



## **TER 2012 Master 2**

***spécialité Informatique à Finalité Professionnelle et Recherche Unifiée  
option Données, interaction et Web***

Faculté des sciences de Montpellier

Jamal Hamed - Fabien Priotto

***Concepts fondamentaux de l'Informatique pour l'étude de la Science du Web***

[PDF](#) [English version](#) [Presentation](#) [Sujet](#)

Version du 4 avril 2012 rectifiée suite au débat qui a suivi la présentation du 28 mars tirée du google document collaboratif.

**Travail d'Etude et de Recherche supervisé par le Pr. Stefano A. Cerri et le Dr. Clément Jonquet.**

- [Avant-propos](#)
- [Introduction](#)
- [Petite chronologie](#)
- [Appel à la création de la Science du Web](#)
- [Objectif de notre travail](#)
- [Des paradoxes](#)
  - [Un Web transformationnel](#)
  - [Avec une mémoire](#)
  - [Basé sur un mode de transport non connecté sans état](#)
    - [HTTP fonctionne en mode déconnecté](#)
  - [HTTP handicape l'évolution du Web et il faudra faire avec...](#)
    - [Il est difficile de coder des applications pour le WEB](#)
    - [HTTP est vulnérable](#)
  - [Les méthodes ignorées de HTTP](#)
    - [Ou l'on parle du facteur confiance](#)
  - [Un langage avec si peu de sémantique](#)
- [le Web 2.0](#)
  - [Des technologies](#)
    - [Des Web services](#)
      - [Quelques briques technologiques manquantes](#)
      - [Des services sans état pauvres en sémantique](#)
      - [Un paradoxe technologique: des méthodes d'accès qui s'opposent](#)
  - [Des applications](#)
    - [La centralisation serait-elle has-been ?](#)
  - [Le Cloud](#)
  - [Les moteurs de recherche](#)
    - [Les faiblesses de la recherche par mots-clé](#)
      - [Un cruel manque de la sémantique](#)
      - [Des lacunes](#)
      - [Des résultats inutiles](#)
      - [Des défaillances](#)
      - [Et un manque de confort](#)
  - [Folksonomies](#)
    - [Les inconvénients des folksonomies](#)
      - [Pas de référence commune](#)
      - [Pas de standard](#)
- [Des usages](#)
  - [L'immanence du Web](#)

- [Notre vision SMA pour le Web](#)
- [Des services dynamiques](#)
- [Des flots](#)
- [Des évaluations retardées](#)
- [Une description qualitative des agents](#)
- [Vers une navigation colorée](#)
- [Canaliser la confiance et l'adhésion](#)
- [Faire passer l'émotion](#)
- [En conclusion](#)
- [En résumé](#)
- [Références](#)
- [Annexes](#)
- [Vocabulaire](#)

*Nous remercions sincèrement Stefano Cerri et Clément Jonquet pour les précieux enseignements et le soutien qu'ils nous ont offerts.*

## Avant-propos

*Appuyés par la démarche de publication des données en plein essor, nous avons souhaité constituer une plate-forme d'observation (tri de logs, extraction, mesures) en choisissant des ressources de collaboration pour analyser les comportements et les communications d'un système opérationnel. Cette étude, nous aurait permis de faire une interprétation des phénomènes observés pour tenter d'établir une généralisation. Malheureusement, il n'a pas été possible d'obtenir comme nous l'espérons les logs de la plateforme EasyChair. Nous avons alors envisagé d'étudier les logs du serveur de listes de diffusion Mailman de la faculté des sciences de Montpellier. Notre analyse [annexe 1], montre que Mailman repose essentiellement sur SMTP.*

*Hors, proposé initialement pour l'acheminement de messages simples définis par le RFC 822<sup>1</sup>, SMTP n'est plus aujourd'hui le principal moyen pour la transmission de messages sur l'Internet. Sans avancer l'idée que SMTP sera bientôt obsolète, nous pensons que ce n'est pas là que nous trouverons de la matière pour notre objectif. Nous ne pouvons donc pas baser notre étude sur l'usage de Mailman centré sur SMTP et nous nous concentrons sur le challenge qui consiste à souligner les paradoxes, les briques manquantes et les concepts importants pour le Web.*

*En conséquence, le résultat de notre travail est essentiellement théorique et se fonde sur notre lecture d'une bibliographie traitant d'Informatique, d'Intelligence Artificielle et de technologie.*

---

<sup>1</sup> Internet Engineering Task Force 1982

# Introduction

Le 19 février 2009, les résultats de Google en termes de suivi planétaire pour la santé ont fait l'objet d'une publication [8] dans la revue spécialisée Nature. Cette publication révélait le fort potentiel des moteurs de recherche à estimer quasiment en temps réel la propagation du virus de la grippe. Pour Google, s'ouvrent de nouvelles perspectives qui découlent d'un usage non anticipé de leur outil de recherche sur le Web (Google Search). Aujourd'hui, il n'existe pas de formalisme commun permettant d'anticiper les évolutions du Web. Celui-ci n'évolue pas de façon conventionnelle. Ce sont les interactions entre les différents acteurs qui en pilotent la croissance.

## Petite chronologie

**1989 - 1992** HTML, HTTP et les URL pour écrire des documents hypertextuels [12]

**1993** L'université de L'Illinois présente Mosaic, le premier navigateur Web populaire

**1995** La guerre des navigateurs a commencé.

**1998** Naissance de Google-Stanford

**2000** Fin de la Dot-com bubble et effondrement des cours boursiers

**2002** Création des premiers réseaux sociaux sur le Web: Friendster et MySpace

**2004** Lancement de Facebook

**2005** Les participants de la conférence à la British Computer Society à Londres [1] examinent les tendances émergentes sur le Web et débattent de la recherche qui sera nécessaire pour exploiter les opportunités.

**2006** L'article "creating a Science of the Web" [2] dans la revue Science lance l'appel aux scientifiques pour "cibler le web comme un objectif principal de l'attention."

**2009** Le moteur de recherche Google s'avère utile en termes de suivi planétaire pour la santé. [8]

**2011** Facebook et Youtube jouent un rôle dans les révolutions arabes.

# Appel à la création de la Science du Web

Le Web constitue un partage de ressources (documents) et est très souvent confondu avec l'Internet qui est un monde dans lequel de nombreux autres services sont associés tels que le courrier électronique, le transfert de fichiers, la messagerie instantanée, l'accès à distance). Nous considérons le Web comme une application de l'Internet. Mais c'est une application exceptionnelle: il s'agit de la plus large construction de connaissances dans l'histoire de l'humanité. Totalement dépendant de l'activité de millions d'agents, qu'ils soient humains ou pas, répartis dans le monde entier, le Web se transforme au fil des ajouts de données, de la fréquentation des sites, de l'extension de services et de la propagation des nuisances.

Un phénomène transformationnel: depuis ses débuts, le Web change les modes de communication, de collaboration et les méthodes d'enseignement [18]. Parce qu'il joue un rôle dans les changements du monde, notamment dans les activités scientifiques, il est devenu le phénomène majeur de ce début de 21<sup>ème</sup> siècle et nous voulons le comprendre.

*"Comprendre le Web est un défi majeur..."*

N. Shadbolt Royal Society, 8 mars 2010

Observer les usages, les tendances et les phénomènes sur le Web.

Comment se fait sa croissance ? Quels sont les principes qui le régissent ?

Comment se transforme-t-il ? Comment modéliser le Web ?

Il faut tracer un programme de recherche qui cible le Web comme un objectif principal de l'attention. C'est le sens de l'appel [1] qui a été lancé depuis la British Computer Society à Londres en Septembre 2005: pour une convergence dans la recherche et la création d'une Science du Web. [1]

C'est de l'émergence de la Science du Web que dépendront les moyens mis en oeuvre pour comprendre des phénomènes importants tels que :

- L'impact du Web sur la façon dont les gens pensent, se comportent et interagissent.
- L'adaptation sociale du Web 2.0 pour qu'il respecte un jour les valeurs sociales (nous voulons un Web social sociable !) et son interpolation dans la théorie des systèmes complexes sociaux.

