

Web et révolution numérique : promesses et écueil

Serge Abiteboul

Sèvres débats, 19 janvier 2011



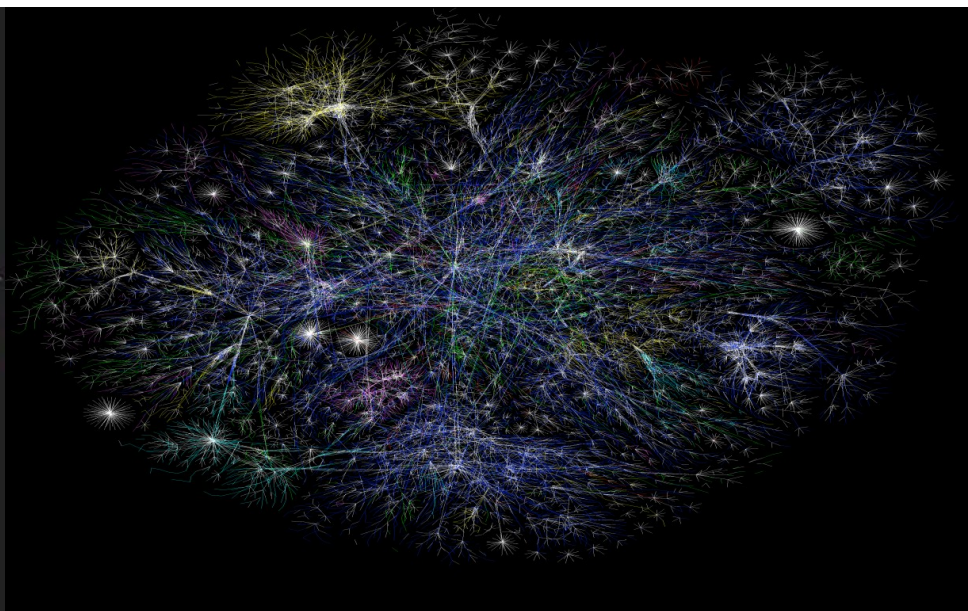
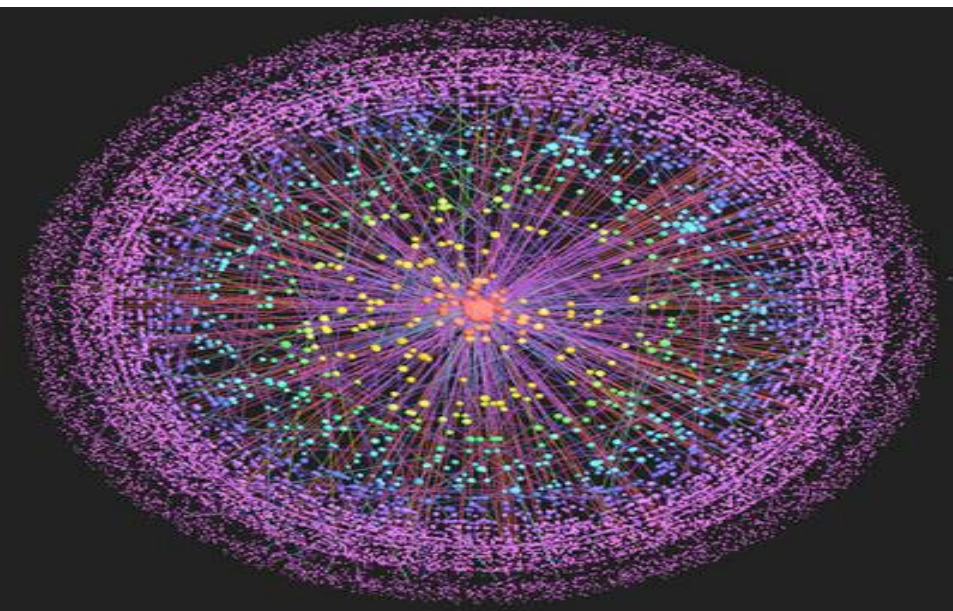
Précaution 1

Je suis un « technicien » du Web pas un sociologue ou un politique

J'ai un avis très biaisé sur les incroyables promesses du Web

Si vous avez des avis très tranchés sur ses risques terrifiants

n'hésitez pas à vous exprimer, à interrompre, à siffler !



Organisation

Le Web

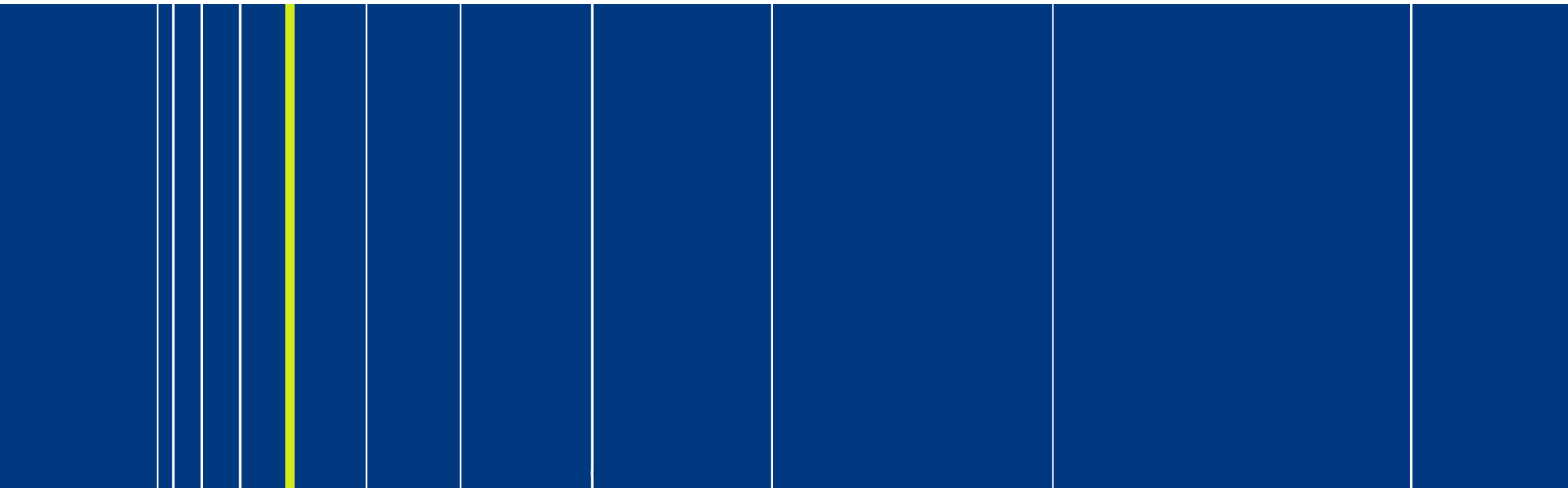
Un zoom sur la technique

Tendances

Le Web pour quoi faire?

