



# HTML

## stream analysis system

### **DTD TAGS Validation**

- Transitional
- Strict

### **TAGS elaboration**

- Attribute replacement
- Content substitution

### **Conformance setting**

- A
- Double-A
- Triple-A

### **Pattern analysis**

### **Link encoding**

**© 2006/2008 Kosmous Srl**

Via G. Scarabelli, 6 - 00157 - Roma

Tel: +39 6 41735400 - Fax: +39 6 41796080

[www.kosmous.com](http://www.kosmous.com)

## il contesto

**L**La crescita incessante di internet come mezzo di comunicazione è un fatto ormai assodato. I più grandi esperti di tecnologia di tutto il mondo sono concordi nel ritenere internet uno strumento che pervaderà sempre di più la nostra vita quotidiana fino a diventare la nostra principale fonte di informazione.

Questo mare magno di siti, portali, blog, forum e quant'altro può in alcuni casi porre dei problemi che necessitano di una sorta di "pulizia" del flusso HTML in transito verso il client.

In quali casi potrebbe essere utile un tale accorgimento?

- In tutti quei casi in cui si volesse migliorare l'accessibilità del documento
- Per eliminare "al volo" dei contenuti ritenuti nocivi per l'utilizzatore
- Per inserire nuovi contenuti nei documenti
- Per eseguire qualsiasi tipo di filtro sui contenuti in transito

E' bene considerare che, ad oggi, sono sempre più diffusi strumenti CMS (Content management system) e che non sempre, purtroppo, tali

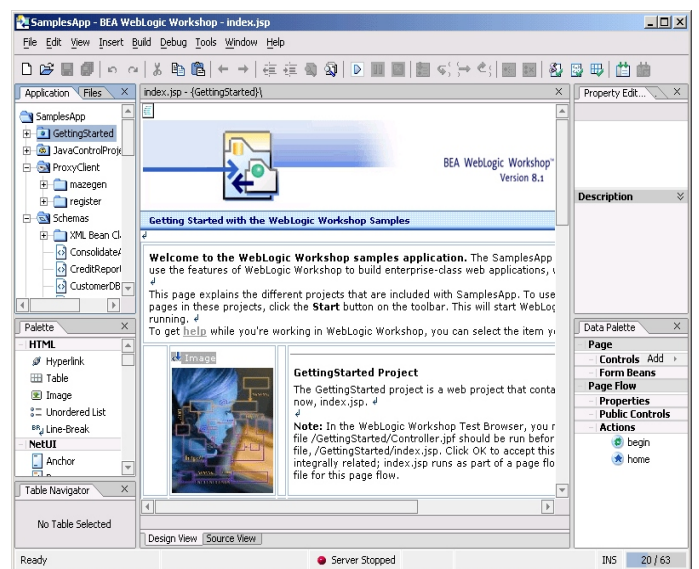
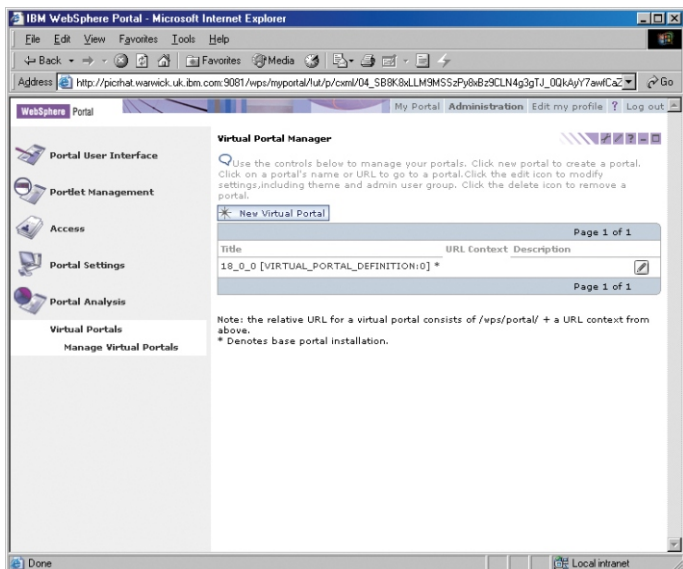
strumenti offrono la possibilità di personalizzare il codice HTML prodotto per la presentazione dei documenti.

Spesso vengono proposti dei veri e propri workaround che consentono di superare questo tipo di problemi sviluppando codice custom. Tali soluzioni limitano buona parte delle funzionalità originali del prodotto vanificando i vantaggi che erano stati inizialmente delineati nell'acquisto dello stesso.

Lo stesso ragionamento può essere fatto per tutti quegli applicativi web realizzati in tempi precedenti alla normativa sull'accessibilità e che richiederebbero ingenti costi di reingegnerizzazione.

