

Introduction à quelques normes du Web

Présenté dans le cadre du cours
“**Modélisation numérique**”
de Seth Van Hooland,
Département SIC,
Université Libre de Bruxelles

*par Christophe Dupriez
dupriez@destin.be*

17/11/2010

1

Informaticien, pas documentaliste...

Introduction = mise à l'étrier, pas plus...

Normes = ensemble de conventions établies par une communauté dans un objectif de pérennité et d'échange... Dépendance vis à vis de l'évolution de cette communauté.

Qui/que servons nous?

- Les banques de données?
- Les bibliothèques?
- Les réseaux d'institutions?
- La littérature?
- Les entreprises?
- Les personnes?



2

La question est un peu existentielle mais comme les normes informatiques sont des fondations de notre travail, il est important que nous partagions une perception de ce travail.

Si on parle de notre métier, les gens nous associent au « nerd » qui plonge dans les réseaux informatiques, au bibliothécaire jaloux de ses documents poussiéreux ou au passionné de littérature seul dans son coin. Personnellement, ma motivation vient de l'apport aux personnes (au public, aux professionnels des institutions et des entreprises).

Ce qui est essentiel, c'est la relation entre l'information et les personnes . Et, comme informaticien, mon boulot est l'automatisation des outils de support à cette relation .

Département SIC - ULB

Sciences de l'Information et de la Communication

Information & Communication

Information & Documentation

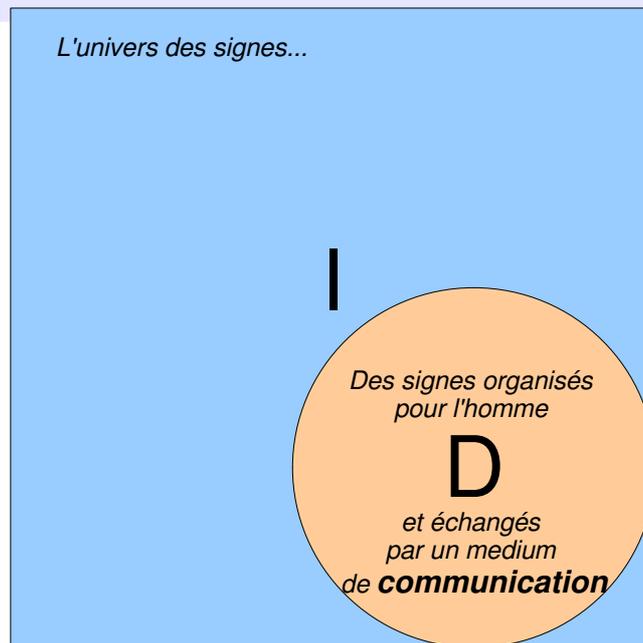
InfoDoc

I&D

3

A l'Université Libre de Bruxelles, le Département Sciences de l'Information et de la Communication a regroupé ce qui s'appelait autrefois InfoDoc. Il forme toujours des spécialistes I&D (Information & Documentation). Notre rôle tourne autour d'une association des concepts d'Information et de Documentation? Pourquoi ceci a été jugé si important depuis bien avant Otlet et l'est toujours aujourd'hui?

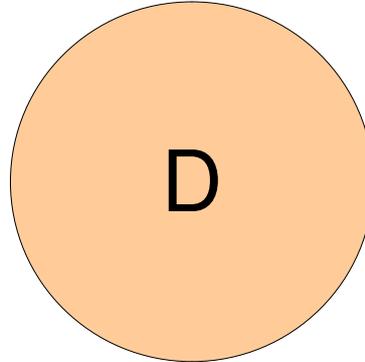
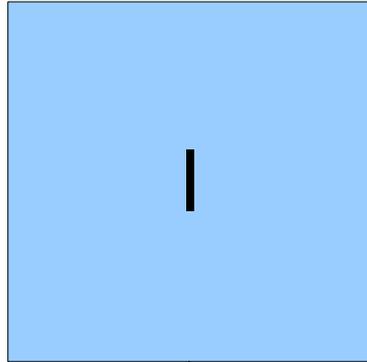
La Documentation est Information?



