



Annotations sur le Web

Yannick Prié
LIRIS – Université Lyon 1
M2R / 2005-2006



Plan

- Introduction
- Annotations dans un web un peu sémantique
- Annotations et méta-données dans le Web sémantique
- en passant
 - Ressource Description Framework (RDF)
 - Web Sémantique

2

Le web, un système d'information global



- Espace de connaissances
 - composé de ressources plus ou moins structurées
 - Documents HTML
 - Images
 - Services
 - ...
 - avec des liens simples et universels
 - ... et ça fonctionne !

3

Extension



- Idée
 - associer aux ressources des informations utiles afin de favoriser leur exploitation
 - annotations / méta-données
- Questions
 - quelles ressources ?
 - des informations
 - utiles pour qui ?
 - utiles pour quoi faire ?

4

Première vision



- Web = espace documentaire ultime (pour l'instant)
 - naturellement distribué et bruité
 - un peu hypertexte
 - essentiellement dédié à la lecture
 - auquel manquent des choses par rapport à l'hypertexte ultime
 - lecteur passif → lecteur actif
 - one to many → many to many
 - interopérabilité des marques
 - repérage communautaire
 - etc.
- Nécessités d'annotations (cf. cours précédent)
 - lecture,
 - enrichissement, liens,
 - discussion – commentaires,
 - etc.

5

Deuxième vision



- Des informations
 - partout, dans tous les sens, incohérentes
 - auxquelles on peut difficilement accéder sans médiateurs automatiques
 - moteurs de recherche
- Nécessité
 - de liens riches
 - de méta-données structurées, sûres, et signifiantes
 - pour les agents logiciels
- Solution : structurer sémantiquement tout ça
 - annotations sémantiques
 - assertions logiques
 - pour un web sémantique

6

Annotation / méta-donnée

- Définition
 - « quelque chose » que l'on ajoute à une ressource Web
- Conséquences
 - cette information doit d'une manière ou d'une autre être structurée (utilisable) et descriptive de la ressource ou de son utilisation
 - la ressource en question doit exister et pouvoir être exploitée sur le Web indépendamment des informations qui lui sont associées

7

Exemples d'annotations

8

Exemples d'utilisation web

- Annoter
 - Construire des cartes de navigation
 - Localiser et étiqueter les informations intéressantes dans un document (termes, concepts)
 - Commenter sa lecture
 - Évaluer le contenu d'un document
 - Décrire une ressource à l'aide du DC
 - Etc.
- Pour
 - Construire automatiquement des index, des tables des matières, ...
 - Recherche d'information : requêtes, filtrage
 - Résumer automatiquement des documents
 - Classifier des documents
 - Supporter la navigation
 - Adapter des documents
 - Etc.

9

Annotation / méta-donnée

- Méta-donnée
 - plutôt attachée à une ressource identifiée en tant que telle sur le web
 - plutôt une pertinence *a priori*
 - plutôt saisie suivant un schéma
 - Exemples
 - description normalisée d'un service web, auteur d'un document, qui permettront de mettre en place des inférences
- Annotation
 - plus située au sein de cette ressource
 - écrite au cours d'un processus d'annotation/lecture.
 - Exemple
 - commentaire libre associé à un fragment d'une page web – quelques mots, un paragraphe – déterminé au besoin.
- Distinction non définitive, pas forcément nécessaire
 - Annotation dans un document → métadonnée sur un serveur
 - A retenir
 - Annotation située
 - méta-donnée plus indépendante, voire ressource en tant que telle.

10

A/MD : Critères (non indépendants)

- types de ressources concernés, plus ou moins fragmentées,
- automatisation plus ou moins marquée de la mise en place,
- structuration plus ou moins forte de leur « schéma »,
- tâches qu'elles soutiennent (ou sous-tendent),
- utilisation qui en est faite par les agents logiciels, plus ou moins « intelligente ».

11

Discussion

- Web vs web sémantique
 - différence subtile, graduelle
 - différentes acceptions de « sémantique » (cf. Cours 1)
- Annotation informelle vs formelle
 - idem : visions caricaturales, mais pratiques
 - formalisme à divers degrés, sémantique à divers degrés
 - web (un peu) sémantique vs web sémantique
- Dans la suite
 - Annotations dans un web (un peu) sémantique
 - Annotations dans le web sémantique
 - Mais d'abord [[RDF](#)]
 - Et puis [[Topic maps](#)]

12

Annotations informelles

- Associer une information à un fragment de page web
 - le plus souvent, du texte.
- Annotation libres
 - la machine n'a jamais accès au contenu « ultime »
- Annotation
 - ancre
 - Lieu de l'annotation, fragment
 - *document entier*
 - *placement*
 - attributs
 - Valeur de l'annotation
 - *contenu*
 - *forme*
 - *etc.*

13

3 grandes fonctionnalités

(Denoue)

- Création des annotations
 - document entier / passage d'un document
- Représentation, stockage, accès
 - dans / hors document
 - local / distant
 - centralisé / distribué
 - contrôle d'accès : privé, public, groupe
- Visualisation
 - à côté du document / dans le document
 - référencé ou non dans le document
- ...évidemment, conséquences sur l'architecture

14

Propriétés attendues...

- ...pour un système d'annotation informelle (annoter / visualiser des annotations de documents sur le web)
 - Léger
 - Efficace
 - Non-intrusif (transparent)
 - Indépendant d'une plateforme
 - « scalable » (supportant l'augmentation de charge)

[VP99] V. Vasudevan & M. Palme, « On Web Annotations: Promises and Pitfalls of Current Web Infrastructure » <http://www.objs.com/survey/annotations-hicss.doc>

15

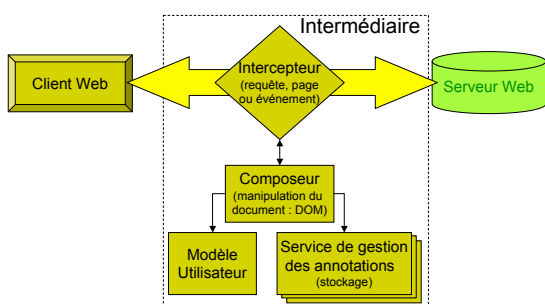
Contraintes architecturales

- Web
 - contraintes réseau : charge
 - langages : HTML / XML / RDF
 - informations réparties
 - variabilité, contrôle simplement local
 - multiples utilisateurs
- aussi bien pour les méta-données que pour les documents

[VP99] V. Vasudevan & M. Palme, « On Web Annotations: Promises and Pitfalls of Current Web Infrastructure » <http://www.objs.com/survey/annotations-hicss.doc>

16

Notion d'intermédiaire



D'après Desmontil / Jacquin (IRIN) d'après [VP99]

17

Dans l'intermédiaire

[VP99]

- Intercepteur
 - de requête (ex. URL)
 - de page (ex. page HTML)
 - d'événement (ex. telle page est en chargement)
- Service de gestion des annotations
 - API de stockage, filtrage, etc.
 - Schéma ?
- Compositeur
 - Stylistique
 - orienté présentation
 - gestion des version
 - gère en plus des versions de documents et annotations
 - sémantique
 - calculs fondés sur la sémantique des annotations
 - web sémantique

18

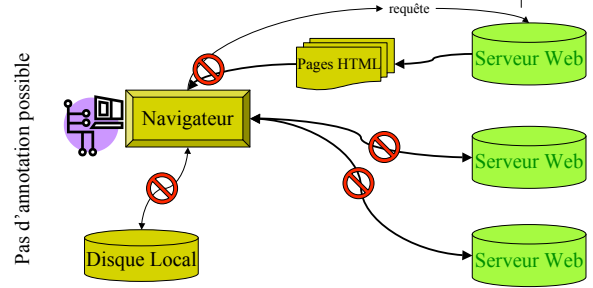
Où placer l'intermédiaire ?



