-new-facebook-graph-search-is-important-for-librarians.html)
- Paul Bruemmer. - Future SEO: Understanding Entity Search -. *Search engine land* 7/10/2013. [en ligne]. Disponible sur : <http://searchengine.com/future-seo-understanding-entity-search-172927>
- Fred Cavazza. - Facebook lance son moteur de recherche mais manque d'inspiration -. *Fred Cavazza.net* 16/01/2013. [en ligne]. Disponible sur : <http://www.fredcavazza.net/2013/01/16/facebook-lance-son-moteur-de-recherche-mais-manque-d-inspiration>
- . - La recherche passe à l'ère sémantique (et sociale) (et pas visuelle) -. *Fred Cavazza.net* 23/05/2012. [en ligne]. Disponible sur : <http://www.fredcavazza.net/2012/05/23/la-recherche-passe-a-ere-semantique-et-sociale-et-pas-visuelle>
- Emmanuelle Delhol. - Comment google transforme le web en or -. *L'usine nouvelle* n° 3342. 29/08/2013. [en ligne]. Disponible sur : <http://www.usinenouvelle.com/article/comment-google-transforme-le-web-en-or-203356>
- Peter Dookhill. - The next generation of search technology -. *TechLife* 09/05/2013. [en ligne]. Disponible sur : <http://www.techlife.net/computing/news/2013/5/the-next-generation-of-search-technology/>
- Olivier Ertzschick. - Sème antique : la guerre du sens est déclarée -. *Affordance info* 3/06/2012. [en ligne]. Disponible sur : [http://affordance.ruptpad.com/moon\\_weblog/2012/06/la-guerre-du-sens-est-declaree.html](http://affordance.ruptpad.com/moon_weblog/2012/06/la-guerre-du-sens-est-declaree.html)
- Thomas Francart. - Linked Open Data Cloud: nouvelle version -. *Spama* 04/09/2014. [en ligne]. Disponible sur : <http://blog.spama.fr/2014/09/04/linked-open-data-cloud-nouvelle-version>
- Paul Ford. *August 2009: How Google beat Amazon and eBay to the semantic web* 26/07/2002. [en ligne]. Disponible sur : [http://www.fram.com/google\\_mike\\_all.html](http://www.fram.com/google_mike_all.html)
- Sean Gallagher. - How Google and Microsoft taught search to understand the web -. *Art technica* 07/06/2012. [en ligne]. Disponible sur : <http://arttechnica.com/information-technology/2012/06/inside-the-architecture-of-google-knowledge-graph-and-microsofts-satori-2>
- Fabien Gandon. *Web sémantique et web social*. Présentation, 2009. 231 p. [en ligne]. Disponible sur : [http://fr.slideshare.net/fabien\\_gandon/web-semantique-et-web-social-1700977](http://fr.slideshare.net/fabien_gandon/web-semantique-et-web-social-1700977)
- Serge Garlati. *Les moteurs de recherche sémantiques*. Présentation, 2014. 47 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/garlati/le-web-semantique-20132014>
- Sylvain Gateau. *Les moteurs de recherche sémantiques. Contextualisation, réinvention et perspectives*. Présentation, 11/2012. 11 p. [en ligne]. Disponible sur : [http://fr.slideshare.net/Cerise\\_sur\\_Le\\_Gateau/les-moteurs-de-recherche-semantique](http://fr.slideshare.net/Cerise_sur_Le_Gateau/les-moteurs-de-recherche-semantique)
- Amy Gersenhues. - What happened to dinosaurs? Google's direct answer gives non-scientific theory from religious site -. *Search engine land* 10/06/2015. [en ligne]. Disponible sur : <http://searchengine.com/what-happened-to-dinosaurs-google-direct-answer-gives-non-scientific-theory-from-religious-site-222934>
- Rob Gonzalez. - Semantic search and the semantic web -. *Cambridge semantic*. [en ligne]. Disponible sur : <http://www.cambridgesemantic.com/semantic-university/semantic-search-and-the-semantic-web>
- Greg. - Cracking the search category problem -. *Blekkio blog* 14/01/2013. [en ligne]. Disponible sur : <http://blog.blekko.com/2013/01/14/cracking-the-search-category-problem>
- Anthony Ha. - Knigine aims to build a natural language-driven app that can answer any question -. *TechCrunch* 30/11/2012. <http://techcrunch.com/2012/11/30/knigine/>
- Gord Hotchkiss. - The language problem: jaguars and the Turing test -. *Search engine land* 9/09/2011. [en ligne]. Disponible sur : <http://searchengine.com/the-language-problem-jaguars-the-turing-test-222965>