4

Si la Documentation est TS de Communication et si la Communication est TS de Information, pourquoi ne parle-t-on pas seulement d'Information? (La relation thésaurale TS: terme spécifique, concept couvert par un concept plus général). *Sommes nous des Gestionnaires de l'Information, modélisateurs s'interposant entre l'utilisateur et l'organisation des données de l'entreprise? Pas pour moi!*

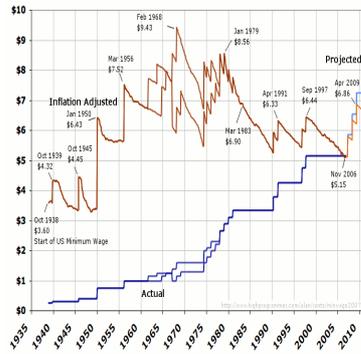
Information & Documentation au même niveau ?



5

Une explication est peut-être que le mot “Information” est surchargé. Il n'a pas le même sens pour le journaliste, le policier, l'historien, l'informaticien et le bibliothécaire...

Des usages différents



Mesurer, Grouper, Calculer,
Visualiser, Comparer,
DÉCIDER



LE KRALANDAIS SANS PEINE Leçon 1

- 1- En Krandalais, je parle comme un pouffion depuis une petite fraction (1)
- 2- Bienvenue ! Vous habitez nichoniquement fort ! Mais ne vous inquiétez pas de rien ! Si vous avez besoin d'une trousse, je ne me débâillera pas ! (2)
- 3- Un moment le monsieur. Un troupeau de chèvres a pénétré dans un ruisseau mais c'était un ruisseau projeté ! Le végétal dans une vallée est-il en feu ? (3)
- 4- Il n'a ni corbeilles ni l'abaque, ni l'ion on sait pas toi, on hab pas l'ion le main tombe dans le ravin. (4)
- 5- Crapouille ! Tu veux le côté travaillé ? Pêche tout pas votre loi pour de me contester avec toi ! Tens, un microphone !
- 6- Mon micro ! Quelle affaire pour des queues de fennel ! (5)
- 7- Tenez bien ! Tu vas voir l'ion tombe pas dans les porcelaines !
- 8- Tenez, tu m'a déposé les boules ! Elles sont en arc-croix ! (6)
- 9- Ha ha ! On peut en croire le destin. Fais pas te moquer de moi ! Tu le dis à la bombe en causer à la même occasion ! (6)

Présentation

- 1- Le pouffion
- 2- Le ruisseau
- 3- Le végétal
- 4- L'ion
- 5- Le microphone
- 6- La bombe

Notes

- (1) La fraction représente plus ou moins 6 heures.
- (2) Kraxion signifie fin, esclave, serviteur ou plus simplement esclave.
- (3) Proverbe très célèbre. Il signifie que l'on doit supporter les conséquences de nos actes.
- (4) Le fennel est la plante la plus cultivée à Kraland.
- (5) Le terme boules est un jeu d'enfant. Il faut souvent utiliser le terme polémique pour désigner le membre en question.
- (6) Tourne grammaticale pour marquer l'insistance.
- (7) La traduction Krandalais peut compter quelques erreurs mineures, mais ce qui est difficile de traduire toutes les subtilités de la langue Kr.



Découvrir, Lier,
Classer, Analyser,
ASSIMILER

On propose donc une distinction entre les *données* et les *documents*, entre l'information structurée et les textes.

