

Un exemple d'intégration de Schema.org : les boutiques Raynette E-business

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le 3 juin 2011, Google/Bing/Yahoo! rendaient public schema.org, un format commun de balisage de données ("Rich Snippets"/"Contenus enrichis"), à intégrer dans les pages HTML. Trois jours plus tard, la plate-forme e-commerce Raynette intégrait ce balisage dans les boutiques en ligne de ses e-commerçants. Pourquoi ? Comment ? Voici le détail et la chronologie de cette mise en place, qui constitue un bon exemple des atouts de ce format sémantique pour obtenir un meilleur référencement à l'avenir...

Rappel sur schema.org

Les moteurs de recherche, dans leur quête d'une meilleure compréhension du contenu des pages web qu'ils indexent, et d'un affichage des résultats plus pointus, se sont accordés sur un format pour baliser les informations des pages web, avec le format décrit sur <http://www.schema.org/>. Schema.org est l'héritier des "rich snippets" de Google, dont il a été souvent question dans cette lettre les mois précédents.

Ce format permet d'indiquer à un robot, à l'intérieur d'une page HTML, les caractéristiques d'un produit à vendre, d'une personne, d'avis clients, d'une recette de cuisine, etc. La liste des types de données balisables est indiquée sur <http://www.schema.org/>.

Pourquoi intégrer le balisage de schema.org sur les boutiques Raynette ?

Affutez vos sens et levez votre nez du côté du référencement... Ne sentez vous pas que le web se dirige lentement mais sûrement vers une structuration des données contenues dans les pages web (le web "sémantique") ?

- Google Panda veut favoriser les contenus originaux, et mettre en queue de peloton les fermes de contenus ou autres contenus dupliqués. Et pour faire le tri entre le bon grain et l'ivraie, de quoi a besoin Google ? Notamment de bien comprendre le contenu des pages web qu'il indexe. Le balisage du contenu des pages web lui est d'une aide précieuse, et est votre ami : il permet de vous faire bien comprendre de Google pour sortir du lot.

- HTML5, qui arrive à grands pas, comprend notamment de nouvelles balises permettant de distinguer les sections de haut et bas de page, de navigation, et la partie de contenu même des pages web. HTML5 permet de baliser les sections d'une page web pour décrire leur contenu. HTML5 a même prévu des attributs pour baliser le contenu des pages. (ex: telle section décrit *-itemscope-* quelque chose du type *-itemtype-* "artiste-en-herbe", et tel texte est sa propriété *- itemprop-* nom, tel autre est l'âge de son capitaine, etc).

Le fait qui a initié cette démarche d'intégrer les Rich Snippets dans les boutiques Raynette est la possibilité de baliser les avis clients, afin que Google puisse les reconnaître, et afficher les fameuses étoiles dans ses résultats "Web" et "Shopping". Avec l'arrivée de Panda en ligne de mire, fournir à Google des informations qualitatives et originales sur les articles de nos e-commerçants nous a paru un atout très sérieux.

Les formats qui s'offrent à nous pour baliser les données sont alors :

- RDFa : ajout de XML.
Inconvénient : Trop lourd...

- Microformats : ajout de classes CSS prédéfinies.
Inconvénient : Depuis quand le CSS est-il fait pour indiquer le type du contenu ? Contre-nature...

- Microdata : Ajout d'attributs html.

Inconvénient : propre et léger mais attributs non reconnus par HTML4/XHTML.

- Schema.org : fonctionne avec Microdata, et maintenant format commun pour les Big 3 (Google/Bing/Yahoo!).

Inconvénient : attributs HTML non reconnus par HTML4/XHTML...

Comme nous aimons les choses légères, propres et carrées, on choisit donc d'office microdata, et tant qu'à faire : schema.org, tout en étant conscient qu'il y a un inconvénient : les attributs HTML utilisés par ce format (itemscope, itemtype, itemprop) ne font pas partie de HTML4/XHTML (mais de HTML5). Cela va donc générer des erreurs aux tests W3C, et les puristes du *W3C-total-respect* vont avoir de l'urticaire. Mais... cela ne gêne en rien les analyseurs HTML qui ne connaissant pas ces attributs vont tout bonnement les ignorer royalement sans rien changer à l'affichage. L'inconvénient nous paraît donc mineur au vu des bénéfices attendus. Nous choisirons donc le tandem microdata/Schema.org, l'affaire est entendue.

Mais tant qu'à utiliser schema.org, autant le faire de façon complète ! On ne va donc pas baliser uniquement les avis client, mais plus franchement toutes les caractéristiques des pages produits. Le but est donc : dans chaque page détail d'article, baliser les différentes informations sur l'article proposé à la vente, avec le format schema.org. Ainsi n'importe quel robot pourra extraire des pages articles, toutes les informations sur l'article (titre, description, prix, image, url, note moyenne des avis client, détails de chaque avis client).

