

# Recherche documentaire et recherche d'information

Trouver sur Internet

Organiser sa veille avec les flux  
RSS

*Pauline Jourdan – URFIST 2014*

# Historique

Du web au web 3.0

Les niveaux du web

# Le web

- 1989 – 1993 : Création du World Wide Web (Tim Berners-Lee)
- 1994 : Yahoo
- 1995 : Altavista
- 1998 : Google
- 2001 : Wikipédia

# Le web 2.0

- 2001 : Explosion de la bulle Internet
- Emergence des grandes innovations
- 2005 : Conférence web 2.0 – Tim O’Reilly (acteur majeur | Open Source)
- 2007 : Démocratisation (plus de 9,5 millions de citations sur Google)
  - Infobésité
  - Web des données (→ openData)

# Le web 3.0

- Tim Berners-Lee 2001 → Web 3.0
- L'ordinateur interprète les résultats et fait le tri avant nous.
- Aujourd'hui les moteurs de recherche lisent des caractères
  
- Une information = 1 triplet :
  - **Sujet** : la ressource à décrire
  - **Prédicat** : un type de propriété applicable à cette ressource
  - **Objet** : La valeur de la propriété

# Le web sémantique

- Inférence : conclusion à partir d'une règle de base

Sujet	Prédicat	Objet
Bill Gates	dirige	Microsoft
Bill Gates	Est né en	1950
Microsoft	Est basé à	Seattle
Seattle	Est dans l'état de	Washington

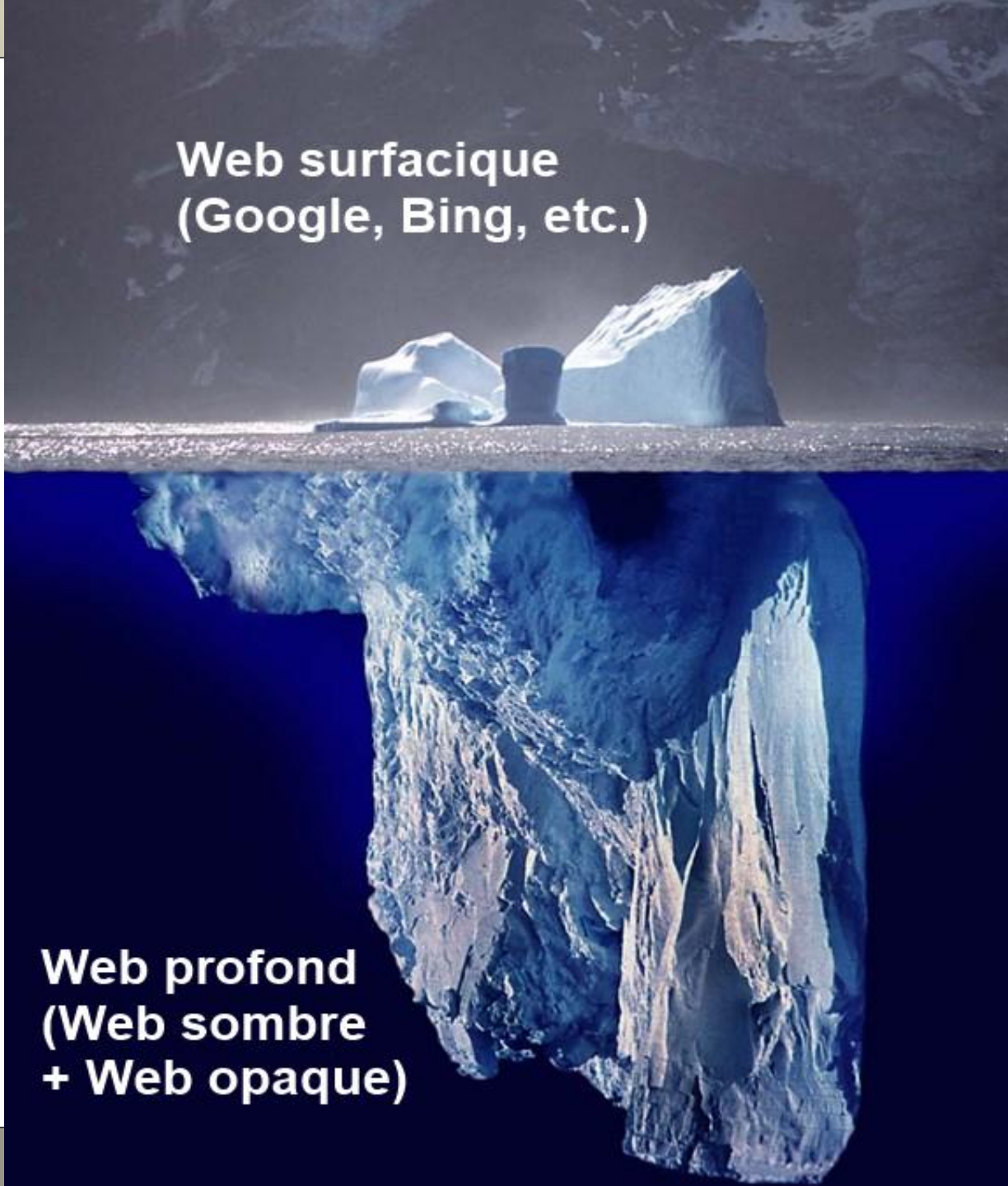
- Bill Gates dirige une entreprise dans l'état de Washington

