



Institut
Mines-Télécom

Améliorer l'indexation de vos pages

Schema.org, rdfa, microdata

Jean-Claude Moissinac



Inspiré de jean.delahousse.net



Balisage des pages web : historique

Langages et Modèles

■ 1989 : HTML (W3C)

- Métadonnées html



■ 2005 : Microformats (CommerceNet)

- hCalendar, hCard, hNews, hAtom...



■ 2008 : RDFa (W3C)

- Ontologies du sémantique web : foaf, DublinCore...



■ 2010 : Open Graph (Facebook)



■ 2010 : HTML5/Microdata (W3C)



Définition

web sémantique = web 3.0

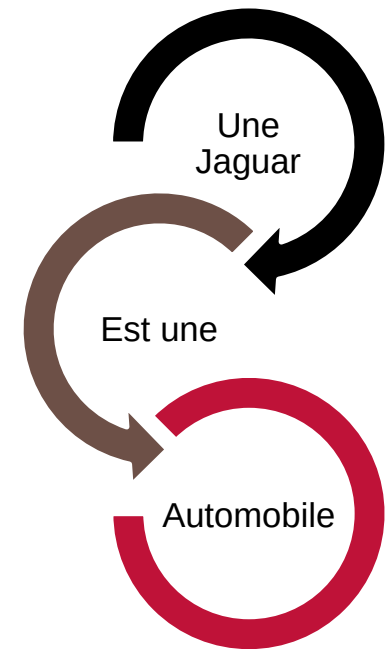
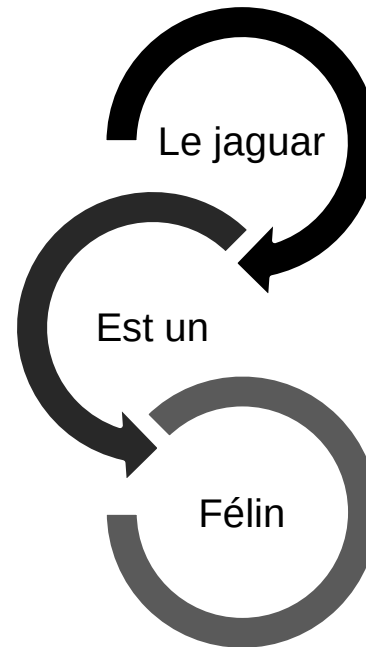
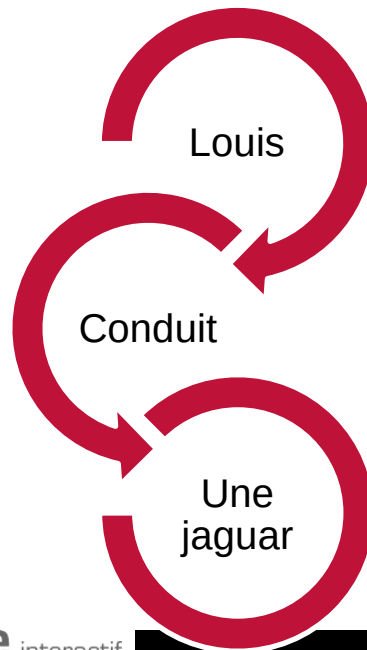
- Tim Berners-Lee
- « Un web de données qui peuvent être traitées directement et indirectement par des machines. »
- Séparer la présentation des données des données elles-mêmes