- Deux grandes catégories
 - Serveurs mandataires (proxy-based systems)
 - Intermédiaires sur les clients (browser-based systems)
- Illustrations

19

Architectures des systèmes d'annotation sur le Web



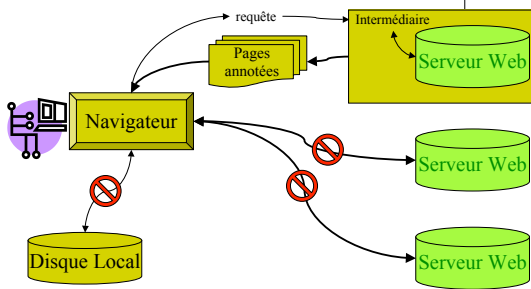
© Desmontil / Jacquin (IRIN) – 2000

20

Architectures des systèmes d'annotation sur le Web



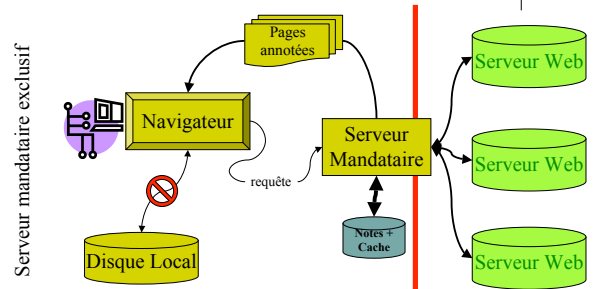
Serveur d'annotations dédié à un site



© Desmontil / Jacquin (IRIN) – 2000

21

Architectures des systèmes d'annotation sur le Web



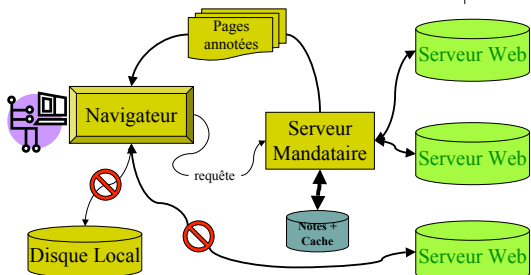
© Desmontil / Jacquin (IRIN) – 2000

22

Architectures des systèmes d'annotation sur le Web



Serveur mandataire facultatif

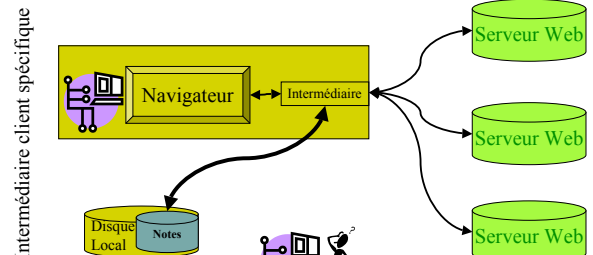


© Desmontil / Jacquin (IRIN) – 2000

Cf. CritLink

23

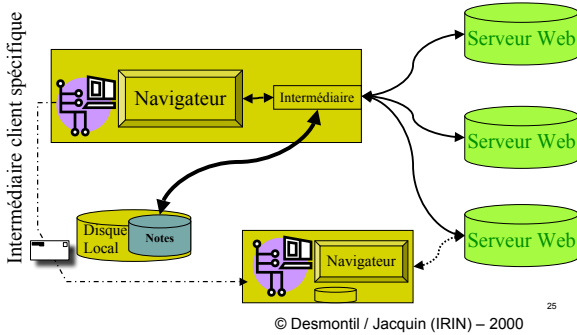
Architectures des systèmes d'annotation sur le Web



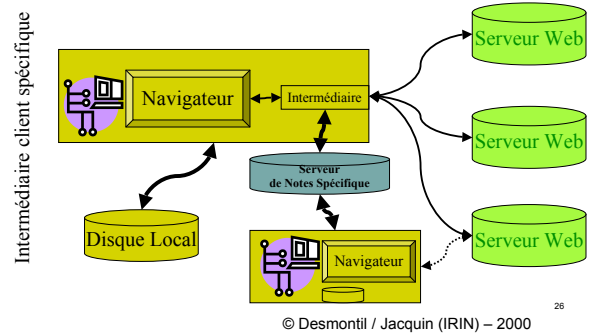
© Desmontil / Jacquin (IRIN) – 2000

24

Architectures des systèmes d'annotation sur le Web



Architectures des systèmes d'annotation sur le Web



Discussion

- L'architecture dépend de la fonction souhaitée
- Besoin de liens bidirectionnels
 - intermédiaires pour gérer le supplément

27

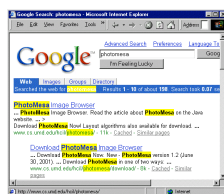
Discussion

- Propriétés proxys
 - Pas d'annotation de documents locaux
 - pas bon pour les espaces de connaissances
 - Lenteur
 - goulot d'étranglement
 - Altération obligatoire du document d'origine
 - pbs techniques : DHTML, javascript, applets, frames, etc.
 - Partage des annotations facile
 - Interface universelle + serveur avec logique applicative
 - utils indépendants du système et du navigateur
 - pas d'application à installer
 - facilité de mise en place
- Client
 - Inverse !
 - IHM plus évoluée (plug-in)

28

Systèmes d'annotation

- Remarques
 - Mosaic avait déjà des fonctionnalités d'annotation en 1994 !
 - Highlight à la Google



29

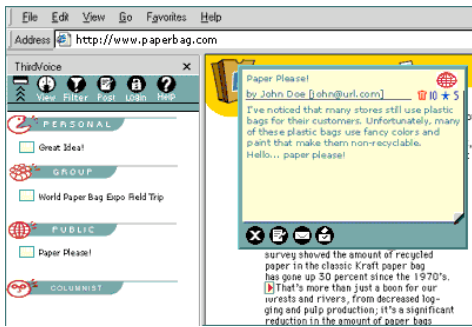
Classifier les systèmes

- Classification
 - Type d'annotations / visualisations
 - Proxy/browser
 - Granularité
 - Partage
 - Stockage externe / interne
 - Discussion / notification
 - etc.
- Voir aussi [Garfunkel 99]

[Garfunkel 99] <http://look.boston.ma.us/garf/webdev/annotate/software.html>

30

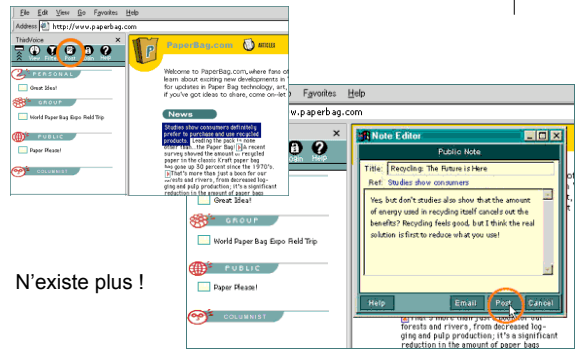
ThirdVoice



[Denoue]

31

ThirdVoice



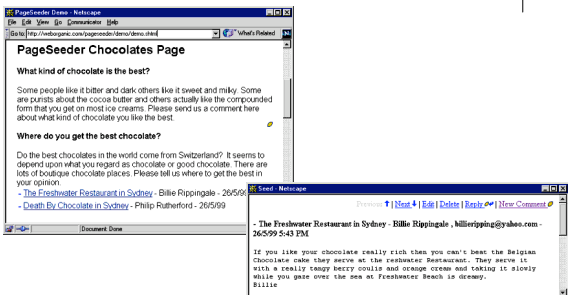
[Denoue]



N'existe plus !