Tutte queste considerazioni evidenziano spesso un problema di opportunità di intervento e/o di valutazione dei costi/benefici per poter prendere una decisione in merito al lavoro già svolto.

Ma sarà veramente necessario buttare tutto e ricominciare da capo? Naturalmente, essendo il problema notevolmente complesso, non esiste una unica risposta a questa domanda.





**A** quasi quarant'anni di distanza dai primi esperimenti di collegamento fra gli elaboratori la rete internet è divenuta un immenso contenitore di documenti scritti in tutte le lingue del mondo.

E' così che la comunità che ogni giorno lavora, studia, si intrattiene sulla rete ha avvertito l'esigenza di tutelare le persone diversamente abili che hanno bisogno di strumenti di supporto per fruire dei contenuti pubblicati sul web.

Nasce così il problema della cosiddetta "**Accessibilità**" alle informazioni esistenti online, ma cos'è esattamente l'accessibilità e cosa si intende con questo termine?

L'accessibilità, in informatica, indica la possibilità per un utente di fruire con facilità di un servizio o di una risorsa. Il termine è comunemente associato alla possibilità anche per persone con ridotta o impedita capacità sensoriale, motoria, o psichica, di fruire dei sistemi informatici e delle risorse software a disposizione.

Con accessibilità si indica anche l'insieme delle tecnologie che permettono ad una risorsa di essere accessibile.

Nel web, un "sito web accessibile" facilita l'accesso ad individui con ogni tipo di disabilità, ma anche ad individui non affetti da patologie

Esistono alcuni standard ed alcune linee guida per definire l'accessibilità. Le linee guida internazionalmente più diffuse per quanto riguarda il web sono le **WCAG**, *Web Content Accessibility Guidelines* e la **WAI**, *Web Accessibility Initiative* (sezione del World Wide Web Consortium).

Negli Stati Uniti è in vigore una legge, seguita come linea guida in tutto il mondo, denominata "Section 508", in Italia tutti i nuovi siti web e gli aggiornamenti in seno alla Pubblica Amministrazione e di enti privati rivolti al pubblico sono obbligati a tenere conto della "Legge Stanca".

## **l'accessibilità**

### **La Legge "Stanca"**

Al termine dell' Anno Europeo dei disabili, il Parlamento italiano ha approvato all'unanimità la "Legge Stanca" (così chiamata in riferimento al suo promotore **Lucio Stanca**, *ministro per l'Innovazione e le Tecnologie*) in materia di accessibilità da parte dei soggetti disabili alle nuove tecnologie digitali e informatiche.

Il voto è stato espresso nella mattina del 17 dicembre 2003 dalla Commissione Lavori Pubblici e Comunicazioni di Palazzo Madama, in sede legislativa.

In applicazione del principio costituzionale di eguaglianza, l'obiettivo della legge è quello di favorire l'accesso dei disabili (il 5% circa della popolazione), cercando di abbattere quelle "barriere virtuali" che limitano la loro partecipazione democratica alla Società dell'Informazione, escludendoli dal mondo dell'istruzione, del lavoro e da una migliore qualità della vita.

La legge si pone quale stimolo a tutta una serie di provvedimenti concreti e necessari che coinvolgono numerosi soggetti, pubblici e privati, per far in modo che le nuove tecnologie si pongano quali strumenti abilitanti e non di esclusione.

La " Legge Stanca" si pone come strumento incentivante nei confronti dei privati, mentre nei confronti della pubblica amministrazione reca degli obblighi sorretti da efficaci sanzioni. È previsto, infatti, che i nuovi contratti stipulati dalla pubblica amministrazione per la realizzazione di siti Internet siano colpiti da nullità qualora non rispettino i requisiti di accessibilità.





# HTML advanced content filter stream analysis system

## soluzioni

**G**li standard fissati dagli organismi internazionali in materia di accessibilità sono molto rigorosi e definiscono sia criteri oggettivi che soggettivi per la validazione di un documento.

La definizione dei criteri oggettivi è stata enunciata dal **W3C** (*World Wide Web Consortium*) attraverso la composizione di regole contenute in cosiddetti **DTD** (*Document Type Definition*). Tali regole definiscono i marcatori (Tag) che possono essere utilizzati in un documento HTML ben formattato e per ognuno di essi stabiliscono quali siano gli attributi consentiti o eventualmente quelli obbligatori.