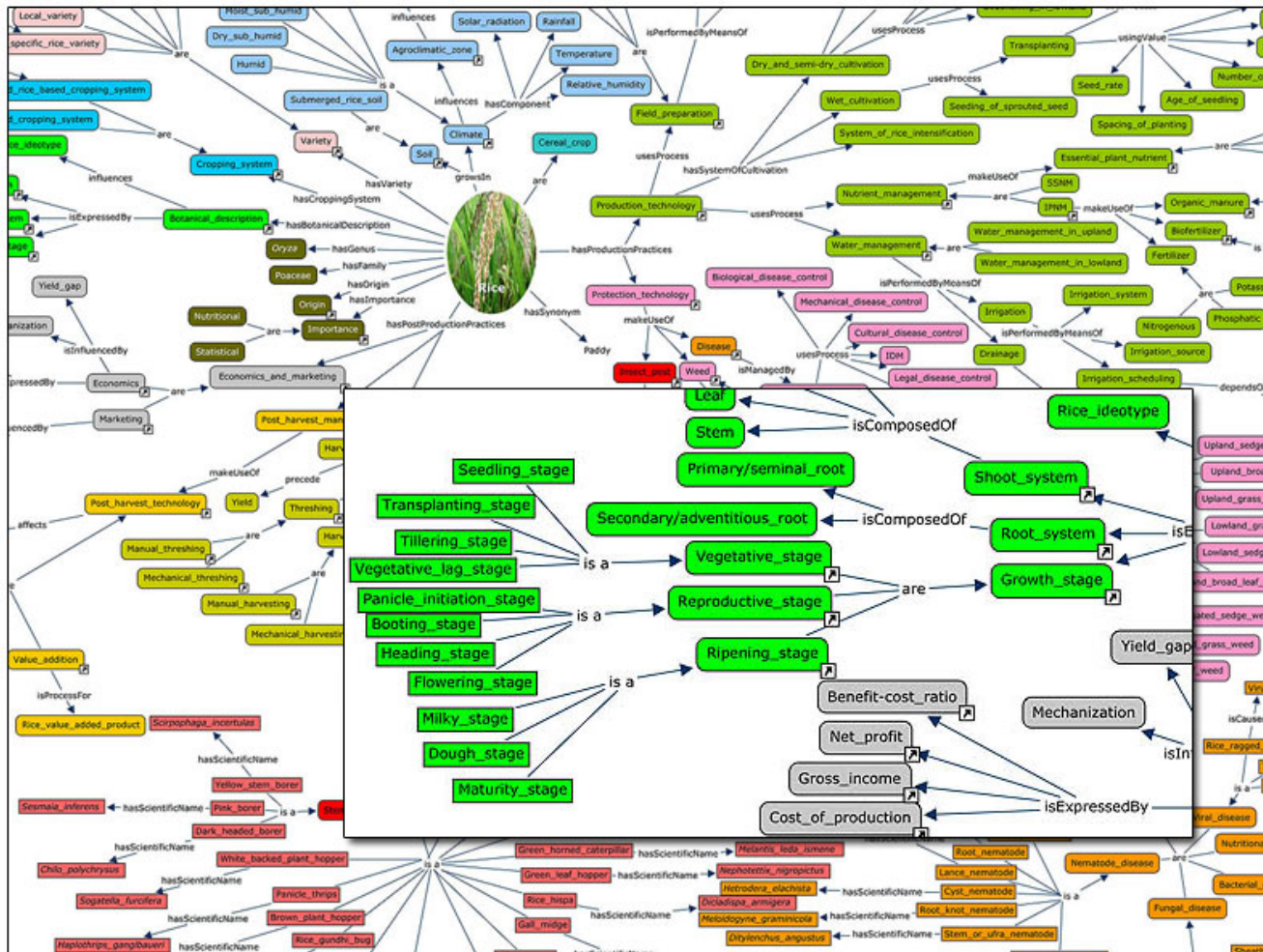
Langage de balisage RDF markup language

- **Les triplets (triples)**

- Sujet
 - Prédicat
 - Objet



ontologie du riz



source
orénoque interactif
solutions de visibilité web

Homonymie

Jaguar ou jaguar

- rel="canonical"
- canocalisation



source

 **orénoque** interactif
solutions de visibilité web

Référencement/seo sémantique

- **Comment utiliser les techniques du web sémantique pour:**
 - Envoyer à Google des informations détaillées sur la signification de votre contenu,
 - D'une manière facilement compréhensible par des algorithmes



Des moteurs de plus en plus sémantiques

BUSINESS CENTER

Mar 24, 2009 5:10 pm

Google Rolls out Semantic Search Capabilities

By Juan Carlos Perez, IDG News



Google has given its Web search engine an injection of **technology**, as the search leader pushes into what the future of search on the Internet.

Ask.com ...

[Microdata chosen over RDFa for semantics by Google, Bing](#)
[ebiquity.umbc.edu/.../microdata-rdfa-g... - États-Unis - Traduire cette page](#)

Amis : Search, Spam, Social Media

2 juin 2011 - Microdata chosen over RDFa for **semantics** by Google, Bing
Tim Finin, 4:32pm 2 June 2011. Tweet. Google, **Bing** and Yahoo! are ...