31

PageSeeder

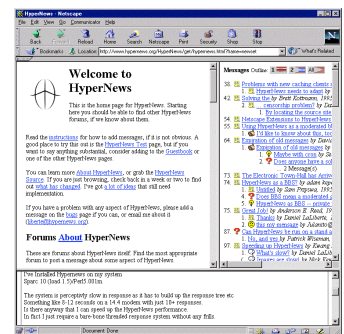


[Denoue]

33

HyperNews

- Commentaires sur pages web
- Textes + types + discussion
- Notion de co-navigation

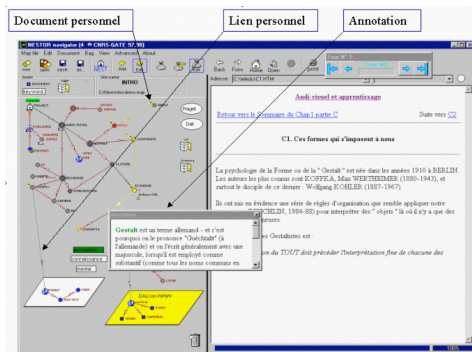


[Denoue]



33

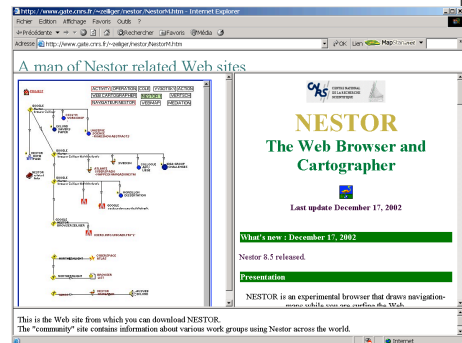
Nestor (Zeiliger)



[Denoue]

35

Nestor



[Denoue]



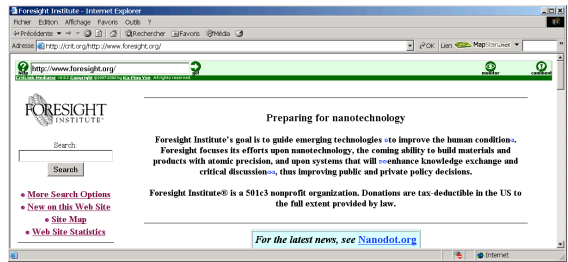
36

Critlink (crit.org)

- Foresight Institute. CritSuite was built to bring to the Web the virtues of the Udanax systems required to support critical discussion. CritSuite is the first 3rd party annotation system for the Web, and is the only one that's
 - open-source & open protocols
 - based on general bi-directional linking
 - allows comments on comments to any depth
 - works within web standards, and therefore on all browsers
 - is inherently uncensorable.
 - lets you ask to be notified when a page you care about is annotated.
- CritSuite
 - CritLink, CritMap, CritMail

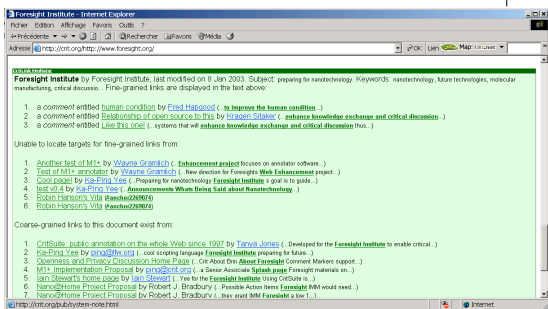
37

Critlink



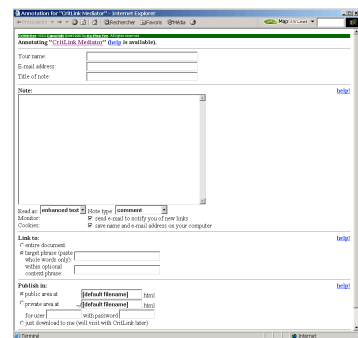
38

Critlink



39

Critlink



40

Autres systèmes

- Systèmes plus ou moins
 - commerciaux
 - en fonctionnement / abandonnés
 - orientés annotation / filtrage / visualisation
 - orientés collaboration / discussion / partage
- Exemples
 - Annotation System for Semantic Web
 - Internet Explorer
 - PageSeeder, Imarkup
 - PHAROS, Web4groups
 - Systèmes / PICS (Platform for Internet Content Selection)
 - Yawas
 - ... et bien d'autres

41

Discussion Développement des systèmes

- Semble fonctionner dans des environnements / communautés limités
 - intranet d'entreprise : systèmes commerciaux
 - chercheurs
 - ...
- Pas d'adoption à grande échelle
 - Deux grands types d'explications
 - Challenge technique
 - Passage à l'échelle
 - Challenge social
 - Et la collaboration ?
 - Si on peut faire la même chose d'une autre manière, pas vraiment de besoin

42

Discussion

Passage à l'échelle ?

| Denoue



- Contact du serveur d'annotation à chaque nouveau document chargé
- Vers des millions d'utilisateurs ?
- Solutions
 - répartition / réplication (cf. newsgroups USENET)
 - ne stocker que les liens vers les annotations
 - tables de hash sur le client
 - ...

43

Discussion

Modification de ressource ?



- Disparition d'une ressource
 - Effacer l'annotation orpheline ?
 - et ses réponses ?
- Modification d'une ressource
 - enjeux du positionnement robuste
 - Voir par exemple
 - Brush, Barger, Gupta, Bigbee, Loehr, Harpe. Robust Annotation Positioning in Digital Documents. CHI2001
 - Denoue, Vignollet. Annotation in the wild. WS 22 / ECAI'02
 - ...

44

Discussion

Confidentialité / sécurité



- Un serveur mandataire exclusif
 - voit tout...
 - reçoit toutes les demandes du navigateur, donc voit passer toutes les demandes de pages
 - stocke tout, y compris les annotations personnelles
 - ... doit garantir la confidentialité
 - Cf. courrier ou agenda sur le web...
- Dans tous les cas : sécurité
 - système de protection des données (login, mot de passe)
 - annotations privées
 - annotations destinées de groupe

45

Discussion

Partage



- Tout le monde n'est pas intéressé par l'annotation de tout le monde
 - à part l'auteur de la page
 - cf. papier : trop d'annotation tue l'annotation
- Liberté de n'importe quoi
 - cf. noms de domaines
 - ne voir que les annotations de confiance?
 - review ?
- Droits d'auteurs d'annotation ?
 - autorisations d'accès

46

Discussion

Ethique et droit d'auteur



- Annotations
 - modification design du site d'origine
 - serveurs mandataires : non contrôlée
 - s'apparente à du graffage
 - ThirdVoice
 - [[Say no to thirdvoice](#)] !
 - Microsoft Smartags
 - mettre des liens vers certains sites en évidence
 - marketing raté
 - Microsoft : contrôle, etc.
 - smarttags par défaut orientés vers Microsoft
 - levée de bouclier pour la liberté
 - web-designers cf. thirdvoice
 - grosses entreprises, procès (ou menaces)
 - Smarttags retirés

47

Discussion

Avenir



- De toute façon fondé sur XML / RDF
- Utilisation Xpointer / langages de requêtes XML
- Sélectionner les annotations que l'on veut voir
 - filtrage / DVP
- Fonctionnalités supplémentaires d'outils de gestion de ses ressources de navigation
 - intégration avec autres outils du web
 - ex. weblogs
 - espace de connaissances, cf. Nestor
 - une application tueuse socialement acceptée ?
 - pour une nouvelle annotation
- Annotations de pages web dans le Web Sémantique
 - Annotea représente une étape

48

Qui a mis du RDF dans mon HTML ?



www.semanticweb.org/lswc

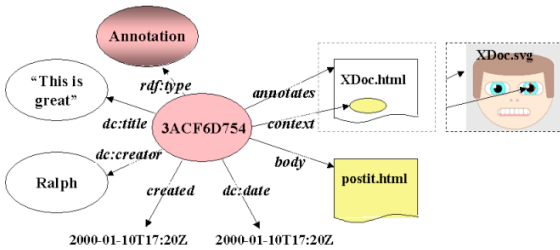
voir aussi les résumés annotés

Annotea



- W3C
- Protocole de transfert d'annotation vers/de un serveur central
- Basé sur RDF
- Annotation de ressources
 - HTML
 - XML : XHTML, SVG, MathML, ...
- Un schéma minimal, extensions possibles
- Implantations
 - Amaya
 - Annotzilla
 - Zope
 - ...