Autant dire que logiquement, bientôt on n'aura même plus besoin de générer des catalogues d'articles, vu que tout est déjà décrit dans chaque page article, mais c'est un autre sujet...

Travaux pratiques : La mise en place du balisage schema.org pour les articles

Type général du contenu de la page

Tout d'abord, une visite sur la liste des types de données disponibles (<http://schema.org/docs/schemas.html>) nous indique qu'il nous faut encapsuler tout la partie détaillant un article par le type "Product".

On entoure donc le code HTML de présentation de chaque article par :

```
<div itemscope itemtype="http://schema.org/Product">...</div>
```

itemscope indique qu'il s'agit d'une section du type (*itemtype*) *Products*.

Cette page : <http://schema.org/Product> liste les informations reconnues pour ce type *Product*.

Note : Ci-dessous, pour simplifier la démonstration de mise en place on vous fait grace du code HTML de mise en forme (classes, id, styles css).

Les propriétés simples de Product

- *Intitulé de l'article* : Pour baliser le nom de l'article, il faut entourer son texte par une balise textuelle HTML (<p>, ou <div> par exemple) contenant l'attribut `itemprop="name"`. On met donc : `<p itemprop="name">Paire de chaussettes grises</p>`

- *Image de l'article* : Pour baliser l'url de l'image de l'article, il suffit d'ajouter à la balise image : `itemprop="image"`. La balise image devient donc : ``

- *Description de l'article* : L'attribut à utiliser est `itemprop="description"`, on transforme donc le code html de la description en : `Ici la description complète, avec des balises
 HTML si besoin et l'encodage des caractères.`

Notez que vous pouvez utiliser un `<div>` ou un `<p>` à la place du ``.

On obtient donc le code HTML suivant qui ne modifie en rien l'affichage, mais qui a l'avantage de bien indiquer aux robots qu'on parle d'un article, dont on balise l'intitulé, l'image, et la description :

```
<div itemscope itemtype="http://schema.org/Product">
<p itemprop="name">Paire de chaussettes grises</p>

<span itemprop="description">Ici la description complète, avec des balises <br /> HTML si
besoin et l'encodage des caractères.</span>
</div>
```

Les propriétés plus complexes de Product

Le type *Product* comprend aussi comme propriétés possible, d'autres type, c'est à dire un sous-ensemble de propriétés.

- Le sous-ensemble offre de vente, nommé *"offer"*, de type *"Offers"*, décrit ici: <http://schema.org/Offer>, qui doit préciser le prix de vente principalement. Il faut encapsuler la section détaillant l'offre de vente par : `<XXX itemprop="offers" itemscope itemtype="http://schema.org/Offer">...</XXX>`

Dans notre cas, on va appliquer cela au formulaire d'ajout au panier qui devient donc:

```
<form action="ajoutaupanier.php" itemprop="offers" itemscope
itemtype="http://schema.org/Offer">...</form>
```

Puis on y intègre les informations sur l'offre de vente : le prix (propriété *"price"*), la devise (propriétés *"priceCurrency"*), le type article neuf (propriété *"itemCondition"*), et l'url de l'article (propriété *"url"*). Pour ne pas altérer du tout l'affichage, on choisit non pas de placer des balises autour de ces informations déjà affichées, mais de mettre cela dans des balises d'informations diverses qui ne sont pas affichées (*"meta"*). Le sous-ensemble présentant l'offre de vente devient donc :

```
<form action="ajoutaupanier.php" itemprop="offers" itemscope
itemtype="http://schema.org/Offer">
<meta itemprop="price" content="10.00" />
<meta itemprop="priceCurrency" content="EUR" />
<meta itemprop="itemCondition" content="http://schema.org/NewCondition" />
<meta itemprop="url" content="http://mondomaine.com/url-page-article.php" />... le
reste habituel du code html d'ajout au panier ici ...</form>
```

On n'altère pas l'affichage, mais toutes les informations sont bien présentes de cette façon.

- Le sous-ensemble note moyenne des avis clients, nommé *"aggregateRating"*, de type *"AggregateRating"*, décrit ici: <http://schema.org/AggregateRating>. Il contient les propriétés *"ratingValue"* (valeur de la note moyenne), *"reviewCount"* (nombre d'avis clients). On ne précisera pas *"bestRating"* car il s'agit de la note maxi possible, qui est 5, et qui est donc standard.

On obtient le code HTML suivant pour ce sous-ensemble :

```
<span itemprop="aggregateRating" itemscope
itemtype="http://schema.org/AggregateRating">

<meta itemprop="ratingValue" content="4" />,<span itemprop="reviewCount">18</span>
avis clients_</span>
```

Notez qu'on met la propriété *"ratingValue"* à nouveau dans une balise non-affichée, puisqu'on n'affiche pas directement la note moyenne, mais plutôt une image contenant le nombre d'étoiles correspondant à ladite note.