Des publics différents

Données structurées
= Traitements
automatisables

I

Mesurer, Grouper, Calculer,
Visualiser, Comparer,
DÉCIDER

Documents
communicables
aux personnes

D

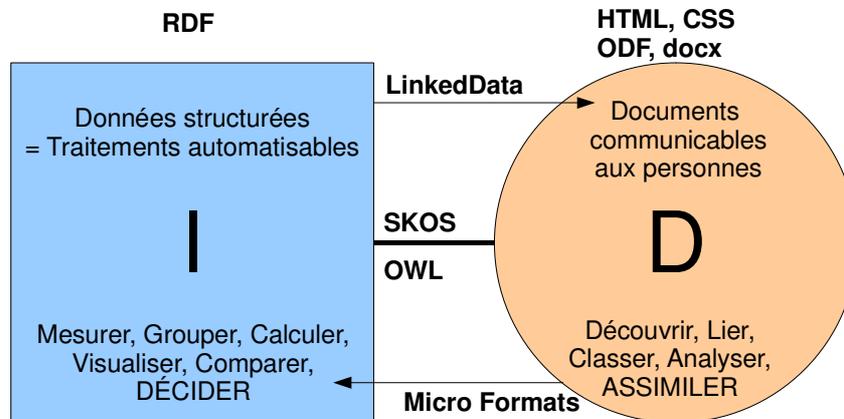
Découvrir, Lier,
Classer, Analyser,
ASSIMILER



Les êtres humains assimilent des documents et décident à partir des données. Les machines ne peuvent (sauf exception!) qu'agir de différentes façons sur base des données.

Des acteurs différents certes mais inter-reliés!
Associer Information et Documentation, c'est réfléchir à leur organisation et à leur automatisation au service des personnes dans leurs activités d'analyse, d'assimilation et de décision...

À chaque besoin, sa norme

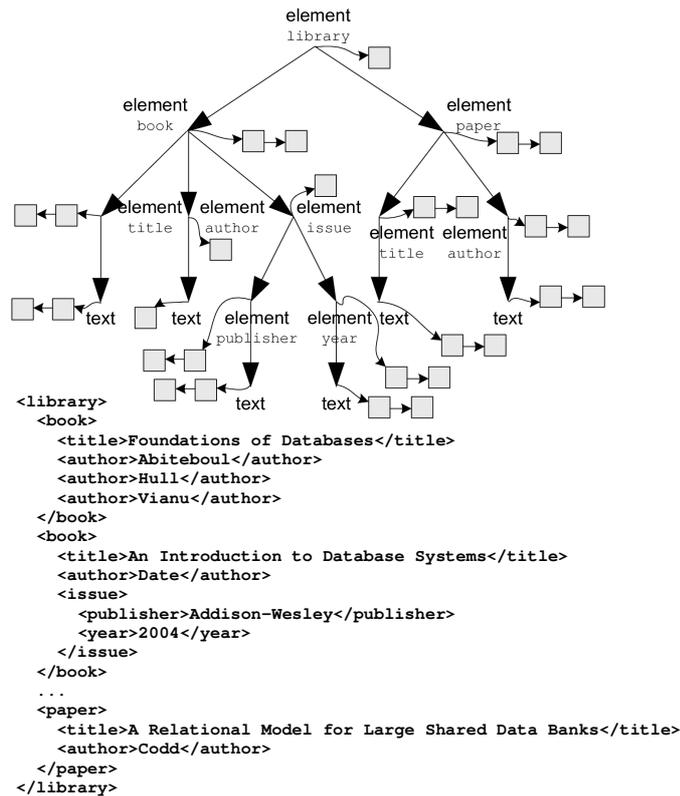


*Les normes ci-dessus s'appuient toutes (sauf CSS) sur le balisage **XML**...
Les **URL** identifient les pages de données accessibles par les réseaux informatiques. Le protocole **http://** s'occupe du transport de ces pages.*

8

Les normes sont des conventions de codage qui encadrent la production des signes transmissibles en fonction du besoin de communication. Il y a un cadre totalement général (XML), il y a une norme très souple pour exprimer les données (RDF), il y a des normes plus spécialisées pour les interactions avec les hommes (HTML, CSS, ODF, docx), il y a des normes pour passer des données aux documents (les URL du LinkedData) et inversement (les MicroFormats qui insèrent du RDF dans les documents), il y a des normes qui organisent l'information en ensembles cohérents exploitables par des applications en interaction avec les hommes: SKOS, OWL.

