

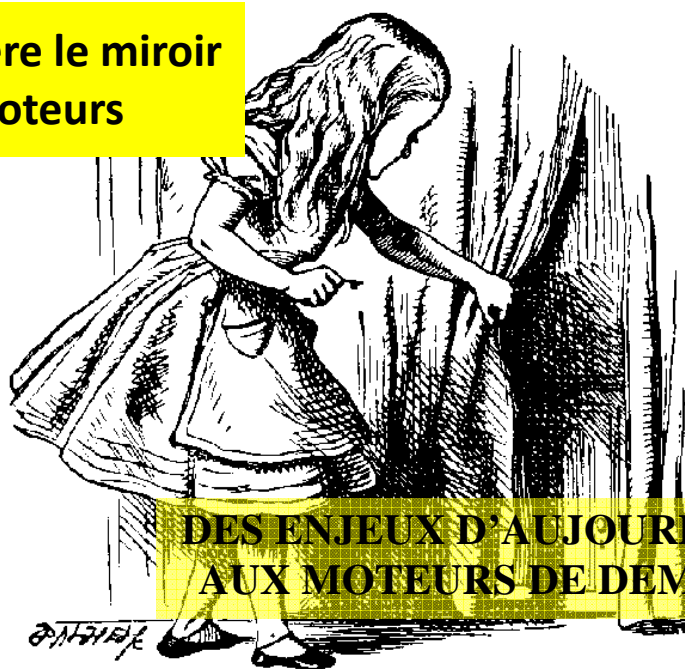
Evolution des moteurs de recherche Stage URFIST Bordeaux. 5 mai 2011



Olivier Ertzscheid // Maître de conférences
// Sciences de l'information et de la communication //
Université de Nantes // IUT La Roche sur Yon
Site : <http://www.affordance.info>



Derrière le miroir
des moteurs



DES ENJEUX D'AUJOURD'HUI
AUX MOTEURS DE DEMAIN.



Le pitch

- Les **moteurs de recherche** occupent une place prépondérante dans nos **accès à l'information** et à la connaissance.
- Ils suscitent de vives interrogations, notamment dans leur capacité à rendre **indexable** des informations relevant indistinctement des **sphères publiques, privées et intimes** des contenus disponibles en ligne.
- Réseaux sociaux et systèmes de recommandation se multiplient.
- Nouveaux usages, nouvelles métriques, nouveaux enjeux.

**L'utilisateur pris dans le graphe :
entre algorithmie ambiante et
nouvelles ingénieries relationnelles**

1934. Paul Otlet. Traité de documentation.

L'univers, l'intelligence, la science, le livre	
<p><u>Les choses</u> L'Univers, la Réalité, le Cosmos</p>	
<p><u>Les intelligences</u> qui perçoivent les choses fragmentairement</p>	
<p><u>La science</u> Reçoit et ordonne en son cadre les pièces de toutes les intelligences particulières</p>	
<p><u>Les livres</u> Transcrivent et photographient la science ainsi l'ordre des connaissances Le Catalogue de livres forme la Bibliothèque</p>	
<p><u>La Bibliographie</u> Inventorie et catalogue les livres La science de toutes les bibliographies forme le répertoire Bibliographique universel</p>	
<p><u>L'Encyclopédie</u> (Texte et image) Résume, clarifie et ordonne le contenu des livres</p>	
<p><u>La Classification</u> Conforme à l'ordre que l'intelligence découvre dans les classes tout ce l'élucide à l'ordonnance de la science des livres de leur Bibliographie et de l'Encyclopédie</p>	

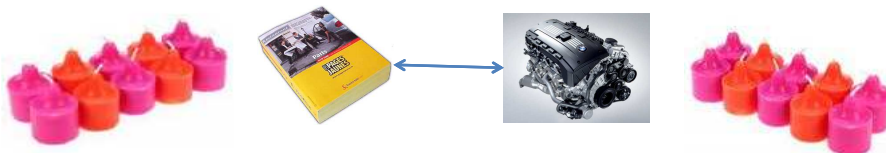
Approche du « traitement documentaire »

2011. La famille documentaire s'agrandit.

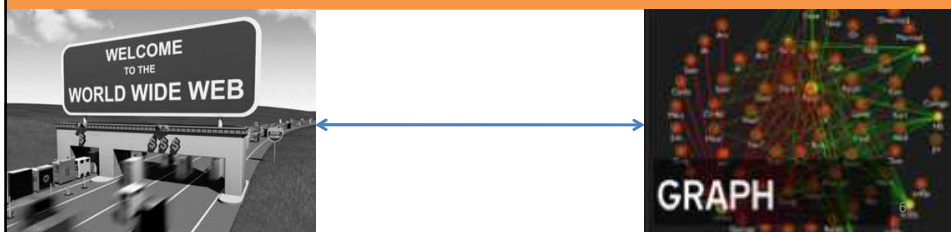


INTRODUCTION

2008. J'ai 10 ans.



Du World Wide Web au Giant Global Graph ?

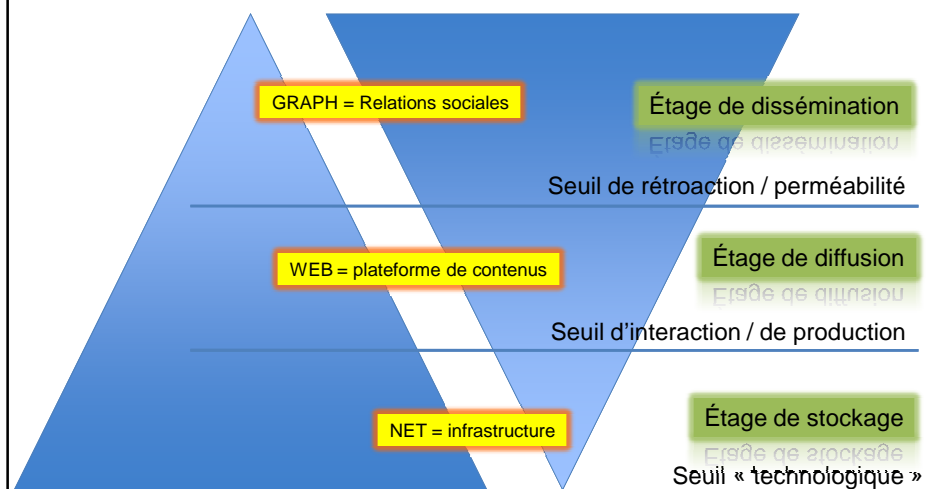


Giant Global Graph ?

- The **less inviting side of sharing is losing some control**. Indeed, at each layer --- Net, Web, or Graph --- **we have ceded some control for greater benefits.**" (...) "Letting your data connect to other people's data (...) is still not about giving to people data which they don't have a right to. It is about letting it be connected to data from peer sites. It is about letting it be joined to data from other applications. **It is about getting excited about connections, rather than nervous.**
- Tim Berners Lee, 21 Novembre 2007.
- <http://dig.csail.mit.edu/breadcrumbs/node/215>