Voulons nous un outil stable, puissant mais voué à accomplir de simple tâches ou au contraire une intelligence artificielle collective ? Cette question est fondamentale: qu'attendons-nous du Web ? Quelles les intentions humaines pour le Web ?

C'est de cette réflexion que dépendra la valeur du Web en termes de bénéfice humain.

Oui, le Web est un phénomène qui doit être traité par une science dédiée. Nous pouvons en apprendre beaucoup. Il s'agit d'un système complexe qui doit être étudié et compris comme il

est important d'étudier le système solaire, l'atmosphère et comment tout cela fonctionne. Nous avons besoin de le penser comme une nouvelle discipline avec une meilleure convergence scientifique afin d'observer et d'étudier les éléments qui le composent (ressources humaines, artificielle, technologique), ainsi que leurs utilisations. Une mobilisation pour être en mesure d'anticiper sur les évolutions du Web, d'en identifier les opportunités et le protéger.

*“Une combinaison de synthèse, d’analyse et de gouvernance.”*

TBL & co. dans “A Framework for Web Science”

## Objectif de notre travail

Nous souhaitons répondre à cet appel avec l'objectif de mettre en évidence les concepts fondamentaux nécessaires pour étudier le Web comme une science. Cet objectif est motivé par la conviction que l'analyse du Web et de son évolution doit faire l'objet d'une science spécifique, interdisciplinaire par nature, pouvant faire appel à des domaines aussi différents que la sociologie, l'informatique ou les mathématiques.

Les enjeux sont d'ordre économiques mais aussi humains: Analyser les interactions en ligne, leurs impacts sur l'architecture du Web ; lutter contre la cybercriminalité ; contre l'addiction, répondre aux questions complexes de vie privée et de propriété intellectuelle. Mais aussi comment contrôler que les valeurs sociales (loyauté, confidentialité, respect des individus et des frontières) soient respectées sur le Web ? [1]

En nous intéressant aux éléments structurels du Web, nous chercherons à faire apparaître ses paradoxes et ses limites. Dans notre approche, nous distinguerons les agents (humains, logiciels, artificiels)<sup>2</sup> des ressources ainsi que leurs usages comme autant de briques qui permettront d'observer le Web avec plus de discernement et mieux appréhender ses changements.

Conscients que nous ne serons jamais exhaustifs et que nos idées sont loin d'être indémontables, nous espérons que ce document sera la cible de critiques constructives et qu'il fournira de quoi alimenter les discussions sur le Web et son devenir.

---

<sup>2</sup> en nous référant aux définitions données par J. Ferber [4]



## Des paradoxes

L'étendue du Web change sans cesse. Chaque interaction (échange de fichier, message, post, tag), le moindre accès à une ressource (même statique) ajout, suppression, modification de données ou tout simplement en inscrivant cette interaction en mémoire (log) la modifie.

## Un Web transformationnel

La naissance des structures sociales a transformé la fonction du Web qui est passée d'un environnement hypertextuel à un environnement interactif très propice à la collaboration. La collaboration et l'interaction au sein du Web 2.0 ont produit un phénomène de transformation. Nous parlerons plus loin de l'immanence du Web.

Humains, logiciels ou artificiels, les agents du Web sont des entités qui sont capables d'agir de manière autonome, communiquer avec d'autres agents et percevoir leur environnement: un utilisateur de Facebook est un agent humain, son navigateur et les serveurs de Facebook sont des agents logiciels. Facebook est un agent artificiel.<sup>3</sup>

L'environnement d'un agent est l'ensemble des liens établis entre les identificateurs des ressources (URI) et leur valeur. Par exemple, l'environnement d'une session Facebook.

Dans les ressources qui constituent l'environnement d'un agent, nous distinguons les services et les données. Nous abordons plus loin la question des services qui est au cœur de notre vision pour le Web.

## Avec une mémoire

A tout moment, un agent peut affecter ou modifier une donnée (par exemple un Google doc) et stocker sa valeur en mémoire via une séquence d'instructions.

Nous appelons **état**, une séquence qui contient l'ensemble des valeurs intermédiaires nécessaires à un calcul (une opération) dans un environnement donné. [28]

Sur le plan technologique, la notion d'état est fondamentale pour aborder les questions sur l'architecture du Web. Nous donnons à la fin de ce document une définition plus mathématique de l'état fondée sur les explications de P. Van Roy et S. Haridi.

Les agents doivent être capables d'accéder aux ressources de manière concurrentielle tout en apprenant du passé. C'est à dire qu'une opération sur le Web doit être interactive et doit mettre en oeuvre des états comme définis précédemment.

---

<sup>3</sup> *Demain, les automobiles, les congélateurs seront-ils des agents intelligents du Web ?*

Notons qu'en ce qui concerne les questions de société, la mémoire du Web soulève des polémiques vis à vis du "droit à l'oubli" qui est fondamental dans une société humaine et qu'il faut considérer comme une question essentielle.

## **Basé sur un mode de transport non connecté sans état**

Le protocole HTTP est construit selon le modèle client-serveur. Une communication via HTTP est constituée par une série de requête/réponse qui sont autant de connexions HTTP. (Connexion tcp/ip - Requête - Réponse - Fermeture de la connexion).

**Ceci en raison du fait que HTTP, le protocole principal du Web fonctionne en mode déconnecté.** La connexion est rompue à la fin de la transaction (et pas de la session ce qui permet de traiter un plus grand nombre de clients)

Ce sont les couches applicatives qui traitent ce problème et donnent au Web sa connectivité.

Paradoxalement, dès les fondements du Web, il manquait donc la notion d'état (HTTP stateless) [12]. On peut dire que l'état est une brique manquante à HTTP.

### **HTTP fonctionne en mode déconnecté**

HTTP produit un même résultat pour une même requête sans jamais tenir compte des évènements précédents. Une fois sa requête transmise, le client se déconnecte du serveur et attend une réponse. Le serveur doit rétablir la connexion après l'avoir traitée et ensuite envoyer une réponse. Dans un échange HTTP, le serveur et le client ne conserve pas de connexion entre chaque requête cliente. Chacun oublie l'autre. Pour cette raison, ni le client ni le navigateur ne peuvent conserver des informations entre les différentes demandes du client.

## **HTTP handicape l'évolution du Web et il faudra faire avec...**

Parce que HTTP est un protocole sans état, il est difficile de mettre en œuvre des sites Web qui réagissent de façon intelligente aux actions de l'utilisateur.

### **Il est difficile de coder des applications pour le WEB**

HTTP n'est pas en mesure de conserver une mémoire de l'identité de chaque client qui se connecte à un site Web et traite donc chaque demande d'une page Web comme une connexion unique et indépendante. Toutefois, pour des applications telles que celles du e-commerce qui accumulent de l'information, il est extrêmement important de savoir ce qui s'est passé précédemment. (par exemple le panier des achats). Pour surmonter ce problème, certaines technologies ont été développées comme ActiveX, Java, JavaScript, les cookies et les sessions. Le W3C n'envisage plus de modification, HTTP étant considéré complet et stable. [30]

### **HTTP est vulnérable**

Le mode déconnecté de HTTP rend le Web très vulnérable à des attaques de type "man-in-the-middle". Ce problème est partiellement traité par la mise en oeuvre de HTTPS [23] mais le Web reste vulnérable. L'OWASP a mis en évidence 10 risques majeurs d'attaque<sup>4</sup>: Sachant que HTTPS ne peut rien contre les menaces de phishing, pharming (DNS), ARP poisoning, Rogue ARP et contre l'usage de faux certificats, de collision MD5, NULL char ou encore du social engineering. [27]

## Les méthodes ignorées de HTTP

Dans ces fondements, le Web peut être vu comme un réseau dont les noeuds sont les pages HTML [19]. Une page opère comme un concentrateur (hub) grâce aux principes des hyperliens et des formulaires. Cette organisation en réseau associée à l'implémentation du protocole HTTP offrait nativement les quatre opérations GET, PUT, POST et DELETE [20]. Hors, le Web a évolué sans l'usage des opérations PUT et DELETE qui n'ont pratiquement pas été exploitées. On peut se demander comment ces opérations natives au protocole central du Web ont pu être ignorées. Selon Elliotte Rusty Harold, ce sont les éditeurs de navigateurs qui ont tout simplement échoué à mettre en oeuvre la spécification [31].