Conclusion

Le Web

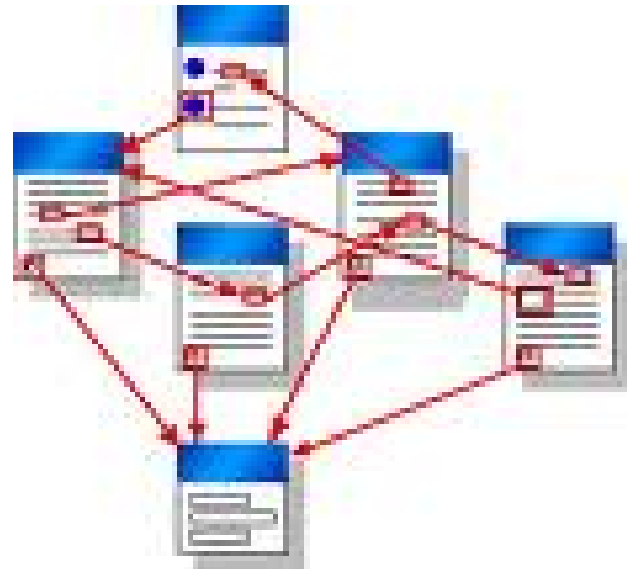


Informatique, information, internet, web

L'informatique est la science de l'information et du traitement de l'information

Internet est un réseau qui permet de transférer des flux d'information entre machines connectées au réseau (TCP)

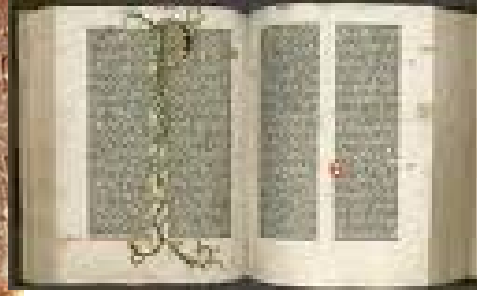
Le World Wide **Web**, littéralement la « toile (d'araignée) mondiale », est un système hypertexte public fonctionnant sur Internet qui permet de consulter, avec un navigateur, des pages accessibles notamment via des moteurs de recherche



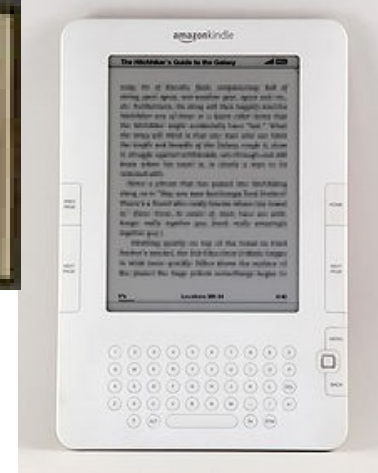
L'accélération de l'histoire



-3500



1450



2000

De la tablette d'argile au Kindle

Révolution industrielle

- Agriculture, la fabrication, le transport
- 18ème et 19ème siècles

Révolution digitale

- **1950**: Informatique
- **1970**: Internet
- **1980**: Email
- **Depuis 1990**: **World Wide Web**, Google, Web 2.0, P2P, Facebook, Blogs – Impact dans tous les domaines

Les dimensions de l'Internet et du Web

1 milliard d'hôtes (adresses IP)

2 milliards d'utilisateurs (11/09 - 29% population)

des milliards d'objets communicants

des centaines de millions de sites Web

1000 milliards de pages (9/08)

500 millions d'utilisateurs sur Facebook (7/10)

Serait le 3ème pays de la planète

plus de 10 milliards de recherches sur le Web/mois (4/08)

Le quantitatif : le volume

ça ne vous parle pas; à moi non plus

1 octet = 8 bits

1 téraoctet = 10^{12} octets

- 200 téraoctets = tous les livres jamais écrits

1 pétaoctet = 10^{15} octets

- 100 pétaoctet: La quantité de données produites par le "collisionneur de particules" du CERN en une minute

1 exaoctet = 10^{18} octets

- 5 exaoctets = un texte avec tous les mots jamais dits

1 zettaoctet = 10^{21} octets

- **$\frac{1}{2}$ zetta le trafic Internet en 2012 – $0.5 \cdot 10^{21}$**
- 66 zetta: l'information visuelle envoyée au cerveau en une année)

L'univers digital double tous les 18 mois

Source: Cisco Visual Networking Index – Forecast, 2007-2011 - Via Michael Brodie

Le quantitatif : la pénétration

Nombre de connections haute vitesse pour 100 habitants

Capacité égale ou supérieure à 144 kbit/s

<http://www.google.com/publicdata>

http://www.google.com/publicdata/explore?ds=l6t8doc0rch3u_&ctype=l&strail=false&nselem=h&met_y=i132&hl=en&dl=en#ctype=m&strail=false&nselem=s&met_s=i132&scale_s=lin&ind_s=false&idim=country:fr&ifdim=country&hl=en&dl=en

Success stories sur le Web

Google: gestion des pages du Web

Facebook: informations personnelles et communautés

Wikipedia: dictionnaire

Amazon, eBay: catalogues de vente sur le Web

YouTube, Dailymotion : vidéos

Twitter: communication, news

Quel est leur point commun ?

Flickr: base de données de photos

iTunes, AresGalaxy, Kazaa, Emule, Batanga, BearShare, etc: musique en ligne

Myspace: pages Web

Gestion d'information sur le Web

Meetic: fiches individuelles

Wikileaks: secrets d'états

Zoom rapide sur la technologie :

1. Les bases de données centralisées
2. Le Web et le parallélisme

La gestion d'information « classique »

Un grand succès de l'informatique du 20ème siècle

- Industrie et recherche académique
- Le modèle relationnel : tableaux à deux dimensions sur des **serveurs centralisés**

Technologie des systèmes de gestion de bases de données

- *Médiateur intelligent*
entre utilisateur & machine
- calcul de prédicats du 1^{er} ordre
- Logique → Algèbre → Optimisation

Nom	Père	Sexe
Adam		M
Eve		F
Caïn	Adam	M
Abel	Adam	M
...		

Pourquoi ça marche?