Net – Web - Graph



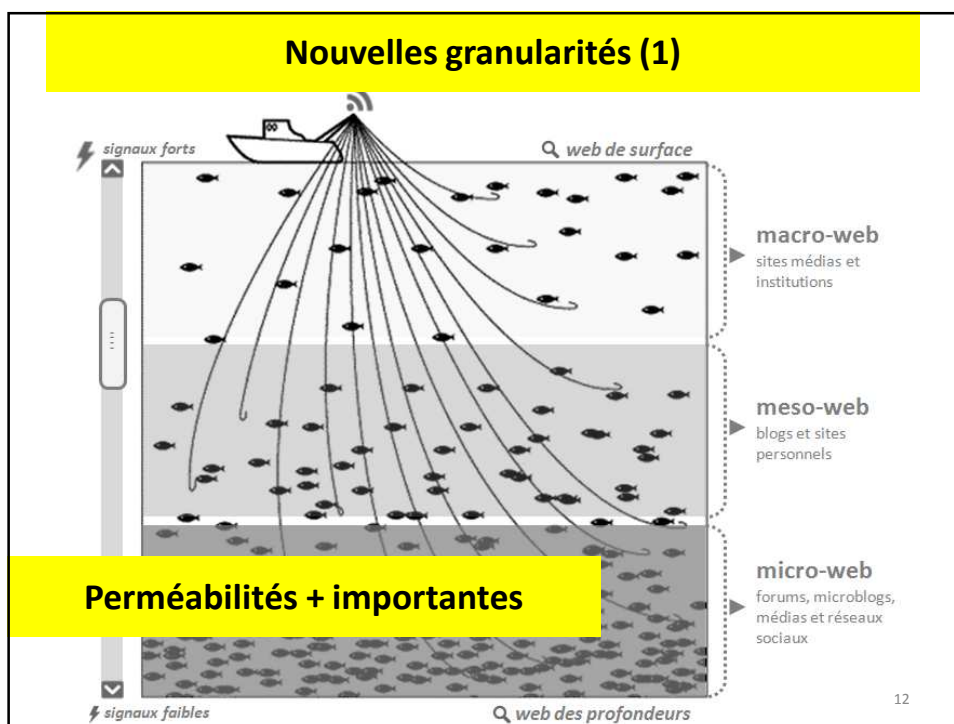
CHARLES AZNAVOUR

LA BOHÊME

PLUS RIEN  ET JE VAIS

- Le web était plus simple : **visible ou invisible**
- la vie était plus simple : **annuaires, les moteurs et les méta-moteurs.**
- l'unité de publication était plus simple : **la page (web).**
- ceux qui publiaient sur le net ne publiaient (généralement) QUE sur le net. Et ceux qui écrivaient dans les grands médias ne publiaient pas sur le net. **Autorité VS Autoritativité**
- ce qui était écrit, restait écrit, restait fixé.

« Je vous parle d'un temps ... » 11



Nouvelles granularités (2)

AVANT



MAINTENANT



INDEXABLE



NON-INDEXABLE



Porosités + importantes

13

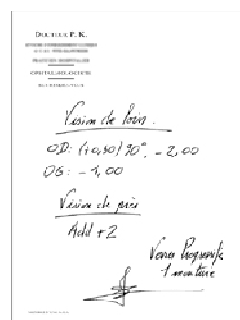
Chapitre Premier

Des machines sociales



Description, restitution, prescription.

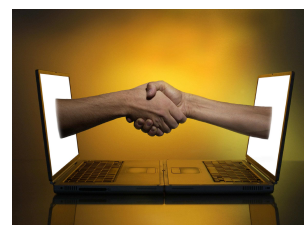
- **Annuaire**
 - le web « décrit »
- **Moteurs**
 - le web « restitué »
- **Aujourd'hui et demain**
 - Le web « prescrit »



15

La recommandation : de toute éternité (numérique)

- **Web 1.0**
 - Annuaire = sites « recommandés » parce que « choisis ».
 - Moteurs = Pagerank : 1 lien = un vote.
- **Web 2.0**
 - « résultats de recherche » + « recommandations » + « choix de reformulation » en accord avec nos choix, nos itinéraires ou nos parcours précédents
- **Demain (web sémantique, socio-sémantique, sémantisé, implicite ...)**
 - Mêmes possibilités mais sans nécessiter un « amorçage ».
 - Savoir ce que nous sommes le plus susceptibles de chercher selon l'heure de la journée, le lieu de notre connexion ou encore notre environnement, notre historique de recherche, nos achats récents, nos recommandations, nos réseaux sociaux ...



16

Chapitre Premier

~~Des machines sociales~~

Des algorithmes ... relationnels & relativistes ...



Algorithmic Night Fever



18

1 homme, 1 méthode, 1 mesure

Abbreviated Journal Title	ISSN	2005	Impact	Intermediary	2006	2007
SCIENCE	0036-8075	381.399	23.025	6.555	695	7.7
CELL	0092-9674	133.930	19.189	6.463	632	6.3
NATURE	0028-0836	280.900	25.663	6.769	602	7.4
NAT CELL BIOL	1467-7701	33.930	16.466	4.463	154	4.2
GENE DEV	0899-8989	49.417	14.066	2.834	297	6.4
PLoS BIOL	1544-4125	33.030	14.436	2.867	160	2.1
NATL CELL	1097-2705	24.700	14.033	3.043	315	4.2
DEV CELL	1099-0807	30.000	13.833	3.176	187	3.9
DEV CELL	1099-0807	30.000	13.833	3.176	187	3.9
GENE DEV	0899-8989	24.945	10.068	2.443	340	4.3
J CELL BIOL	0021-9525	89.403	10.121	3.906	271	6.2
PNAS	0027-8149	7.016	10.098	3.849	187	7.4
J NATL ACAD SCI USA	0027-8149	37.037	8.642	3.760	236	7.1
EMBO REP	1469-2211	6.294	8.275	3.702	133	3.5
DEV CELL	1099-0807	4.846	7.764	3.570	147	6.8
DEV CELL	1099-0807	4.846	7.764	3.570	147	6.8
FASEB J	0895-6626	33.021	6.712	3.241	449	5.8
DISCOVER	1539-1019	3.102	6.612	2.790	120	7.4
NATL CELL	1099-0807	3.038	6.562	3.472	162	4.0
J CELL BIOL	0021-9525	80.710	6.405	2.961	463	3.2
NATL ACAD SCI USA	0027-8149	7.016	6.331	3.741	193	6.1
J Biol Chem	0021-9924	436.600	6.006	2.100	436	6.0
PLoS	1553-6760	4.961	5.111	3.020	217	4.2

1 algo, DES variables, 1 boîte noire « Matching » versus « Watching »

PageRank « original » + Algorithme de pertinence + Monétisation publicitaire

« And that, in simple terms, is how you increase your ranking on search engines. »

« This Google Algorithm always works for me! »

