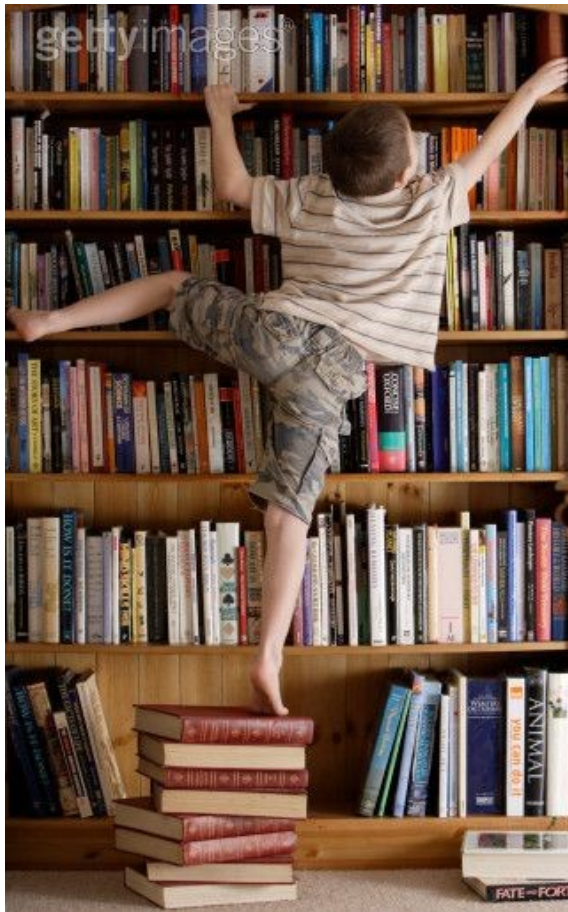


# Le Web sémantique, une méta-utopie ?



Web 3.0 ?  
Web des objets ?  
Web du futur ?

Jean DEMARTINI  
PhD, PMP



# Introduction

- On va parler du Web et s'intéresser surtout à son contenu.
  - On ne s'intéresse pas à toute la machinerie qui permet de le faire fonctionner.
  - On ne s'intéresse qu'à l'usage qu'on en a ou qu'on peut en avoir.
- Une petite révision pour vérifier qu'on parle, en gros, de la même chose !
  - en langage commun, pas de jargon,
  - des idées (assez) bien connues au début.
- On se posera ensuite quelques questions dérangeantes.

# Quel est l'enjeu ?

- Le Web est **pervasif** !
  - persavif (franglais) : *ce qui s'infiltré irrésistiblement partout, même à notre insu. Comme de l'eau !*
  - Personne ne peut (n'a envie d') y échapper :
    - déclarer ses impôts ... et même les payer,
    - faire ses achats,
    - obtenir une information administrative,
    - organiser son travail, ses loisirs,
    - écouter de la musique, regarder des vidéos, etc.
- Le Web est en train de devenir notre boussole obligée,
  - pourvu qu'elle indique bien le nord !

# Le Web ?

- Une **énorme quantité** (plus de 100 milliards) de documents enregistrés quelque part dans le monde et accessibles par un simple clic :
  - certains contiennent des choses directement **interprétables par des humains** :
    - textes
    - images (Flickr, ...)
    - vidéos (Youtube, Dailymotion, Zattoo, ...)
    - musique (Last.fm, ...)
  - d'autres contiennent, en plus, des **références** à d'autres documents : **documents hypertexte**.
    - références : citations, digressions, explications, illustrations ...

# Document Hypertexte

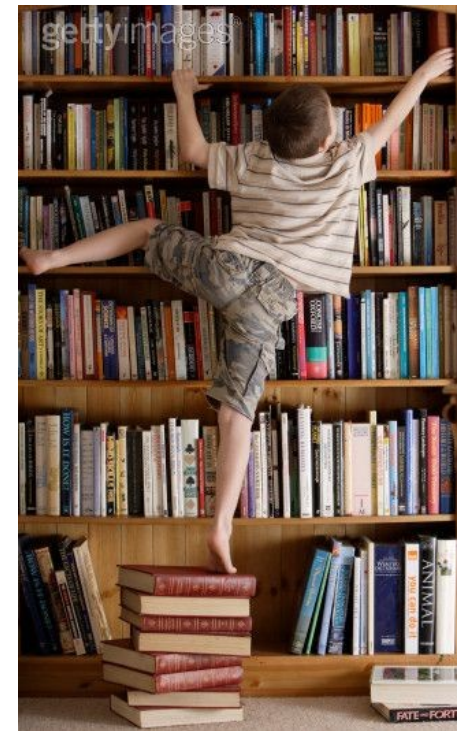
- Document qui fait référence à d'autres documents.
  - Une référence s'appelle un **lien hypertexte**.
  - Chaque document a un **identifiant\* unique**.
- Quand on active un lien hypertexte, le document référencé s'affiche ...
  - web 1.0 :
    - en s'insérant dans le document affiché ou
    - en remplaçant le document affiché ou
    - à coté du document affiché
  - web 2.0 :
    - en remplaçant une portion du document affiché