Une nouvelle fonctionnalité ou l'amélioration d'une fonction existante

- Plus facile d'écrire des applications
- Bonne efficacité

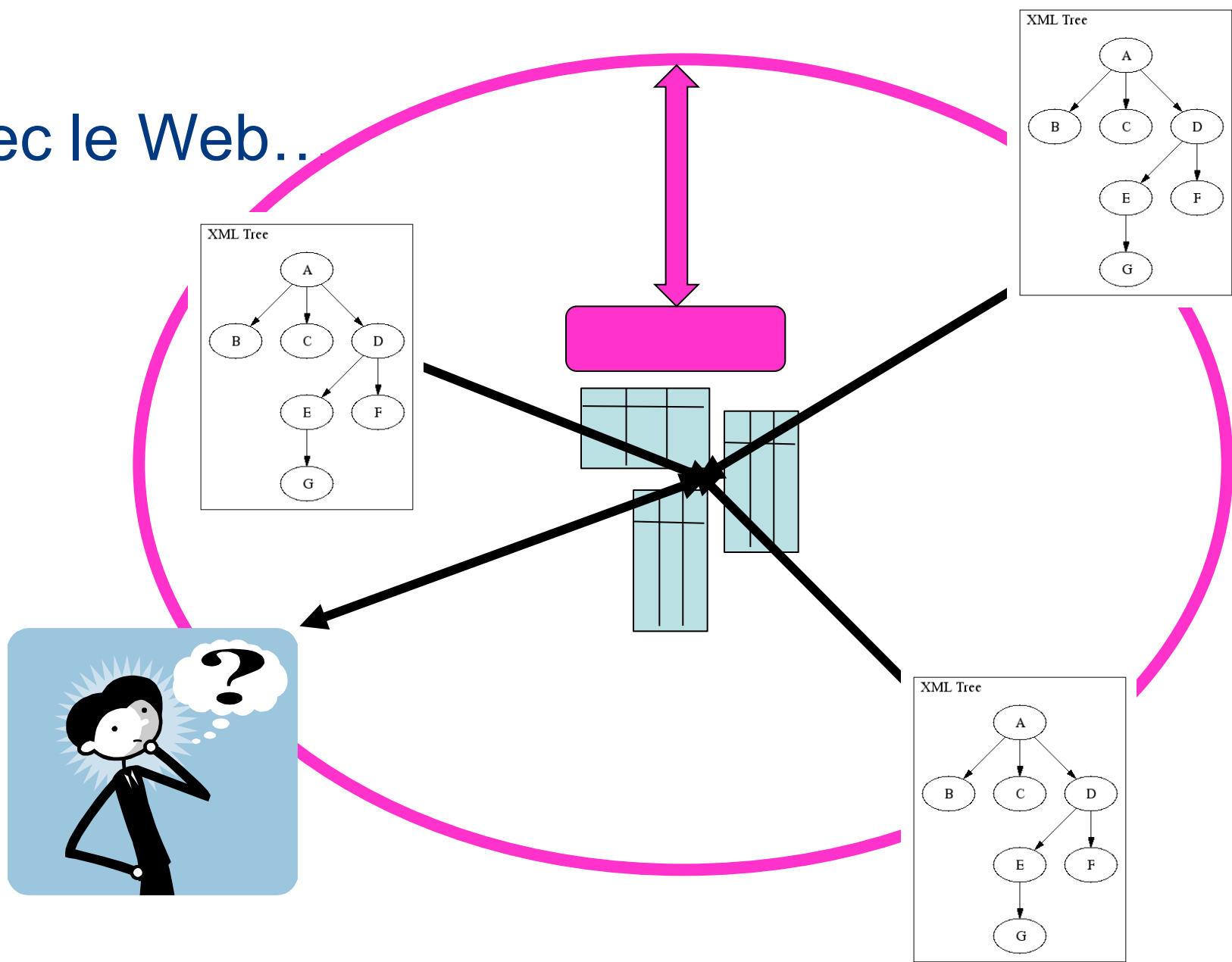
Des modèles mathématiques simples

Des algorithmes intelligents pour optimiser l'usage des ressources
(processeurs, *mémoire*, *disque*, réseau)

Un engineering solide

Et tout cela en tenant compte des progrès des machines

Avec le Web...



Ce qui a changé avec le Web

L'information résidait sur des îles avec des formats, des langages de programmation, des applications, des systèmes d'exploitations différents, maintenant

un accès uniforme et universel à l'information

Des standards universels pour échanger des information entre:

- Bases de données, systèmes de fichier, serveurs Web, PC, téléphones intelligents, senseurs, objets communicants, domotique, voitures, jouets, machines outils, etc.
- N'importe quel objet ou logiciel connecté au réseau avec de l'information à partager

Changement de dimensions: Millions de serveurs, téraoctets de données, milliards d'objets communicants

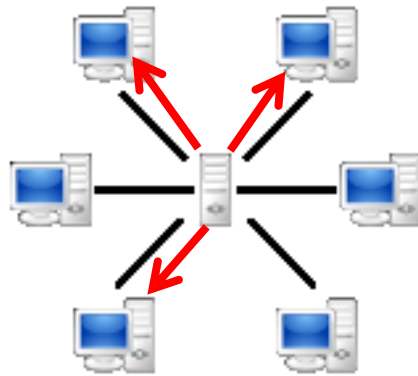
Et surtout parallélisme entre les machines

Pair-à-pair et parallélisme

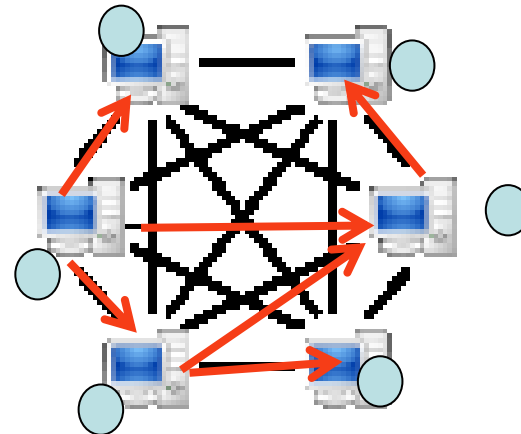
Utiliser les ressources du « réseau » sans passer forcément par des serveurs

- Communication: Skype
- Calcul: seti@home
- Stockage: emule

Puissa



e: (



nance, etc.

Exemple: Indexeur du Web

Gérer un index qui pour chaque mot donne la liste des pages qui le contiennent

Mot	Numéro de page	num	url
...		1	www.inria.fr
Sèvres	34,56,223,9900,111111...	2	www.bnf.com
...		3	www.inria.fr/~bhe
Ville	56,778,6560,9900,9999...	4	www.inria.fr/a/b
...			...
Zèbre	9890,11122290...		
...			