[Google, Yahoo! and Bing Announce Schema.org - semanticweb.com/google-yahoo-and-bing-annou...](#) - Traduire cette page

2 juin 2011 - In a collaborative effort reminiscent of sitemaps.org, Google, **Bing** have announced the launch of schema.org. Perhaps the most significant ...

Bing vs Google, the quiet semantic war

Posted on **July 16, 2010** by lewisshepherd

On Wednesday night I had dinner at a burger joint with four old friends; two work in the intelligence community today on top-secret programs, and two others are technologists in the private sector who have done IC work for years. The five of us share a particular interest besides good burgers: semantic technology.



Associated Press

Google's Amit Singhal, shown in 2009, sees better matches for queries.

Google isn't replacing its current keyword-search system, which determines the importance of a website based on the words it contains, how often other sites link to it, and dozens of other measures. Rather, the company is aiming to provide more relevant results by incorporating technology called "semantic search," which refers to the process of understanding the actual meaning of words.

Amit Singhal, a top Google search executive, said in a recent interview that the search engine will better match search queries with a database.

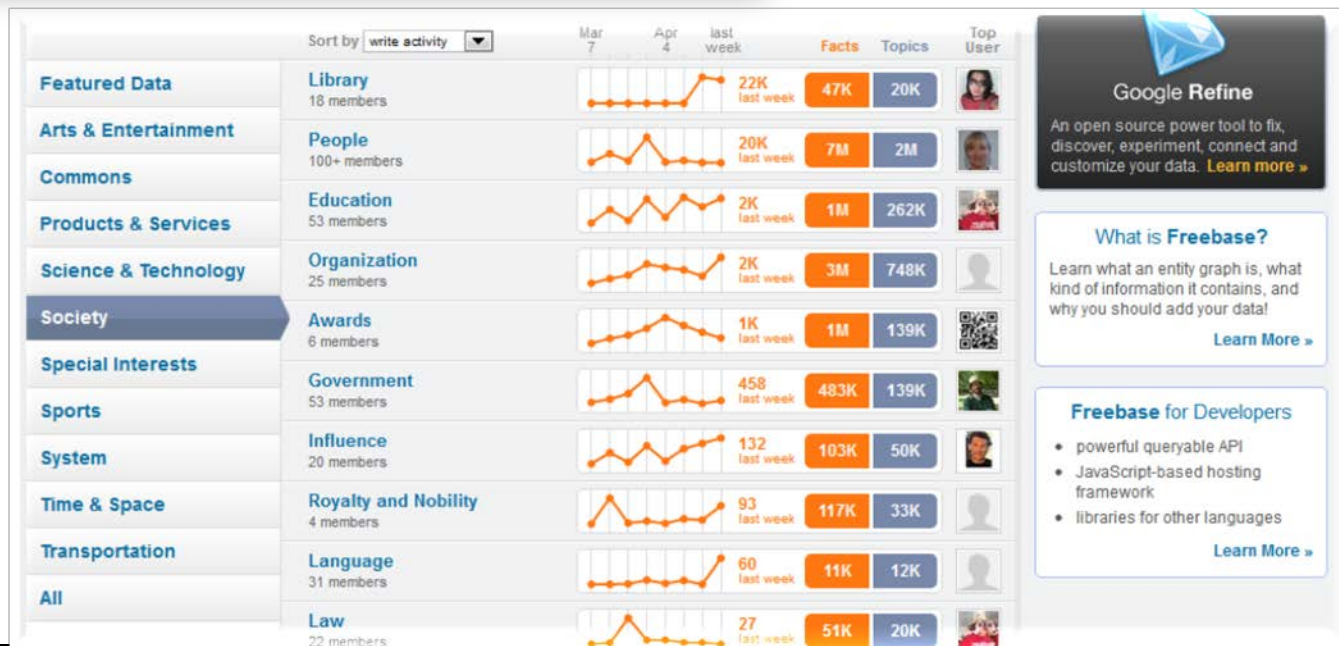
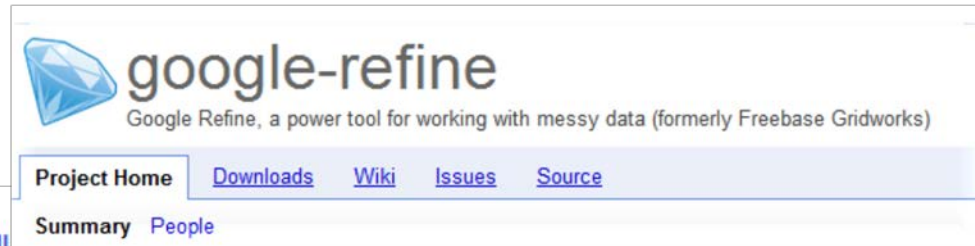