# Les niveaux du web

- Le web surfacique
  - Indexé par les moteurs de recherche
- Le web profond (Deep web)
  - Web sombre (Darknet)
    - Souvent illégal
  - Web opaque
    - Potentiellement indexable
  - Web abyssal (légende ?)

**Web surfacique  
(Google, Bing, etc.)**

**Web profond  
(Web sombre  
+ Web opaque)**



# La recherche documentaire sur Internet

Les annuaires

Les moteurs de recherche

Les équations

La recherche avancée

# Les annuaires / répertoires de signets

- Sélection par des professionnels
- Nombre de ressources limité
  
- Les signets de la BNF : <http://signets.bnf.fr/>
- Open Directory Project : <http://www.dmoz.org/>
- Signets du CERIMES : <http://www.signets-universites.fr/>
- L'annuaire d'Aldébaran : <http://aldebaran.revues.org/index.html>

# Les moteurs de recherche

- Grande quantité d'information indexée
- Rapidité d'exécution
- Bruit informationnel, redondance
- Aucune assurance fiabilité
  
- Généralistes
  - Google, Yahoo, Bing, Qwant, Exalead

# Les moteurs de recherche

- Spécialisés

- <http://diderot.fr/> (lettres et sciences humaines)
- <http://scholar.google.fr/>

- Sémantiques

- <http://www.wolframalpha.com/>

# Les moteurs de recherche

- Web profond
  - <http://www.base-search.net/> (académique)
  - <https://www.torproject.org/> (darkweb)
- Métamoteurs :
  - <https://ixquick.com/fra/>

# Exercice

- Entrez les mots « Nelson Mandela » dans les différents moteurs de recherche :
- Google, Bing, Qwant, Yahoo, Wolfram Alpha, Exalead et Ixquick.
- Sur Google Scholar ? Base Search ?

# Les équations



Recherche Google

J'ai de la chance

Paramètres de recherche

Recherche avancée

Historique Web

Aide sur la recherche

Envoyer des commentaires

Confidentialité et conditions d'utilisation

Paramètres

Google.com

→ Filtrer et affiner vos résultats de recherche

# A mémoriser

- « expression » : expression exacte
- Site: rechercher sur un site ou un domaine (ex : site:gouv.fr)
- \* : remplace un mot dans une expression (exacte ou non)
- Expression1 OR expression2 : une des expressions
- -mot : mot est exclu de la recherche

# Recherche avancée



Recherche Google

J'ai de la chance

Paramètres de recherche

Recherche avancée

Historique Web

Aide sur la recherche

Envoyer des commentaires

# La recherche avancée

- Permet de personnaliser sa requête :
  - Expression
  - Langue
  - Localisation
  - Date
  - Site / domaine
  - Type de fichier (PDF, MP3, DOC, etc.)
  - Licence d'utilisation

Vérifier ses sources

# Vérifier ses sources

- Qui : l'auteur
- Quoi : qualité du document, du site
- Quand : repérer la date
- Où : origine de l'information
- Pourquoi : objectifs
- Comment : présentation

# Vérifier ses sources

- Décortiquer l'URL :
  - <http://www.education.gouv.fr/cid24150/que-faire-apres-le-baccalaureat.html>
  - Protocole/domaine/dossier/page
- Extensions de domaine :
  - .fr, .org, .gouv.fr, .com, etc.
- Domaines connus :
  - ac-XXX, u-XXX, univ-XXX, chu-XXX, etc.
- Service *Whois* <http://www.domaintools.com/>

# L'auteur

- Qui est-il ?
- Lire « mentions légales » et « qui sommes nous »
- A défaut : whois
- Est-il reconnu dans le domaine ?
  - Autres publications...
- Organisation / association reconnue ?