- Des dizaines de milliards de requêtes de recherche par mois
- Des milliards de pages
- L'index est du même ordre de grandeur que les pages indexées

Exemple: Indexeur du Web (2)

Choisir quelles pages montrer d'abord – Pagerank

- PageRank d'une page: probabilité pour qu'un surfeur aléatoire se retrouve sur une page
- Probabilité est plus forte pour le monde.fr que pour la page Web de Madame Michu
- Donc la page du monde est plus importante et on la montre d'abord
- Calculé avec des produits de matrices gigantesques

Solution Google

- Parallélisme et mémoire
- Des fermes avec des milliers de machines
- « truc »: mettre l'index en mémoire (la mémoire est plus rapide)

Digression – Comment on y est arrivé

Une nouvelle fonctionnalité ou l'amélioration d'une fonction existante
(aussi la baisse de prix des machines et surtout des mémoires)

Des modèles mathématiques simples

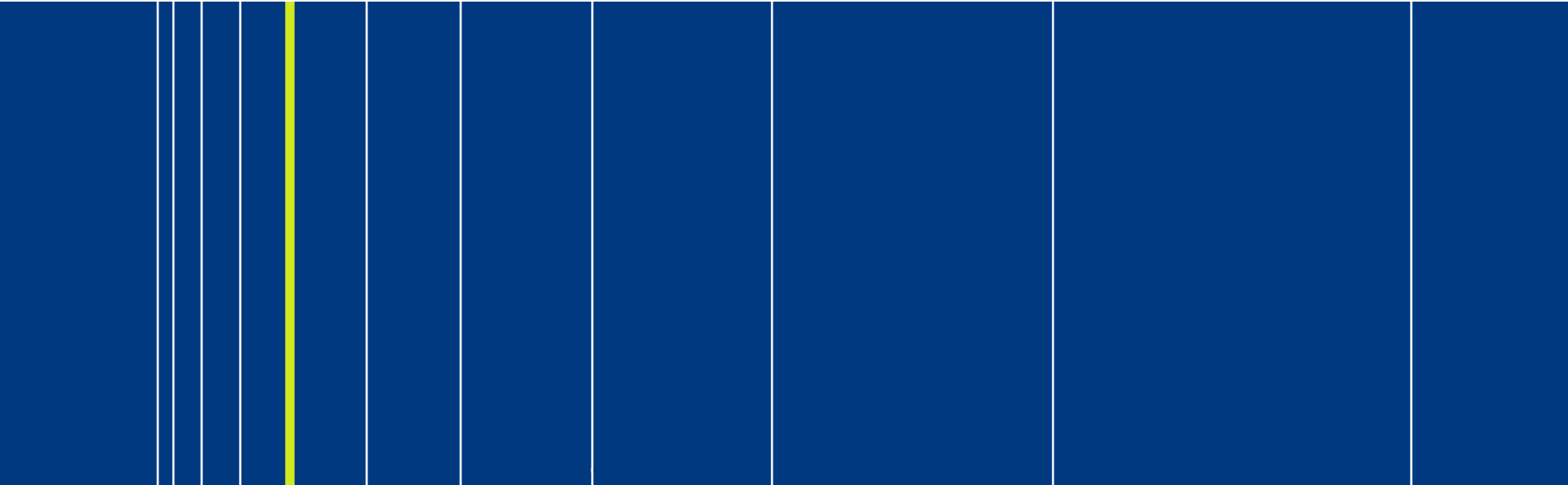
Des algorithmes intelligents pour optimiser l'usage des ressources
(processeurs, **mémoire**, disque, **réseau**)

Un engineering solide

Et tout cela en tenant compte des progrès sur les machines

Même chose que pour le modèle relationnel

Tendances





Les enfants animateurs ont besoin de vous



publicité
(comme sur le Web 😊)

Tendance: Web sémantique

Publier des connaissances au lieu de publier du texte

Un programme peut utiliser ces connaissances

- Aider l'utilisateur
- On peut poser des questions précises et obtenir des réponses précises

Problèmes durs

- Obtenir ces connaissances
- Extraction: linguistique et multilinguisme

Gordon Brown, 2010 : *Underpinning the digital transformation that we are likely to see over the coming decade is the creation of the next generation of the web - what is called the **semantic web**, or the web of linked data”*

Tendance – Web 2.0 (juste du buzz ?)

Le Web n'est pas juste fait pour obtenir des données

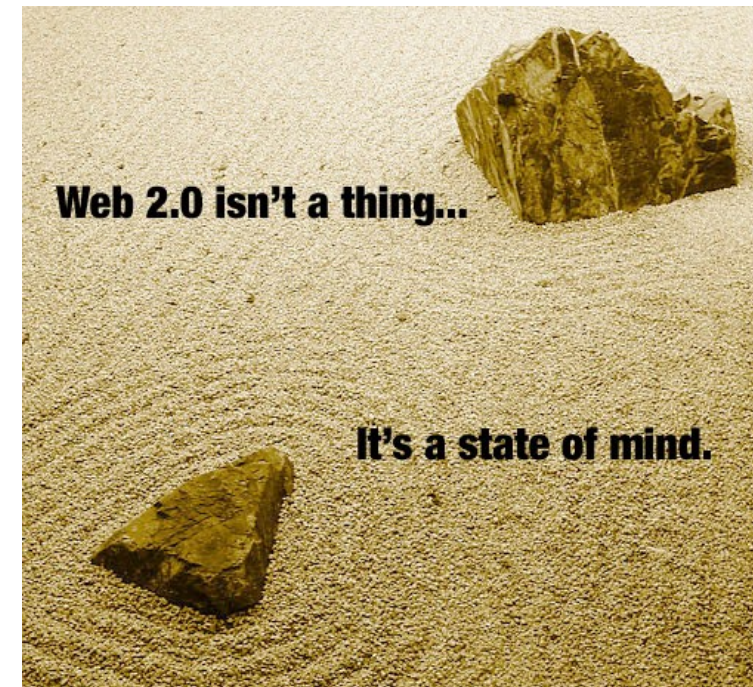
Mot clé: interaction, communauté

Pub: Wiki, Wikipedia, Twitter

Développement d'applications:
mashups

Multimédia

- musique, photos, vidéos...



