

Métadonnée

Cours « Recherche d'Information et Graphes de Données »

Nicolas Delestre

Plan...

- 1 Quelques constats
- 2 Vocabulaire : métadonnées, schéma, profil d'applications
- 3 Deux exemples de schéma de métadonnées
 - Dublin Core
 - Schema.org
- 4 Diffusion et utilisation des métadonnées
 - Diffusion au sein des pages HTML
 - Pour les pages
 - Pour une partie d'une page
 - Utilisation par les moteurs de recherche
- 5 Conclusion

Google's Algorithm Is Lying to You About Onions

<http://daringfireball.net/linked/2017/03/07/google-onions>

« Not only does Google, the world's preeminent index of information, tell its users that caramelizing onions takes "about 5 minutes" - it pulls that information from an article whose entire point was to tell people exactly the opposite. A block of text from the Times that I had published as a quote, to illustrate how it was a lie, had been extracted by the algorithm as the authoritative truth on the subject. » (Tom Scocca)

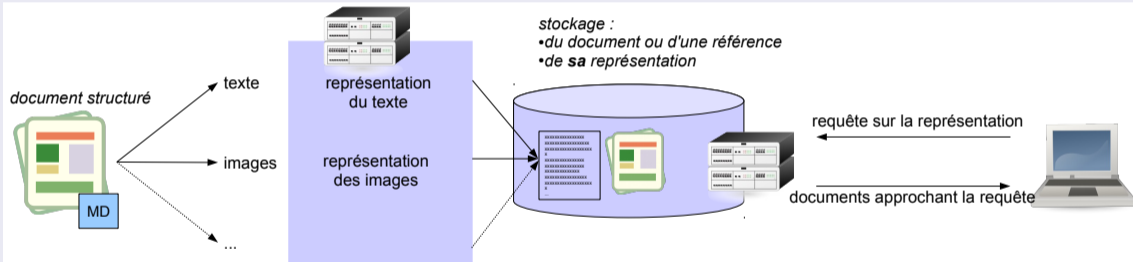
Quelques constats 2 / 3

Pourquoi ?

- Les moteurs de recherche indexent des documents à destination des humains et non des machines
 - Non prise en compte du contexte (à l'indexation et au moment de la recherche)
 - Les moteurs de recherche utilisent les données (mots) pas l'information (relation entre les mots, les concepts)
-
- Les moteurs de recherche devraient demander des précisions concernant la recherche (interaction entre l'utilisateur et le moteur de recherche)

Quelques constats 3 / 3

Rôle des métadonnées



Définitions

« Une métadonnée (du grec *meta* "après" et du latin *data* "informations") est une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique). » (Wikipédia)

« Les métadonnées sont, dans le cadre du Web sémantique, des données significatives qui permettent de faciliter l'accès au contenu informationnel d'une ressource informatique, une notice de contenu intégrée en quelque sorte (dans l'en-tête des documents HTML côté code source ou en tant que fichier XML autonome par exemple). » (Wikipédia)

Métadonnée (en pratique)

- C'est une donnée sur une donnée :
 - clairement définie
 - avec une arité (attention au LangString)
 - a un type :
 - type énuméré (vocabulaire fermé)
 - type prédéfini (entier, date, etc.) avec formalisme de représentation

Schéma de métadonnées

- Ensemble de métadonnées défini par un organisme, une entreprise, etc.
- Exemples de schéma de métadonnées
 - Dublin Core
 - Learning Object Metadata



Profil d'application

- Définition

« ...métadonnées issues d'un ou plusieurs schémas de métadonnées combinés afin d'améliorer et d'optimiser leur utilisation dans un cadre particulier » [?]

« ... est d'adapter et de combiner des schémas existants afin d'obtenir un nouveau schéma pour une application particulière tout en gardant l'interopérabilité avec le ou les schémas de base » [?]

- Un profil d'application est un schéma de métadonnées
- Exemples de profil d'application :
 - CanCore
 - LOM-FR

Le Dublin Core 1 / 2

Qu'est ce ?

- Géré par le (*Dublin Core Metadata Initiative*) : <http://dublincore.org/index.shtml>
- Schéma de métadonnées permettant de décrire des documents (numériques ou physiques)
 - 15 métadonnées forment le DCMES (*Dublin Core Metadata Element Set*). C'est aujourd'hui une norme (ISO 15836)
 - 55 (avec les 15 précédents) forment le *DCMI Metadata Terms*

Les métadonnées du DCMES (Wikipédia)

Élément	Commentaire
Title	Titre principal du document
Creator	Nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document
Subject	Mots-clefs, phrases de résumé, ou codes de classement
Description	Résumé, table des matières, ou texte libre
Publisher	Nom de la personne, de l'organisation ou du service à l'origine de la publication du document



Le Dublin Core 2 / 2

Les métadonnées (suite)