Schéma annotea de base



<http://www.w3.org/Talks/2001/1021-KCAP/>

Annotation en RDF



```
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:a="http://www.w3.org/2000/10/annotation-ns#"
xmlns:d="http://dublincore.org/">
<a:Annotation about="http://.../annotation/3ACF6D754">
<d:creator>Ralph</d:creator>
<d:title>This is great</d:title>
<a:annotates resource="http://.../XDoc.html" />
<a:body resource="http://.../body/postit.html" />
</a:Annotation> </RDF>
```

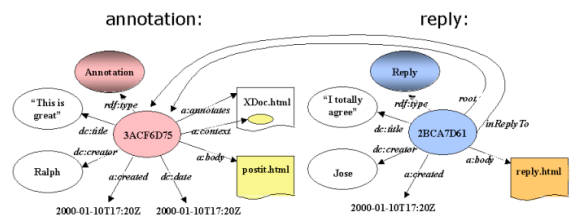
<http://www.w3.org/Talks/2001/1021-KCAP/>

Amaya



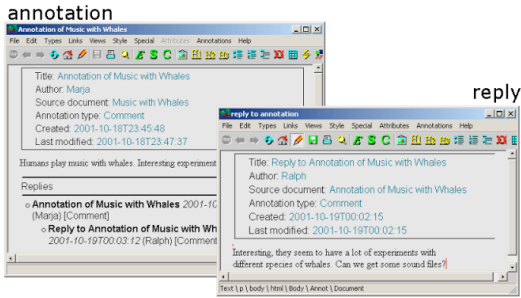
<http://www.w3.org/Talks/2001/1021-KCAP/>

Extension répondre à une annotation



<http://www.w3.org/Talks/2001/1021-KCAP/>

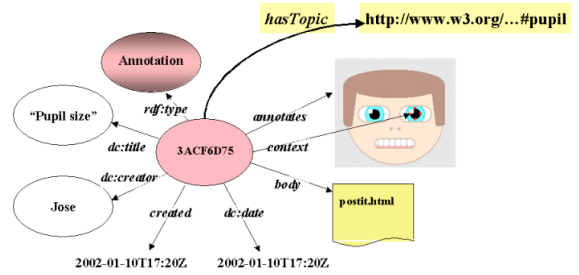
Amaya : réponses



<http://www.w3.org/Talks/2001/1021-KCAP/>

55

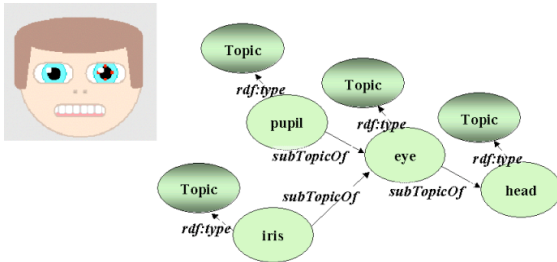
Extension préciser une annotation



<http://www.w3.org/Talks/2001/1021-KCAP/>

56

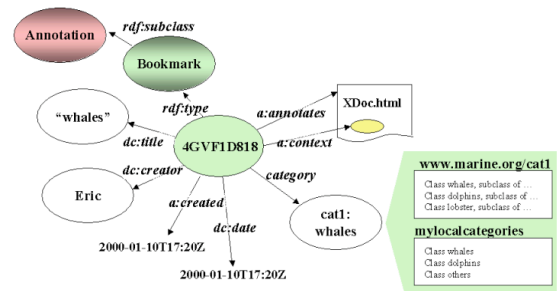
Extension Classifier les topics



<http://www.w3.org/Talks/2001/1021-KCAP/>

57

Extension étendre le schéma



<http://www.w3.org/Talks/2001/1021-KCAP/>

58

Annotea : requêtes



- Algae
 - Langage de requête pour RDF
 - Requêtes sur les annotations avec leurs propriétés / valeurs
- Exemple

(namespace (a <http://www.w3.org/2000/10/annotation-ns#>) ask '(a::annotates ?annot <http://example.com/page.html>) (?property ?annot ?value)) :collect '(?annot ?property ?value))

<http://www.w3.org/Talks/2001/1021-KCAP/>

59

Annotea : commentaire



- En cours
 - Annotation de ressources évolutives / révision d'annotation
 - Spécialisation pour applications
 - IHM (!)
 - filtrage des annotations
 - autres schémas...
- De l'informel vers le formel
 - Passage au niveau connaissances
 - Web sémantique

60

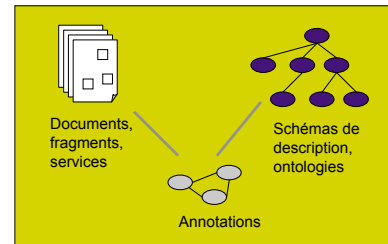
Web sémantique



- Au fait, c'est quoi le [[web sémantique](#)] ?
- Résumé
 - Des langages pour décrire des ressources de façon plus ou moins formelle de telle sorte que des machines puissent (également) utiliser ces ressources.
 - Pour rendre de nouveaux / meilleurs services aux utilisateurs du web.

61

Annotations « sémantiques »



Différents niveaux de connaissances

62

Deux grandes catégories d'usage



- Recherche d'information
 - toute méta-donnée informatique liée à une ressource représente de fait un index pour cette ressource, et peut être utilisée comme tel
 - Rechercher des annotations-concepts
 - Utiliser les ontologies pour transformer les requêtes / réponses
 - Répondre à des questions
- Composition
 - de documents, de fragments de documents, de services en vue de construire de nouveaux documents (cours à la carte, catalogues, réponses à une question, etc.) ou de nouveaux services (hôtel + spectacle)
 - Adaptation
 - Ex. construire un cours automatiquement
 - Utilisateur pas forcément conscient de l'utilisation de la ressource / MD
- Mais toujours
 - Consultation des annotations
 - Discussion, etc.

63

Annotations « sémantiques »



- Annotations
 - formelles, explicites, permanentes
- Outils d'annotation basés sur les ontologies
 - « Ontology based annotation tool »
- Critères pour les annotations et les systèmes
 - Types de ressources concernées
 - Automatisation plus ou moins marquée de la mise en place
 - Structuration plus ou moins forte des « schémas » de description
 - Tâches soutenues (sous tendues), utilisation plus ou moins formelle qui en est faite
 - Etc.

64

Annoter sur le WS



- Manipuler
 - Documents et fragments
 - Connaissances de description
 - Ontologies (concepts, relations et règles)
 - Tous schémas de structuration de connaissances
- Pour mettre en place des annotations
 - Localisation de la ressource
 - document, fragment
 - Description des annotations comme connaissances
 - instanciation de concepts

65

Des langages



- Langages de description des documents
 - Monde documentaire
- Formalismes et langages de représentation des connaissances
 - Ontologies / instances
 - Monde de la représentation de connaissances
 - logique de description, graphes conceptuels, frames
- Pour le web sémantique
 - HTML, XML, RDF, RDFS, DAML+OIL, OWL, SHOE, ...

66

Une ingénierie des connaissances documentaires



- Pour une application
 - le bon langage
 - expressivité / possibilités d'inférences
 - une bonne représentation des connaissances
 - exprimée dans ce langage sur un certain domaine
 - ontologie des moteurs à essence, des parties de cours en ligne
 - Dublin Core, ...
 - une application d'annotation
 - annotation automatique
 - annotation manuelle / semi-automatique
 - intervenants spécialistes
 - un système complet
 - sur le web
 - distribution des annotations et des calculs
 - rendant un vrai service

67

Pas évident



- Dans la suite
 - SHOE
 - WebKB
 - KAON
 - Applications diverses
 - Discussion

68

SHOE : modèle



- Simple HTML Ontology Extensions
- Eléments supplémentaires pour décrire dans des pages web
 - des ontologies
 - des annotations
 - des inférences
- Ontologies SHOE
 - classification
 - relations typées
 - inférences (sans négation)
 - liens entre ontologies : extension, versions
- Une ontologie de base

<http://www.cs.umd.edu/projects/plus/SHOE/>

69

SHOE : modèle



- Annotations SHOE
 - entités
 - instances de concepts
 - relations entre entités
 - notion de claim
 - assertions locales à la page
- Exemples
 - [Ontologie](#)
 - [Annotations](#)

70

SHOE : applications



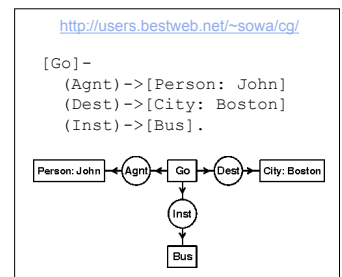
- Interface graphique d'annotation
- Robot pour récupérer les assertions, et les stocker dans une base
- Requêtes à cette base
 - Interface graphique de construction de requêtes.
- Remarques
 - SHOE plus ou moins arrêté (aucune applet ne marche)
 - Export des ontologies SHOE vers DAML.