Tendance – Web 2.0 – les réseaux sociaux

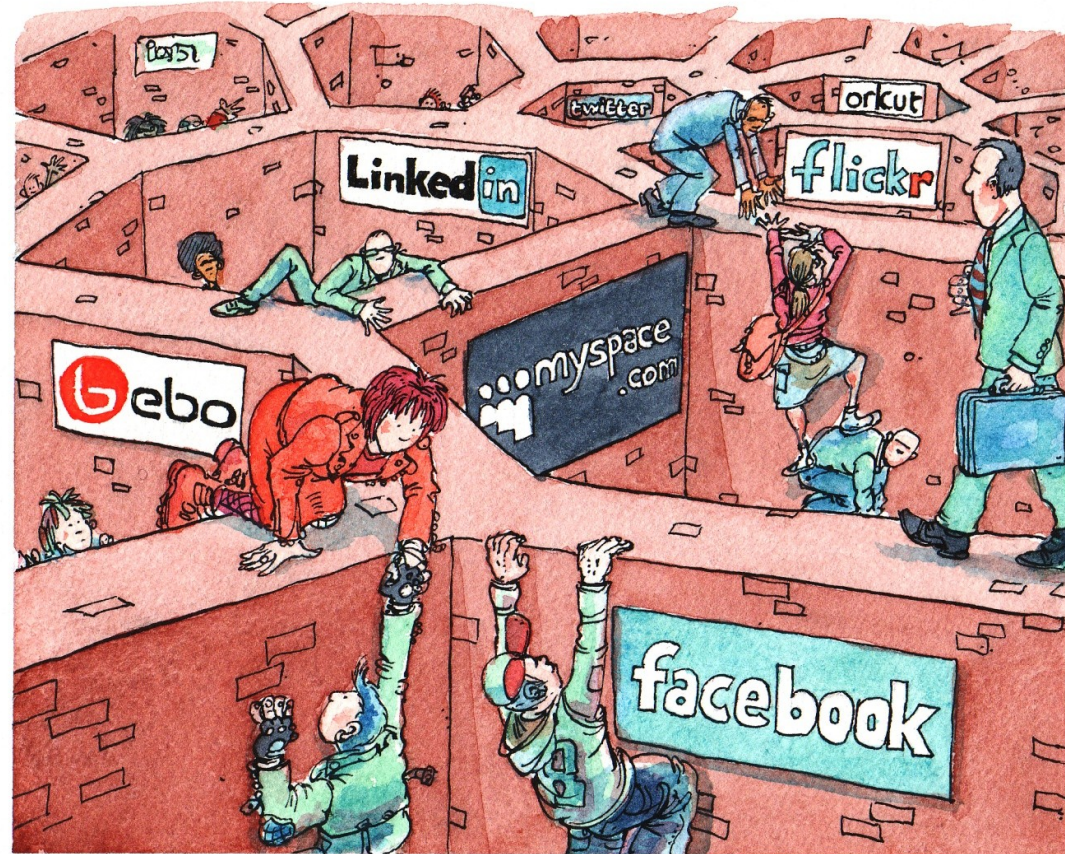
Partage & échange
d'information

Les jeunes se sont
appropriés le Web
avec Facebook

Les entreprises...

Les associations...

Les séniors...



Tendance – et d'autres

Web des smart phones & mobilité

- Iphone
- Un peu en confrontation avec le « world wide web »
- Les applis payantes deviennent la norme

Web des objets et intelligence ambiante

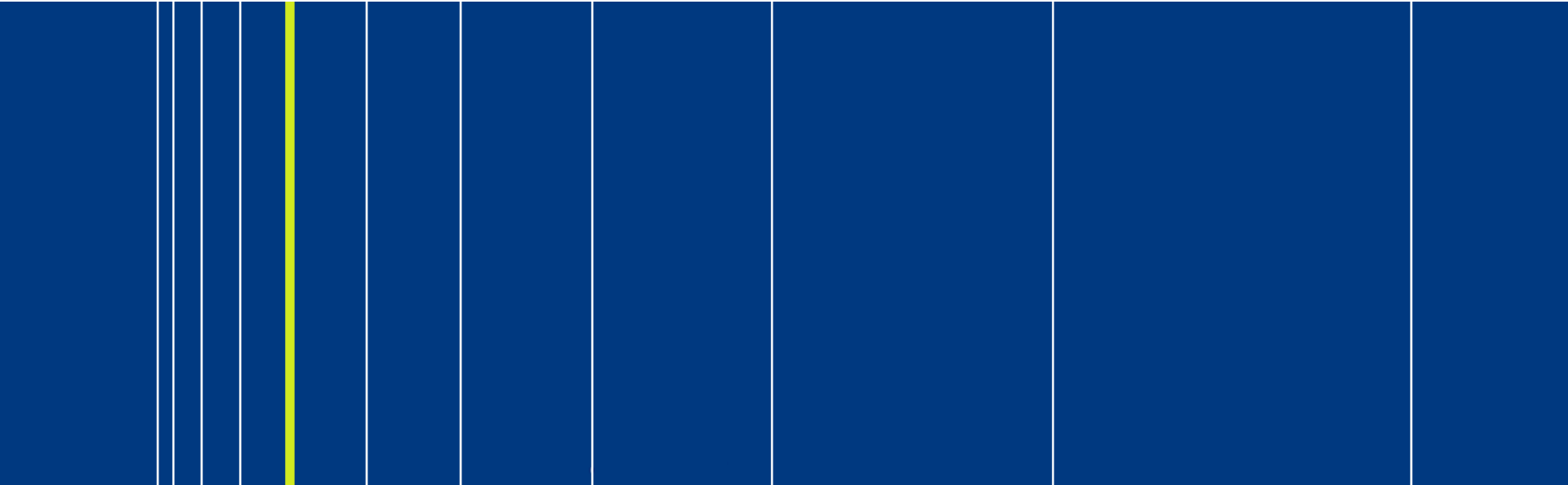
- De plus en plus d'objets communicants
- Objets avec des étiquettes genre RFID

Web des jeux, du 3D et des mondes virtuels

- World of Warcraft

Le Web pour quoi faire?

Promesses et écueil





Un projet : La fabrique Gaupillat à Meudon/Seine

publicité

Le Web : Pour quoi faire?

Plein de dimensions: Economie, santé, culture, gouvernement, sciences, écologie, etc.

Avec le Web

- « Ils » changent leur vie, changent le monde
- « Ils » veulent tout contrôler (Chine, Tunisie, US, France, etc.)
- « Ils » veulent d'autres modèles commerciaux (Apple, Téléphonie, etc.)
- « Ils » veulent des îles protégées à eux (Facebook, Apple, etc.)

Et ceux qui ne veulent pas du Web et ceux qui n'y ont pas accès

Pour un monde meilleur ou pire?

On va parler ici plus des écueils du Web que de ses succès

1. C'est normal, on parle moins de ce qui marche
2. On imaginerait mal de se passer d'Internet

Tentative d'organisation

On va parler de 3 sujets (si on a le temps)

- Toujours plus de connaissances ? Toujours trop ?
- Pour des utilisateurs meilleurs ou pires ?
- Anarchie, démocratie, contre-démocratie, dictature ?

Pour chaque thème

- J'introduits
- On s'empoigne
- Les organisateurs nous font passer au point suivant quand ils le souhaitent



Toujours plus de connaissances ? Toujours trop ?