Élément	Commentaire
Contributor	Nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document. Chaque contributeur fait l'objet d'un élément Contributor séparé
Date	Date d'un évènement dans le cycle de vie du document
Type	Genre du contenu
Format	Type MIME, ou format physique du document
Identifier	Identificateur non ambigu : il est recommandé d'utiliser un système de référencement précis, afin que l'identifiant soit unique au sein du site, par exemple les URI ou les numéros ISBN
Source	Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question. Il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur URI
Language	La langue du document
Relation	Lien avec d'autres ressources. De nombreux raffinements permettent d'établir des liens précis, par exemple de version, de chapitres, de standard, etc.
Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle
Rights	Droits de propriété intellectuelle, Copyright, droits de propriété divers



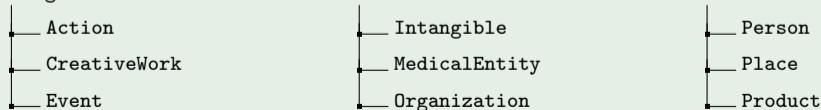
Schema.org 1 / 3

Qu'est-ce ?

- <http://schema.org>
- Proposé par Google, Yahoo, Bing et Yandex
- Schéma de métadonnées généraliste (version 28.1 de novembre 2024)
 - environ 800 classes (organisées en hiérarchie - héritage -)
 - environ 1450 propriétés
 - plusieurs centaines valeurs énumérées

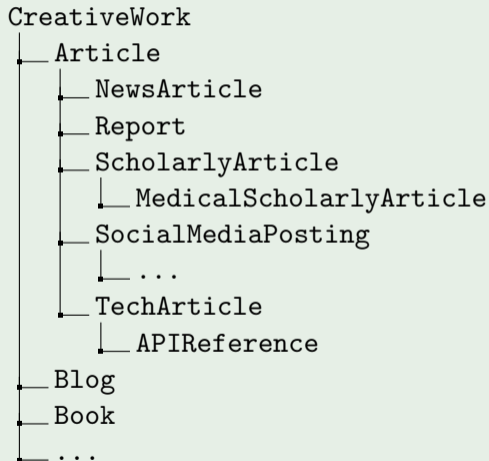
Premier niveau de la hiérarchie

Thing



Schema.org 2 / 3

Un extrait de la hiérarchie de classes



Schema.org 3 / 3

Propriétés

- Les propriétés sont attachées aux classes :
 - **Thing** additionalType, alternateName, description, image, name, potentialAction, sameAs, url
 - **Person** additionalName, address, affiliation, alumniOf, award, ...
- Les valeurs des propriétés peuvent de type simples (Datatype) ou de classes :
 - Datatype : Boolean, Date, DateTime, Number (deux sous types : Float et Integer), Text (un sous type URL), Time
 - ex : additionalName : Text
 - ex : address : PostalAddress
- Une sous classes peut utiliser les propriétés des super classes

Métadonnées pour les pages (HTML5)

- Métadonnée par défaut : balise *title*
- Métadonnées additionnelles, utilisation de la balise *meta* avec les attributs :
 - name** pour le nom de la métadonnée : *author*, *description*, *generator*, *keywords*
 - content** pour la valeur

Exemple : <http://www.w3.org/> (2015)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
<!-- Generated from data/head-home.php, ../../smarty/{head.tpl} -->
<head>
<title>World Wide Web Consortium (W3C)</title>
...
<meta name="description" content="The World Wide Web Consortium (W3C) is an international community where
Member organizations, a full-time staff, and the public work together to develop Web standards." />
...
</head>
```



JSON-LD

- Notation JSON avec des clés réservées (commençant par un @ pour le Linked Data)
- Format recommandé par Google

Exemple (<http://schema.org/Movie>)

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Movie",
  "name": "Pirates of the Carribean: On Stranger Tides (2011)",
  "actor": [{"@type": "Person", "name": "Johnny Depp"}, {"@type": "Person", "name": "Penelope Cruz"}],
  "aggregateRating": {
    "@type": "AggregateRating", "bestRating": "10", "ratingCount":
"200", "ratingValue": "8", "reviewCount": "50"
  },
  "author": [{"@type": "Person", "name": "Ted Elliott"}, {"@type": "Person", "name": "Terry Rossio"}],
  "description": "Jack Sparrow and Barbossa embark on a quest to find the elusive fountain of youth, only to
  discover that Blackbeard and his daughter are after it too.",
  "director": {"@type": "Person", "name": "Rob Marshall"}
}
</script>
```



RDFa

- Mécanisme proposé par le w3c
- Définit la syntaxe, ne spécifie pas les schémas à utiliser
- Deux versions : RDFa v1.0 pour XHTML 1.0 et HTML 4.01 et **RDFa v1.1 pour HTML5** (deux versions HTML+RDFaLite et HTML+RDFa)
- Propriétés : vocab, typeof, property, etc.