71

WebKB (Martin / Eklund)



- Web Knowledge Base – <http://www.webkb.org/>
- Basé sur les graphes conceptuels
 - Sowa
 - Treillis
 - concepts
 - relations
 - Instances
 - graphes conceptuels
 - Inférences
 - recherche de graphe



72

WebKB

- Ontologie générale construite à partir de Wordnet.
- Annotation de ressources web
 - Dans les pages en RDF
 - Sur un serveur
- Requêtes sur le serveur
 - Recherche d'information
 - Interface très complexe
- Utilisé ?

73

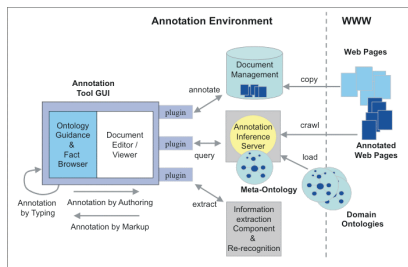
KAON

- The **K**arlsruhe **O**ntology and Semantic Web Tool Suite <http://kaon.semanticweb.org>
- Outils pour
 - Gestion des ontologies
 - RDF / relationnel
 - Génération de portail basé sur les ontologies
 - Construction d'ontologie
 - manuelle + traitement de corpus
 - Robot collecteur de triplets RDF sur le web
 - Gérer tout ces outils : KAON server
- Utilisé dans un nombre respectable de projets

74

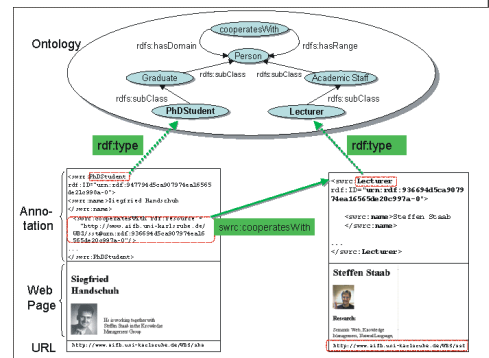
CREAM

- CREAM (CREATING Metadata)
 - Framework pour l'annotation / édition
 - Notion de méta-données relationnelle (ie. basées sur les ontologies)



75

Approche CREAM



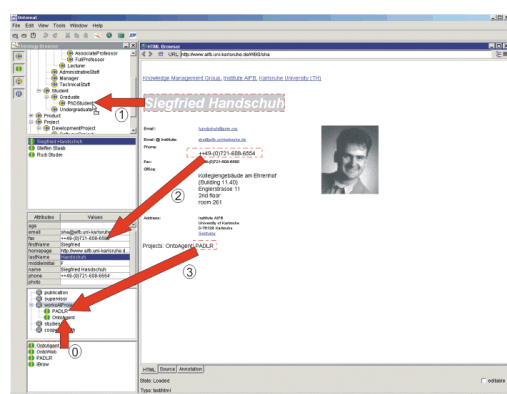
76

Ontomat

- Ontomat
 - <http://annotation.semanticweb.org/tools/ontomat>
 - Implémentation de CREAM pour création / annotation de pages web
 - Navigateur HTML
 - Navigateur Ontologie
 - Instances DAML+OIL sur fragments

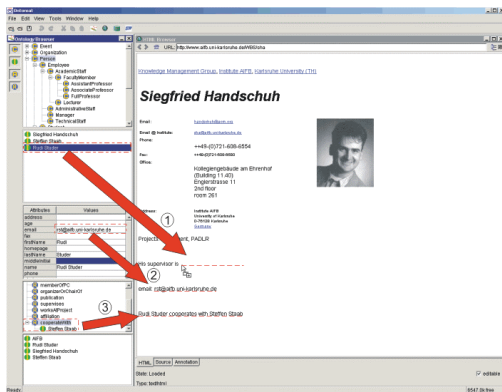
77

Ontomat



78

Ontomat



79

Ontoprise

- Ontobroker <http://ontobroker.semanticweb.org/>
 - application : KA2
- <http://www.ontoprise.de/>
 - OntoBroker
 - OntoCollect / OntoAnnotate
 - version commerciale de Ontomat
 - SemanticMiner
 - OntoEdit
 - OntoOffice



80

RDFweb

- <http://rdfweb.org/>
- Schémas RDFS / descriptions RDF
- Robots → serveur / base de données
- Applications
 - FOAF (Friend of a Friend)
 - Description de personne (propriétés DAML) + propriété SeeAlso vers d'autres personnes
 - Recherche de personnes, de liens entre personnes, etc.
 - Co-Depiction project
 - foaf:depiction
 - possibilité d'annotation de fragments d'images SVG



81

Outils d'annotation

- COHSE (Conceptual Open Hypermedia Services Environment)
 - <http://www.ecs.soton.ac.uk/~tmb/cohse/annotator/>
 - Ontologie DAML
 - Fragments de pages web HTML
 - Serveurs d'annotation dédiés.
 - Génération d' « hypermédias ontologiques »
 - proposition de liens...
- SMORE : Semantic Markup, Ontology and RDF Editor
 - <http://www.mindswap.org/~aditka/editor2.shtml>
 - Schémas RDF
 - Ontologies extérieures ou construction d'ontologies
 - Annotation et création de pages Web
 - Moins orienté connaissance que KAON
 - Annoter d'abord, construire l'ontologie plus tard
 - Portail pour informations liées aux annotations
 - Version pour annotation d'image et de courrier électroniques



82

Autres projets / schémas

- RDF Pic (<http://jigsaw.w3.org/rdfpic/>)
 - Images / RDF / Dublin Core étendu (schéma subject : portrait, architecture, etc. + schéma technique (appareil, film...))
 - Serveurs dédiés, stockage images + MD (jigsaw)
- Questions sur les annotations en RDF
 - MedCertain
 - Annotation de ressources médicales sur le web
 - Validation / confiance
 - EARL
 - Evaluation And Report Language
 - Assertions
- Artequakt
 - Génération automatique de biographies à partir du web
- GDA (Global Document Annotation)
 - Schéma XML
 - propriétés linguistiques, conversationnelles, etc.
 - Sur pages HTML



83

Web sémantique et P2P

- Edutella <http://edutella.jxta.org>
 - échange de méta-données sur un réseau Pote à Pote
 - première application
 - échange de ressources entre universités
 - actuellement
 - modèles de données
 - langages de requêtes (5 de complexité croissante)
 - construit au dessus de JXTA
 - plateforme P2P standard soutenue par Sun
 - différents niveaux d'abstraction
- SWAP
 - Semantic Web and Peer to peer



84

Discussions



- Sur la sémantique du lien d'annotation
 - Annoter une ressource ou décrire le monde ?
- Sur la notion d'auto-indexation
 - annotations *a posteriori* sur les ressources
 - vs conception de page / intégration de méta-données
 - est-on expert de la description de ses documents ?
- L'ontologie dirige ?
 - annoter en utilisant une ontologie
 - annoter en construisant/étendant
 - annoter d'abord, formaliser ensuite
 - SMORE vs OntoMat

85

Mise en place



- Automatique
 - dépend des genres documentaires
- Semi-automatique
 - méthodes statistiques pour proposer des méta-données
 - Amilcare
 - Système adaptatif d'extraction d'information
 - Apprentissage
 - Klarity (www.klarity.com.au)
 - champs DC
- Manuelle

86

Discussion



- Annotation par l'usage
 - MUNETTE
- Maîtriser l'ontologie et ses conséquences (inférences) pour pouvoir annoter
 - Qui est expert ?
 - Qui voit les méta-données ?
 - Mais elle est où l'ontologie ?
 - Portail banalisé
- L'expérience musicale
 - mp3

87

Remarques



- Vu d'un point de vue technique
 - l'annotation est un *mécanisme* pouvant exprimer beaucoup d'applications de gestion de BD et de documents
 - différentes manières de voir les choses
- La différence entre une application du WS et une en dehors peut simplement provenir de la façon dont elle est présentée

88

Conclusion



- Annotation sur le web (sémantique)
 - toujours pas de système diffusé et utilisé à grande échelle
 - Trop tôt ?
 - Trop grand ?
 - d'autres communautés que celle du WS s'intéressent à la question, et ont des choses intéressantes à dire
 - niveau méta-données
 - niveau collaboration
- RDF largement utilisé
 - joue son rôle
 - fragments XML / fragments documentaires
 - robots (crawler)
 - parcours du web
 - extraction de triplets RDF
 - petit début des moteurs de recherche basés sur les triplets
- RDFS / OWL

89

Conclusion



- Approches de l'annotation sur le WS
 - Annotea
 - approche locale, un schéma après l'autre
 - KAON
 - approche globale
- L'annotation informelle va s'y glisser
 - ou rester à XMLS ?
- Google, Yahoo, Amazon vont-ils tuer le WS ?