XML

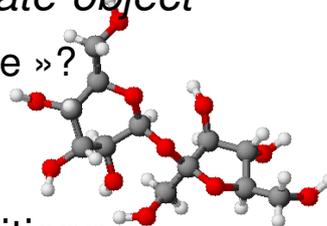


9

XML, eXtensible Markup Language: Structurer au moyen de <balises> une séquence d'informations... Une balise de début se marque <nom> et une balise de fin se marque </nom>. Les structures peuvent s'emboîter comme des poupées russes. C'est donc un système souple et organisé de « contenants » pour les données. C'est aussi une méthode de codage « extensible » car on peut rajouter des zones d'information supplémentaires sans déranger les traitements de celles qui étaient déjà là. Les langages Relax NG ou XML-Schema permettent de définir les balises permises et des règles de validation de leur contenu.

RDF/RDFSchema

- 1 énoncé = 1 sujet, 1 verbe, 1 complément
RDF statement = subject-predicate-object
 - Pourrait on faire plus « élémentaire »?
 - Ou moins? *TopicMaps...*
- Réseau d'interrelations:
plusieurs sources peuvent s'additionner.
- RDFSchema permet de définir:
 - les verbes permis (« predicate »),
 - des « types » pour les sujets (« subject »)
et les compléments (« object »)



10

RDF (Resource Description Framework) est basé sur XML. Il permet de déclarer des « atomes » d'information: un « sujet », une relation, une valeur ou un autre « objet » (autre « objet » qui pourra être le sujet de différentes relations). Un même sujet pourra avoir de nombreuses relations (semblables ou différentes) avec différentes valeurs ou autres objets. On peut représenter ainsi des réseaux aussi touffus que désiré pour décrire des ensembles de données. De plus on peut fusionner des réseaux de différentes sources (les objets devant alors être identifiés par des conventions similaires). Les définitions en RDF-Schema permettent de mettre de spécifier des règles de validation.

Énoncés RDF:

Notation N3:

```
@prefix : <http://www.example.org/> .  
:john    a          :Person .  
:john    :hasMother :susan .  
:john    :hasFather :richard .  
:richard :hasBrother :luke .
```