Notre intuition vient de l'idée que le Web, en tant que réseau de communication, ne se restreint pas à une seule autorité. Dans ces conditions, la **confiance** ne pouvait pas être un élément fort dans le développement du Web. Les opérations de suppression et d'écriture ont rapidement été confiées au serveur (script CGI) alors que les opérations PUT et DELETE sont conçues pour être pilotées par le client. [20]

## Où l'on parle du facteur confiance

Nous pensons que la **confiance** est un facteur déterminant pour l'évolution du Web. On l'a vu, c'est l'absence de confiance qui a orienté l'usage des opérations du protocole HTTP. Nous verrons plus loin qu'il est possible d'envisager des moyens pour canaliser la confiance sur le Web. L'enjeu est de taille car nous pensons que la confiance et l'adhésion sont des briques dont l'absence pourrait faire échouer le Web sémantique.

Quelles sera la valeur des ressources du Web si ceux qui les exploitent (les éditeurs tels que facebook, google, youtube), ceux qui les constituent (les internautes, les entreprises et institutions) ne trouvent pas une démarche commune pour partager un vocabulaire, adopter des standards (l'OpenID par exemple) et exposer leurs API pour en permettre l'utilisation ?

## Un langage avec si peu de sémantique

De son côté, le langage HTML n'incluait presque pas de sémantique hormis celle proposée par les balises meta [24]. C'est un inconvénient majeur pour le développement du Web sémantique.

---

<sup>4</sup> Injection, XSS, Broken Authentication and Session Management, Insecure Direct Object References, CSRF, Security Misconfiguration, Insecure Cryptographic Storage, Failure to Restrict URL Access, Insufficient Transport Layer Protection, Unvalidated Redirects and Forward

Aujourd'hui on voit bien qu'il faut oeuvrer pour enrichir le contenu des ressources du Web et permettre d'établir des relations entre elles. Le retard en termes de sémantique du langage HTML est une autre brique manquante du Web. Nous parlons de retard car les efforts se poursuivent notamment depuis HTML 4 pour donner aux pages Web plus de sens. [17]

## le Web 2.0

### Des technologies

Web 2.0 est une expression utilisée pour décrire la structure en réseau social d'un ensemble d'applications qui permettent à leurs utilisateurs d'interagir les uns avec les autres, partager l'information et modifier le contenu Web (web site content).

Dans notre travail, nous examinons, avec un regard d'informaticien, les éléments qui nous paraissent les plus importants pour la transformation du Web.

### Des Web services

*Quelques briques technologiques manquantes*

#### Des services sans état pauvres en sémantique

Dans la construction des systèmes distribués tels que le Web, les services sont des composants logiciels standardisés et interopérables qui effectuent certaines tâches, comme renvoyer certaines informations (messages) ou accéder à une base de donnée. Selon les normes SOA, un service peut être décrit, découvert et exécuté:

WSDL permet de décrire le service (donner sa fonction).

SOAP permet de décrire les méthodes pour accéder aux composants du service (par échanges de messages). SOAP est basé sur les protocoles HTTP et UDDI pour publier un service (identification, découverte). Mais la description des services Web est limitée. SOA repose sur les Remote Procedure Call (RPC), qui ne savent pas s'adapter à l'activité de l'utilisateur: pas de mémoire (stateless), pas de session utilisateur, pas de maîtrise de la conversation (il s'agit d'une simple interaction de type question/réponse) et une communication purement synchrone.

L'inconvénient majeur des services Web est d'ordre sémantique [5]: WSDL permet de définir une interface de service et de mise en œuvre du service, mais seulement à un niveau syntaxique. Il n'y a aucun moyen de transférer la sémantique de transaction à travers un appel SOAP.

SOAP est basé sur le protocole HTTP qui, on l'a vu est sans connexion et sans état (ne conserve pas de mémoire, n'entretient pas les relations avec l'utilisateur, incapable de gérer une

interaction longue durée de vie comme une conversation). De ce fait, SOAP est également un protocole sans état et aucun message SOAP n'est lié à un autre message.

UDDI ne prend pas en charge les descriptions sémantiques pour les recherches de fonctionnalités. De plus ces recherches ne peuvent être fondées que sur des mots-clés seulement.

#### *Un paradoxe technologique: des méthodes d'accès qui s'opposent*

En ce qui concerne les services, deux visions s'opposent [7]. L'une les envisage sous l'aspect RPC-centric, c'est à dire offrant un ensemble de méthodes pouvant être invoquées à distance via des appels de procédures distantes. L'autre aborde les services comme des agents pour l'échange de documents. (document-centric view).

Dans les deux cas, ce sont des documents XML qui sont transmis pour décrire les objets correspondants. Cependant, il y a une différence conceptuelle importante:

- Le point de vue RPC voit les documents XML comme accessoires au calcul distribué. Les documents XML ne sont que des représentations des objets sur lequel le calcul principal a lieu.
- Le point de vue document-centric considère les documents XML comme les principales représentations et usages du calcul distribué. Chaque composant du service lit, produit, stocke et transmet les documents. Dans cette approche, les fichiers XML sont au coeur du calcul informatique.

Le point de vue RPC correspond donc à une mince couche de services Web sur une application existante. L'application détermine quelles fonctionnalités sont à la charge du service.

L'approche document-centric voit les services Web plus naturellement et les considère comme un moyen d'établir des relations entre les objets.

Dans la construction des systèmes distribués tels que le Web, SOAP est donc par nature opposé au principe REST. Dans une architecture REST, un composant lit ou modifie une ressource en utilisant une représentation de cette ressource (méta-data, RDF, URI). Une ressource est nommée (sur le Web, il s'agit de son URI) et peut évoluer avec le temps. Les agents disposent de sa représentation qui peut être enrichie par des meta-données.

Hors, avec la progression du Web sémantique, c'est ce style d'architecture que l'on trouve de plus en plus sur le Web. Nous voyons là un paradoxe d'ordre technologique.

## Des applications

Les principales applications du Web 2.0 sont les **réseaux sociaux**, les **sites de partage de ressources** (video, musique etc.), les **services Web**, les **moteurs de recherche**, les **sites de médias** (culture, sports, actualités etc.), les **sites Wiki**, les **blogs** sans oublier les **techniques d'étiquetage social** (folksonomies).

*La centralisation serait-elle has-been ?*

Sur le Web tout ordinateur devrait pouvoir partager des ressources, que ce soit un serveur Web, un pc ou un téléphone mobile. bien que les évolutions technologiques permettent d'envisager des applications riches (notamment avec html 5), une architecture de type point à point (peer-to-peer) serait par nature plus adaptée aux échanges entre agents. Cependant, le développement du Cloud est un élément déterminant pour la puissance du Web 2.0.

## Le Cloud

Historiquement, l'informatique a connu divers modèles architecturaux tels que les systèmes Mainframe avec terminaux, les réseaux de pc et les stations de travail, le client-serveur.... Aujourd'hui, le large Clouding s'impose via un Internet très distribué. En termes de société, les gens sont déjà familiarisés avec le data Clouding.

La plupart des grands centres de données du Cloud appartiennent à des groupes commerciaux alors que le Web est axé sur le collaboratif et le partage. Néanmoins, la croissance en termes d'utilisateurs est forte (par exemple, Animoto.com +3000 utilisateurs/heure). L'évolutivité sera une notion de première classe dans la prochaine génération de systèmes de base de données. Pour les entreprises, la souplesse du Clouding offre l'alternative du « pay by use » à la place du coûteux « provisioning for peak ».

