

Contenuti portabili con DocBook

Marco Deseri

Docbook è un'applicazione di xml per descrivere testi strutturati, come libri e articoli tecnici. Consente una gestione dei contenuti orientata alle modalità d'uso dei visitatori del sito. DocBook definisce la struttura e la semantica di un documento, separandole completamente dalla forma: opportuni fogli di stile producono output in formati diversi (pdf, tex, html e rtf) dall'aspetto personalizzabile.

La flessibilità di Xml

Prima di occuparci nel dettaglio di DocBook, è meglio chiarire cos'è xml e quali vantaggi offre, in particolare per la gestione dei contenuti.

Xml è un linguaggio di markup, simile ad html, ma molto più flessibile. Attraverso xml è possibile separare la forma dal contenuto. Per un sistema di publishing, questa separazione offre vantaggi considerevoli e apre una serie di possibilità:

- Creazione di documenti portabili;
- Creazione di versioni diverse dello stesso documento, sia esteticamente, sia come formato, attraverso fogli di stile xslt: il file xml può essere trasformato in un pdf, in un file html, in un file rtf;
- Riutilizzo di parti del documento per scopi diversi, come la news syndication (pubblicare l'elenco degli ultimi titoli e relativi abstract);
- Possibilità di modificare l'aspetto dei documenti, senza intervenire sul contenuto;

Xml e Html

A differenza di html, che è un linguaggio espressamente concepito per descrivere documenti web, xml può essere usato per descrivere *qualsiasi tipo di risorsa*, perché consente di definire tipologie di documento e di popolarle con elementi definiti dall'utente.

La struttura di html è abbastanza rigida ed è stata definita dal w3c: un documento html è composto da alcuni elementi, gerarchicamente ordinati.

Esiste un elemento radice, da cui derivano rami, che a loro contengono altri elementi. L'elemento radice è html e contiene gli elementi head e body. Per ogni elemento sono definiti degli elementi figli, gli unici che all'interno del documento possono apparire a quel livello di profondità. Per esempio, l'elemento head può contenere gli elementi title e meta, ma non può contenere table. Invece, table può comparire all'interno dell'elemento body.

Questa struttura rigida è descritta nella DTD (document type definition), di cui esistono diverse versioni, leggermente diverse tra loro: la più ampiamente usata è html 4.01 transitional, mentre la più recente è xhtml 1.1. Sono le DTD che determinano la forma che può assumere html.

Con Xml si lavora ad un livello più astratto: chiunque può scrivere una propria DTD, definendo elementi personalizzati e i rapporti tra loro. Xml non definisce nulla dell'aspetto di un documento, ma ne descrive la struttura. Xml è in realtà un *meta-linguaggio*, cioè un linguaggio attraverso cui si possono definire altri linguaggi: nel caso di xml, i linguaggi derivati si chiamano *applicazioni xml*.

DocBook

Docbook è un'applicazione di xml: una DTD che descrive gli elementi che compongono un libro, e la loro organizzazione. Oltre alla *DTD*, che costituisce uno standard e garantisce compatibilità e portabilità dei documenti, docbook è accompagnato da una serie di *strumenti per il publishing*, quasi tutti free. In particolare, sono disponibili fogli di stile per convertire il codice xml in (x)html, pdf, rtf, tex, htmlhelp. Questi fogli di stile sono personalizzabili e il loro codice è disponibile a chiunque per modifiche anche sostanziali.

Vantaggi:

-

DocBook consente di *standardizzare la produzione* di testi, facilitando l'editing. I documenti docbook, almeno da un punto di vista tecnologico, hanno lo stesso formato e la stessa impostazione. Avere a che fare con autori diversi pone il problema della standardizzazione dei contenuti, perché, anche di fronte a style-guide precise, è inevitabile qualche dimenticanza da parte degli autori, o una preferenza particolare da parte degli editor. In fase di revisione e di editing è importante trovarsi di fronte a un documento ben leggibile, formattato secondo le proprie preferenze. Con docbook è possibile: ogni correttore può avere un foglio di stile personalizzato e trasformare il docbook a piacimento: interlinea, font preferito, margini del foglio.

-

DocBook rende i contenuti comprensibili e *manipolabili dalle macchine*. Identifica per ogni articolo (o ogni libro) il significato delle parti che lo compongono. In questo modo, anche una macchina è in grado di estrarre informazioni: il nome dell'autore, l'abstract, la bibliografia. In html questo scenario è impossibile, perché il markup non dice niente del significato: possiamo sapere che un blocco di testo è un paragrafo, non che in quel blocco di testo è scritto il nome dell'autore del documento.

Problemi aperti

Il rapporto tra *forma e contenuto* è particolarmente importante nella scrittura su Internet: l'efficacia di un testo non dipende solo dalla qualità del contenuto, o dall'esattezza della forma. L'armonia con gli altri elementi e l'equilibrio della pagina hanno pari importanza. La separazione tra forma e contenuto si scontra con la necessità di verificare l'impatto della pagina nel suo contesto.

Tuttavia, questo problema è aggirabile, almeno in parte. In primo luogo, perché un buon web writer dovrebbe essere in grado di produrre testi con paragrafi brevi, espressioni evidenziate opportunamente ed elenchi puntati per esprimere rapidamente concetti.

Oltre a questo, esiste un modo abbastanza semplice per verificare come apparirà la pagina, basta far avere ad ogni web writer il foglio di stile che produce il codice html. All'atto pratico, la separazione tra forma e contenuto non sarà mai totale; un articolo adattato ad un foglio di stile potrebbe deludere con uno completamente diverso. Allo stesso modo, se i contenuti sono stati pensati per il web, su carta potrebbero apparire inconsistenti. Questo perché la lettura funziona in maniera diversa su schermo e su carta, e l'unico modo per avere contenuti coerenti con il medium è quello di scriverli appositamente.

Conclusioni

Nella maggior parte dei casi, non ci sono le risorse per redigere contenuti specifici per ogni medium, o per renderne disponibili versioni diverse: un testo sintetico e scorrevole per il web, e un pdf più completo e approfondito per il download.

DocBook offre un compromesso accettabile, perché consente di offrire formati diversi dello stesso documento con pochissimo sforzo. Offrire oltre alla classica pagina html una versione pdf dei contenuti, spogliata della navigazione e dei banner, comodamente archiviabile, è un servizio molto utile, anche senza l'inserimento di contenuti ad hoc.

Nell'articolo *Contenuti orientati agli utenti* (<http://www.i-use.it/approfondimenti/contenuti/1/>), si parla delle diverse modalità di fruizione, con un'attenzione particolare all'influenza del tempo sul comportamento degli utenti: DocBook è una risorsa che consente di offrire una risposta a queste diverse esigenze.

Sul web

-

La risorsa più completa su DocBook è il sito del libro *DocBook: The Definitive Guide*.

-

Una guida piuttosto tecnica sull'installazione e la messa a punto di una suite di editing DocBook: *Installing And Using An XML/SGML DocBook Editing Suite*

-

Su *DocBook Open Repository* si trovano documentazione e strumenti open source per l'editing DocBook