Notation XML:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:ns="http://www.example.org/#">  
  <ns:Person rdf:about="http://www.example.org/#john">  
    <ns:hasMother rdf:resource="http://www.example.org/#susan" />  
    <ns:hasFather>  
      <rdf:Description rdf:about="http://www.example.org/#richard">  
        <ns:hasBrother rdf:resource="http://www.example.org/#luke" />  
      </rdf:Description>  
    </ns:hasFather>  
  </ns:Person>  
</rdf:RDF>
```

11

On voit ici un petit réseau d'énoncés RDF. Les « triplets » sujet-relation-objet peuvent être notés en « N3 » (assez lisible) ou en XML (nettement moins lisible mais plus robuste à des extensions futures).

Les normes pour les machines...

- **XML**: eXtensible Markup Language
Structures hiérarchisées de Contenants pour les Éléments d'information
- **RDF**: Resource Description Framework
Déclarations formant un réseau de relations entre "objets" et entre ces objets et les "valeurs" de leurs attributs.
- **OWL**: Web Ontology Language
Application RDF pour représenter des relations logiques
- **LinkedData**: Mode d'utilisation de RDF où les "objets" sont identifiés par des URL où:
 - les humains trouvent de la documentation sur ces objets
 - les machines trouvent du RDF (autres données liées...)

12

OWL permet d'organiser les informations et la logique de leurs relations pour effectuer des raisonnements automatiques sur des triplets RDF comme « les hommes sont mortels » + « Socrate est un homme » et d'en déduire le triplet « Socrate est mortel ».

LinkedData a été proposé par Tim Berner Lee:
http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

Les normes pour les humains

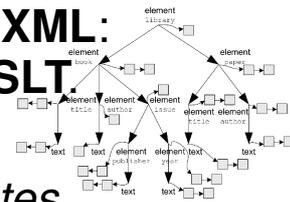
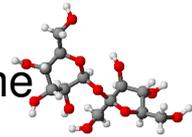
- **HTML**: HyperText Markup Language
Inventé pour structurer les éléments de texte d'un document. Devint un langage de mise en page. En train de revenir à sa vocation d'origine...
- **CSS**: Cascading Style Sheets
Indiquer les attributs de mise en page pour chaque élément structurel d'un document.
- *Formats "ouverts" pour la bureautique basés sur XML:*
 - **ODF**: OpenDocument Format = ISO
= IBM, Oracle
 - **docx**: Office Open XML = ECMA
= Microsoft

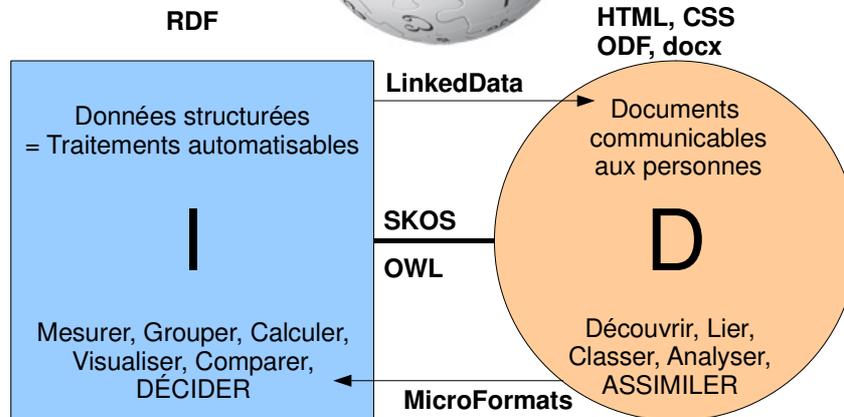
Les normes qui lient humains et machines...

- **LinkedData**
- **SKOS**: Simple Knowledge Organisation System *Application RDF qui simplifie la création de terminologies structurées (thesaurus, liste d'autorité)*
 - Lier les termes de différentes langues « humaines » et les concepts, symboles traités par les machines...
- **MicroFormats**: *partie de la norme HTML 5 permettant d'insérer des données en RDF dans un document*: <http://www.w3.org/TR/html5/microdata.html>

Séparer Données et Présentation

- Un document **HTML** peut être (ou ne pas être) « structuré »: son balisage représente alors la sémantique de chaque élément du texte.
- Données structurées en réseaux **RDF**:
extraction, séquençage par un programme **SPARQL** et production de **XML**.
- Données structurées en séquences **XML**:
mise en forme par un programme **XSLT**.
- **CSS**: typographie et mise en page
Différents fichiers CSS pour différentes présentations des mêmes données.





La place des Encyclopédies?

Par exemple, WikiPedia



- Une organisation permettant d'identifier des sujets individuels (*pages de désambiguïsation..*)
- Un document par sujet pour pouvoir **assimiler** rapidement le contexte autour de celui-ci.
- Des portails thématiques avec une structuration des sujets sous-jacents (*InfoBox*: tableau de **données** comparables entre les sujets d'un même thème)
- Une mise-à-jour collaborative, une méritocratie?

Quelles normes pour WikiPedia ?



- Documents **HTML+CSS**:
<http://fr.wikipedia.org/wiki/Bruxelles>
 - **LinkedData**: autorité pour identifier et documenter le concept « Bruxelles »?
- **MicroFormat**: inclure des énoncés RDF qui représentent les données contenues dans le texte?
- **SKOS**: Qu'apporteraient les relations thésaurales à WikiPedia?
<http://fr.wikipedia.org/wiki/Rosaceae>

18

Seuls HTML+CSS sont réellement en usage
aujourd'hui: le Web Sémantique a encore du
chemin à faire!



DBpedia = Wikipedia en RDF

- Énoncés **RDF** ayant Bruxelles comme sujet:
<http://dbpedia.org/page/Brussels?sid=2939&sp=1>
- Extraits automatiquement de Wikipedia!
Principalement des « InfoBox »
- *Où est mort Jacques Martin?*
 - *Quel Jacques Martin?*
 - La gestion des « identifiants » est **essentielle** !
 - *Le retour des Autorités? Indépendantes et Pérennes?*
Mercatiques? Démocratiques? Méritocratiques? ...

Wikipedia Infobox → RDF