Lire en numérique

2010 Michel Houellebecq : Oui, je pourrais résumer mon année comme ça : j'aurai passé la moitié de mon temps à me demander si j'allais acheter un iPad.



2006 Jean-Noël Jeanneney : Le livre ne va pas disparaître comme jadis l'araire. Car il a une personnalité que n'a pas l'écran, offre du plaisir au toucher, à l'odorat (d'où d'ailleurs l'échec des livres électroniques téléchargés). Et une commodité d'usage: on peut feuilleter un livre plus rapidement qu'un texte sur ordinateur.



(*) L'**araire** (du latin *aratrum*) est un instrument aratoire du IV^e millénaire av. J.-C. qui apparut en Mésopotamie. Tracté par un animal, il scarifie le sol, l'*arairage* qu'il effectue est donc superficiel. *Wikipédia*

Livres/journaux numériques

Les jours du papier sont comptés car le numérique

- C'est tellement mieux pour les journaux
- C'est tellement mieux pour les photos
- C'est tellement mieux pour les dictionnaires qui viennent avec
- C'est tellement plus pratique par exemple en vacances
- C'est tellement mieux pour les forêts
- Cela finira par coûter moins cher...

Seul point noir : plus compliqué de passer un livre à un copain

Le problème c'est le prix du livre...

Les bibliothèques étaient quasi gratuites, pourquoi pas la bibliothèque universelle ?

Pas seulement accéder aux connaissances Aussi publier des connaissances

Académique: on trouve quasi tout sur le Web

Encyclopédie: Wikipedia

Personnel: Pages Web, Blogs, Tweets, & tout ce qui reste à inventer

Révolution : ***n'importe qui peut publier***

des trucs qui intéressent le monde entier
 ou juste quelques potes
 ou peut-être personne...

Autres sujets autour de ça

Campus et laboratoires virtuels

- ≠ Campus d'Orsay & Syndrome BNF

Publication des données publiques

- Data.gov (US), data.gouv.uk (UK), data-publica.com (France)
- Fin 2010
- Rendre accessible les données publiques

Position	Annuaire	pays	url	Jeux de données
1.	data.gov	Etats-Unis	http://www.data.gov	307 100
2.	infochimps	Etats-Unis	http://infochimps.com	10 276
3.	data.gov.uk	Grande Bretagne	http://www.data.gov.uk	4 600
4.	data publica	France	http://www.data-publica.com	1 653
5.	ckan	Monde	http://www.ckan.net	1 564
6.	datamarket	Islande	http://datamarket.com	1 126
7.	data.govt.nz	Nouvelle Zélande	http://www.data.govt.nz	410
8.	stat.fi	Finlande	http://www.stat.fi/	279
9.	data.australia.gov	Australie	http://data.australia.gov.au/	190
10.	Nosdonnées.fr	France	http://www.nosdonnees.fr	13

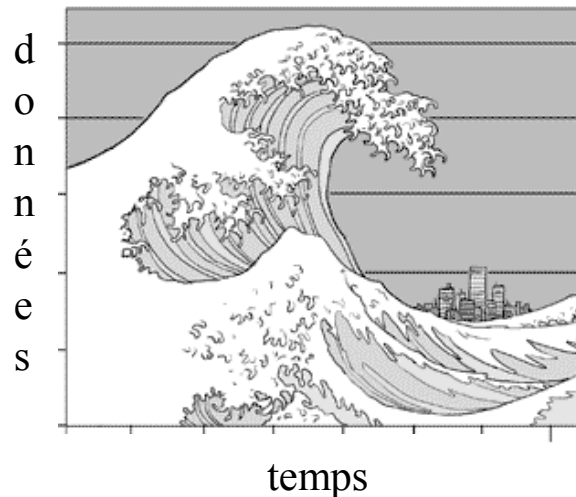
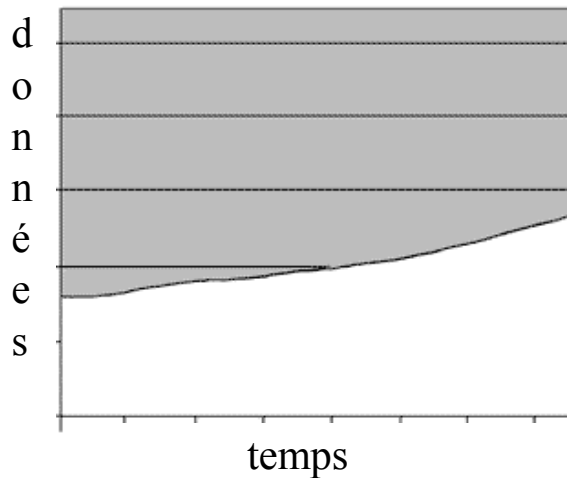
Un risque majeur : Se noyer sous un océan de données