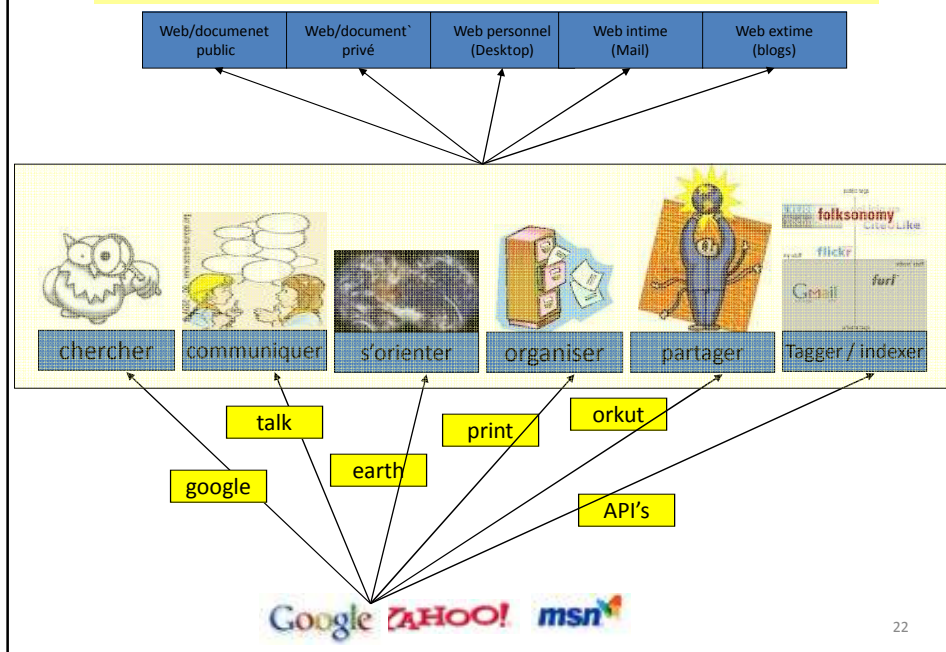
Google

L'algorithme ET la main de l'homme

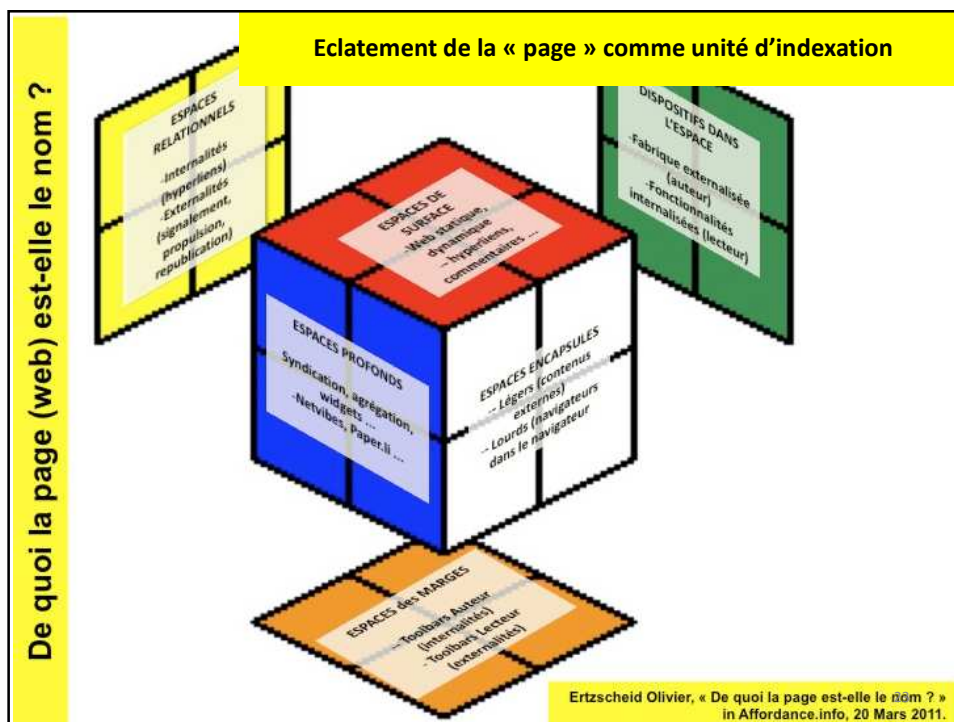
- | | |
|--|--|
| <ul style="list-style-type: none"> • Nous n'intervenons jamais manuellement sur le contenu des pages de résultat <ul style="list-style-type: none"> – Gimmick de la firme • (mais) Les publicités sont du contenu <ul style="list-style-type: none"> – (« ads are content » Ormid Kordestani) • (et) Nous intervenons manuellement sur l'achat des mots-clés publicitaires <ul style="list-style-type: none"> – (épisode de l'achat du mot-clé « abortion » par le Christian Institute) | <ul style="list-style-type: none"> • Nous n'intervenons jamais manuellement sur le contenu des pages de résultat <ul style="list-style-type: none"> – Gimmick de la firme • (mais) Nous modifions manuellement l'algorithme de pertinence <ul style="list-style-type: none"> – (400 modif en un an, Google Bombing) • (et) L'algorithme de pertinence contrôle l'affichage des pages de résultats <ul style="list-style-type: none"> – (mot-clé « abortion » par le Christian Institute : registre émotionnel interdit) |
| <ul style="list-style-type: none"> • Donc nous intervenons manuellement sur le contenu. | <ul style="list-style-type: none"> – Donc nous intervenons manuellement sur le contenu. |

21

Confusion des pratiques informationnelles



22



Moteurs de prestidigitation et d'escamotage

- Escamotage des contenus
- Moteurs : trous noirs du web ?
- Syndrome Netvibes : le web « en creux »
- Read/Write Web



Chapitre Premier

~~Des machines sociales~~

~~Des algorithmes ... relationnels & relativistes ...~~

Économie « concentrée » de l'attention



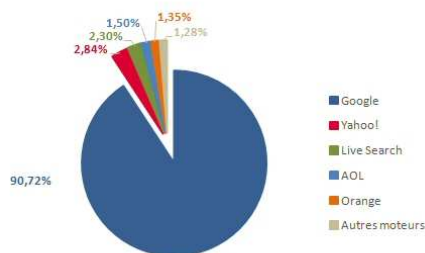
Question d'économie(S)

1. Attention economy :
 - la captation de l'attention comme modèle, la monétisation des accès comme enjeu
2. Abundance economy
 - La longue traîne comme contexte, le filtrage comme enjeu
3. Gift economy :
 - le don comme idéal, la gratuité comme modèle
4. Reputation economy
 - Babélisation des expertises comme « milieu », localisation d'expertise comme enjeu.
5. Trust economy
 - La foule comme support, la modélisation de la confiance comme vecteur

Très très très concentré

comScore Core Search Report*			
July 2008 vs. June 2008			
Total U.S. – Home/Work/University Locations			
Source: comScore qSearch 2.0			
Core Search Entity	Share of Searches (%)		Point Change Jul-08 vs. Jun-08
	Jun-08	Jul-08	
Total Core Search	100.0%	100.0%	NA
Google Sites	61.5%	61.9%	0.4
Yahoo! Sites	20.9%	20.5%	-0.4
Microsoft Sites	9.2%	8.9%	-0.3
Ask Network	4.3%	4.5%	0.2
AOL LLC	4.1%	4.2%	0.1

TOP 5 des moteurs de recherche en parts de visites
Juillet 2008



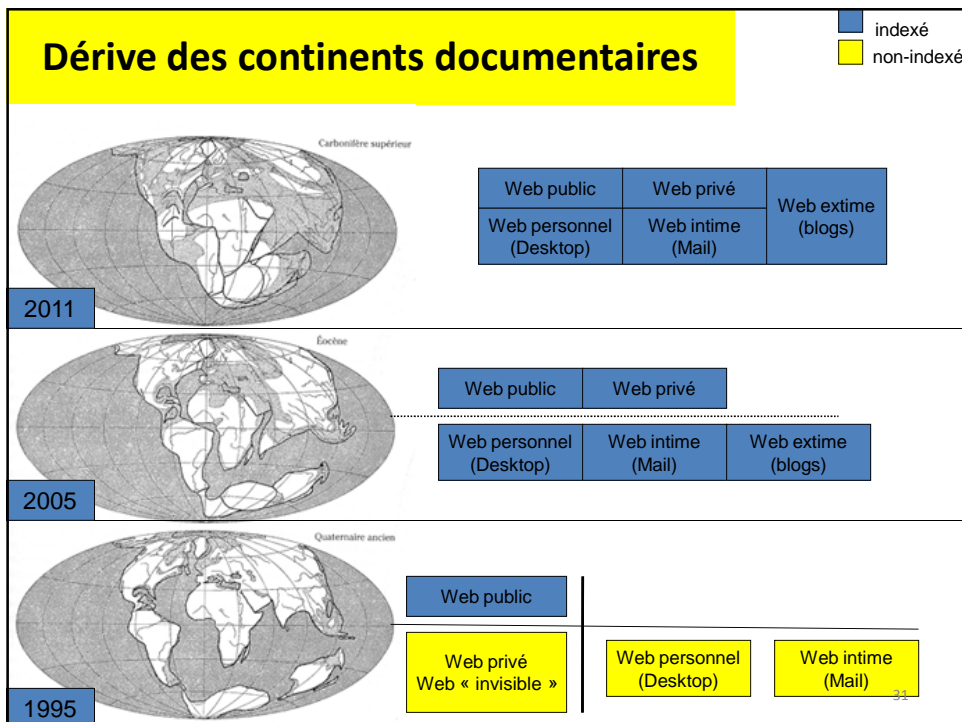
XITI Monitor
www.xiti.com

29

Chapitre Second

DE LA DÉRIVE DES CONTINENTS DOCUMENTAIRES ...
 À LA RECHERCHE UNIVERSELLE ...
 ET A SES IMPLICATIONS SUR LA NATURE DOCUMENTAIRE DE L'HOMME.
 ... OU COMMENT LE WEB DE MÖBIUS LASSE PLACE AU SYNDROME DE
 BLANCHE NEIGE.