En termes d'architecture, le Cloud Computing offre plus d'évolutivité, et permet notamment l'élasticité des systèmes. Contrairement aux SGBD traditionnels, la technologie du « key-value Data Storing » possède un potentiel à grande échelle ainsi qu'une meilleure tolérance aux pannes. Ce modèle de données s'articule sur la notion de clé comme unique identifiant avec pour granularité l'association clé-valeur. Il existe aujourd'hui différentes conceptions de ce modèle : Big table (Google), PNUTS (Yahoo !), Dynamo (Amazon). Le routage des informations s'effectue par une approche hiérarchique, cartographique, (explicit storing of mapping) ou encore via hash tables distribuées. L'accès aux informations de grande taille est facilité par la mise en ligne de metadata légères (who has what) via de Data tablets. Sur le plan technique, le Chubby file constitue la racine de cette arborescence. La concurrence entre les processus est gérée au travers de vecteurs asynchrones. De nouveaux procédés permettent d'optimiser l'accès aux données (table level partitionning, co-locate items call together, data fusion) <sup>5</sup>[32].

---

<sup>5</sup> Des exemples de technologies actuelles :

## Les moteurs de recherche

Avec l'explosion du Web 2.0, un constat s'impose: la quantité d'informations grandit chaque jour d'avantage et parmi ce flot incessant de données, la recherche par mots-clef, se montre de moins en moins précise et pertinente. C'est à cause des faiblesses de cette technologie:

### *Les faiblesses de la recherche par mots-clé*

#### Un cruel manque de la sémantique

Les moteurs ne comprennent pas la question, comme l'expliquait déjà Nova Spivack en 2008 [11]. De plus, les résultats ne sont pas fondés sur des associations sémantiques.

Il existe des approches plus sophistiquées et plus efficaces comme celle qui se fondent sur les taxonomies, les thésaurus ou les ontologies: aujourd'hui, un internaute qui recherche sur Google des informations concernant "les armées de Napoléon" se heurte rapidement aux limites de la recherche par mots-clés. Effectivement, celle-ci ne prend pas en compte le sens du groupe nominal "quoi de qui" qui possède pourtant une structure grammaticale et hiérarchisée. Le terme "armées de Napoléon" peut-être vu comme une instance de la superclasse "armée" spécialisée par "Napoléon". De plus, la recherche par mot clé ne se base pas sur un lexique thématique et manque cruellement d'expertise sur le domaine ciblé qui lui gagnerait à être spécialisé (principe de l'annuaire). Le projet SygAnnot [Annexe 2] et son extension RDF [Annexe 3] a montré qu'il est possible grâce aux techniques du TALN d'étiqueter le texte d'un document par un balisage lexical et thématique. Ce balisage permet de structurer la recherche en établissant une hiérarchie entre les termes par référence à une ontologie thématique.

Pour illustrer notre propos, précisons qu'aujourd'hui pour une recherche sur les mots-clés "HTTP DELETE", google renvoie environ 1 580 000 000 résultats. Si la recherche était thématique et hiérarchisée, il suffirait de préciser THEME="Web Science" pour faire une recherche pertinente. C'est la force de l'indexation thématique.

#### Des lacunes

Les moteurs ne capitalisent pas sur la satisfaction des utilisateurs.

Les données qui ne se présentent pas sous le format HTML ne peuvent être trouvées.

Pas de traitement des synonymes pour plus de résultat et des homonymes pour une meilleure pertinence.<sup>6</sup>

#### Des résultats inutiles

la recherche par mot clé renvoie beaucoup de résultats inutiles. C'est une autre faiblesse qui est

---

- Le G-Store qui offre des garanties sur les données transactionnelles.

- ElasTraS qui permet des transactions à l'aide de bases de données logiques contenue dans des partitions.

Amr El Abbadi, ElasTraS: An Elastic Transactional Data Store in the Cloud [http://www.usenix.org/event/hotcloud09/tech/full\\_papers/das.pdf](http://www.usenix.org/event/hotcloud09/tech/full_papers/das.pdf)

<sup>6</sup> Les mots qui ont le même sens que les mots-clés saisis par l'utilisateur devraient figurer dans les résultats de la recherche et inversement, les pages contenant des homonymes ne devraient pas s'y trouver.

aussi une conséquence du manque d'efficacité.

Nous trouvons aussi des lacunes moins fondamentales:

#### **Des défaillances**

Beaucoup de moteurs de recherche ne supportent pas les pages multi-cadres (Frames)

#### **Et un manque de confort**

L'usage des moteurs de recherche manque de confort (correcteur d'orthographe, proposition de recherche plus qu'une simple autocomplétion, proposition des URL les plus visités pour la période en cours).

### **Folksonomies**

Une folksonomie est un système de classification, création et gestion de balises (tags) pour annoter et catégoriser des ressources telles que les images, les pages Web et autres documents. Elle constitue une méthode facile pour organiser les informations et offre des perspectives d'évolution importantes pour les recherches sur le Web.

#### *Les inconvénients des folksonomies*

Sur le Web, les folksonomies ont des capacités de recherche limitées par rapport à des systèmes basés sur les taxonomies, les thesaurus ou les ontologies. Chaque utilisateur crée des balises et les affecte librement aux ressources.

#### **Pas de référence commune**

Dans le principe, pour donner du sens à une ressource, chaque humain peut annoter selon sa propre perception (tagging) . Ces annotations sont des relations sémantiques qui ne reposent sur aucune référence commune:

#### **Pas de standard**

Les balises sémantiques ne sont pas créées selon un standard.

- multiples variations syntaxiques d'un même tag (Multiple Syntactic Variation)
- les liens sémantiques sont la propriété des éditeurs
- pas d'ontologie commune
- pas de thesaurus commun
- langues multiples et fautes d'orthographe (semanticweb / semanticweb / semanticweb).
- singulier ou pluriel (semanticweb / semanticwebs).
- usage de différents séparateurs (web sémantique / web-sémantique);
- combinaison de plusieurs balises (semanticweb / semantic-web)
- synonymes (différentes balises pour la même signification)
- homonymes (les mêmes balises utilisées avec des significations différentes)



La réduction des variations syntaxiques des balises permettrait d'améliorer la qualité des folksonomies et des recherches basées sur cette approche. Pour la santé du Web sémantique, il s'agit d'un problème important à prendre en considération.

## Des usages

Selon le CEFRIO au Québec, les principales familles des usages du Web 2.0 pour les organisations [26] sont les suivantes: **partager** et **collaborer**, **converser**, **rechercher** et **collecter**, **diffuser**, **veiller**.

Toujours selon le CEFRIO, l'usage des outils du Web 2.0 pour une administration 2.0 apporte les améliorations suivantes: une relation à l'utilisateur différenciée et individualisée, un fonctionnement interne collaboratif, des partenariats complémentaires et une efficacité transparente. [26]

Bénéfices tangibles	Bénéfices intangibles
Réduction du nombre de courriels	Meilleure gestion des connaissances
Réduction des coûts de voyage	Satisfaction des employés
Réduction des coûts téléphoniques	Meilleure capacité d'innovation
Réduction des budgets TI Compétitivité	renforcée
Stratégies de marketing plus efficaces	Relation client améliorée
Développement de produits plus rapide	Veille économique
Fidélisation de clients	Meilleure intégration et communication
Meilleure performance individuelle	Renforcement de marque
Développement professionnel des individus	Fidélisation des collaborateurs

Tableau: les bénéfices de l'usage des outils du Web 2.0 pour les organisations selon le CEFRIO

Tim O'Reilly a introduit le terme Web 2.0 comme un socle d'échanges entre les utilisateurs (intelligence collective) et les applications en ligne. Selon lui , la transformation qui s'est opéré entre le Web 1.0 et le Web 2.0 serait illustrée par le tableau suivant [25]:

Web 1.0	Web 2.0
DoubleClick	Google AdSense
Ofoto	Flickr
Akamai	BitTorrent
mp3.com	Napster
Britannica Online	Wikipedia
personal websites	blogging
evite	upcoming.org and EVDB
domain name speculation	search engine optimization
page views	cost per click
screen scraping	web services
publishing	participation
content management systems	wikis
directories (taxonomy)	tagging ("folksonomy")
stickiness	syndication

Tableau: illustration du passage aux applications Web 2.0 selon Tim O'Reilly and John Battelle (initial brainstorming )

On voit que les applications 2.0 s'appuient plus sur le partage entre utilisateurs, leur implication et sont moins centralisées.