# Bibliographie

## Moteurs de réponse, recherche prédictive et recherche conversationnelle

- Olivier Andrieu - Du moteur de recherche aux moteurs de réponses et de recommandations -. *Abondance*. 02/05/2013. [en ligne]. Disponible sur : <http://www.abondance.com/actualites/20130502-18520-du-moteur-de-recherche-aux-moteurs-de-reponses-et-de-recommandations.html>.
- John Battelle - The database of intentions is far larger than I thought -. *John's Battelle searchblog* 05/03/2010. [en ligne]. Disponible sur : [http://battellemedia.com/archive/2010/03/the\\_database\\_of\\_intentions\\_is\\_far\\_larger](http://battellemedia.com/archive/2010/03/the_database_of_intentions_is_far_larger).
- CNETTV. *Google Now vs Siri: the results speak for themselves*. 16/05/2013. [en ligne]. Disponible sur : <http://www.youtube.com/watch?v=e41v10FTTGo>.
- Coreight - Tous les trucs incroyables que l'on peut faire avec la recherche vocale de Google -. *Coreight.com* 11/07/2013. [en ligne]. Disponible sur : <http://coreight.com/contests/fonctions-incroyables-recherche-vocale-google>.
- Peter Dackrill - The next generation of search technology -. *Teclife* 09/05/2013. [en ligne]. Disponible sur : <http://www.teclife.net/computing/nem/2013/5/the-next-generation-of-search-technology>.
- Yves Eudes - Les prodiges de Sophie -. *Le Monde Technologies*. 16/03/2012. [en ligne]. Disponible sur : [http://www.lemonde.fr/technologies/article/2012/03/16/les-prodige-de-sophie\\_1666947\\_431965.html#sthash=AL3282825](http://www.lemonde.fr/technologies/article/2012/03/16/les-prodige-de-sophie_1666947_431965.html#sthash=AL3282825).
- Olivier Ertzscheid - Moteurs de prédiction et psychosociologie dans-centrique -. *Affordance info* 16/04/2013. [en ligne]. Disponible sur : [http://affordance.rpspad.com/mon\\_weblog/2013/04/prediction-psychosociologie-dans-centrique.html](http://affordance.rpspad.com/mon_weblog/2013/04/prediction-psychosociologie-dans-centrique.html).
- Le web : entre stratégies prédictives et technologies injonctives. Présentation, 53 p. 2013. [en ligne]. Disponible sur : [http://affordance.rpspad.com/mon\\_weblog/2013/04/neutraites-de-internet.html](http://affordance.rpspad.com/mon_weblog/2013/04/neutraites-de-internet.html).
- Twitter : l'oiseau-lyre du bruit du monde -. Gabriel Gallezot et Nicolas Pellissier (dir.). *Twitter, un monde en tout petit* Paris : l'Harmattan, 2013. 260 p. p. 115-124. Texte disponible sur : [http://affordance.rpspad.com/mon\\_weblog/2013/04/twitter-oiseau-lyre-bruit-monde.html](http://affordance.rpspad.com/mon_weblog/2013/04/twitter-oiseau-lyre-bruit-monde.html).
- Paul Ford. *Amazon and Ebay to the semantic web*. 26/07/2002. [en ligne]. Disponible sur : [http://www.fraai.com/google\\_makes\\_all.html](http://www.fraai.com/google_makes_all.html).
- Aaron Friedman - The future of search engines is context -. *Search engine land*. 06/04/2015. [en ligne]. Disponible sur : <http://searchengineland.com/future-search-engines-context-217550>.
- Sean Gallagher - How Google and Microsoft taught search to understand the web -. *Arx technica* 07/06/2012. [en ligne]. Disponible sur : <http://arxtechnica.com/information-technology/2012/06/inside-the-architecture-of-google-knowledge-graph-and-microsoft-lumina-2/>.
- Fabien Gandon. *Web sémantique et web social*. Présentation, 2009. 231 p. [en ligne]. Disponible sur : [http://fr.slideshare.net/fabien\\_gandon/web-semantique-et-web-social-1700977](http://fr.slideshare.net/fabien_gandon/web-semantique-et-web-social-1700977).
- Sarah Kissler - Forget Google Instant: in the future search engines will read your habits -. *Mashable*. 31/03/2012. [en ligne]. Disponible sur : <http://mashable.com/2012/03/31/wolfram-alpha-future-search/>.
- Jessica Lee - OK Google: the end of search as we know it -. *Search engine watch* 16/05/2013. [en ligne]. Disponible sur : <http://searchenginewatch.com/article/2268726/OR-Google-The-End-of-Search-as-We-Know-It>.
- Guillaume Mian - Un algo prédictif financé par le ministre de la Défense américain sur le Web -. *Sixtrophie*. 26/05/2013. [en ligne]. Disponible sur : <http://sixtrophie.fr/actualites/2013/05/26/algorithmes-predictifs-ministere-de-la-defense>.
- Farhad Manjoo - Where no search engine has gone before -. *Slate* 11/04/2013. [en ligne]. Disponible sur : [http://www.slate.com/articles/technology/2013/04/google\\_has\\_a\\_serious\\_search\\_objection\\_it\\_wants\\_to\\_build\\_the\\_star\\_wreck\\_computer.html](http://www.slate.com/articles/technology/2013/04/google_has_a_serious_search_objection_it_wants_to_build_the_star_wreck_computer.html).
- Matt McGee - Google's Schmidt - next great stage - of search is autonomous, personal -. *Search engine land* 07/09/2010. [en ligne]. Disponible sur : <http://searchengineland.com/schmidt-great-stage-search-is-autonomous-personal-30014>.
- Christopher Mims - This is what comes after search -. *Quartz* 04/05/2014. [en ligne]. Disponible sur : <http://qz.com/205589/context-this-is-what-comes-after-search/>.
- Brian Proffitt - Forget searching for content - content is about to start searching for you -. *Read write* 25/04/2013. [en ligne]. Disponible sur : <http://readwrite.com/2013/04/25/forget-searching-for-content-soon-content-will-be-searching-for-you/#h1m9c91vedL769H0Y>.
- How the internet of things will revolutionize search -. *Read write* 26/04/2013. [en ligne]. Disponible sur : <http://readwrite.com/2013/04/26/how-the-internet-of-things-will-revolutionize-search/#h1m9c91vedL769H0Y>.
- Next-generation search: software bots will anticipate your needs -. *Read write* 08/05/2013. [en ligne]. Disponible sur : <http://readwrite.com/2013/05/08/future-of-search-software-bots-anticipate-your-needs/#h1m9c91vedL769H0Y>.
- Leo Mirani Quartz - The search engine of the future will be all-seeing, all-knowing 'watcher in the sky' -. *Neogov*. 21/01/2015. [en ligne]. Disponible sur : <http://www.neogov.com/big-data-2015-01-search-engine-future-will-be-all-seeing-all-knowing-watcher-the-103410>.