Recherche Universelle : le principe

Google Universal Search Explained

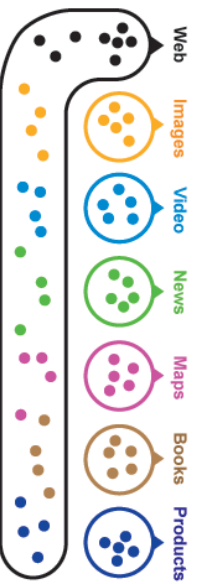
Old Google Model

Previously, a web search on Google would only return web pages and document contents. Items such as images and video were not included.



New Google Model: Universal Search

A web search on Google can now include multiple types of information such as images, video, news, maps, books, and products.



© 2007 Elluminate, Inc. | www.elluminate.com

33

GYM universelle

The screenshot shows a Google search interface with the query 'GYM universelle'. The results are categorized into several sections:

- Images:** A grid of images showing people in various poses, likely related to fitness or gymnastics.
- Maps:** A map showing the location of 'Gym Universelle' in Paris, France.
- Books:** A list of books related to the search, including titles like 'Gymnastique Universelle' and 'Gymnastique Universelle pour les enfants'.
- Products:** A list of products, including items like 'Gymnastique Universelle' and 'Gymnastique Universelle pour les enfants'.
- Web:** A list of websites, including 'Gym Universelle' and 'Gymnastique Universelle'.

 The search results are displayed in a clean, organized layout with clear labels for each category.

Recherche universelle : après les autoroutes ...



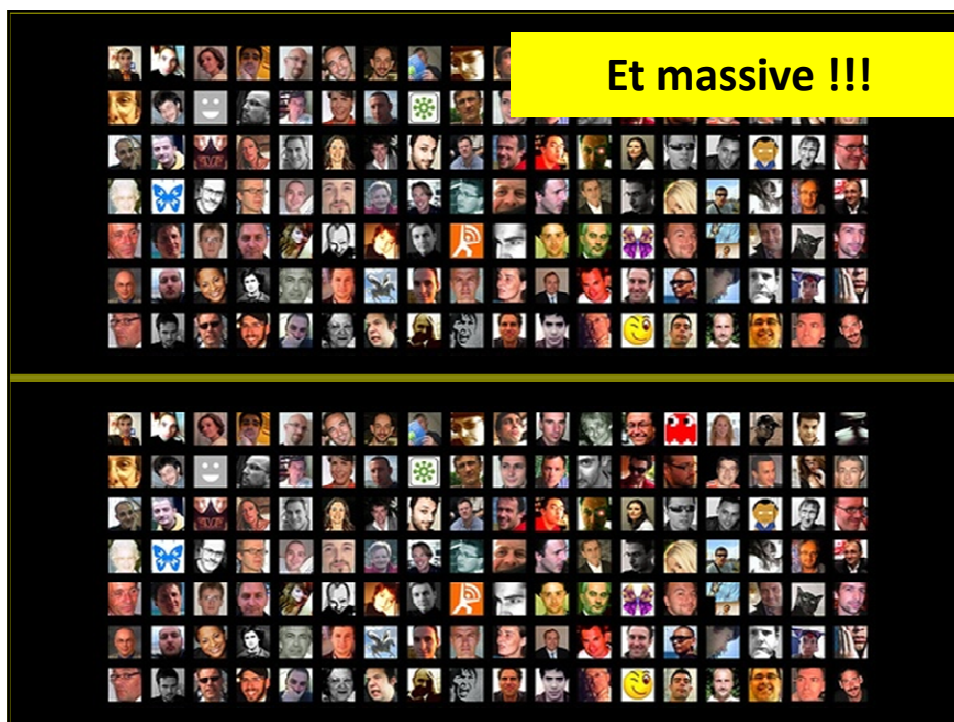
Les « hypermarchés » de l'information ?

35



Participation active ...

36



Une
« universalité »
très
très
très
« personnalisée »

ALGORITHMIQUE
AMBIANTE



38

Mon nom est personne

- Personnalisation :
 - « Invisible »
 - « Persistante »
 - « Participative »



39

Moteur, mon beau moteur ...

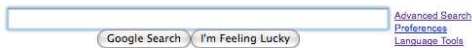


Le syndrome de Blanche-Neige

© <http://www.radidesigners.com/miroir1.htm>

40

L'illusion de la page blanche



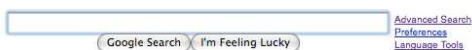
[Google is turning 10](#). See what's happened in the past 10 years.

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#) - [Go to Google France](#)

©2008 - [Privacy](#)

41

Le syndrome de Blanche-Neige



[Google is turning 10](#). See what's happened in the past 10 years.

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#) - [Go to Google France](#)

©2008 - [Privacy](#)



42

Attention !!!!!



43

Pêché originel

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) olivier.ertzscheid@gmail.com [My Account](#) | [Sign out](#)



Google Search I'm Feeling Lucky

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

[Google is turning 10](#). See what's happened in the past 10 years.

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#) - [Go to Google France](#)

©2008 - [Privacy](#)



44

Moi ...

Google Search I'm Feeling Lucky

45

Cocktail explosif

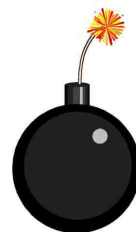
Personnalisation

- + « universalité » des recherches
- + Migration dans les nuages
- + Massification
- + Communautarisation
- + Économie de l'attention
- + Confusion des pratiques

.....

BOUM !

- Ce que l'on savait déjà :
 - Explosion de la masse documentaire
- Ce que l'on ne savait pas :
 - Bombe à fragmentation
- Fragmentation
 - Des documents (world wide web)
 - Des usages
 - Des synchronicités (world life web)
 - Des modalités
 - de diffusion
 - d'appropriation
 - d'édition
 - ...



47



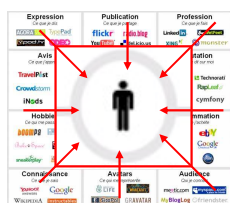
48

Question : quel est le plus grand dénominateur commun et/ou le plus petit commun multiple de nos identités documentaires numériques ?



Nous sommes le PPCM de nos identités numériques

Ils sont le PGDC de nos identités numériques



49

Base de donnée des intentions (John Battelle)

FIELDS IN THE DATABASE OF INTENTIONS AS OF EARLY 2010 (V2)		
FIELD	SIGNAL	CURRENT PLAYERS (SAMPLE)
"The Purchase"	What I Buy	amazon ebY Walmart
"The Query"	What I Want	Google Yahoo! bing
"The Social Graph"	Who I Am	facebook myspace Google
	Who I Know	facebook
"The Status Update"	What I'm Doing	twitter facebook Google
	What's Happening	facebook Google
"The Check-in"	Where I Am	foursquare Yelp Gowalla

source = 808kimedia.com

50

Le carré magique des usages

SHOPPING (digital cameras)	6	HEALTH (Diabetes)	10
13	25	11	1
TRAVEL (United 875)	3	LOCAL (French Restaurant)	32

51

Moralité ?