## L'immanence du Web

Le Web change parce des changements naissent au sein de ses propres objets. La puissance d'une application n'est pas inhérente à la seule application mais à son environnement. Ce sont les utilisateurs qui lui donnent un potentiel de changer le monde. L'idée générale derrière la notion d'immanence, c'est que la cause du développement d'un objet (Facebook, par exemple) se produit à l'intérieur de cet objet. Le Web est immanence dans le sens où la puissance n'est pas dans Facebook ou YouTube, mais dans l'environnement de Facebook et YouTube [3]. Par exemple, **si les utilisateurs des réseaux sociaux n'avaient pas diffusé des photographies de personnes, la reconnaissance faciale sur le Web n'aurait pas eu le même développement.**

L'immanence est un concept fondamental pour le Web. Sa transformation est principalement induite par des échanges humains individuels ou communautaires qui constituent une intelligence collective humaine. [25]

## Notre vision SMA pour le Web

Dans notre démarche, nous avons une vision tournée vers les systèmes multi-agents (SMA). Celle-ci est liée à la notion de génération de services dynamiques (DSG) et à l'utilisation d'une grille au sens du "Cloud computing". Basé sur une architecture orientée service (SOA), le Cloud constitue un mécanisme pour traiter le Web avec des propriétés d'environnement et d'état. [5]

### Des services dynamiques

Le processus d'obtention des services dynamiques se réalise via les conversations entre agents. En DSG, les fournisseurs de services et les consommateurs sont deux agents (humain ou artificiel) membres d'une communauté qui peuvent avoir des conversations. Ainsi la génération dynamique de services implique deux types d'agent: les entités humaines et les entités artificielles. Le DSG fournit au Web une solution et un mécanisme pour fournir de réels environnements de collaboration en permettant à toutes sortes d'agents de communiquer et de générer des services.

Les communications entre agents et les flux d'entrée/sortie des messages peuvent être vus comme des flots (streams) traités par évaluation paresseuse.<sup>7</sup>

### Des flots

Les flots offrent une approche alternative à la notion d'état. Les flux d'information sont perçus comme des structures de données intelligentes pour modéliser les systèmes à état et leur évolution par le biais de séquences. Celles-ci constituent le fractionnement dans le temps pour les systèmes modélisés comme par exemple le moment où les éléments de la séquence sont évalués ou encore celui où la donnée est construite ou transformée.

### Des évaluations retardées

Un agent peut retarder la production du message suivant pour attendre la réaction de son interlocuteur à son message précédent. Ce principe d'évaluation retardée (paresseuse) s'applique sur les flots d'information du Web. Il est incontournable pour la collaboration au sein d'un SMA.

### Une description qualitative des agents

Aujourd'hui, la description des agents logiciels est portée par des moyens limités tels que les balises méta de HTML (W3C HTML meta elements).

---

<sup>7</sup> L'opération de sauvegarde d'un document Google est un bon exemple pour illustrer l'efficacité de ces concepts.

On remarque qu'il n'existe pas de moyens pour donner une **description qualitative** des agents. Quel est son niveau de fiabilité en terme de sécurité ? (Facebook serait plus orange que vert clair...), en termes de robustesse ? (on peut envisager un indice de taux de pannes), son niveau d'utilité ?

Dans notre vision SMA, nous considérons les agents qui communiquent par un ensemble de relations (via des opérations qui on l'a vu se font par séquences à état), avec des ressources qui leurs sont propres (leurs objets) et qu'ils sont susceptibles de partager. Le tout constitue un environnement.

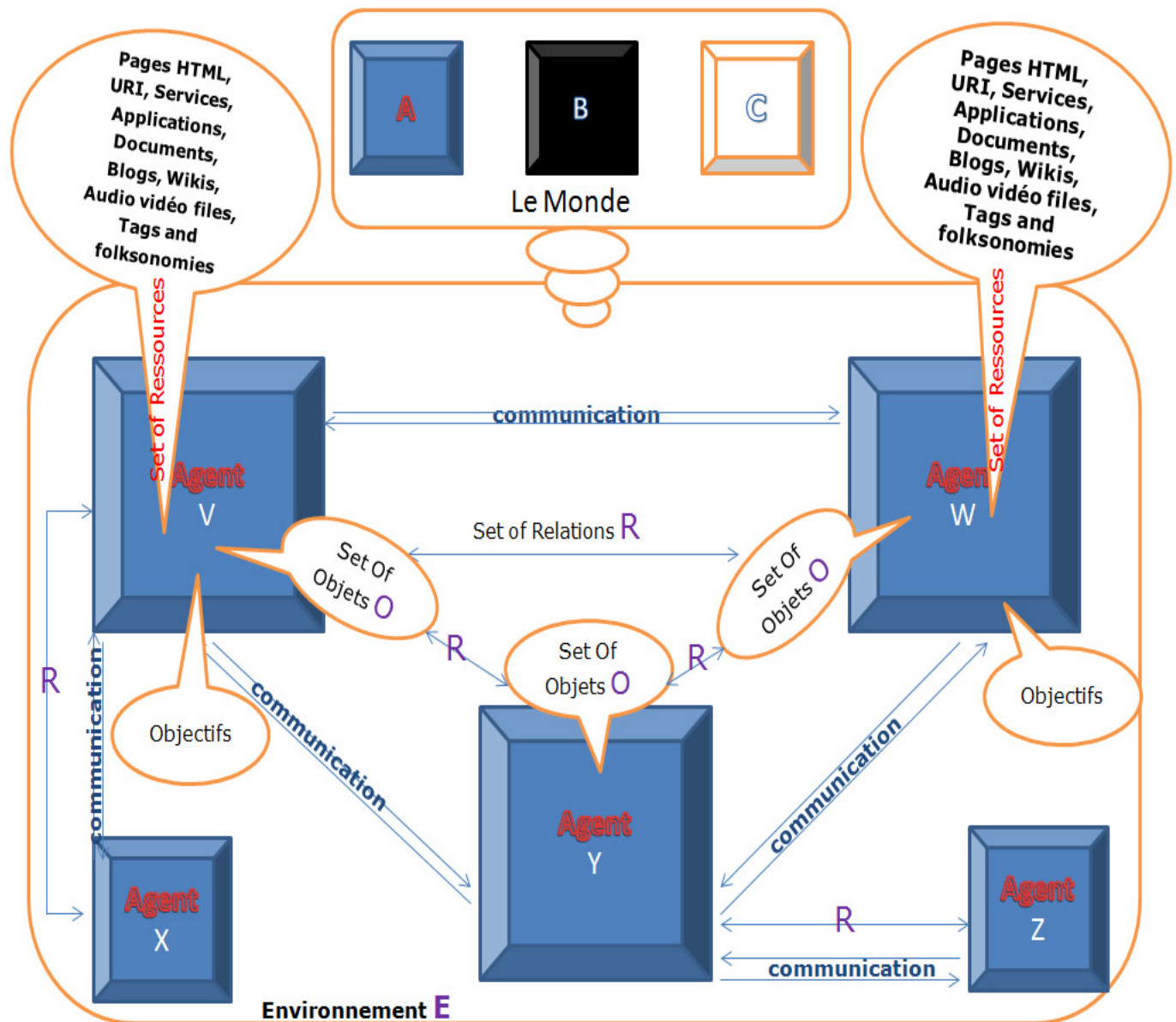


schéma: notre vision SMA pour le Web

L'agent humain Z (un internaute) communique via un agent logiciel Y (un navigateur) qui est

connecté à un agent virtuel (un service au sens Agora, un site Web, une e-application) et accède à différents objets (ressources dédiées au sens Agora).

Pour Y et Z, nous pouvons envisager une graduation qui permettrait de leur attribuer des **indices** de confiance, d'intérêt, de proximité ou tout autre valeur permettant d'apprécier leurs qualités. Celui-ci serait calculé dynamiquement selon les attributs de l'agent (FQDN, URL, le contenu des entêtes HTTP), le référencement, l'éventuel certificat employé (X509 EV, validité, authentification) et en tenant compte d'informations utiles pour ajuster la valeur telles que les logs, les statistiques de runtime et les relations sémantiques qui le concernent (tags).