## Bibliographie

- Social Media Club. - Réseaux sociaux : de la veille en temps-réel à l'analyse prédictive -. *ZDNet.fr*: 04/04/2013. [en ligne]. Disponible sur : <http://www.zdnet.fr/actualites/reseaux-sociaux-de-la-veille-en-temps-reel-a-l-analyse-predictive-39739011.htm>.
- Danny Sullivan. - FAQ : All about the new Google "Hummingbird" algorithm -. *Search engine land*. 26/09/2013. [en ligne]. Disponible sur : <http://searchengineland.com/google-hummingbird-172816>.
- . - Google's impressive - conversational search - goes live on Chrome -. *Search engine land*. 22/05/2013. [en ligne]. Disponible sur : <http://searchengineland.com/google-impressive-conversational-search-goes-live-on-chrome-160445>.
- . - Impressive : the Wolfram|Alpha - fact engine -. *Search engine land*. 1<sup>er</sup>/05/2009. [en ligne]. Disponible sur : <http://searchengineland.com/wolfram-alpha-fact-engine-13431>.
- Owen Thomas. - Google is turning search into the planet's biggest anticipatory system -. *Readwrite*. 15/05/2013. [en ligne]. Disponible sur : [http://readwrite.com/2013/05/15/google-search-anticipatory-system-1013fa9e1m?\\_e91COpQv9zOj](http://readwrite.com/2013/05/15/google-search-anticipatory-system-1013fa9e1m?_e91COpQv9zOj).
- Christina Warren. - What Apple's acquisition of Siri means for the future of mobile search -. *Mashable*. 23/04/2010. [en ligne]. Disponible sur : <http://mashable.com/2010/04/23/siri-mobile-search/>.

# Bibliographie

## Conclusion

- Elisabeth Bargas. - Le web est mort, vive le mobile ! -. *Renaissance numérique* 20/11/2014. [en ligne]. Disponible sur : <http://www.renaissance-numerique.org/travaux/706-le-web-est-mort-vive-le-mobile>.
- John Battelle. - What is search now? Dijoined -. *John Battelle's searchblog* 31/08/2012. [en ligne]. Disponible sur : <http://battellemedia.com/archives/2012/08/whats-search-now-dijoined.php>.
- Boris Beaudé. *Les fins d'Internet* [Limoges] : Ed. FYP, 2014. 95 p.
- Tim Berners-Lee. - Long live the web. A call for continued open standards and neutrality -. *Scientific American* 22/11/2010. [en ligne]. Disponible sur : <http://www.scientificamerican.com/article.cfm?id=long-live-the-web>.
- Phil Bradley. - Is search traffic dropping? -. *Phil Bradley's weblog* 02/05/2013. [en ligne]. Disponible sur : [http://philbradley.typepad.com/phil\\_bradleys\\_weblog/2013/05/is-search-traffic-dropping.html](http://philbradley.typepad.com/phil_bradleys_weblog/2013/05/is-search-traffic-dropping.html).
- Anil Dash. - The web we lost -. *Dashes. A blog about making culture* 13/12/2013. [en ligne]. Disponible sur : <http://dashes.com/aug/2013/12/the-web-we-lost.html>.
- Chris Dixon. - The decline of the mobile web -. *Chris Dixon's blog* 07/04/2014. [en ligne]. Disponible sur : <http://cdixon.org/2014/04/07/the-decline-of-the-mobile-web/>.
- Stuart Dredge. - The five types of music discovery -. *The Guardian* 19/03/2014. [en ligne]. Disponible sur : <http://www.theguardian.com/technology/2014/mar/19/music-discovery-spotify-apps-facebook>.