# Le Web est chaotique !

- L'ensemble des documents accessibles par le Web n'est structuré que par les liens hypertexte que les rédacteurs ont placés dans les documents.
  - et chacun peut être aussi bien rédacteur que lecteur.
- Chaque lecteur structure son propre monde en mémorisant les **marques-page** qui l'intéressent.
  - en fait, en créant un document qui contient toutes les références qui l'intéressent.
    - enregistré sur sa propre machine ... mais pas forcément !
    - éventuellement accessibles par le Web.

# Alors comment s'y retrouver ?

- Par certains aspects, le Web ressemble à une gigantesque bibliothèque. Mais :
  - Qui publie les livres ?
  - Où sont les rayonnages ?
  - Où est le bibliothécaire ?
  - Où est le catalogue ?
- Et bien, le Web n'a ni éditeur, ni rayonnage, ni catalogue, ni bibliothécaire, ni contrôleur.
  - C'est bien ce que irrite certains !
    - le pouvoir de l'information ?



# Sémantique ?

- **Sémantique** : étude du **sens** qu'on associe aux différentes formes d'expressions constituant un langage.
  - Un document ne nous intéresse que s'il a du sens.
  - L'association entre une forme d'expression et un sens est une simple convention culturelle.
    - Le sens des mots n'est pas totalement défini, il évolue dans le temps, n'est pas tout à fait le même partout, ni pour tout le monde. Est-ce gênant ?
- Quand on organise des documents, les classe-t-on par leur sens ou par leur forme ?
  - Plutôt par leur forme, non ?

# Organiser par le sens ?

- Cette question a-t-elle un sens ?
  - Classer des mots peut suffire à classer leur sens.
    - "chaise", "vache", ... : OK !
    - "vérité", "fidélité", ... : Hmm ?
  - Peut-on déduire le sens d'une phrase uniquement à partir de celui des mots qu'elle contient ?
    - " Le saumon *remonte* la rivière pour aller pondre. " : OK !?
    - " Ainsi, nous trouvons que toutes nos idées simples et impressions simples se ressemblent les unes les autres ; et comme les idées complexes et impressions complexes sont formées à partir d'elles, nous pouvons affirmer en général que ces deux espèces de perceptions se correspondent exactement. " (David Hume) : Euh ??

# Oui, mais le sens vu par qui ?

- Lorsqu'un auteur crée et publie un document, il lui a donné du sens.
  - On peut au moins l'espérer !
- Lorsqu'un lecteur consulte un document, il lui trouve un certain sens,
  - après plus ou moins de travail.
- Mais, est-ce bien le même ?
- Si on organise des documents pour en faciliter l'accès, vaut-il mieux le faire :
  - selon le point de vue de l'auteur ?
  - ou selon le point de vue du lecteur ?

# Web + Sémantique = ?

- Le Web pose un problème nouveau : le nombre des documents accessibles est tellement grand qu'on n'arrive même pas à imaginer à quel point il est grand.
- Ce nombre écrase toutes les méthodes de classement auxquelles on est habitué.
- Faut-il, qui peut, qui doit construire le catalogue ou l'index du Web ?
- Comment construire un catalogue, un index ?
  - Capable de fonctionner à une telle échelle.

# Stratégies jacobines\*

- Pour construire un catalogue :
  - un **ensemble d'experts** définissent les **catégories officielles autorisées** à décrire le monde.
  - les **auteurs** placent leurs documents dans ce système de catégories.
    - si c'est mal fait, on sait qui est responsable !
- Pour construire un index :
  - un ensemble d'experts établissent la liste des **mots-clé officiels autorisés** à décrire le monde.
  - les **auteurs** placent ces mots-clé dans une annexe de chacun de leurs documents.
    - l'utilisation d'un mauvais mot-clé condamne à l'invisibilité.

# Stratégies anarchistes\*

- Pour construire un index :
  - chaque lecteur construit **son propre index** en définissant **ses propres mots-clé** et en les associant aux documents qui l'intéressent.
  - chaque lecteur a la possibilité de rendre **public** son index personnel.
- L'index global se construit donc tout seul par la réunion des index de chaque lecteur.
  - uniquement par l'usage qu'on a du Web.
- On ne cherche pas construire un système de catégories universelles !

# Les experts !



- Un regroupement de gens très compétents, qui font un très bon travail technique.
- Le W3C produit des standards **libres** et **ouverts**.
  - Tous les documents officiels sont **publics** et **gratuits**.
- Ce n'est cependant pas une raison pour ne pas réfléchir !