Des moteurs de plus en plus sémantiques

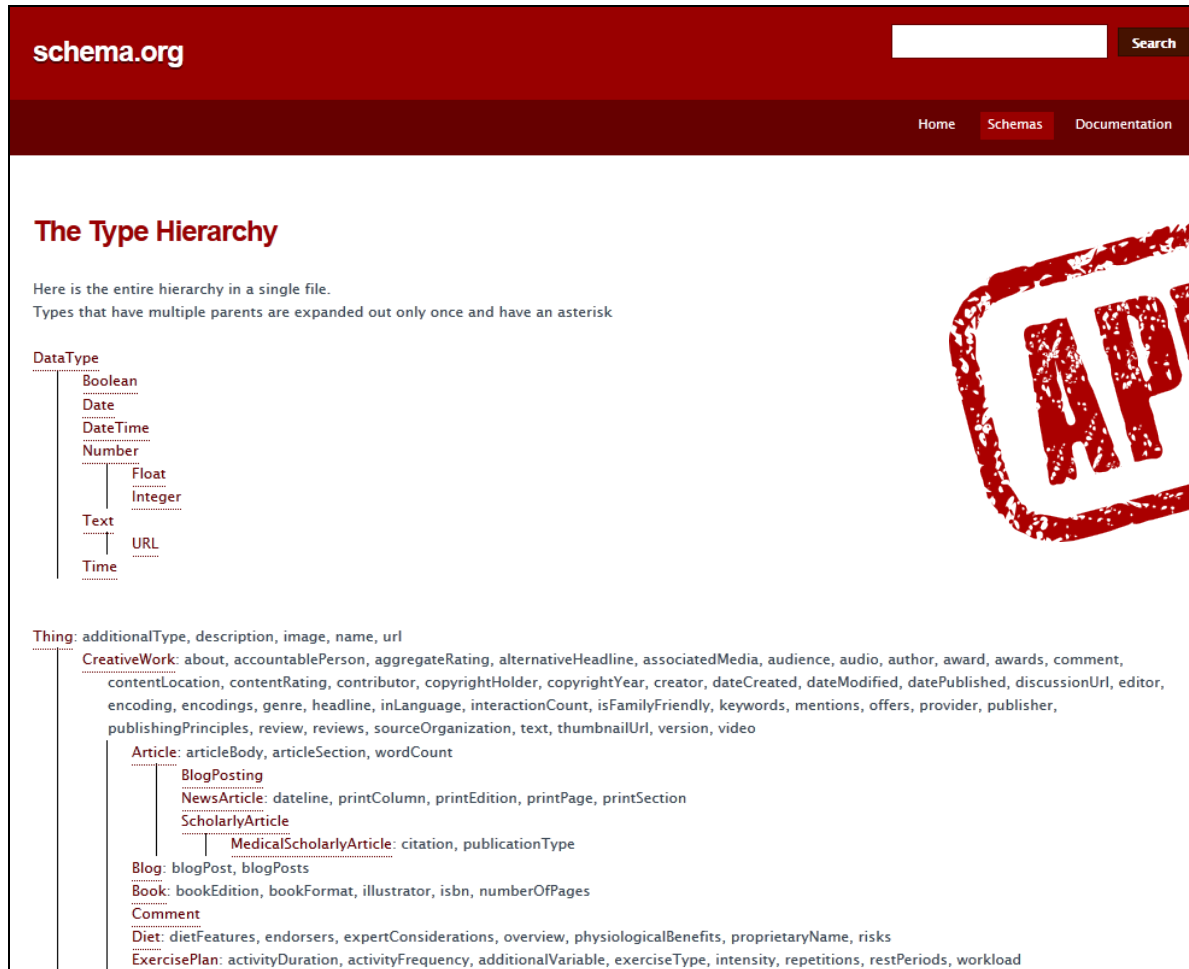
Google rachète **Metaweb**, spécialiste des données sémantiques

www.zdnet.fr > News > Internet

19 juil. 2010 – Editeur de la base de données ouvertes Freebase, **Metaweb** se spécialise dans l'enrichissement de sites à partir de données ouvertes. Google ...