```
{{Infobox Ville
|nom="{{lang|fr|Bruxelles}}" {{fr icon}}
|autres noms="{{lang|nl|Brussel}}" {{nl icon}}
...|blason=Belgium brussels iris.svg
|drapeau=Flag Belgium brussels.svg
|image=[[Image:TE-Collage Brussels.png|250px]]
|légende=Montage des différentes vues de
Bruxelles.
|pays=Belgique
...|gouverneur=[[Hugo Nys]]
...|latitude=13
|longitude=50.85
|altitude=4.35
|population={{formatnum:1120000}}
|année_pop=2010
|superficie= 162
...|
source1=http://portail.irisnet.be/en/home/home.s
html
}}
```

```
dbpedia-owl:areaTotal
* 161400000.000000 (xsd:double)
dbpedia-owl:elevation
* 13.106400 (xsd:double)
dbpedia-owl:foundingDate
* 1989-06-18 (xsd:date)
dbpedia-owl:leaderName
* dbpedia:Hugo_Nys
dbpedia-owl:populationAsOf
* 2008-11-01 (xsd:date)
dbpedia-owl:populationDensity
* 6508.524087 (xsd:double)
dbpedia-owl:populationMetro
* 1830000 (xsd:integer)
dbpedia-owl:populationTotal
* 1080790 (xsd:integer)
dbpedia-owl:thumbnail
*
http://upload.wikimedia.org/wikipedia/commons/t
humb/9/96/Collage_Bruxelles.jpg/200px-
Collage_Bruxelles.jpg
```

Strates de compétence

1. Documenter, communiquer:
HTML(structuré) + CSS + URL
2. Représenter des données structurées:
XML, Relax NG (XML-Schema)
3. Méta-données sur les documents;
Réseaux d'information:
RDF, RDF-Schema (TopicMaps)
4. Organiser les sujets, l'indexation: SKOS
5. Modéliser: OWL (logiciel Protégé)
6. Traiter les informations: SPARQL, XSLT...

Pour aller un peu plus loin...

- <http://rdfabout.com/>
- <http://www.slideshare.net/Figoblog/linked-data-and-why-we-librarians-should-care>
- <http://www.slideshare.net/antoineisaac/skos-and-linked-data>
- <http://blog.foolip.org/2009/08/23/microformats-vs-rdfa-vs-microdata/>
- <http://www.transcendingcss.com/>
- <http://www.slideshare.net/ChristopheDupriez/destin-iskoucl>
- <http://www.slideshare.net/ABLVienna/poolparty-thesaurus-management-isko-uk-london-2010>
- <http://www.slideshare.net/NigelECS/government-linked-data-a-tipping-point-for-the-semantic-web>

Sources des images...

- Alan De Smet
<http://www.highprogrammer.com/alan/rants/minwage2007/>
- <http://www.kraland.org/main.php?page=2%3B1%3B1%3B4%3B1>
- Cell Phone Digest
http://www.cellphonedigest.net/news/2009/07/qualcomm_verizon_machinetomach.php
- IStockPhoto #11822201, #9001584, #937355
- WikiPedia
- DBpedia