De plus en plus facile de publier

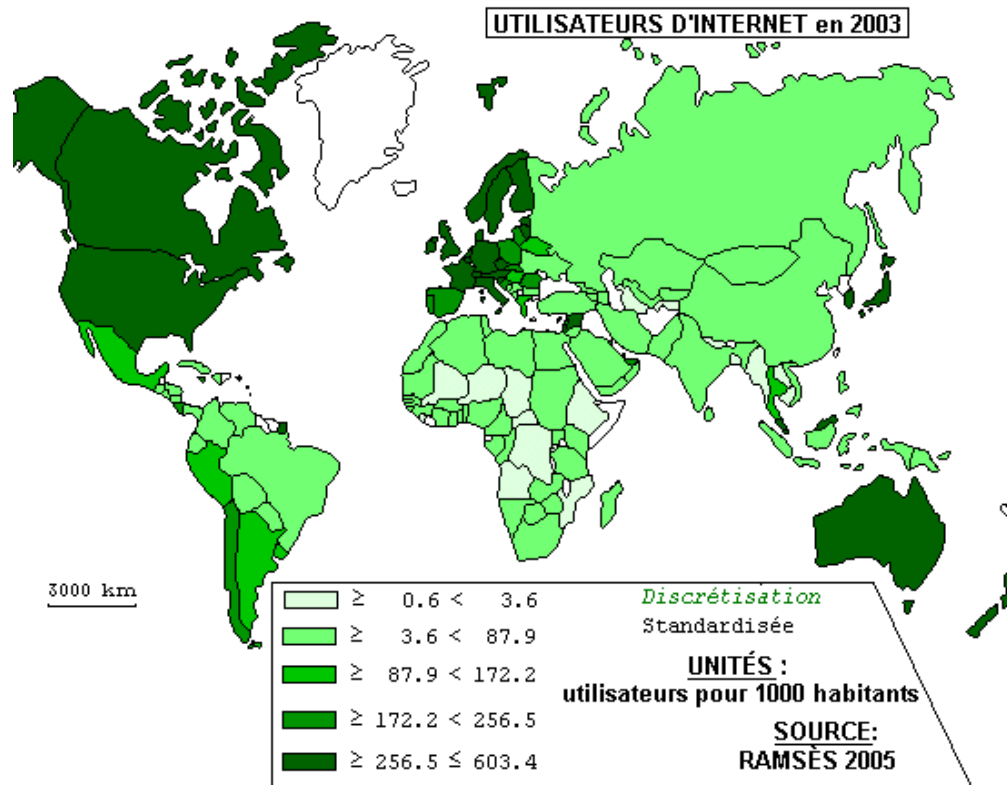
De plus en plus de données publiées

De plus en plus difficile de trouver l'information

De plus en plus difficile de l'avoir à temps



Fin des frontières ou fracture numérique



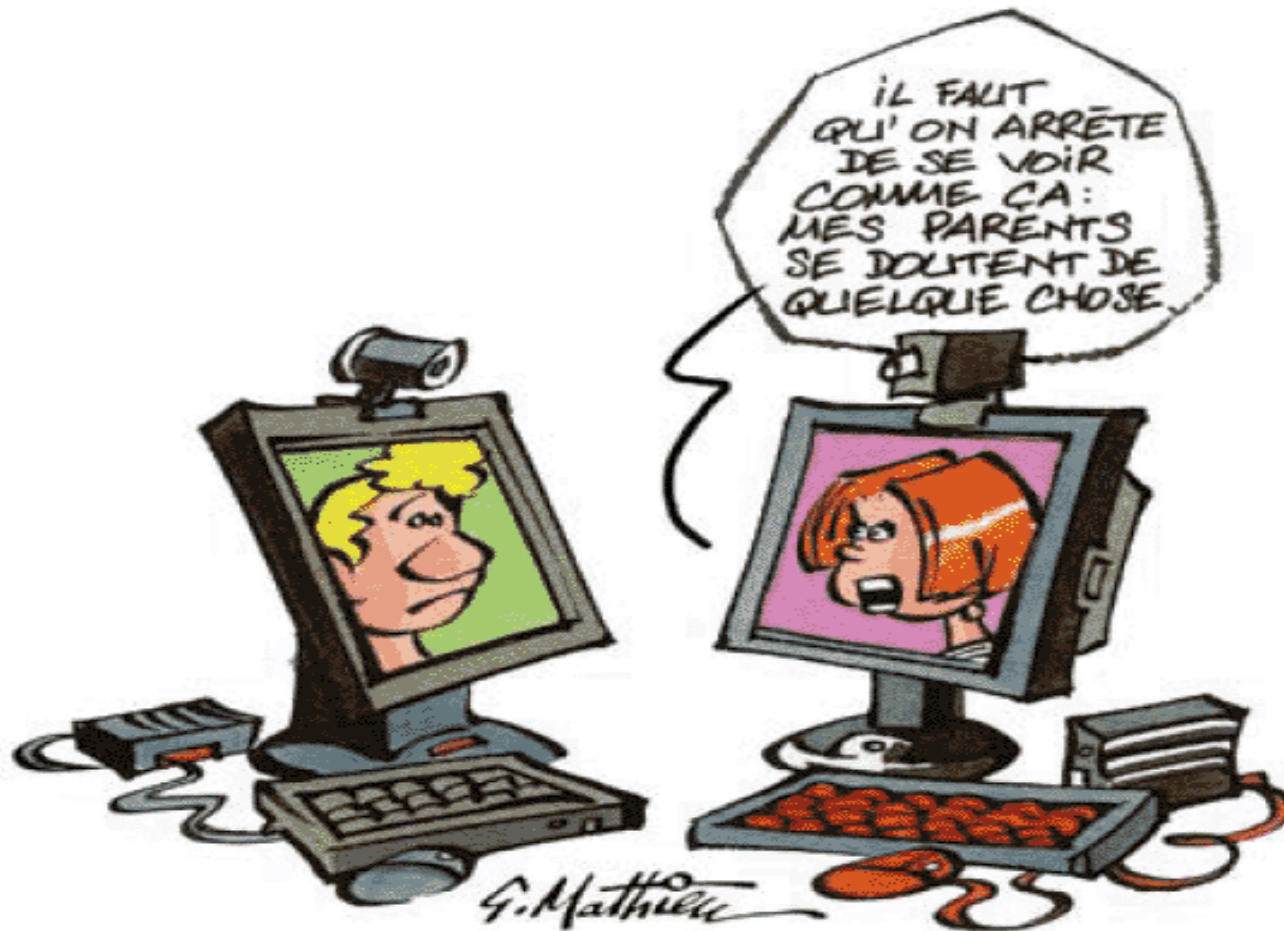
Nord/sud, Urbain/rural, Jeune/sénior

CREDOC 2009: En France, 40% de la population n'utilise jamais l'informatique

A vous ?



Pour des utilisateurs meilleurs ou pires ?



*Pour en savoir plus sur moi,
allez voir mon site Web*

Web et vie sociale

On passe notre vie devant un écran

- Les jeunes surtout mais bientôt les seniors
- On est vraiment plus productifs ?
- Y-a-il un risque de perdre contact avec la vraie vie ?
- Et si la vie virtuelle était mieux ? Où est le problème ?



Info globale – info ciblée

- On a accès à de plus en plus d'info
- On peut choisir de plus en plus finement ce qu'on lit
- On risque de ne plus parler qu'avec ceux qui partagent nos avis ?

Droit à une vie privé et droit à l'oubli

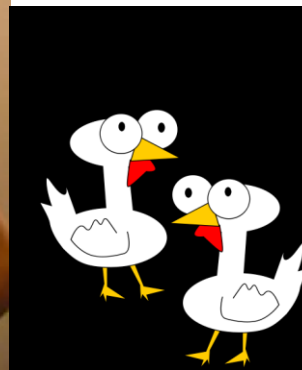
- On risque d'abandonner tout contrôle sur la vie privée ?
- Plus de vie privée, c'est acceptable si on le sait... ou pas ?

On pense mieux ou on ne pense plus

2007 Michel Serre: Les nouvelles technologies nous ont condamnés à devenir intelligents !

Is Google making us Stoopid ?

- E.g., Bonus à la popularité et pagerank



A vous ?



Les dossiers du
Canard
 enchaîné

"Je te vois !"

**Filés
 Fichés
 Fliqués**

**Comment nous
 sommes tous
 sous surveillance**

Anarchie,
 démocratie,
 contre-démocratie,
 dictature ?