- Pour les moteurs ...
- Pour les réseaux sociaux

Indexable, cherchable,
Profilable, « monétisable » ...



« un document »

52

« L'antilope qui court dans les plaines d'Afrique ne peut être considérée comme un document...»



« Mais si elle est capturée... et devient un objet d'études, on la considère alors comme un document. Elle devient une preuve physique. » Suzanne Briet

53

Autarcithécaires

- Temps 1 : Matching
- Temps 2 : Watching
- Recherches « personnalisées »
- Temps 3 : Warming (cocooning)
- Moteurs « à la carte »
- Résultats « à la carte »
- Pose la question de la CARDINALITE de la recherche
 - **Cherchons-nous pour découvrir** ce que nous ne savons pas (au risque de l'erreur, de l'inutile, du futile). Cherchons nous la différence ? Espace de navigation inconnu
 - **Cherchons-nous pour retrouver** ce que nous ou nos « amis » connaissent déjà ? Cherchons-nous de l'aliénant ? Temps de cerveau disponible
 - **Cherchons-nous pour acheter** (cf carré des pages)



54



Moteur de recherche « de voisinage »

Intelius.com

Everyblock.com / <http://chicago.everyblock.com/news/>

Adresse / numéro de sécurité sociale / historique déménagements et changements d'adresse depuis 10 ans
 Date de naissance / Casier judiciaire / Diverses amendes / Dates mariages et divorces / le nom et n° tel voisins
 Montant du bien immobilier qu'il possède (sa maison), sa date d'achat et le nom de celui qui lui a vendu
 la taille du "lot" (cadastre) sur lequel elle (sa maison) est implantée
 une vue satellite de chez lui
 Âge moyen des femmes et des hommes habitant dans son voisinage
 Nombre de meurtres, de viols, de vols de véhicule et autres larcins dans son quartier
 Composition ethnique du quartier / Niveau d'éducation et les langues parlées
 les permis qu'il possède (permis de conduire mais aussi permis bateau, avion, moto, etc...)



INTELIOUS
Live in the know.™

<http://www.Intelius.com>

[<< Return to Home](#)

Select A Product or Special Offer

People Search Report

Includes all 2 Records for **Olivier Ertzscheid** in **San Diego, CA.**

Report includes when available:

- ✓ Full Name
- ✓ Age & DOB
- ✓ Relatives
- ✓ Avg. Income
- ✓ Address
- ✓ Phone Number
- ✓ Address History
- ✓ Home Value

24 Hour Pass

Unlimited People Search Reports by Name or Address for 24 hours.

Report includes when available:

- ✓ Full Name
- ✓ Age & DOB
- ✓ Relatives
- ✓ Avg. Income
- ✓ Address
- ✓ Phone Number
- ✓ Address History
- ✓ Home Value

Background Report

Includes all 2 records for **Pierre Ertzscheid** in **San Diego, CA.**

Report includes: (when available)

- ✓ Full Name
- ✓ Age & DOB
- ✓ Relatives
- ✓ Avg. Income
- ✓ Criminal Check
- ✓ Liens
- ✓ Aliases
- ✓ Neighbors
- ✓ Marriage
- ✓ Address
- ✓ Phone Number
- ✓ Address History
- ✓ Property
- ✓ Bankruptcies
- ✓ Judgments
- ✓ Lawsuits
- ✓ Death Records
- ✓ Divorce

[View Sample](#)

"Nous cherchons dans des milliards de données publiques pour vous aider à trouver ce que vous cherchez. Nous analysons des listes de vente, des commandes sur catalogue, des abonnements à des magazines, des enregistrements d'adresse, des enregistrements de propriété immobilière, des arrêts de cour (...) et toute une série d'autres enregistrements publics et de sources publiques pour vous."

\$0.95

51% off Special Price!
[Learn More](#)

Add to Cart

With FREE Identity Protect Trial

\$14.95

25% off Special Price!
[Learn More](#)

Add to Cart

With FREE Identity Protect Trial

\$39.95

\$10 off Special Price!
[Learn More](#)

Add to Cart

With FREE Identity Protect Trial 56

Moteurs => Pré-scripteurs => Scripteurs

- Moteurs ??
- Dé-scripteurs
- Lecteurs (lectures industrielles)
- Pré-scripteurs
- Scripteurs
 - (<http://actu.abondance.com/2011/02/google-change-les-titres-de-pages-dans.html>)

La question du lien est celle du vote. Du cheminement

$$PR(A) = \frac{1-d}{C(A)} + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

- Algorithmie du lien
 - Avènement du Pagerank
- Technologies sociales
 - Web 2.0 + réseaux sociaux

- Ingénieries relationnelles

- Le degré zéro du lien
- Like, share, recommend, "+1" : chimères d'hyperliens.
- n'inaugurent aucun chemin sauf celui du retour vers un entre-soi.

La question du « like » est celle de « l'appréciation »

59

Chapitre troisième

SI LOIN ... SI PROCHE.
RÊVES ET RÉALITÉS MOTORISÉS.

WEB-OS ...
REMIXÉ ...
IMPLICITE ...
SEMANTIQUE ...
SYNCHRONE ...

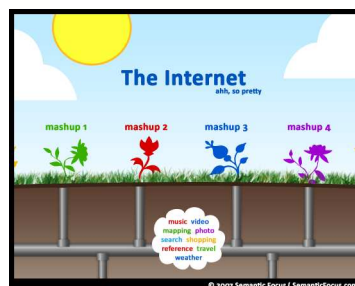


Vers les WebOS



WEB REMIXÉ

- Mixage / Mash-ups
- « tout est dans tout »
- « tout est en ligne »
- Empilement applicatif
- Brassage logiciel
- Hybridation des usages



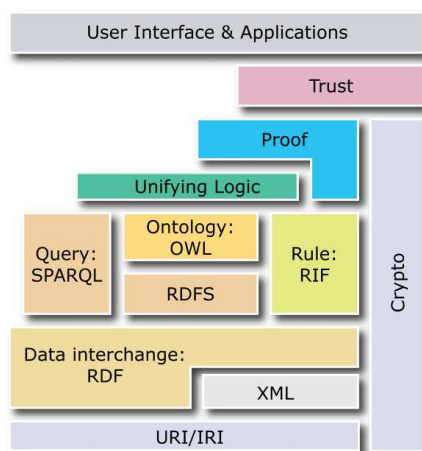
Web implicite

- « *Le chemin plus que le lien* »
- Comportements et requêtes « recommandés »
- « Myware » : « je suis partout ».
- « Everyware » : le « tout interface »

63

Rêve sémantique : Visions

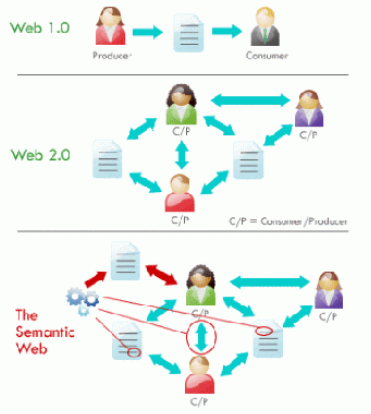
- A l'origine : Technique, protocoles, ontologies.