## Vers une navigation colorée

La valeur des indices serait surtout tirée de l'historique et renvoyée aux navigateurs (vers un plugin par exemple). Ceux-ci ajusteraient le code couleur de leurs barres d'adresse ou de bordure. L'utilisateur bénéficierait ainsi d'un facteur visuel pour mieux "sentir" sa navigation. La sensibilité de l'internaute serait ainsi améliorée grâce à une alternance de couleur (du rouge "alerte" au "vert confiance"). On pourrait alors parler d'une navigation colorée ou dit autrement, d'indications ouvrant la possibilité d'une évaluation logicielle, comme une espèce de PageRank pour la confiance, à l'instar du Trust Rank qui a été introduit comme un facteur de confiance à l'usage des bots. Cela donnerait probablement naissance à la e-discrimination mais ce phénomène tendrait alors à montrer que les conduites sur le Web social se rapprochent de celles constatées dans la réalité.

L'idée centrale est de donner aux agents une valeur qui permettrait d'évaluer leurs qualités. Pour l'interaction entre agents humains, les qualités intéressantes sont entre autres la fiabilité, l'intimité (ou la familiarité), la popularité, la courtoisie.

La difficulté pour évaluer les indices est bien sûr directement liée à la nature de la qualité que l'on souhaite évaluer.

En supposant que Z est un gestionnaire de communautés dans un réseau social, par exemple un cyber-formateur qui gère une communauté pour sa pédagogie (une VO selon Agora), le CEFRIO [26] donne les indicateurs suivants classés par familles d'usages:

**Partager et collaborer:** nombre d'amis, de fans, de followers, de communautés gérées, nombre et pertinence des contributions, nombre et pertinence des commentaires, niveau de sentiment<sup>8</sup>, taux d'engagement<sup>9</sup>.

**Rechercher et collecter:** nombre et pertinence des résultats de recherche

**Diffuser:** nombre de communiqués, nombre et pertinence des commentaires internes, nombre

---

<sup>8</sup> Le « niveau de sentiment » fait référence à l'impact personnel des propos se disant à l'égard d'une personne.

<sup>9</sup> Le « taux d'engagement » fait référence à la récurrence des visiteurs.

de formation proposées, nombre de membre par formation, satisfaction des membres, nombre et pertinence des commentaires

**Surveiller:** nombre et pertinence des contributions, nombre et pertinence des commentaires, nombre et pertinence des idées recueillies, niveau de sentiment, taux d'engagement.

Pour les objets du Web, nous pouvons citer entre autres les attributs décrivant la taille, le type d'encodage, l'adresse réseau.

Il reste à imaginer quelles entités seraient susceptibles d'évaluer cet indice et de quels droits...

## **Canaliser la confiance et l'adhésion**

L'enjeu est de taille car nous pensons que la confiance et l'adhésion sont des briques dont l'absence pourrait faire échouer le Web sémantique. Il faut canaliser ces sentiments sur le Web.

## **Faire passer l'émotion**

Pour donner plus de sens au Web, nous pouvons envisager de retranscrire l'émotion exprimée par l'internaute via une association entre les relations sémantiques de données et l'environnement graphique du navigateur. Aujourd'hui, un utilisateur sur un réseau social clique sur "J'aime" et le lien sémantique entre la donnée et cet utilisateur est créé. Mais la ressource reste la même. La table [URI, nbre de personnes qui aiment] est la propriété de Facebook. Le lien "aime" est crédité sur le compte Facebook, c'est à dire pour le compte de celui qui aime mais pas du côté de celui qui est aimé (l'URI). On peut parler d'un lien "*sémantique propriétaire*".

Pour que la sémantique soit étendue à l'ensemble du Web, il faudrait que les serveurs qui mettent les URI à disposition des agents soit capables de comptabiliser les liens sémantiques (tags). Nous parlons alors d'une vision "HTTP ILIKE". Pour que la valeur sémantique soit encore plus forte, il faudrait que les serveurs permettent l'annotation directe des ressources. Une URI désignerait alors une donnée et l'ensemble des liens qui la caractérisent. C'est la vision "HTTP ANNOTATE".

## **En conclusion**

En termes de technologie, construit sur la base de recommandations, le Web est effectivement fait de limites, de paradoxes et de contradictions. Parce que le Web se transforme et possède un potentiel sans égal pour changer le monde, il a besoin d'une science dédiée dans laquelle il sera nécessaire de considérer un ensemble de concepts et de phénomènes qui font sa force et ses faiblesses.



## En résumé

- Des fondamentaux
  - Le Web est transformationnel et immanent.
  - La vision SMA pour le Web
  - La mémoire, l'état, les services, les flots et l'évaluation retardée
  - Architectures et Cloud computing
- Des paradoxes
  - Un Web basé sur un protocole sans état
  - HTTP handicape l'évolution du Web
  - HTML, un langage avec si peu de sémantique
- Des limites et des faiblesses technologiques
  - Web services
  - Moteurs de recherche
  - Folksonomies
- Les usages et la transformation du Web 2.0
- Des idées
  - Améliorer les moteurs de recherche
  - Evaluer qualitativement les agents
  - Colorer la navigation
  - Retranscrire l'émotion
  - Canaliser la confiance

# Références

- [1] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. J. Weitzner. *Creating a Science of the Web*. Science, 313(5788):769–771, August 2006.
- [2] Stefano A. Cerri. *Dynamic typing and lazy evaluation as necessary requirements for Web languages*. In European Lisp User Group Meeting, ELUGM'99, Amsterdam, Netherlands, June 1999.
- [3] P.I Dugénie and S. A. Cerri. *The Principle of Immanence in Event-Based Distributed Systems*. In *Reasoning in Event-Based Distributed Systems*, volume 347 of Studies in Computational Intelligence, pages 239–256. Springer, 2011.
- [4] Jacques Ferber. *Les Systemes Multi-Agents, vers une intelligence collective*. InterEditions, Paris, France, 1995.
- [5] Clement Jonquet. *Dynamic Service Generation: Agent interactions for service exchange on the Grid*. Phd thesis, University Montpellier 2, Montpellier, France, November 2006.
- [6] P. Van Roy and S. Haridi. *Concepts, Techniques, and Models of Computer Programming*. MIT Press, March 2004.
- [7] M. P. Singh and M. N. Huhns. *Service-Oriented Computing, Semantics, Processes, Agents*. John Wiley & Sons, 2005.
- [8] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski & L. Brilliant *Detecting influenza epidemics using search engine query data Nature* 457, 1012-1024 2009[9]
- [9] W3C <http://www.w3.org/TR/ws-gloss/> Web services glossary
- [10] Dilip C. Naik, *INTERNET - Standards and Protocols*. Microsoft Press, 1998
- [11] Michael Arrington (adaptation: Alain Eskenazi) 26 avril 2008 "*La recherche par mots-clefs atteint-elle ses limites?*"
- [12] W3C Tim Berner Lee <http://www.w3.org/Protocols/HTTP/HTTP2.html> Basic HTTP as defined in 1982  
Archive document
- [13] Pierre Lévy. *L'intelligence collective. Pour une anthropologie du cyberspace*. La Découverte, Paris, 1994.
- [14] P. Dugénie, S. A. Cerri, P. Lemoisson et A. Gouaich. Agora UCS Ubiquitous Collaborative Space LIRMM: Université Montpellier 2 & CNRS
- [15] Microsoft.com. "*Windows History: Internet Explorer History*". Microsoft.com. 2003-06-30. Retrieved 2009-11-11.
- [16] W3C <http://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html> HTTP Header Field Definitions
- [17] W3C <http://www.w3.org/TR/WCAG20-TECHS/G115.html> Using semantic elements to mark up structure
- [18] Nigel Shallbolt. Why study the Web ?, Royal society 2010-03-08
- [19] J. Kleinberg and S. Lawrence *The Structure of the Web*, Science 30 November 2001 Vol. 294 no. 5548 pp. 1849-1850 DOI: 10.1126/science.1067014 PERSPECTIVE NETWORK ANALYSIS
- [20] W3C Tim Berners-Lee <http://www.w3.org/Protocols/HTTP/Asimplemplemented.html> Basic HTTP as defined in 1992
- [21] Zoltán Gyöngyi, « *Combating Web Spam with TrustRank* », Proceedings of the International Conference on Very Large Data Bases, vol. 30, 2004, p. 576
- [22] IETF <http://www.ietf.org/rfc/rfc2828.txt> Internet Security Glossary RFC
- [23] IETF <http://www.ietf.org/rfc/rfc2818.txt> HTTP over TLS
- [24] W3C T. Berners-Lee, D. Connolly, MIT/W3C [http://www.w3.org/MarkUp/html-spec/html-spec\\_toc.htm](http://www.w3.org/MarkUp/html-spec/html-spec_toc.htm) Hypertext Markup Language - 2.0 September 22, 1995
- [25] Tim O'Reilly What Is Web 2.0 <http://oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- [26] CEFRIIO et Atelya expert conseil *Livre blanc Les usages du Web 2.0 dans les organisations*
- [27] OWASP [https://www.owasp.org/index.php/Top\\_10\\_2010-Main/](https://www.owasp.org/index.php/Top_10_2010-Main/) Top 10 Application Security Risks - 2010
- [28] J. Kolbitsch H. Maurer *The Transformation of the Web: How Emerging Communities Shape the Information we Consume Journal of Universal Computer Science*, vol. 12, no. 2 (2006), 187-213
- [29] W3C *The CSS saga* <http://www.w3.org/Style/LieBos2e/history/>
- [30] W3C *W3C has closed the HTTP activity* <http://www.w3.org/Protocols/>
- [31] Elliott Rusty Harold, *Why REST Failed* <http://cafe.elharo.com/web/why-rest-failed/>
- [32] Amr El Abbadi, *Data Management In the Cloud*, Les colloquiums du LIRMM - Campus St Priest Montpellier – 18/10/2011