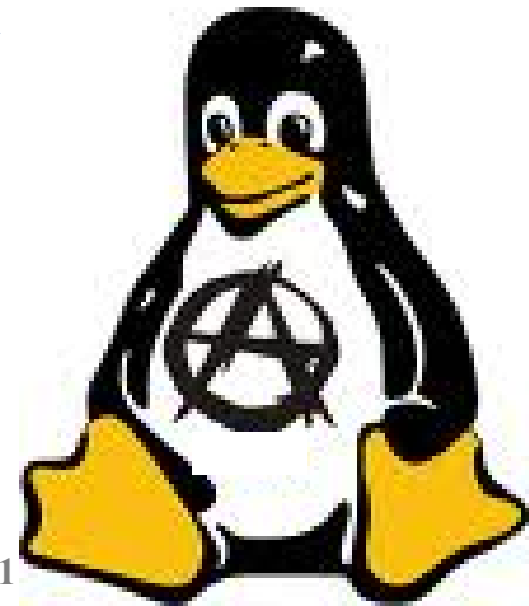
Anarchie et liberté

John Perry Barlow, 1990: *Gouvernements du monde industrialisé, géants fatigués de chair et d'acier, je viens du cyberspace, nouvelle demeure de l'esprit. Au nom de l'avenir, je vous demande, à vous qui appartenez au passé, de nous laisser tranquilles. Vous n'êtes pas les bienvenus parmi nous. Vous n'avez aucun droit de souveraineté là où nous nous rencontrons.*

Laboratoires:

les communautés de logiciel libre

Wikipedia



Démocratie traditionnelle

Le Web est un super outil militant pour structurer, organiser, échanger des informations, débattre, réduire la distance entre la direction et la base

Activité militante

- De Barak Obama ou Ségolène Royal
- Révolution Facebook en Tunisie (plutôt que Jasmin)

Concentration des influences

- La distribution des liens suit une loi de puissance: un petit nombre de sites reçoivent presque toutes les pointeurs

Contre-démocratie (ou meilleure démocratie)

Réseaux de militants très individualistes et actifs et moins structurés que dans les parties traditionnels

Comme les mouvements altermondialistes

Rôles: ***contrôle, surveillance, dénonciation, notation***

Nouvelles formes de démocratie

Des citoyens qui ne se contentent pas d'élire des représentants

Des citoyens qui veulent vraiment participer aux affaires de la cité

C'est la démocratie qu'on veut ?

Cela ne se passe pas sans combats

Les guerres du Web

Wikileaks : publication de “secrets d'états”

Attaques légales ou pas contre Wikileaks

Réaction d'internautes: données disponibles, attaques contre ceux qui nuisent à Wikileaks: Amazon, Visa et une banque qui avait gelé le compte de Wikileaks



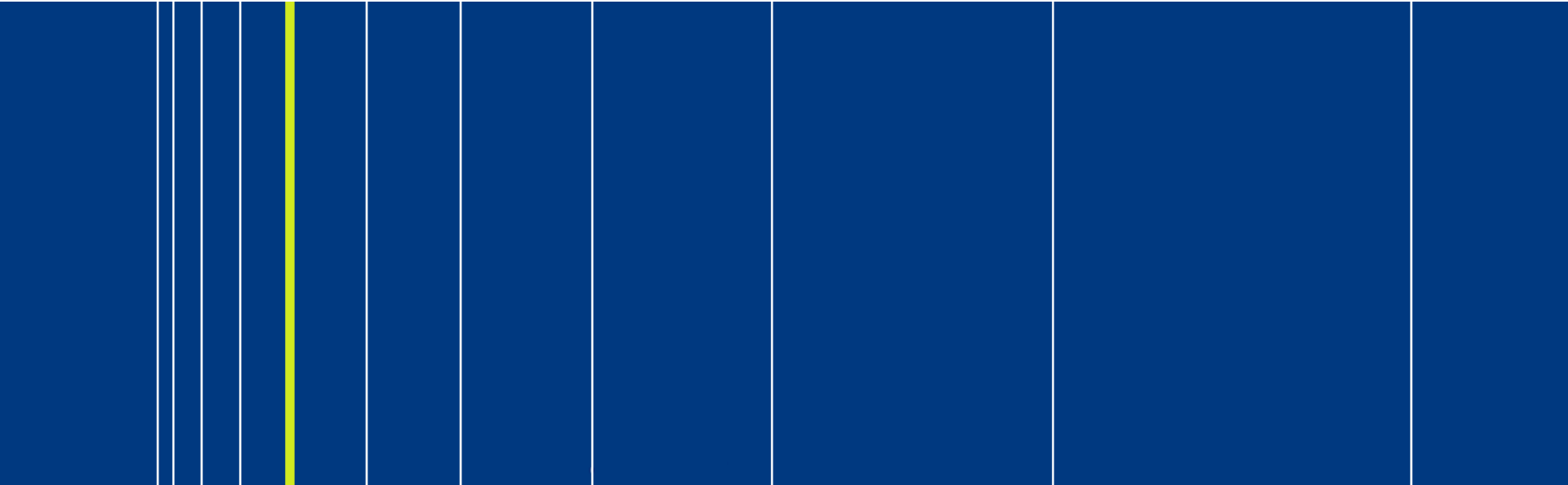
Pour ou contre

- ce n'est pas si simple – tout publier?
- pénalise surtout les gouvernements démocratiques

A vous ?



Conclusion



Prévisions - Technologie

Sur le Web, soyons modestes !

Ça va ralentir – on a fait le plus facile & on arrive aux sujets durs

- Linguistique, gestion de connaissances
- Vérification et qualité de service

Ça va continuer à bouger – créativité humaine

- Web sémantique : Web de connaissances plutôt que de texte
- Web 2.0 : un Web plus interactif, plus communautaire
- Web des objets, du pair-a-pair, des téléphones intelligents...

Rappel : pour le Web, on a toujours sous-estimé la créativité humaine

Prévisions – Société

Quand à prédire ce que nous
allons en faire...

Sans moi...

Rappel : on a souvent
sous-estimé les risques des
sciences



Quelques pointeurs

Serge Abiteboul, <http://www-rocq.inria.fr/~abitebou/news.html>

Site Web interstices.info, découvrir la recherche en informatique

Gérard Berry, Collège de France <http://www.college-de-france.fr>

Tim O'Reilly, *What Is Web 2.0*, <http://oreilly.com/>

Jean-Pierre Corniou, blog, <http://jeanpierrecorniou.typepad.com/>

Peter Denning, *Au delà de la pensée informatique*

Tim Berners-Lee, *Weaving The Web*

David Harel, Yishai A. Feldman, *Algorithmics: the spirit of computing*

Jean-Paul Delahaye, *Pour la Science*, <http://www.pourlascience.fr>