- Olivier Ertzscheid. - Documentation haute fréquence -. *Affordance.info* 25/03/2012. [en ligne]. Disponible sur : [http://affordance.typepad.com/mon\\_weblog/2012/03/documentation-haute-frequence.html](http://affordance.typepad.com/mon_weblog/2012/03/documentation-haute-frequence.html).
- . *Du web ouvert aux jardins fermés*. Présentation, 30/03/2015. 109 p. [en ligne]. Disponible sur : <http://fr.slideshare.net/olivier.montauze/webouvertendf>.
- . Le monde d'après (Facebook et Google) -. *Affordance.info* 22/01/2014. [en ligne]. Disponible sur : [http://affordance.typepad.com/mon\\_weblog/2014/01/le-monde-apres-facebook-et-google.html](http://affordance.typepad.com/mon_weblog/2014/01/le-monde-apres-facebook-et-google.html).
- . Vers une sanctuarisation du search? -. *Affordance.info* 12/12/2011. [en ligne]. Disponible sur : [http://affordance.typepad.com/mon\\_weblog/2011/12/sanctuarisation-du-search.html](http://affordance.typepad.com/mon_weblog/2011/12/sanctuarisation-du-search.html).
- Dinah Galligo. - Le web a 25 ans ... Survivra-t-il toutes ses déviances? -. *Prospectives* 12/03/2014. [en ligne]. Disponible sur : <http://blog.science-po.fr/prospectives/2014/03/12/le-web-a-25-ans-survivra-t-il-toutes-ses-deviances>.
- William Gibson. - Google's earth -. *The New York Times* 31/08/2010. [en ligne]. Disponible sur : [http://www.nytimes.com/2010/09/01/opinion/01gibson.html?\\_r=46](http://www.nytimes.com/2010/09/01/opinion/01gibson.html?_r=46).
- Aaron Harris. - How Google is killing organic search -. *Tutorpiece blog* 1<sup>er</sup>/07/2013. [en ligne]. Disponible sur : <http://blog.tutorpiece.com/post/54349646327/death-of-organic-search>.
- Hal Hodson. - Google's fact-checking bot build vast knowledge bank -. *New Scientist* 20/08/2015. [en ligne]. Disponible sur : <http://www.newscientist.com/article/mg23329932.700-googles-factchecking-bot-builds-knowledge-bank.html?full=true>. [initialement paru en version papier dans le n°2993, sous le titre "Welcome to the oracle"].
- Gord Hotchkiss. - John Battelle on the future of search. *Search engine land* 19/03/2010 [en ligne]. Disponible sur : <http://searchengineland.com/john-battelle-on-the-future-of-search-3338>; suivi de - John Battelle on the future of search: part two -. *Ibid* 16/04/2010. [en ligne]. Disponible sur : <http://searchengineland.com/john-battelle-on-the-future-of-search-part-two-40094>.
- Frédéric Marinet. - Tendances et innovation : les outils de recherche d'information -. *Aculligence.com* 12/04/2014. [en ligne]. Disponible sur : <http://www.aculligence.com/2014/04/12/tendances-et-innovation-les-outils-de-recherche-d-information>.
- Joshua Topolsky. - What Google just announced is a bombshell -. *BloombergBusiness* 28/05/2015. [en ligne]. Disponible sur : <http://www.bloomberg.com/news/articles/2015-05-28/what-google-just-announced-is-a-bombshell>.
- Ian Sefferman. - The convergence of mobile apps & search is happening. Get ready! -. *Marketingland* 30/05/2014. [en ligne]. Disponible sur : <http://marketingland.com/get-ready-for-the-convergence-mobile-apps-34249>.
- Jennifer Van Grove. - How tablets are changing the way we search -. *Mashable* 13/09/2011. [en ligne]. Disponible sur : <http://mashable.com/2011/09/13/tablets-changing-search/>.