# Annexes

- [annexe 1] Analyse des logs de Mailman <http://informe.free.fr/confonsciweb/analyseLogsMailman.html>
- [annexe 2] SygAnnot, anotateur automatique lexical A. Ibrahim, M. Laclau, T. Khelhout, F. Priotto <http://informe.free.fr/sygannot/>
- [annexe 3] F. Priotto SygAnnot, RDF <http://informe.free.fr/sygannotrdf>

# Vocabulaire

*A la fin de ce mémoire, nous tentons de formaliser un vocabulaire en nous référant aux ressources bibliographiques pour les notions représentatives de notre vision pour le Web.*

## **accès**

Condition nécessaire pour interagir avec un système, le manipuler, l'utiliser, acquérir des connaissances sur ce système, et/ou obtenir une représentation des ressources de ce système. [RFC 2828]

## **agent**

Toute entité autonome réelle ou virtuelle qui évolue dans un environnement et capable de percevoir cet environnement, agir sur cet environnement et interagir avec d'autres agents. [DEMAZEAUX 96].

En terme de technologie, un agent est un programme agissant au nom d'une personne ou une organisation. Cette définition correspond à la notion d'agent logiciel dans l'architecture du Web.

Selon la kénétique de Jacques Ferber [4], on appelle agent une entité physique ou virtuelle:

- qui est capable d'agir dans un environnement,
- qui peut communiquer directement avec d'autres agents,
- qui est mue par un ensemble de tendances (sous la forme d'objectifs individuels ou d'une fonction de satisfaction, voire de survie, qu'elle cherche à optimiser),
- qui possède des ressources propres,
- qui est capable de percevoir (mais de manière limitée) son environnement,
- qui ne dispose que d'une représentation partielle de cet environnement (et éventuellement aucune),
- qui possède des compétences et offre des services,
- qui peut éventuellement se reproduire,
- dont le comportement tend à satisfaire ses objectifs, en tenant compte des ressources et des compétences dont elle dispose, et en fonction de sa perception, de ses représentations et des communications qu'elle reçoit.

Ferber divise cette définition en distinguant les agents qui possèdent des capacités cognitives avec la capacité à anticiper sur les événements futurs et à s'y préparer.

Pour affiner cette définition, Ferber sépare aussi les agents purement communicants qui évoluent dans un environnement informatique.

Sur le Web, nous trouvons une majorité d'agents purement communicants, humains ou

logiciels, aux capacités cognitives variables. Un agent est un acteur qui peut jouer tour à tour les rôles de fournisseur de service et d'utilisateur.

### **AGORA UCS**

AGORA Ubiquitous Collaborative Space est une plate-forme développée dans le but d'expérimenter les principes de collaboration. [14]

Agora UCS distingue 4 mécanismes fondamentaux pour l'interaction au sein d'un système: la cooptation, l'implication, la délégation et l'habilitation (4 interaction mechanisms).

Une communauté vue sous le modèle AGORA est composée de membres qui peuvent avoir des droits différents sur les services (autorisations). Elle se décrit selon un certain nombre de propriétés (identifiant, description etc.) La création d'une nouvelle communauté doit être faite dans le cadre d'une autre communauté. Toutes les données appartenant à une communauté doivent être stockés, entretenus et indexée (histoire, mémoire). La trace des changements, des journaux d'événements et aussi des sessions de collaboration permettent de fournir une évaluation de la fréquence d'accès aux ressources, qui est utilisé pour le mécanisme d'implication. Une communauté peut personnaliser son environnement. Cet environnement est lié à un un conteneur de services. La communication entre les membres d'une communauté et les services est réalisée par l'intermédiaire des notifications utiles pour le déclenchement d'événements. [3]

Facebook est un conteneur. Le service des groupes Facebook permet aux membres d'une communauté de créer de nouvelles communautés ou certains membres possèdent plus de droits que d'autres.

### **asynchrone**

Une interaction est dite asynchrone lorsque les messages associés sont par ordre chronologique.

### **authentification**

L'authentification est le processus consistant à vérifier que le partenaire potentiel dans une conversation est susceptible de représenter une personne ou une organisation.

### **autorisation**

Le processus qui permet de déterminer si un sujet est autorisé à accéder à une ressource particulière selon des types d'accès déterminés. [STG]

### **bot**

Logiciel d'exploration automatique du Web.

Les bots de google sont des agents logiciels réactifs purement situés.

### **composant**

Un composant est un objet logiciel, destiné à interagir avec d'autres composants, encapsulant une certaine fonctionnalité ou un ensemble de fonctionnalités. Un composant possède une

interface clairement définie et est conforme à un comportement prescrit commun à tous les composants dans une architecture donnée.

### **contrôle d'accès**

Protection des ressources contre les accès non autorisés. Il s'agit d'un processus par lequel l'utilisation des ressources est réglementée selon une politique de sécurité et est permise par les entités du système agréées conformément à cette politique [RFC 2828]

### **confidentialité**

Assurer qu'une information sera gardée secrète, avec un accès limité aux personnes appropriées. [NSA Glossaire]

### **configuration**

Une collection de propriétés qui peuvent être modifiées. Une propriété peut influencer le comportement d'une entité.

### **conversation**

Une conversation consiste à maintenir un certain état au cours d'une interaction qui implique de multiples messages et/ou plusieurs participants.

Des conversations sont établies entre services web.

### **concurrence**

un ensemble d'activités qui s'exécutent de façon indépendante sur des ressources communes.

### **cooptation**

Selon Agora, la cooptation est un mécanisme au sein d'une communauté qui permet à ses membres de nommer de nouveaux membres. [14]

### **délégation**

Selon Agora, le mécanisme de délégation permet à un membre de transférer une partie de ses droits à un autre membre (le délégué). [14]

### **déviance**

Éloignement d'un usage par rapport à l'usage normal.

### **découverte de service**

Mécanisme qui permet la localisation de la description d'un service Web lié à des ressources et qui répond à certains critères fonctionnels. Il consiste à faire correspondre un ensemble de critères fonctionnels avec un ensemble de descriptions de ressources. L'objectif est de trouver un service Web approprié.

### **découverte de services**

Un service de découverte est un service de qui permet de récupérer les agents des services

Web description de la ressource liée.

### **description de service**

Une description du service est un ensemble de documents qui décrivent l'interface et la sémantique d'un service Web.