Google approved: schema.org



schema.org

Home Schemas Documentation

The Type Hierarchy

Here is the entire hierarchy in a single file.
Types that have multiple parents are expanded out only once and have an asterisk

DataType

- Boolean
- Date
- DateTime
- Number
 - Float
 - Integer
- Text
 - URL
- Time

Thing: additionalType, description, image, name, url

CreativeWork: about, accountablePerson, aggregateRating, alternativeHeadline, associatedMedia, audience, audio, author, award, awards, comment, contentLocation, contentRating, contributor, copyrightHolder, copyrightYear, creator, dateCreated, dateModified, datePublished, discussionUrl, editor, encoding, encodings, genre, headline, inLanguage, interactionCount, isFamilyFriendly, keywords, mentions, offers, provider, publisher, publishingPrinciples, review, reviews, sourceOrganization, text, thumbnailUrl, version, video

Article: articleBody, articleSection, wordCount

- BlogPosting**
- NewsArticle**: dateline, printColumn, printEdition, printPage, printSection
- ScholarlyArticle**
 - MedicalScholarlyArticle**: citation, publicationType
- Blog**: blogPost, blogPosts
- Book**: bookEdition, bookFormat, illustrator, isbn, numberOfPages
- Comment**
- Diet**: dietFeatures, endorers, expertConsiderations, overview, physiologicalBenefits, proprietaryName, risks
- ExercisePlan**: activityDuration, activityFrequency, additionalVariable, exerciseType, intensity, repetitions, restPeriods, workload

source
orénoque interactif
solutions de visibilité web

Juin 2011: création d'une
modélisation commune pour
le contenu des pages web

Un modèle adapté aux
langages de balisage
Microdata et RDFa

“Schema.org est une collaboration entre Bing, Google,
Yahoo! Et Yandex pour faciliter la tâche des webmaster
quand ils nous procure des données qui nous permettent
de mieux diriger les utilisateurs vers leurs sites.”

<http://schema.org/docs/faq.html>



Les attributs et relations pour Person

Thing > Person

A person (alive, dead, undead, or fictional).

Property	Expected Type	Description
Properties from Thing		
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
url	URL	URL of the item.
Properties from Person		
additionalName	Text	An additional name for a Person, can be used for a middle name.
address	PostalAddress	Physical address of the item.
affiliation	Organization	An organization that this person is affiliated with. For example, a school/university, a club, or a team.
alumniOf	EducationalOrganization	An educational organizations that the person is an alumni of.
award	Text	An award won by this person or for this creative work.
awards	Text	Awards won by this person or for this creative work. (legacy spelling; see singular form, award)
birthDate	Date	Date of birth.
children	Person	A child of the person.
colleague	Person	A colleague of the person.
colleagues	Person	A colleague of the person (legacy spelling; see singular form, colleague).
contactPoint	ContactPoint	A contact point for a person or organization.
contactPoints	ContactPoint	A contact point for a person or organization (legacy spelling; see singular form, contactPoint).
deathDate	Date	Date of death.
email	Text	Email address.
familyName	Text	Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the Name property.
faxNumber	Text	The fax number.
follows	Person	The most generic uni-directional social relation.
gender	Text	Gender of the person.
givenName	Text	Given name. In the U.S., the first name of a Person. This can be used along with familyName instead of the Name property.
homeLocation	Place or ContactPoint	A contact location for a person's residence.
honorificPrefix	Text	An honorific prefix preceding a Person's name such as Dr/Mrs/Mr.
honorificSuffix	Text	An honorific suffix preceding a Person's name such as M.D. /PhD/MSCSW.
interactionCount	Text	A count of a specific user interactions with this item—for example, 20 UserLikes , 5 UserComments , or 300 UserDownloads . The user interaction type should be one of the sub types of UserInteraction .
jobTitle	Text	The job title of the person (for example, Financial Manager).
knows	Person	The most generic bi-directional social/work relation.
memberOf	Organization	An organization to which the person belongs.
nationality	Country	Nationality of the person.
parent	Person	A parent of this person.