64

Rêve sémantique : Visions

- La promesse d'un web plus « intelligent » / « intelligent » / « intelligent »

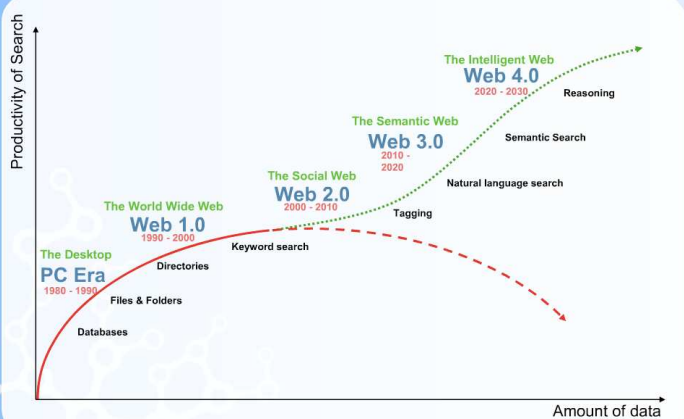


65

Rêve sémantique : Visions

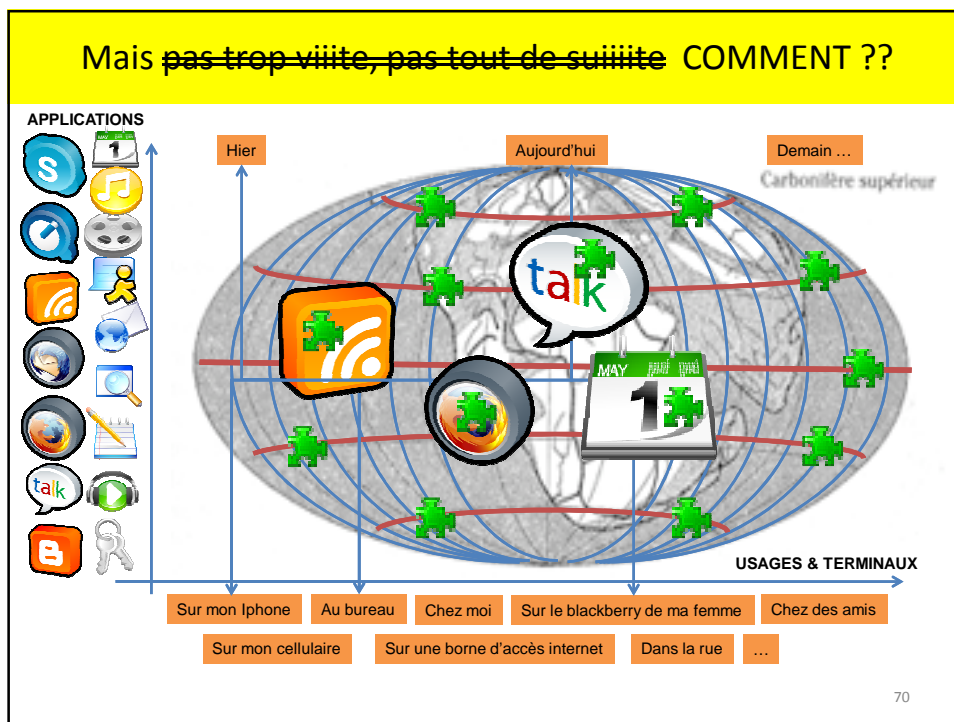
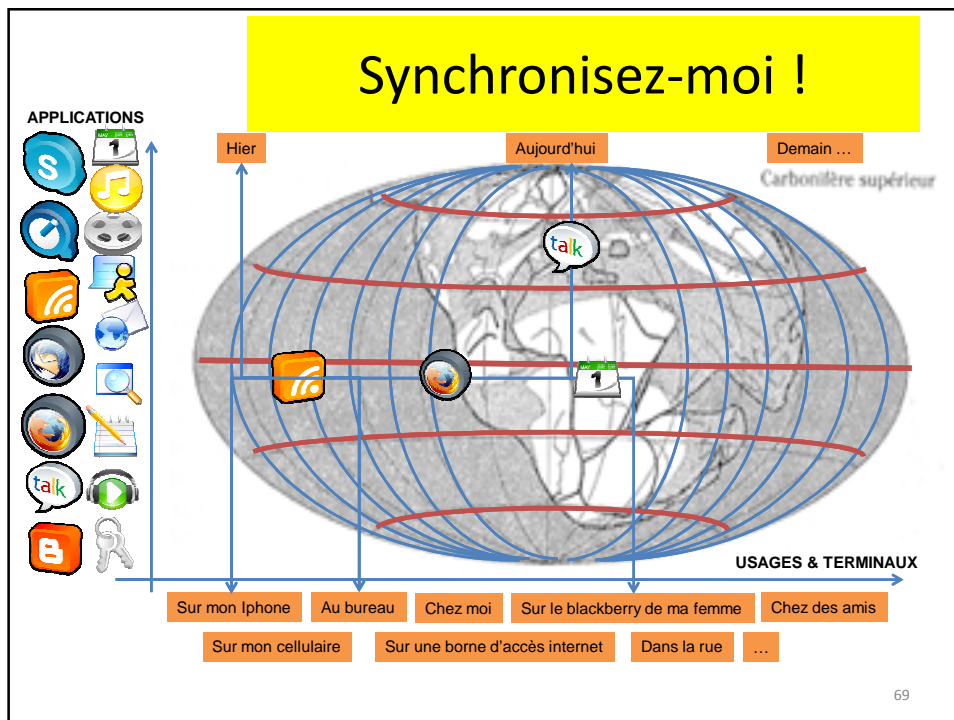
En diachronie ...

Beyond the Limits of Keyword Search

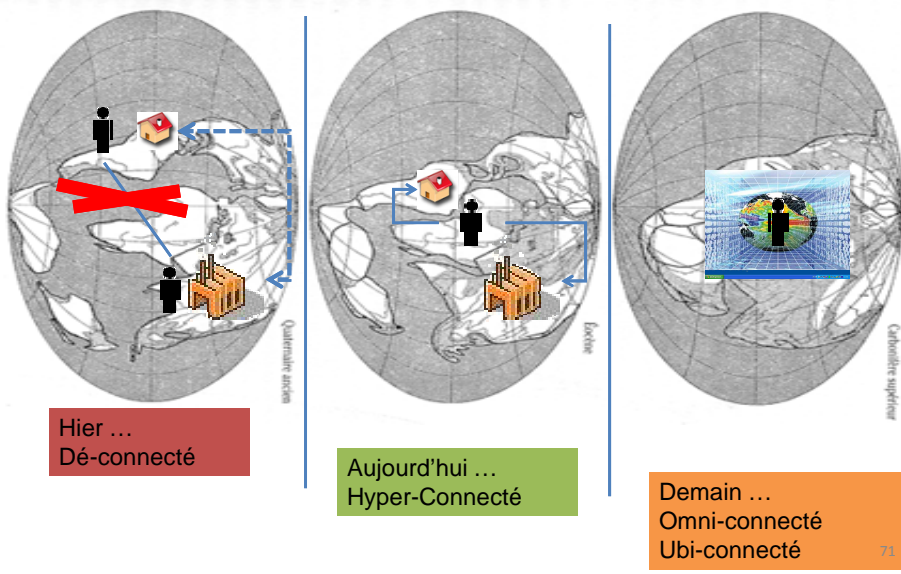


66

Radar Networks 6



Sinon ...