90

Conclusion



- Dans tous les cas, un monde XML
- Description de document XML
 - DTD / schéma XML
 - figement de méta-données
- Méta-données XML
 - RDF
 - extension document pour certains usages
- Où est la structure du document ? Où s'arrête le document ?

91

Remerciements – emprunts



- Laurent Denoue (FXPal)
- E. Desmontil / C. Jacquin (IRIN)

92

Resource Description Framework

Pierre-Antoine Champin
champin@lisi.univ-lyon1.fr



Plan

- RDF (modèle)
 - Resource
 - Description
- RDF (syntaxes)
- RDF Schema
- DAML+OIL



94

Ressource



- « Ressources : moyen d'action, réserve de force, d'habileté, etc. »
- Initialement, RDF devait décrire les ressources du Web (méta-données)
- Plus généralement, tout objet identifié impliqué dans une tâche peut être considéré comme une ressource

95

Uniform Resource Identifier



- URL : <http://www.w3.org/RDF>
 - sémantique « opérationnelle »
- URN : urn:ietf:rfc:2414
 - sémantique « conventionnelle »
- URI : généralisation des URLs et URNs

- Préférence du W3C pour les URLs...
...ce qui suscite quelques [polémiques](#).

96

Notation des URIs

- *Qname* en "Notation 3" : préfixe:nom ou préfixe est associé à un URI

my_pref:my_name = <http://www.my-host.org/my_path#my_name>
avec
@prefix my_pref : <http://www.my-host.org/my_path#>

- Préfixes standards

@prefix rdf : <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix rdfs : <http://www.w3.org/2000/01/rdf-schema#>
@prefix dc : <http://purl.org/dc/elements/1.1/>
@prefix pac : <http://www710.univ-lyon1.fr/~champin/>



97

Description (1)

- Une ressource est décrite par ses **propriétés** (couples attribut-valeur)

cette présentation :
auteur → Pierre-Antoine Champin
version → 1.1

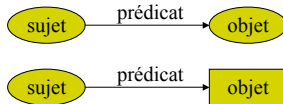
- Triplet : sujet prédicat objet.



98

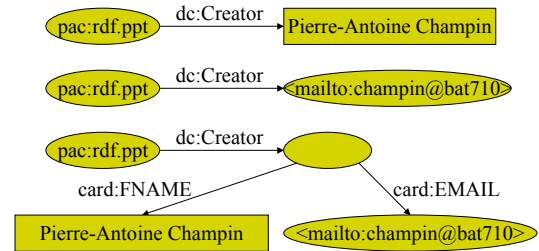
Description (2)

- En RDF,
 - sujet = URI
 - objet = URI ou littéral
 - prédicat = URI



99

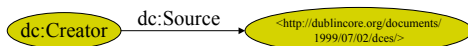
Description (3)



100

Réflexivité (1)

- Tous les éléments du langage RDF sont réifiés, donc peuvent être décrits / annotés avec RDF
- Les propriétés sont des ressources



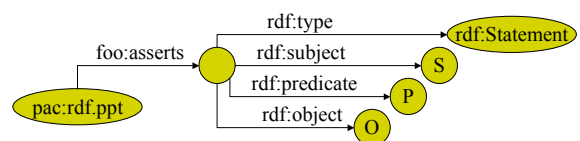
101

Réflexivité (2)

- Les triplets sont des ressources



peut être décrit par



102

Plan

- RDF (modèle)
- RDF (syntaxes)
 - XML/RDF
 - Notation 3
 - Xlink, etc.
- RDF Schema
- DAML+OIL

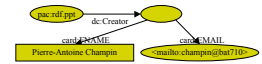


103

Syntaxe XML/RDF

- Recommandation RDF (M&S)
<http://www.w3.org/TR/REC-rdf-syntax>

```
<rdf:Description rdf:about="http://www710.univ-lyon1.fr/rdf-pres.ppt">
  <dc:Creator>
    <rdf:Description>
      <card:EMAIL rdf:resource="mailto:champin@bat710"/>
      <card:FNAME> Pierre-Antoine Champin </card:FNAME>
    </rdf:Description>
  </dc:Creator>
</rdf:Description>
```

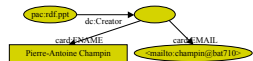


104

Syntaxe Notation 3 (N3)

- Proposée par Tim Berners-Lee
<http://www.w3.org/DesignIssues/Notation3.html>

```
pac:rdf-pres.ppt dc:Creator
[ card:EMAIL <mailto:champin@bat710>;
  card:FNAME "Pierre-Antoine Champin" ].
```



105

Syntaxe Xlink (etc.)

- « Moissonner » du RDF avec Xlink
<http://www.w3.org/TR/xlink2rdf/>
- Idée d'un interpréteur RDF plutôt que d'un analyseur RDF (étendre les DTDs et schémas XML)



106

Plan

- RDF (modèle)
- RDF (syntaxes)
- RDF Schema
 - Propriétés
 - Classes
- DAML+OIL



107

RDF Schema

- Candidate recommendation (27/3/00)
<http://www.w3.org/TR/rdf-schema>
- Décrire des vocabulaires RDF avec RDF
 - Hiérarchie de propriétés (prédicats)
 - Hiérarchie de classes

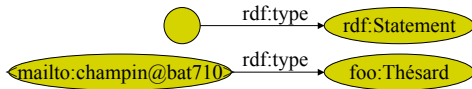


108

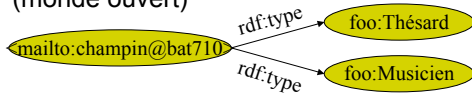
rdf:type



- Relation entre une ressource et une classe



- Possibilité d'avoir plusieurs types (monde ouvert)

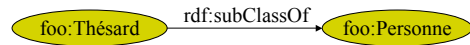


109

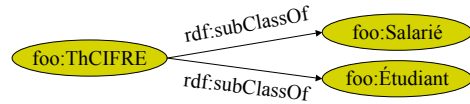
rdfs:subClassOf



- Relation d'inclusion entre classes



- Possibilité d'avoir plusieurs super-classes



110

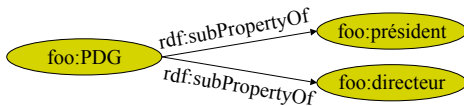
rdfs:subPropertyOf



- Relation d'inclusion entre propriétés (au sens d'inclusion de relations)



- Possibilité d'avoir plusieurs super-propriétés

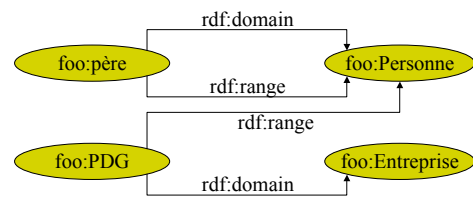


111

rdfs:domain, rdfs:range



- Contrainte de type sur le sujet et l'objet d'un prédicat



112

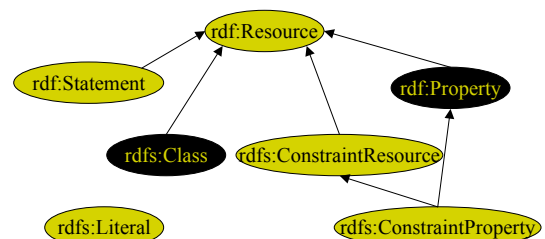
Autres propriétés



- rdfs:label
- rdfs:comment
- rdfs:seeAlso
- rdfs:isDefinedBy
rdfs:subPropertyOf rdfs:seeAlso