- Les sous-ensembles avis client, nommés *"reviews"*, de type *"Review"*, décrit ici:

<http://schema.org/Review>. Il permet de décrire de façon complète l'avis d'un client avec sa note (sous-sous-ensemble nommé "reviewRating" de type "Rating"), la date de l'avis client, son intitulé et le commentaire éventuel).

Voici un exemple type du code HTML produit :

```
<li itemprop="reviews" itemscope itemtype="http://schema.org/Review">
<!-- Sous ensemble pour la note -->
<span itemprop="reviewRating" itemscope itemtype="http://schema.org/Rating">

<meta itemprop="ratingValue" content="5" />
</span>
<!-- titre de l'avis client -->
<span itemprop="name">Super chaussettes de qualité</span>
<!-- Auteur de l'avis client -->
Par : <span itemprop="reviewer">Marcel Dupont</span>
<!-- date de l'avis client -->
le <time itemprop="dtreviewed" datetime="2010-12-31">31 décembre 2010</time>
<br />_<!-- Commentaire du client -->
<span itemprop="description">Ces chaussettes sont très agréables,<br />
mon chat veut les mêmes.</span>
</li>
```

Notes :

- l'ensemble "Rating" est un sous-ensemble de "Review", qui est lui-même un sous-ensemble de "Product". Tout s'emboîte dans les boîtes :-)
- on utilise à nouveau une balise "meta" (information non affichée) pour indiquer la note.
- pour que la date soit bien compréhensible par l'analyseur HTML, on utilise la balise "time" (balise non conforme HTML4/XHTML par le W3C, mais ne gêne pas l'affichage), en mettant cette date au format attendu par schema.org (YYYY-MM-DD) dans un attribut "datetime", puis on l'affiche ensuite comme on veut (31 décembre 2011).
- La description peut contenir les balises HTML classiques de mise en forme, et des caractères html-encodés, aucun souci.
- On crée autant de sous-ensemble avis client, qu'il y a d'avis à lister.

Résumé des travaux pratiques

On a donc ci-dessus :

- Encapsulé toutes les informations sur le produit à vendre dans un type "Product".
- Balisé les propriétés simples comme le nom, l'image, la description de l'article.
- Balisé des sous-ensembles d'informations comme l'offre de vente, la note moyenne des avis client, et chaque avis client.

Exemples réels et en ligne de pages articles balisées ainsi

- Dans <http://www.boutique-abondance.com/>, page de détails sur la vente de l'étude sur Google Panda: <http://www.boutique-abondance.com/livres-et-etudes/71-google-panda-comprendre-analyser-agir.php>

Et voici ce que l'outil Google de test des Rich Snippets extrait :

<http://www.google.com/webmasters/tools/richsnippets?url=http%3A%2F%2Fwww.boutique-abondance.com%2Flivres-et-etudes%2F71-google-panda-comprendre-analyser-agir.php>

On y retrouve bien les données que nous avons balisées, Google sait donc maintenant bien de quoi parle cette page.

- Sur le site <http://www.rascol.com/>, page de détails d'un mannequin de couture. Cette page comprend en plus des avis client : <http://www.rascol.com/mercerie/mannequin-couture/8323.php>

Et voici ce que l'outil Google de test des Rich Snippets extrait :

<http://www.google.com/webmasters/tools/richsnippets?url=http%3A%2F%2Fwww.rascol.com%2Fmercerie%2Fmannequin-couture%2F8323.php>

Par rapport à l'exemple précédent, on y trouve donc en plus, les avis des clients qui ont bien

été extraits (note moyenne, et détail de chaque avis).

Voici Google outillé pour faire ressortir ces e-commerçants avant les autres, au vu des avis clients.

Conclusion

Un robot venant visiter une page web balisée comme montré dans cet article sait donc qu'il s'agit d'une page détail d'article, dispose de ses caractéristiques, et des avis des clients. On peut supposer que les avis clients ainsi récupérés, même s'ils ne sont pas encore affichés par Google, vont lui servir à juger de la qualité de la page, dans le cadre du fameux Panda. Un avantage sérieux, et une technique d'avenir.

Et maintenant, à vous de mettre en place le balisage de schema.org dans vos boutiques en ligne ! (ou alors de mettre aussi votre boutique chez RAYNETTE, et vous bénéficierez de cela automatiquement :-)

Sébastien Joncheray, gérant de la société Raynette, <http://www.raynette.fr>.

A propos de Raynette :

Raynette E-Business est une plateforme e-commerce : le e-commerçant choisit uniquement les modules qui lui sont nécessaires, pour que sa boutique soit strictement adaptée à ses besoins.

L'offre comprend la prise en charge de toute la technique (logiciel e-commerce / nom de domaine/ mail / maintenance / mise à jour / hébergement / mise en place paiement CB....), l'assistance et le conseil.

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://blog-abonnes.abondance.com/2011/07/un-exemple-dintegration-de-schemaorg.html>