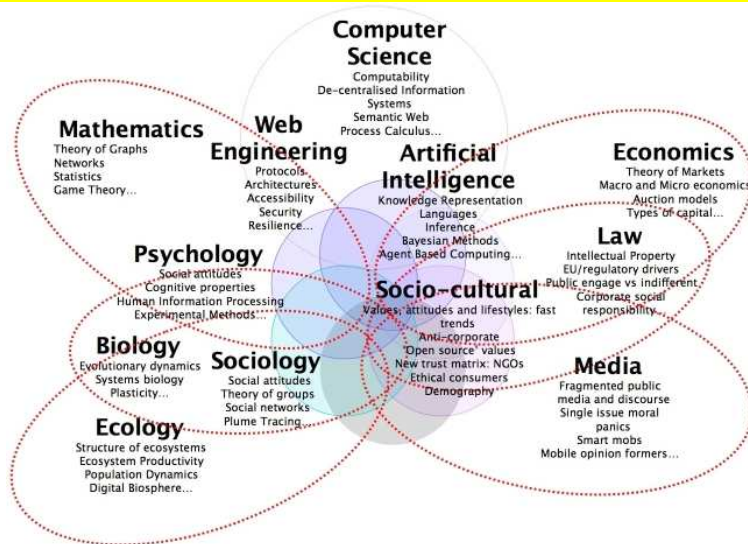
CONCLUSION



POUR DEMAIN ...

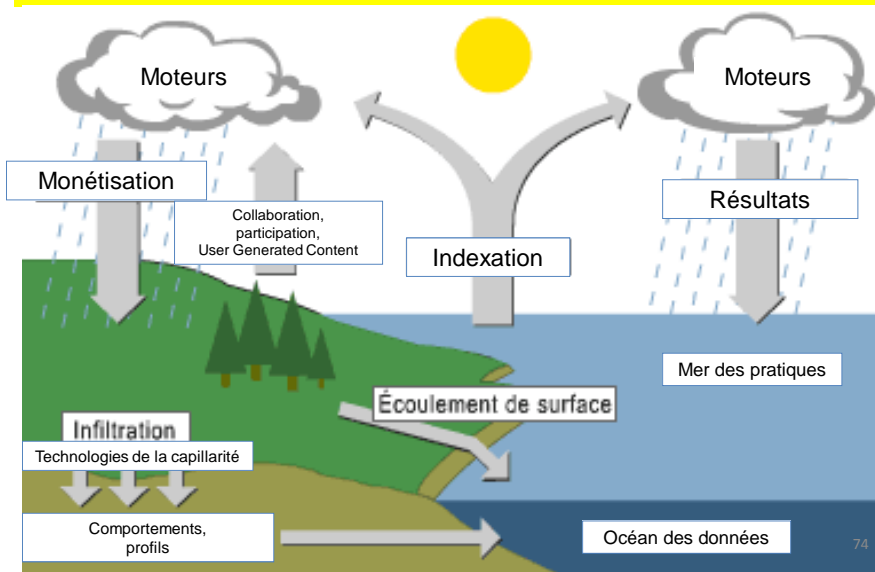
- Quelle Science du Web ?
- Quel écosystème informationnel ?
- Quelle alternative motorisée ?
- Quelle écologie politique ?

Web Science /::/ <http://webscience.org/>



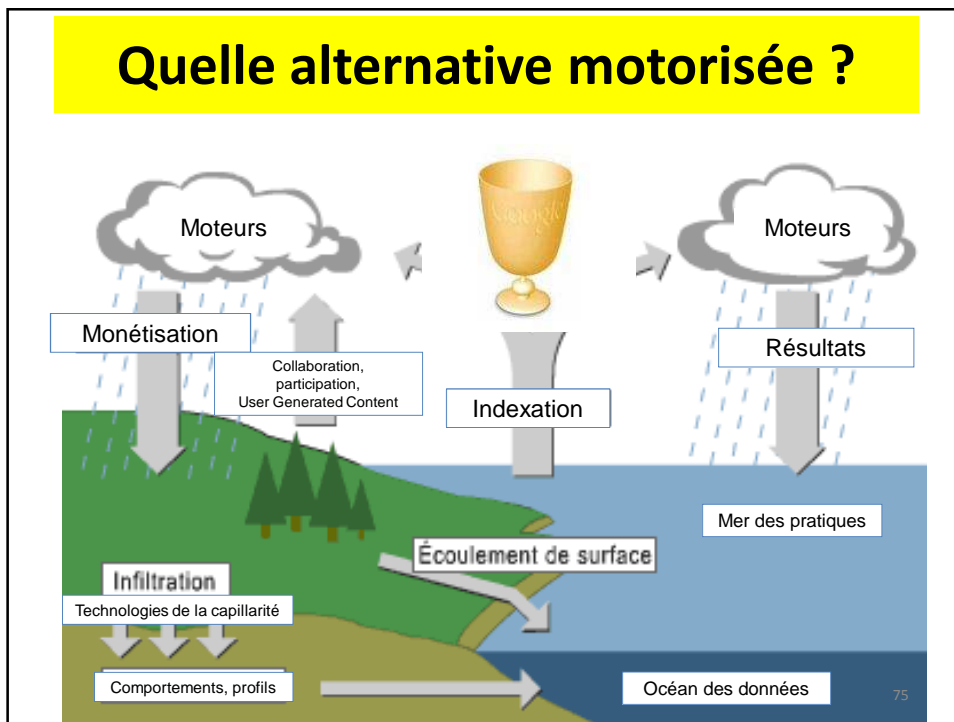
73

Quel écosystème informationnel ?