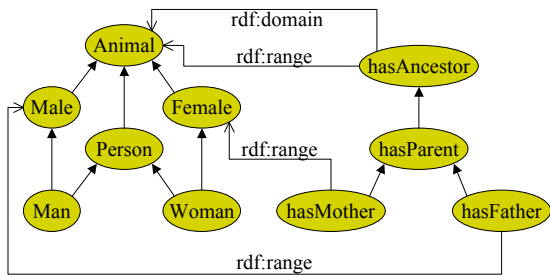
113

Classes



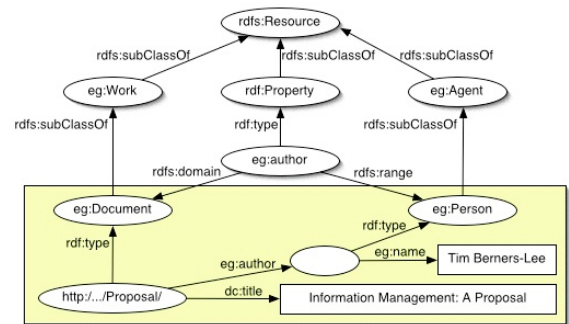
114

Exemple de Schéma RDF



115

Autre exemple



Plan

- RDF (modèle)
- RDF (syntaxes)
- RDF Schema
- DAML+OIL
 - Présentation
 - Exemple

117

DAML+OIL

- OIL : projet européen
Langage de représentation d'ontologies basé sur les logiques de description
- DAML : projet américain (DARPA)
Représentation objet de connaissances à l'aide de RDF
- <http://www.daml.org/>

118

Ajouts DAML+OIL

- Disjonction de classe :
 - Male est disjointe de Female
- Restriction de classe :
 - Person est restreinte aux instances dont les parents sont des personnes
- Définition plus fine des propriétés
 - hasAncestor est transitive
 - hasFather est à valeur unique
- Cardinalités pour les propriétés
- Egalité entre classes
- ...

119

OWL (Ontology Web Language)

- <http://www.w3.org/TR/owl-features/>
(29/07/2002)
- W3C – révision de DAML+OIL
- Intégration d'ontologies
- OWL Lite
- Full OWL
- ...

[Retour]

URLs vs. URNs



- Les URLs sont beaucoup plus répandues que les URNs.
- Les variations d'architectures peuvent être masquée par les protocoles mis en œuvre.
<http://www.w3.org/TR/REC-rdf-syntax>
- ... ce qui suppose un déploiement très discipliné des URLs !...

121

URLs et identification



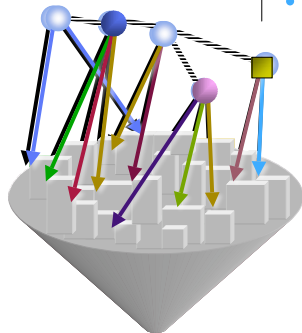
- <http://www710.univ-lyon.fr/~champin>
 - Une page HTML ?
 - Mon CV ?
 - Mon site web ?
 - Moi ?
- Indiqué par les méta-données accompagnant les données.
- Quelle URI pour la page HTML ?

[[Retour](#)]

Topic maps



- Principe
 - Couche sémantique ajoutée à un ensemble d'informations liées
- Eléments (TAO)
 - Topics
 - Associations
 - Occurrences



© Michel Biezunski 123

Topic maps



- ISO/IEC 13250:2000
- XTM
 - Format d'échange de TM en XML
- Topic maps vs? RDF
- Informations
 - <http://xml.coverpages.org/topicMaps.html>

[[Retour](#)]

ThirdVoice Les webmasters en colère



- Privacy
 - Traque sur Internet
- Cyber Trespassing
 - Utilisation de la bande passante du site annoté
- Graffiti (and vandalism)
- Exclusion
 - Exclusion de ceux qui ne peuvent utiliser ThirdVoice
- Note Content
 - 44 % seulement lié à la page
- No Choice
 - Pas de choix pour le créateur
- Copyright
 - Des créateurs de site bafoués



<http://www.worldzone.net/internet/pixelsntv/>

[[Retour](#)]

Web sémantique

Yannick Prié



Le Web sémantique : philosophie (N. Chung)



- Un web actuel extrêmement complexe
 - Un seul type de lien
 - Nécessité de médiateur
 - ...
- Objectifs
 - Revenir à une structure sémantiquement riche de type base de données orientée ressources...
 - tout en conservant des méta-données de base dans les documents,...
 - tout en conservant un caractère distribué, souple, évolutif, extensible, interopérable,...
 - avec une standardisation minimale.

127

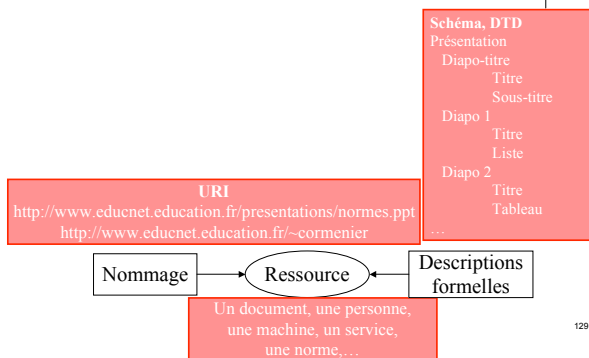
Alors on aura... (P. Laublet)



- Un immense espace d'échanges de ressources *entre machines* permettant à des utilisateurs d'accéder à de grands volumes d'informations et à des services variés (répondre à des requêtes, global computing, etc.)
 - Identification, exploitation et combinaison intelligente et automatique de ressources (documents, services)
 - Traitements automatisés des données et des connaissances
 - Gestion de l'hétérogénéité
 - Gestion de l'interaction avec l'utilisateur

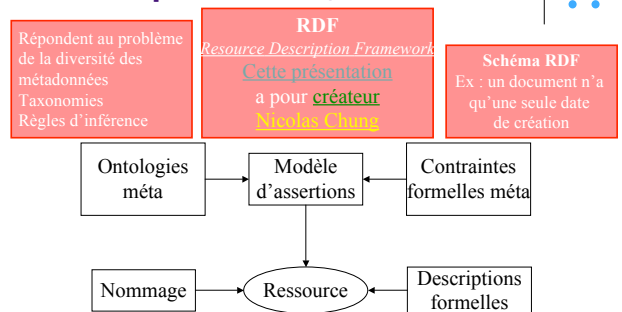
128

Le Web sémantique : les briques (P. Chung)



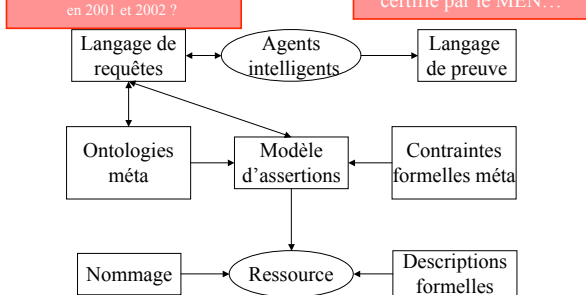
129

Le Web sémantique : les briques (P. Chung)



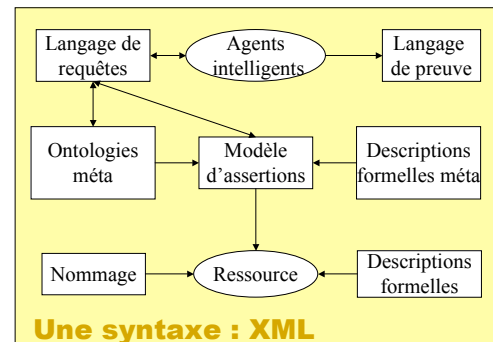
130

Le Web sémantique : les briques (P. Chung)