# Le W3C et le Web sémantique ...

- Une constatation immédiate :
  - On ne peut pas concevoir une machine universelle capable d'extraire le sens qu'on peut associer à n'importe quel document produit pour des humains.
    - Déjà que les humains ont de la peine à se comprendre entre eux.
- Comment contourner ce problème :
  - Demander à chaque auteur (un humain) de rédiger une annexe qui décrit le sens (pour les humains) qu'il associe à son document (produit pour des humains).
  - Ça tourne en rond, non?

# Un langage pour décrire le sens ?

- Ce langage ne peut donc pas être une langue naturelle,
  - une de celles qui servent à rédiger les documents eux-même.
- C'est alors un langage artificiel qui doit :
  - être universel pour que toutes les machines puissent parler le même,
  - être simple et rigoureux pour pouvoir être interprété mécaniquement,
  - être logique pour permettre des déductions,
  - assez riche pour pouvoir parler des choses dont on parle avec les langages naturels.

# Un langage de Métadonnées

- Ce langage va utiliser des **métadonnées** et permettre de construire des expressions à partir de ces métadonnées.
- La **nécessité des métadonnées** est admise confusément depuis que les hommes se parlent. Il faut pouvoir :
  - dire concrètement ce qu'on a à dire,
  - mais aussi :
    - dire comment on peut dire ce qu'on a à dire,
    - dire comment on peut dire comment on peut dire ce qu'on a à dire,
    - etc.



données



métadonnées

# Un langage standard

- Pour être universel, ce langage doit être **standard**. Le W3C est chargé de le définir ainsi que tous les outils de travail associés.
- Et ce sera très bien fait !

**Mais est-ce une si bonne idée ?**

- Ce langage doit :
  - tenir compte de toutes les cultures,
    - ou conduire à éliminer les cultures minoritaires.
  - être capable évoluer,
    - ou conduire à interdire les changements techniques, culturels, moraux ou spirituels non planifiés.

# Le Web Sémantique selon le W3C

- C'est une solution technique très bien conçue.
- Cependant, elle ne peut fonctionner que si on l'applique à un domaine fermé, bien défini, assez restreint.
  - un domaine où les auteurs ont intérêt à être honnêtes !
- Elle trop facile à falsifier si on l'applique à un domaine ouvert, mal défini, trop large.
  - un domaine où les auteurs n'ont pas forcément intérêt à être honnêtes, car le sens attribué à un document peut être une forme de publicité ou de propagande.
  - pensez au produit médiocre que la commission européenne autorise à nommer "chocolat".

# Finalemment ...

- Pour aider les humains à communiquer entre eux, **la solution anarchiste** présente deux avantages importants :
  - elle privilégie le lecteur,
  - elle est difficile à falsifier.
    - à moins d'un complot mondial.
- Mais fonctionne-t-elle bien ? Oh oui !



# Le plan B

- Quel dommage de ne pas utiliser le système sémantique du W3C, il est si bien conçu !
- S'il ne semble pas souhaitable de l'utiliser sans restrictions pour aider à la communication entre les humains,
- par contre, il sera très utile pour aider à la communication entre les machines !
  - Le langage des machines est très simple.
  - Les machines ne sont pas malicieuses.
  - Et de plus en plus de machines sont amenées à communiquer entre elles.

# Le Web Sémantique selon l'ETSI



**European Telecommunications Standards Institute**

- Un autre groupe d'experts s'est chargé d'harmoniser les communications entre des machines : M2M (Machine to Machine)
  - car les machines elles-même ont besoin d'utiliser le Web pour communiquer :
    - mesures météo, mesures d'impact environnemental publiées tout autour du monde, ...

# Standards libres

- On peut critiquer le standard du W3C en cours d'élaboration ... parce qu'il est **libre** !
- En le libérant, le W3C permet à chacun de l'examiner et accepte la critique.
- Un standard libre n'est pas forcément bon mais il permet le débat public et l'amélioration.
- Que penser d'un standard qui ne dévoile ce qu'il est réellement qu'aux membres d'une communauté très fermée ?
  - et pas forcément altruiste.

# Colophon

- Ce document a été produit en utilisant l'application libre et gratuite **OpenOffice**.
- Il est publié sous licence GFDL (GNU Free Documentation Licence) :

Cher lecteur, vous avez le droit de faire ce que vous voulez de ce document, sauf empêcher un autre lecteur de faire ce qu'il veut de ce document, sauf empêcher un autre lecteur de faire ce qu'il veut de ce document, sauf empêcher un autre lecteur de faire ce qu'il veut de ce document, sauf empêcher un autre lecteur de faire ce qu'il veut de ce document,

...

Merci pour votre attention.

Des questions ?