Balisages reconnus

■ Microdata

- <https://html.spec.whatwg.org/multipage/microdata.html>

■ Rdfa

- <http://www.w3.org/TR/xhtml-rdfa-primer/>

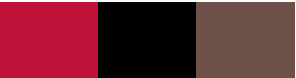
Les opportunités pour votre visibilité web

span



Exemple de balisage microdata

```
<!DOCTYPE html>
<html itemid="http://givigsense.eu/foaf/moissinac.htm">
  <head>
    <title>Jean-Claude Moissinac</title>
    <meta itemprop="inLanguage" content="fr">
  </head>
  <body itemscope itemtype="http://schema.org/WebPage">
    <div itemscope itemtype="http://schema.org/Person">
      <span itemprop="name">Jean-Claude Moissinac</span>
      <span itemprop="jobTitle">Maître de Conférences</span>
      
      <a href="https://twitter.com/jcmoissinac" itemprop="url">twitter.com/jcmoissinac</a>
      <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
        <span itemprop="addressLocality">Paris</span>
        <span itemprop="postalCode">75000</span>
        <span itemprop="addressCountry">France</span>
      </div>
      <div itemprop="memberOf" itemscope itemtype="http://schema.org/Organization">
        <span itemprop="name">Telecom ParisTech</span>
      </div>
    </div>
  </body>
</html>
```



L'échange

- **Les moteurs demandent aux entreprises plus de données structurées, plus de connaissances réutilisables par des systèmes automatisés**

- **En échange ils offrent aux entreprises**
 - Une meilleure indexation des contenus, améliorant la pertinence des résultats de recherche
 - une meilleure visibilité de leurs contenus par la génération de vues résumées (RichSnippets) aidant les utilisateurs à choisir dans les résultats de recherche
 - une prise en compte des contenus dans les filtrages à facettes

Outil de test des données structurées

The screenshot shows the Google Structured Data Testing Tool interface. At the top, the Google logo is on the left, and the user name 'Louis Durocher' with a profile picture and a '+ Partager' button are on the right. Below the header, the title 'Outils pour les webmasters' is displayed. The main content area is titled 'Outil de test des données structurées'. On the left, there is a sidebar with 'Accueil', 'Outil de test des données structurées' (highlighted), and 'Ressources d'aide : Dépannage', 'Centre d'aide'. The main area has two tabs: 'URL' (selected) and 'HTML'. Below the tabs is a text input field containing the URL 'http://www.ebay.com/itm/Canon-EOS-Rebel-T2i-550D-18-MP-Digital-SLR-Camera-Ki' and a red 'APERÇU' button. Below the input field, there are two sections: 'Résultats de recherche Google' (highlighted with a red bar) and 'Recherche personnalisée Google'. Under 'Résultats de recherche Google', there is an 'Aperçu' section showing a search result for 'Canon EOS Rebel T2i 550D 18 MP Digital SLR Camera Kit w EF s 18 ...' with a green link to the eBay page, a star rating of 4.5 (114 votes), and a price of 203,05 \$US. Below the preview, a message states: 'L'extrait de la page s'affichera ici. Nous ne pouvons pas afficher le texte de votre page Web, car le texte dépend de la requête saisie par l'utilisateur.'

source

Rich Snippets Testing Tool Beta

Use the Rich Snippets Testing Tool to check that Google can correctly parse your structured data markup and display it in search results.

Test your website

Enter a web page URL to see how it may appear in search results:

Examples: [Applications](#), [Authors](#), [Events](#), [Movie](#), [Music](#), [People](#), [Products](#), [Products with many offers](#), [Recipes](#), [Reviews](#), [TV Series](#)

Google search preview

Banana Banana Bread Recipe - Allrecipes.com



allrecipes.com/recipe/banana-banana-bread/ - [Cached](#)

★★★★★ 6778 reviews - 1 hr 20 mins - 229 cal

The excerpt from the page will show up here. The reason we can't show text from your webpage is because the text depends on the query the user types.

Note that there is no guarantee that a Rich Snippet will be shown for this page on actual search results. For more details, see the [FAQ](#).

Extracted Author/Publisher for this page

Page does not contain authorship markup. [Learn more](#).