Exemple HTML+RDFaLite (<http://www.w3.org/TR/html-rdfa/>)

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title>Example Document</title>
  </head>
  <body vocab="http://schema.org/">
    <p typeof="Blog">
      Welcome to my <a property="url" href="http://example.org/">blog</a>.
    </p>
  </body>
</html>
```

Un exemple : <https://delestre.pages.insa-rouen.fr/siteweb/>

```

<!doctype html>
<html lang="fr">
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
  ....
  <script type="application/ld+json">
  {
    "@context": {
      "sch": "http://schema.org"
    },
    "@graph": [
      {
        "@id": "#me",
        "@type": "sch:Person",
        "sch:givenName": "Nicolas",
        "sch:familyName": "Delestre",
        ...
      }
    ]
  }
  </script>
</head>
<body >
  <div id="w">
    <header class="clearfix" >
      <div id="info">
        <h1>Nicolas Delestre</h1>
  
```

pour les machines

pour les humains

Nicolas Delestre

Maître de conférences en Informatique (27ème section)

INSA Rouen Normandie
 BP 08 - Avenue de l'Université
 F-76001 Saint-Etienne du Rouvray
 France
 nicolas.delestre@insa-rouen.fr



Je travaille à l'INSA Rouen Normandie depuis 2009. Je suis membre du département Informatique et Technologie de l'Information (ITI) et de l'équipe MIND du Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes (LITIS).

NOUVEAU TEST		
Détectés	0 ERREUR	0 AVERTISSEMENT 7 ÉLÉMENTS
Book	0 ERREUR	0 AVERTISSEMENT 1 ÉLÉMENT
Course	0 ERREUR	0 AVERTISSEMENT 6 ÉLÉMENTS

INSA
ROUEN NORMANDIE



Rich Snippet 1 / 2


Qu'est ce ?

- Informations ajoutées dans la page des résultats
- Peut être des images ou du texte



Utilisation des métadonnées de www.marmiton.org par Google


ACCUEIL > RECETTES > TOMATES FARCIES FACILE

Tomates farcies facile

 caroline_3528
plus de recettes >



Plat principal Sans gluten

★★★★★ 137 commentaires  11k ajoutés au carnet  380 partages



40

Temps 1h20 Personnes 4 Très facile Ben marché

 Je sauvegarde  Je partage avec mes amis

Ingrédients **Préparation** Imprimer

Nombre de personnes: 4 +

TEMPS TOTAL : 1h20

Préparation : 25 min Cuisson : 1h

Recette de Tomates farcies facile - Marmiton



www.marmiton.org > Recettes ▾

★★★★★ Note : 4,5 - 137 avis - 1 h 20 min

Nombre de personnes. -. +. 500. g de chair à saucisse. 4. tomates (ou 8 petites). 3. oignons. 2. gousses d'all. Thym. Persil. Beurre. Poivre. Sel. J'ajoute à ma liste de courses ...

Vous avez consulté cette page de nombreuses fois. Date de la dernière visite : 18/02/18

Recette de Tomates farcies facile et rapide - L'atelier des Chefs



<https://www.atelierdeschefs.fr> > ... > Recettes de cuisine > Recettes viande ▾

★★★★★ Note : 3,7 - 9 395 votes - 1 h 20 min

Découvrez cette recette de Tomates farcies expliquée par nos chefs.

Tomates farcies maison : la meilleure recette



cuisine.journaldesfemmes.fr > Recettes > Viandes > Recettes agneau ▾

★★★★★ Note : 4 - 3 avis - 15 min

La recette facile et rapide pour réussir ses tomates farcies, à base de tomates, jambon blanc, chair à saucisses, huile d'olive... recette pour 4 personnes, prête en 15 minutes, et inratable.

Rich Snippet 2 / 2

Classes prises en compte par Google

- Product
- Recipe
- Review
- Event
- SoftwareApplication
- Video
- NewsArticle

Conclusion

- Les métadonnées permettent à des programmes de récupérer et d'interpréter des informations structurées
- C'est grâce à ces métadonnées que les moteurs de recherche peuvent ajouter des *Rich Snippets*
- Mais :
 - ces métadonnées ne sont pas vraiment liées
 - mélanger métadonnées et document peut ralentir les processus de chargement, d'extraction et d'interprétation des documents (le navigateur charge les métadonnées contenues dans les pages alors qu'il ne va pas en tirer parti. . .)
 - il n'y a pas de langage permettant d'interroger ces métadonnées (à l'image du SQL pour le monde relationnel)

Recette de Tomates farcies facile - Marmiton



www.marmiton.org > Recettes ▾

★★★★★ Note : 4,5 - 137 avis - 1 h 20 min

Nombre de personnes. -. +. 500. g de chair à saucisse. 4. tomates (ou 8 petites). 3. oignons. 2. gousses d'ail. Thym. Persil. Beurre. Poivre. Sel. J'ajoute à ma liste de courses ...

Vous avez consulté cette page de nombreuses fois. Date de la dernière visite : 18/02/18