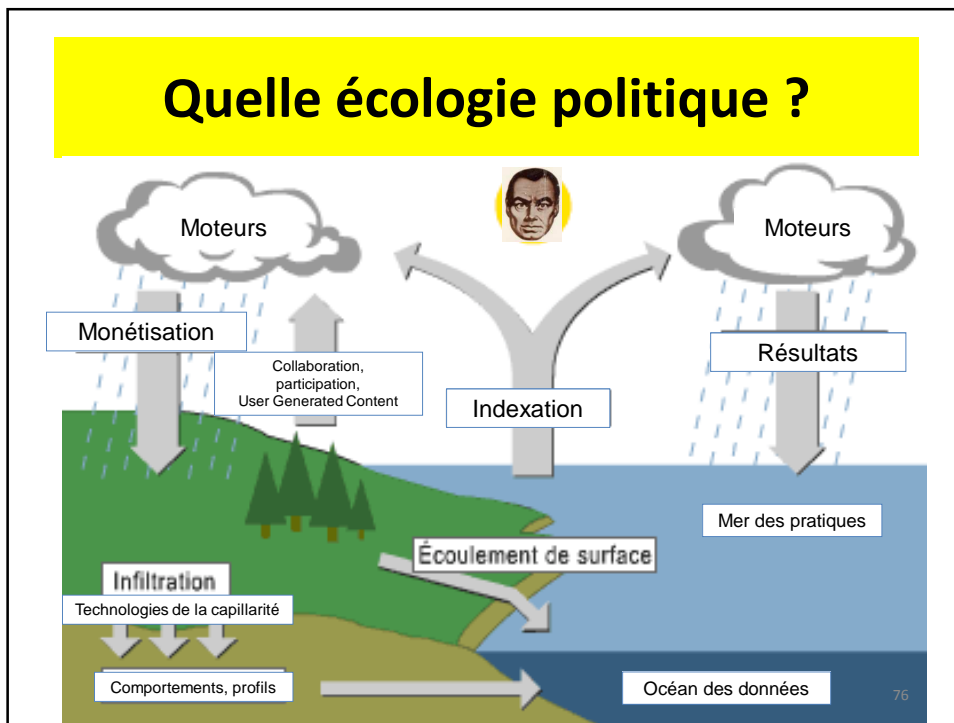


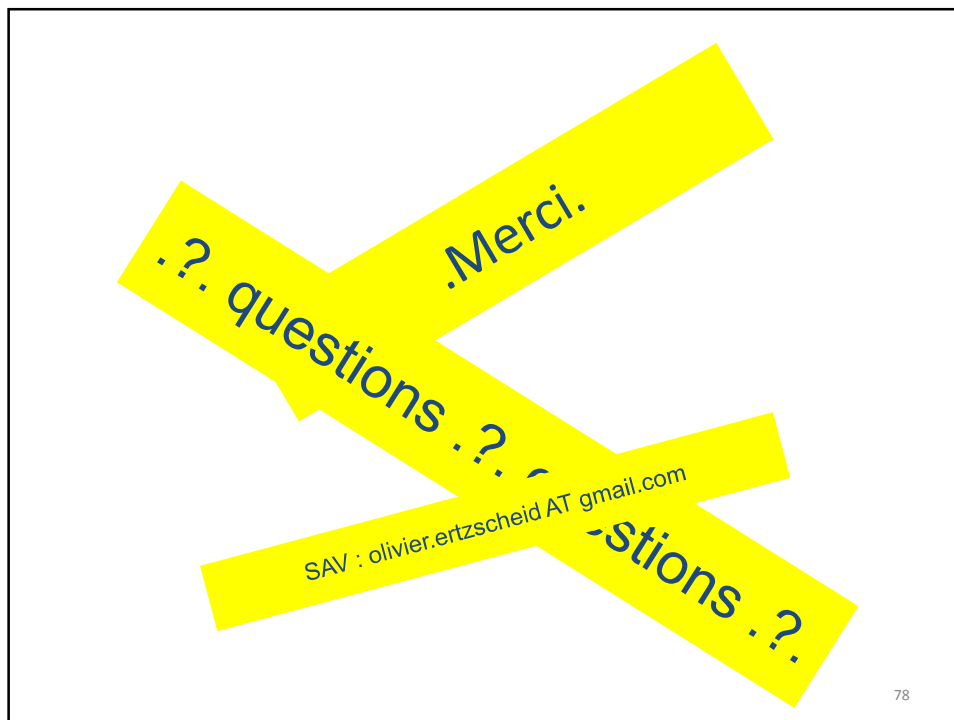
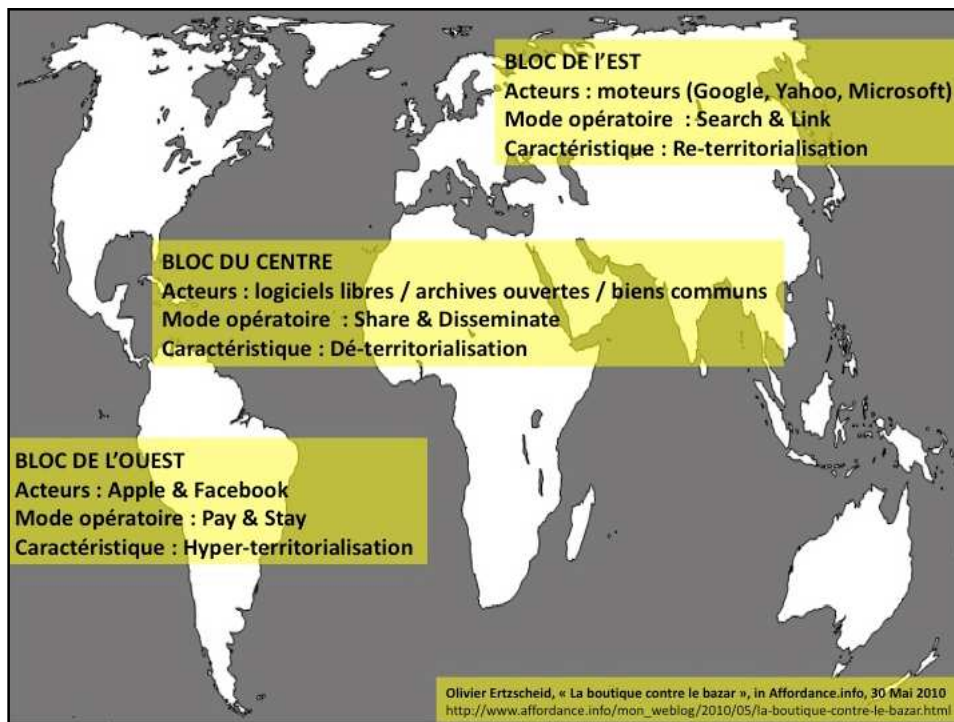
74

Quelle alternative motorisée ?



Quelle écologie politique ?



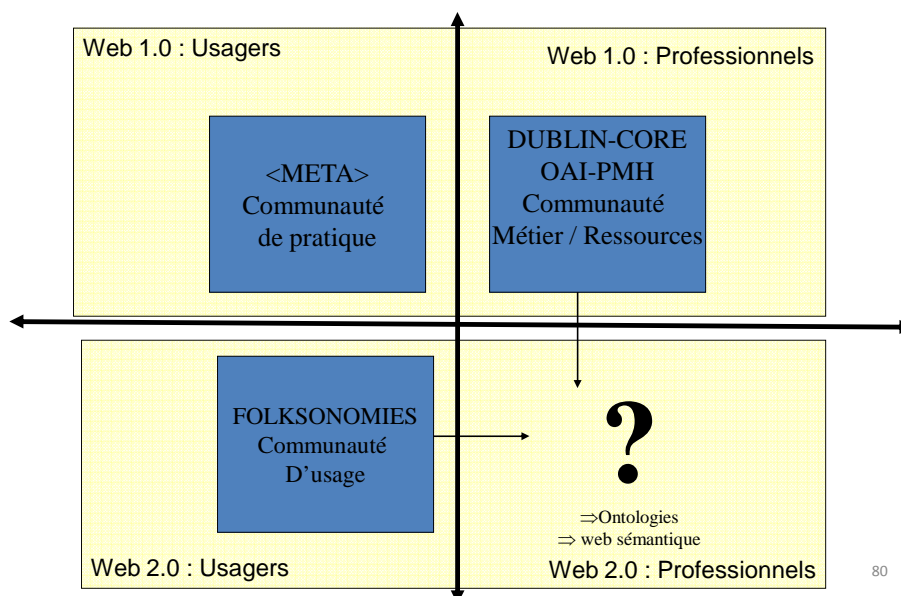


Le thesaurus comme trésor de guerre.

- *De l'indexation « ouverte »*
 - **Indexeurs sans le savoir.**
 - **L'échec des balises <META>**
 - **Standardisation et communauté métier.**
- *à l'indexation marchande ...*
- *Jusqu'à l'indexation sociale*

79

Indexation métier ... marchande ... sociale ... structurée ...



Indexation ... marchande

- Thesaurus ... est un « thesaurus »

Indexation ... sociale

- **Folksonomies : le retour de la communauté comme indexeur.**
- **Le mot plutôt que l'absence du mot.**

81

Le mot plutôt que son absence

The image shows two side-by-side screenshots. The left screenshot is from Del.icio.us, displaying a tag cloud of popular tags such as 'blog', 'design', 'music', 'software', 'reference', 'news', 'vide', 'web2.0', 'tools', 'art', 'webdesign', 'programming', 'fre', 'photography', 'linux', 'web', 'mac', 'css', 'shopping', 'travel', 'game', 'tutorial', 'blogs', 'imported', 'business', 'howto', 'search', 'education', 'food', 'technology', 'javascript', 'politics', 'tips', 'photo', 'fun', 'goog', 'online', 'flash', 'research', 'tv', 'books', 'windows', 'science', 'home', 'opensource', 'recipes', 'osx', 'ajax', 'photoshop', 'photos', 'health', 'development', 'humor', 'app', 'internet', 'media', 'culture', 'community', 'funny', 'java', 'incarnation', 'history', 'sarcu'. The right screenshot is from Google, showing the search bar, navigation links (Web, Images, Groups, News, Froogle, Maps, more), and the Google logo.

82