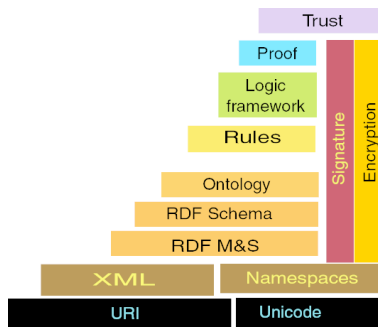
131

Le Web sémantique : les briques (P. Chung)



132

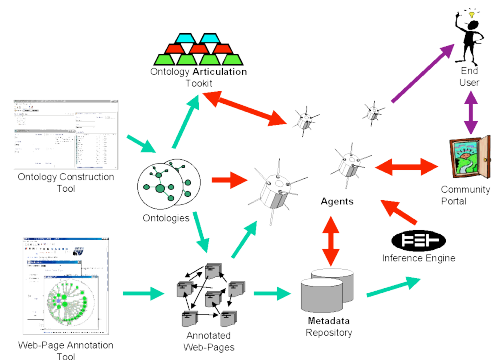
Le gâteau (TBL)



<http://www.w3.org/2002/Talks/09-ics-sweb-tbl/slide19-0.html>

133

The big picture (SemanticWeb.org)



134

Grands axes de réflexion

(Action spécifique Web sémantique – CNRS)

- Langages pour le Web sémantique
- Ontologies pour le Web sémantique
- Méta-données et annotations
- Intégration de sources de données hétérogènes
- Services dans le Web sémantique
- Adaptation et personnalisation

135

Les langages pour le Web sémantique

Les besoins

- Un langage de description des ressources du Web : contenu sémantique, fonctionnalités, concepts manipulés par un agent logiciel Web
- Un formalisme avec une *sémantique formelle* bien définie limitant les erreurs d'interprétation et permettant d'effectuer des traitements automatisés
- Un standard pour des raisons d'interopérabilité

Problème

- Quel(s) standard(s) ?
Compromis expressivité / complexité, simplicité

Existant

- Travaux du W3C : RDF, RDF-S (RDFCore), DAML+OIL, OWL
- Autres langages : SHOE, DAML-ONT, OIL, DAML+OIL, etc.

136

Les ontologies pour le Web sémantique

Les besoins

- Un vocabulaire structuré et homogène pour l'expression de requêtes et pour la description des contenus et des fonctionnalités
- Une connexion entre des agents logiciels (sources d'information, serveurs d'information, hétérogènes – *sémantique formelle*) et entre des utilisateurs humains (*sémantique interprétative*) et des agents logiciels

Les problèmes

- Définir une ontologie unifiant des schémas disparates de sources
- Construire une ontologie (méthodologie, gérer la coopération et les différents points de vue, gérer l'évolution)
- Trouver, définir et représenter simplement des correspondances sémantiques entre ontologies (à automatiser)

137

Les ontologies pour le Web sémantique

Travaux en Ingénierie des connaissances

- Conceptualisation universelle (SUMO) / partagée par une (plus petite) communauté et spécifique à une tâche
- Nature, niveau et langages de description des connaissances ontologiques
- Méthodes de construction à partir de textes (ASSTICOT Structuration de ressources terminologiques à partir de textes)
- Méthodologies propres au développement d'ontologies (METHONTOLOGY [Fernandez et al. 1997], méthode basée sur l'analyse de corpus [Aussenac-Gilles et al. 2000], méthode basée sur l'analyse différentielle [Bachimont 2000], etc.)
- Éditeurs d'ontologies (Ontolingua/Chimæra [Farquhar et al. 1995], Protege 2000 [Noy et al. 2000], OILEd [Bechhofer et al. 2001], WebOnto, OntoEdit [Maedche et al. 2000], WebODE [Aripirez et al. 2001], DOE [Troncy et al. 2002], etc.)

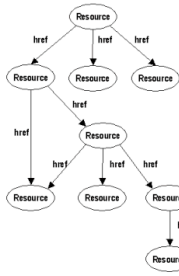
Existant

138

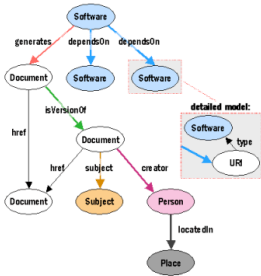
Méta-données et annotations



Le Web aujourd'hui



Le Web Sémantique



W3C Semantic Web Activity
Marja-Riita Koivunen and Eric Miller, Nov. 2001

139

Méta-données et annotations



Les besoins

- Des descriptions attachées à une ressource identifiée en tant que telle (méta-donnée) ou située au sein de la ressource (annotation)
- Des descriptions du contenu ou de la fonctionnalité des ressources du Web

Les problèmes

- Quel formalisme de représentation ?
- Quel vocabulaire utiliser ?
- Quelles descriptions pour quelles ressources ? Pour quelles utilisations ?
- Comment automatiser ces descriptions ?
- Comment établir les liens entre ces descriptions et les données ?

Existant

Technologie Web du W3C : RDF, RDF-S

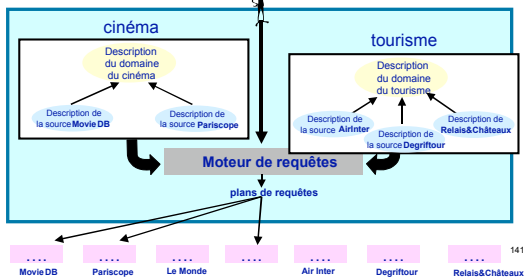
140

L'intégration de sources d'information



L'approche médiation centralisée

(PICSEL : <http://www.fr.fr/~picssel>)

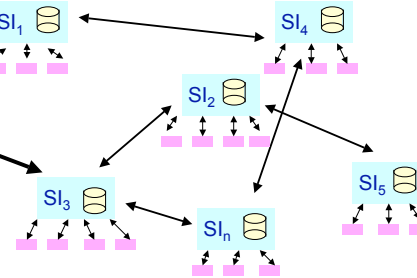
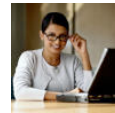


141

L'intégration de sources d'information



L'approche médiation décentralisée



142

Les services



Les besoins

- Vers des services complexes et non prédéfinis
- Automatisation et interopérabilité des services

Les problèmes

- Spécification des e-services
- Publication, découverte et sélection
- Contrôle et exécution
- Composition
- Infrastructure de médiation
- Sécurité des accès

Existant • Émergence de normes : SOAP, WSDL, UDDI, DAML-S

143

Personnalisation / adaptation



Les besoins

- Prise en compte de la diversité des utilisateurs, de leurs besoins, attentes, compétences et centres d'intérêts particuliers, avec ou sans modèle utilisateur / utilisation

Les problèmes

- Modélisation de l'utilisateur ou des utilisations
- Aide à la formulation de requêtes
- Filtrage des informations pertinentes en fonction de données utilisateurs
- Création de services et documents adaptatifs
- Protection de la vie privée
- Ergonomie des services et des documents
- Accès aux ontologies via le langage de l'utilisateur (requêtes, indexation)

Existant

- Travaux en ergonomie (sociologie, psychologie cognitive, ergonomie cognitive, ...)
- Travaux du W3C sur la gestion des données utilisateur

144

Le Web Sémantique : conclusion



- Le Web Sémantique est une série de standards en cours d'élaboration
- Points d'interrogation nombreux
 - le WS n'existe-t-il pas déjà ?
 - usages ?
 - HTML simple → succès du web, quoi pour le WS ?
 - ontologies
 - tension entre le caractère global du Web et la limitation nécessaire d'une ontologie.
 - deux directions complémentaires :
 - une ontologie minimaliste pour tous documents
 - des ontologies riches mais limitées à un secteur.

145

Liens



- Action Spécifique Web Sémantique
<http://www.lalic.paris4.sorbonne.fr/stic/as5.html>
- W3C – Semantic Web
<http://www.w3.org/2001/sw/>
- SemanticWeb.org
<http://www.semanticweb.org/>
- Ontoweb
<http://www.ontoweb.org/>

146

Remerciements – emprunts



- Nicolas Chung
« Ressources éducatives numériques : normes et standards de métadonnées »
- Philippe Laublet
« AS 32 – Web Sémantique »

[[Retour](#)]