### **document**

Toutes les données qui peuvent être représentés sous une forme numérique. [UEB Glossaire]

### **entité systèmes**

Un élément actif d'un système informatique / réseau. Par exemple, un processus automatisé ou un ensemble de processus, un sous-système, une personne ou un groupe de personnes qui incorpore un ensemble distinct de la fonctionnalité. [RFC 2828]

### **environnement**

Dans un système informatique, c'est l'ensemble des liens entre identificateurs et variables en mémoire à un instant donné.

Dans un SMA, chaque agent évolue dans un environnement limité, celui qu'il est capable de percevoir. Sur le Web, les agents humains, tels que les membres d'une communauté ou artificiels tels que les navigateurs Web sont clairement différenciés des ressources qui sont des éléments de l'environnement dans lequel ils évoluent.

### **état**

Sur le Web, une ressource est identifiée par son URI qui lui donne une affectation unique.

RESSOURCE:  **$X_1$**

Un programme en exécution est un calcul qui met en oeuvre une paire  **$(ST, s)$**  où  **$ST$**  est une pile d'instructions sémantiques et  **$s$**  est une mémoire à affectation unique.

L'ensemble des ressources perçues par un agent constitue l'environnement de l'agent.

ENVIRONNEMENT:  **$E\{X_1, \dots, X_n\}$**

Les agents effectuent des opérations qui affectent les ressources. Ces opérations sont constituées d'instructions (code) pour un environnement donné.

Les opérations constituent des instructions sémantiques car elles changent la valeur sémantique des ressources. Par exemple un ajout (nouveau doc), une édition (modification du doc), suppression, Annotation ILIKE (Tag), annotation HTTP ANNOTATE(commentaire).

Une instruction sémantique est donc une paire qui associe une séquence (de code) à un environnement.

INSTRUCTION SEMANTIQUE :  **$ST=(s, E)$**

Nous pouvons définir un état d'exécution pour un calcul donné comme une paire **(ST, mem)** ou **mem** est une mémoire (à affectation unique).

De même, si l'on considère **En**, l'ensemble des ressources disponibles sur le Web  $\{X1, \dots, Xn\}$  et **STn**, celui des instructions sémantiques en cours, nous pouvons définir un état pour le Web noté **{STn, En}**.<sup>10</sup> [6]

### **grille**

Infrastructure virtuelle constituée d'un ensemble de ressources informatiques potentiellement partagées, distribuées, hétérogènes, délocalisées et autonomes.

### **habilitation**

Selon Agora, l'implication est le mécanisme qui gère les décisions à allouer progressivement des droits d'accès aux services et aux ressources pour les membres d'une communauté. [14]

### **implication**

Selon Agora, l'implication est un mécanisme qui fournit une estimation de la participation de chaque membre à des activités communautaires. [14]

### **interface de service**

Une interface de service est la façade abstraite qu'un service expose. Elle définit les types de messages, les modèles d'échange de messages et les conditions à satisfaire pour interagir avec le service.

### **identificateur**

Un identificateur est un nom sans ambiguïté pour une ressource.  
Les objets du Web sont identifiés par leur URI.

### **kénétique**

J. Ferber propose en 1994 le terme de kénétique pour désigner la science émergente des organisations artificielles s'intéressant aux interactions entre entités, quelle que soit leur nature (physique, biologique, sociale etc.). D'un point de vue théorique, cette discipline émergente consiste à dégager les principes et les mécanismes conduisant à l'auto-organisation d'un système par une analyse fine des différentes formes d'interactions comme la coopération, la compétition, la collaboration, etc., et par la définition de modèles opérationnels de ces interactions en décrivant le fonctionnement des agents et des systèmes multi-agents. D'un point de vue expérimental, la kénétique entend construire des univers artificiels, peuplés d'artefacts collectivement organisés de manière à mieux appréhender les composants nécessaires à la description, à l'auto-organisation et à l'évolution des sociétés d'agents.[S. Caldéroni]

### **licence**

---

<sup>10</sup> Le Web est ainsi considéré comme un système séquentiel.

Contrat par lequel le titulaire des droits d'un logiciel autorise un tiers pour son usage ou sa distribution.

### **message**

Un message est l'unité de base envoyé d'un services Web à un autre.

### **moteurs de recherche**

Application permettant de retrouver des ressources (pages web, forums Usenet, images, vidéo, fichiers, etc.) associées à des mots quelconques.

### **objet**

Pour le Web, nous considérons que les objets sont les éléments qui constituent l'environnement dans lequel évoluent les agents. (données, services, documents et tous les autres types de ressources).

### **opération**

Un ensemble de messages liés à une seule action d'un service Web.

### **OpenID**

système d'authentification décentralisé qui permet l'authentification unique, ainsi que le partage d'attributs.

### **port**

Dans l'acheminement d'un message, la notion de port permet, de distinguer les différents interlocuteurs logiciels.

### **SMA**

Système Multi-Agents.

Les 4 dimensions qui composent un système Multi-Agents sont les agents qui y évoluent, l'environnement, l'interaction entre les entités et les structures organisationnelles. [DEMAZEAUX 96]

Notre approche est de considérer le Web 2.0 comme le plus grand SMA que l'homme est créé. Selon Ferber, un SMA est un système composé des éléments suivants:

1. Un environnement E, c'est-à-dire un espace disposant généralement d'une métrique.
2. Un ensemble d'objets O. Ces objets sont situés, c'est-à-dire que, pour tout objet, il est possible, à un moment donné d'associer une position dans E. Ces objets sont passifs, c'est-à-dire qu'ils peuvent être perçus, créés, détruits et modifiés par les agents.
3. Un ensemble A d'agents, qui sont des objets particuliers ( $A \subseteq O$ ), lesquels représentent les entités actives du système.
4. Un ensemble de relations R qui unissent des objets (et donc des agents) entre eux.
5. Un ensemble d'opérations Op permettant aux agents de A de percevoir, produire,



consommer, transformer et manipuler des objets de O.

6. Des opérateurs chargés de représenter l'application de ces opérations et la réaction du monde à cette tentative de modification, que l'on appellera les lois de l'univers.

### **Trust Rank**

Trust Rank qui signifie "indice de confiance" est le nom d'un algorithme qui donne le degré de confiance accordé à une page web ou un site web et peut ensuite être utilisé pour faire partie de l'un des facteurs pris en compte pour classer les sites dans les résultats de recherche d'un moteur de recherche.

### **URI**

Chaine de caractère identifiant une ressource sur le Web (identifiant uniforme de ressource)

### **usage**

Désigne les règles usuelles et communément répandues pour se servir d'un service.

### **sémantique de service**

La sémantique d'un service est le comportement attendu lors de l'interaction avec ce service. La sémantique exprime un contrat entre le fournisseur et le demandeur.

### **session**

Une interaction durable entre les entités d'un système, impliquant souvent un utilisateur, caractérisée par le maintien d'un certain état de l'interaction pendant toute sa durée. [LSPAAT Glossaire]

### **SOAP**

Simple Object Access Protocol, est un protocole de spécification pour échanger des informations structurées dans la mise en œuvre de services Web dans les réseaux informatiques . Elle s'appuie sur XML pour son format de message.

### **service Web**

Un service Web est un système logiciel conçu pour permettre l'interopérabilité **de machine à machine** sur un réseau. Il dispose d'une interface décrite dans un format machine (spécifiquement WSDL). D'autres systèmes peuvent interagir avec le service Web de la manière prescrite par sa description en utilisant des messages SOAP, généralement transmis via le protocole HTTP.

### **synchrone**

Une interaction est dite synchrone quand les agents qui y participent doivent être disponibles pour recevoir et traiter les messages.

### **WSDL**

Décrit une interface publique d'accès à un service Web, notamment dans le cadre d'architectures de type SOA. c'est une description fondée sur XML qui indique comment communiquer pour utiliser le service.

Le WSDL sert à décrire :

1. le protocole de communication (SOAP RPC ou SOAP orienté message)
2. le format de messages requis pour communiquer avec ce service.
3. les méthodes que le client peut invoquer.
4. la localisation du service.

Une description WSDL est un document XML qui commence par la balise <definitions> et qui contient les balises suivantes :

1. <binding> : définit le protocole à utiliser pour invoquer le service web.
2. <port> : spécifie l'emplacement actuel du service.
3. <service> : décrit un ensemble de points finaux du réseau.