Extracted rich snippet data from the page

hrecipe

fn = Banana Banana Bread

yield = 1 - 9x5 inch loaf

author = Shelley Albeluhn

photo = <http://images.media-allrecipes.com/userphotos/140x140/00/25/41/254186.jpg>

nutrition

calories = 229

sodium = 223mg

fat = 8.8g

protein = 3.8g

cholesterol = 56mg

totaltime

value = PT1H20M

value in minutes = 80

cooktime

value = PT1H5M

value in minutes = 65

preptime

value = PT15M

value in minutes = 15

ingredient

value = 2 cups all-purpose flour

ingredient

value = 1 teaspoon baking soda

ingredient

value = 1/4 teaspoon salt

Etendre le modèle en fonction de ses besoins

- Un mécanisme d'extension permet d'enrichir les classes et propriétés présentes dans Schema.org pour construire un schéma tenant compte des spécificités du service
- Etendre une classe : Person
 - Person/Engineer
- Etendre la propriété : musicGroupMember
 - musicGroupMember/leadVocalist
 - musicGroupMember/leadGuitar

- **Un groupe de travail dirigé par Dan Brickley pour l'évolution de Schema.org**
- **Des propositions d'évolutions provenant de différentes communautés**
 - <http://www.w3.org/wiki/WebSchemas/SchemaDotOrgProposals>
- **Une mailing list de discussion**
 - <http://lists.w3.org/Archives/Public/public-vocabs/>
- **Des outils de test du balisage**
 - <http://www.google.com/webmasters/tools/richsnippets>

Une ligne directrice pour les évolutions

- **Objectif : Structurer les pages web pour améliorer la recherche**
- **La question n'est pas : comment décrire une voiture ? Mais comment je peux améliorer le balisage de pages qui parlent de voitures**
- **Simplifier le travail des webmasters et éditeurs**
- **Avoir un consensus entre les moteurs de recherche**

Un seul objectif final : des services de recherche plus efficaces

Microdata Generator

<http://microdatagenerator.org/>

Microdata Generator

Easy to use Microdata Generator for adding Schema.org tags to your site. Send add on suggestions to info@microdatagenerator.org

8 6
Tweet J'aime

Home

Business Name

Street Address

City

State

Zipcode

Phone

Fax

Email Clickable? ☐ Yes

URL Clickable? ☐ Yes

Payment Accepted ☐ Cash ☐ Check ☐ Credit Card ☐
Invoice ☐ PayPal

Example Display

Orénoque interactif inc.
Email: info@orenoque.com
Phone: 514 844-5049
Fax: 514 844-5049
Url: <http://www.orenoque.com>
2, Place du Commerce, Bureau 100
Montréal, QC H3E 1A1

Days Open: ☐ Monday ☐ Tuesday ☐ Wednesday ☐ Thursday ☐ Friday ☐ Saturday ☐ Sunday

Geo: (How to get)
Lat:
Long:

Your Code

```
<div itemscope itemtype="http://schema.org/LocalBusiness">
<div itemprop="name">Orénoque interactif inc.</div>
<div>Email: <span itemprop="email">info@orenoque.com</span></div>
<div>Phone: <span itemprop="telephone">514 844-5049</span></div>
<div>Fax: <span itemprop="faxNumber">514 844-5049</span></div>
<div>Url: <span itemprop="url">http://www.orenoque.com</span></div>
```

source



Wordpress Schema Creator Plugin

The screenshot shows the WordPress.org Plugin Directory page for the 'Schema Creator' plugin. The page header includes the WordPress logo, the URL 'WORDPRESS.ORG', and a search bar. Navigation links for 'Showcase', 'Themes', 'Plugins', 'Mobile', 'Support', 'Get Involved', 'About', 'Blog', 'Hosting', and 'Download' are present. The 'Plugins' tab is active. Below the header, the 'Plugin Directory' section includes a login form with 'Username' and 'Password' fields, a 'Log in' button, and links for '(forgot?)' and 'or Register'. On the left sidebar, there are links for 'Extend Home', 'Plugins', 'Developer Center', 'Themes', 'Mobile', 'Ideas', 'Kvetch!', and a 'Search Plugins' section with a search input and a 'Search' button. Below the search section are 'Popular Tags' and a link to 'More »'. The main content area features a large blue banner with the text 'SCHEMA-CREATOR' and 'Schema Creator by Raven'. Below the banner, it says 'Insert schema.org microdata into WordPress pages and posts.' and provides a 'Download Version 1.03' button. At the bottom of the main content area, there are tabs for 'Description', 'Installation', 'FAQ', 'Screenshots', 'Changelog', 'Stats', 'Support', and 'Developers'. The 'Description' tab is selected. A small code snippet is visible on the left side of the banner:

```
<div class="review_body" id="review-5">
</div>
```