# Désinformation, rumeur, etc.

- <http://www.legorafi.fr/>
- <http://www.hoaxbuster.com/>
- Sur Twitter on partage, mais on ne lit pas toujours (ZDNET) : <http://www.zdnet.fr/actualites/sur-twitter-on-partage-mais-on-ne-lit-pas-toujours-39788115.htm>
- Chef Raoni : <http://ecologie.blog.lemonde.fr/2013/08/22/hoax-ecolos-24-les-larmes-du-chef-raoni/>

# Désinformation, rumeur, etc.

- Créer un faux message Facebook :  
<http://thewallmachine.com/>
- Un faux tweet :  
<http://lemmetweetthatforyou.com/>
- Recherche inversée d'images :
  - <http://www.tineye.com/>
  - Google image

# Exercice

- Vérifier les différentes sources :
- <http://www.ogm.org/>
- <http://www.ogmdangers.org/>
- <http://guide-ogm.greenpeace.fr/>
- <http://ogm.gouv.fr/>
- <http://ogmenmedecine.over-blog.com/>

# Exercice

- Trouver une source fiable pour le sujet :
  - Projet HAARP
- Comparer :
  - <http://www.sante.gouv.fr/ivg>
  - <http://www.ivg.net/>
    - Lancer un Whois (<http://whois.domaintools.com/>)
    - Rechercher « Philippe Marie Sos détresse » sur Google

# Exercice

- Quels types de sites ?
  - <http://www.agoravox.fr/actualites/economie/article/l-argent-par-les-fenetres-147083>
  - [http://www.lemonde.fr/vous/video/2014/01/31/a-la-poste-les-usagers-impriment-en-3d\\_4357411\\_3238.html](http://www.lemonde.fr/vous/video/2014/01/31/a-la-poste-les-usagers-impriment-en-3d_4357411_3238.html)

- <http://www.economie.gouv.fr/economie/auto-bonus-malus-ecologique>
- <http://jlhuss.blog.lemonde.fr/>
- [http://forum.aufeminin.com/forum/societe2/ f14701\\_societe2-France-65-des-francais-rejettent-le-capitalisme.html](http://forum.aufeminin.com/forum/societe2/ f14701_societe2-France-65-des-francais-rejettent-le-capitalisme.html)

Organiser sa veille

# Deux méthodes

- **Méthode *Pull*** : on va chercher l'information.
  - Efficace mais chronophage
  
- **Méthode *Push*** : l'information arrive automatiquement.
  - Gain de temps mais risque de bruit informationnel.

# Outils

- Newsletters | lettre d'information
- Google Alertes
  - <http://www.google.fr/alerts?hl=fr>
  - La requête doit être suffisamment délimitée
- Flux RSS

# Les flux RSS

- Really Simple Syndication
  - Ressource mise à jour automatiquement (titres)
  - En lien avec une source
  - Métadonnées interprétées automatiquement
- Utilisation d'un agrégateur
  - Sur le web
  - Intégré (navigateur, client mail...)
  - Logiciel externe

# Exemples

- Netvibes : <http://www.netvibes.com/fr>
- Feedly : <http://feedly.com/index.html#welcome>
- Mozilla Firefox

- Repérer un flux :



- Ex Netvibes : <http://www.netvibes.com/ufr-stc#Accueil>
- [http://www.netvibes.com/mshlorraine#MSH\\_Lorraine](http://www.netvibes.com/mshlorraine#MSH_Lorraine)

# Exercice

- Créer un compte Netvibes
- Alimenter son Dashboard avec :
  - Au moins un flux RSS
  - Différentes sources
- Guide Netvibes form@doct Université européenne de Bretagne : <http://guides-formadoct.ueb.eu/content.php?pid=224615&sid=1862622>