# Bibliographie

## Suivre l'actualité de ces questions (*sélection*)

[francophone]

Olivier Andrieu. *Abondance*. [en ligne]. Disponible sur : <http://www.abondance.com/>, et notamment un blog : <http://www.abondance.com/actualites/>.

*L'atelier BNP Paribas Group*. [en ligne]. Disponible sur : <http://www.atelier.net/>.

Thomas Bertrand et Chris Hede. *Moteurzone*. [en ligne]. Disponible sur : <http://www.moteurzone.com/>.

Centre de documentation et d'information interuniversitaire en sciences sociales et Inforpro. *EchosDoe*. [en ligne]. Disponible sur : <http://www.echosdoe.net/>.

Christophe Deschamps. *Outils froids*. [en ligne]. Disponible sur : <http://www.outilsfroids.net/>.

Olivier Duffez. *WebRankInfo*. [en ligne]. Disponible sur : <http://www.webrankinfo.com/> et forum : <http://forum.webrankinfo.com/>.

Olivier Ertzscheid. *Affordance.info*. [en ligne]. Disponible sur : <http://affordance.typepad.com/>.

Béatrice Foenix-Niou. *Le blog de recherche-estivée.com*. [en ligne]. Disponible sur : <http://blog.recherche-estivee.com/>. Compte Twitter : <https://twitter.com/bfoenis>.

*Goopilation. The Google compilation*. [en ligne]. Disponible sur : <http://www.goopilation.com/>.

*Les infostratégies*. Didier Brodat et Fabrice Molinaro, dir. publ. [en ligne]. Disponible sur : <http://www.les-infostrategies.com/>.