 and another on the right:

```
{
  "schema": "review",
  "url": "http://domain.com/review/",
  "name": "Widget",
  "description": "Review of Widget",
  "author": "John Henshaw",
  "pubdate": "2011-06-28",
  "user_review": "5"
}
```

Bing Webmaster Tools Markup Validation Tool



The screenshot shows the Bing Webmaster Tools Markup Validation Tool interface. At the top, there is a navigation bar with links for WEB, IMAGES, VIDÉOS, CARTES, ACTUALITÉS, SHOPPING, WEBMASTER, and PLUS. The Bing logo is on the left, and a search bar with a magnifying glass icon is in the center. On the right, there are links for 'Connexion' and a settings gear icon. Below the navigation bar, the main content area has a dark gray background. It features the title 'Valideur de balisage Bing' followed by a horizontal line. Below the line, there is a paragraph of text in French explaining the tool's purpose: 'Vérifiez le balisage que vous avez ajouté à vos pages avec le valideur de balisage. Obtenez un rapport à la demande affichant les balisages qui ont été découverts, y compris celui de HTML Microdata, Microformats, RDFa, Schema.org, et OpenGraph. Pour commencer, il suffit de vous connecter ou de vous inscrire sur les Outils pour webmasters de Bing.' At the bottom of this section, there is an orange button labeled 'Connexion' and a link that says 'Inscrivez-vous avec votre compte Microsoft'.

WEB IMAGES VIDÉOS CARTES ACTUALITÉS SHOPPING WEBMASTER PLUS

bing Beta

Connexion ⌵ ⚙

Valideur de balisage Bing

Vérifiez le balisage que vous avez ajouté à vos pages avec le valideur de balisage. Obtenez un rapport à la demande affichant les balisages qui ont été découverts, y compris celui de HTML Microdata, Microformats, RDFa, Schema.org, et OpenGraph. Pour commencer, il suffit de vous connecter ou de vous inscrire sur les Outils pour webmasters de Bing.

Connexion Inscrivez-vous avec votre compte Microsoft

Références

références

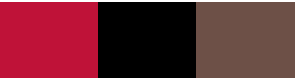


Références (1)

- **Définition**
- [Web sémantique – Wikipedia](#)
- **Vue d'ensemble et introduction**
 - [SEO, the Semantic Web and Information Discovery](#)
 - [Schema.org – Why You're Behind if You're Not Using It...](#)
- **Formats de données et vocabulaires**
 - Ontologie et vocabulaire pour le e-commerce: [Good Relations](#)
 - [Open Graph Protocol](#)
 - [Le concept du Open Graph Protocol](#)
 - [RDFa](#)
 - [Microformats](#)
 - [Schema.org](#)
 - [Les nouveaux schémas pour schema.org proposés par le W3C](#)
- **Knowledge Graph**
 - [Google Knowledge Graph](#)
 - L'impact du Knowledge Graph pour les marques: [What the New Google Plus Knowledge Graph Update Means For Brands](#)

source

 **orénoque** interactif
solutions de visibilité web

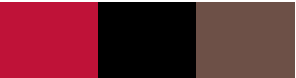


Références (2)

■ Outils

- [Facebook Object Debugger](#)
- [Raven Schema Creator](#)
- [Local Business Microdata Generator](#)
- [Plugin WordPress Schema Creator](#)
- [Guide d'utilisation de Drupal pour le web sémantique](#)
- [Outil de test des données structurées de Google](#)
- [Valdateur de balisage de Bing Webmaster Tools](#)

source



Références (3)

■ Plus

- [W3C Web sémantique](#)
- [Google Rich Snippets](#)
- Recherche Knowledge Graph de type carroussel: [Jazz Musicians](#)
- Recherche Knowledge Graph: [Marie Curie](#)
- [rel=publisher](#)
- [rel author](#)

source