*InternetActu*. [en ligne]. Disponible sur : <http://www.internetactu.net/>.

*Le journal du net*. [en ligne]. Disponible sur : <http://www.journaldu.net/>.

*Net sources*. François Libman, dir. publ. Paris : Bases publications, 1996.

*Presse citron*. [en ligne]. Disponible sur : <http://www.presse-citron.net/>.

*Techeurisch*. [en ligne]. Version française disponible sur : <http://fr.techerunch.com/>.

*ZDNet.fr*. [en ligne]. Disponible sur : <http://www.zdnet.fr/>.

Zorglook. Compte Twitter. [en ligne]. Disponible sur : <https://twitter.com/Zorglook>.

# Bibliographie

## Suivre l'actualité de ces questions (*sélection*)

[anglophone]

John Battelle. *John Battelle's searchblog* [en ligne]. Disponible sur : <http://batellamedia.com>.

Karen Blakeman. *Karen Blakeman's blog* [en ligne]. Disponible sur : <http://www.kba.co.uk/wordpress/>.

Phil Bradley. *Phil Bradley's weblog* [en ligne]. Disponible sur : <http://philbradley.sponad.com>. Scoop it : <http://www.scoop.it/internet-search>.

Wendy Boswell. *Websearch* [en ligne]. Disponible sur : <http://websearch.about.com>.

Tara Calishain. *Research Buzz* [en ligne]. Disponible sur : <http://researchbuzz.me>.

Rubén Gómez. *All G testing* [en ligne]. Disponible sur : <http://allcodelearning.blogspot.fr>.

Goven Harris. *InternetNews* [en ligne]. Disponible sur : <http://www.websearchaide.ca/netblog/>.

Mashable [en ligne]. Disponible sur : <http://mashable.com>.

ReadWrite [en ligne]. Disponible sur : <http://readwrite.com>.

Search Engine Journal [en ligne]. Disponible sur : <http://www.searchenginejournal.com>.

Search Engine Land [en ligne]. Disponible sur : <http://searchengineland.com>.

Search Engine Watch. Jonathan Allen, dir. publ. [en ligne]. Disponible sur : <http://searchenginewatch.com>.

Techcrunch [en ligne]. Disponible sur : <http://techcrunch.com>, version européenne : <http://techcrunch.com/europe/>.

WQJ.D [en ligne]. Disponible sur : <http://online.wsj.com/news/technology>.

ZDNet [en ligne]. Disponible sur : <http://www.zdnet.com>.

Marcus F. Zillman. *White papers* [en ligne]. Disponible sur : <http://www.zillman.us/white-papers/>, nombreux *white papers* mis à jour régulièrement dont moteurs de recherche scientifiques, web invisible, recherche sociale...

## Crédits

Toutes images : © Norman Rockwell  
avec, par ordre d'apparition,



*Computer scene* - Dessin, ca. 1960, [Norman Rockwell Art Collection Trust](#)



*Artist facing blank canvas - Deadline* - *Saturday Evening Post*, 08/10/1938



*Boulevard Haussmann in Paris* - *Saturday Evening Post*, 30/01/1932



*The catch* - *Country Gentleman*, 03/05/1919



*Tough Call (Bottom of the sixth, game called because of rain)* - *Saturday Evening Post*, 23/04/1949



*Full treatment* - *Saturday Evening Post*, 18/05/1940



*Fortune teller (Girl reading palm)* - *Saturday Evening Post*, 12/03/1921



*Grocer with girl's bonnet (The model, thoughtful shopper)* - *Saturday Evening Post*, 03/05/1924



*Antique Dealer (Brass merchant)* - *Saturday Evening Post*, 19/05/1934



*Armchair General (Man charting war maneuvers)* - *Saturday Evening Post*, 29/04/1944



*Connoisseur* - *Saturday Evening Post*, 13/01/1962



*Framed (Man carrying frame)* - *Saturday Evening Post*, 02/03/1946



*Expense account* - *Saturday Evening Post*, 30/11/1957



*Ship Ahoy (Setting one's sight)* - *Saturday Evening Post*, 19/08/1922



*Bookworm* - *Saturday Evening Post*, 14/08/1926