Oltre alla definizione dei DTD il W3C ha composto un documento chiamato **“Web Content Accessibility Guidelines”** che enuncia le linee guida per la scrittura di un codice accessibile gestendo tre diversi livelli di priorità chiamati **“Conformance level”**:

- A (*Priority 1 checkpoints*)
- Double-A (*Priority 1, 2 checkpoints*)
- Triple-A (*Priority 1, 2, 3 checkpoints*)

Gli sviluppatori che intendano produrre documenti completamente accessibili devono attenersi alle norme definite per il **“Conformance level Triple-A”**. La normativa in vigore nello stato Italiano ritiene quest’ultimo livello di conformità l’obiettivo da raggiungere nello sviluppo di siti e portali nella Pubblica Amministrazione.



Di fronte a quanto appena illustrato come intervenire? Quali sono le possibili soluzioni?

Com’è ovvio la migliore soluzione consisterebbe nell’avere il massimo controllo sul codice HTML prodotto potendo così intervenire a monte per modificare template, inclusioni e quant’altro.

Purtroppo spesso questo non è del tutto possibile, vuoi perché lo strumento utilizzato non lo consente, vuoi perché il lavoro è stato già in gran parte sviluppato e le modifiche in questione comporterebbero la pressoché totale riscrittura delle applicazioni.

Quali soluzioni possono essere adottate allora in questi casi?

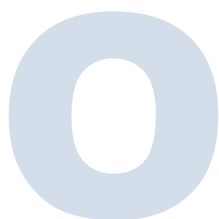
In prima istanza generalmente ci si rivolge al produttore del software chiedendo una patch che risolva il problema, ma non sempre il produttore in questione ha interesse a risolvere le problematiche legate magari agli ordinamenti specifici di paesi dove non ha un grande mercato.

Spesso quindi la soluzione scelta comporta la sostituzione dello strumento fino ad allora utilizzato o la riscrittura di un ambiente che offra le medesime funzionalità del prodotto originale ma con produzione di codice accessibile.





# HTML advanced content filter stream analysis system



**Orange-box** è la soluzione che costituisce una buona alternativa in tutti quei casi in cui il rapporto dei costi/benefici precedentemente illustrati è decisamente sfavorevole.

In tutti i casi in cui l'intervento diretto sui codici non può essere pianificato è possibile inserire un filtro fra il codice generato dall'applicativo ed il browser web che ha richiesto il documento.

Applicare una simile soluzione comporta una attenta valutazione dei parametri di performance del filtro come ad esempio:

- Il filtro è capace di gestire connessioni multiple richieste dai client o tende a "strozzare" le comunicazioni in un solo punto causando il cosiddetto "collo di bottiglia"?

- E' affidabile dal punto di vista della sicurezza?

- Gestisce le sole richieste in GET o può gestire anche le richieste inviate in POST con tanto di eventuali moduli o documenti allegati?

## Orange-box

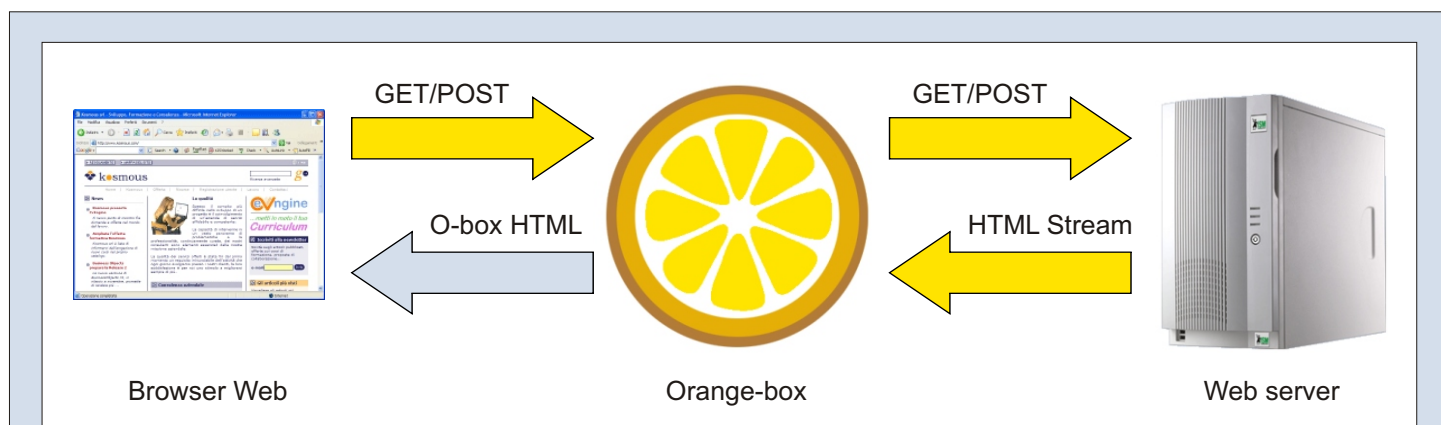
- Gestisce correttamente gli header scambiati fra il client ed il server web per il mantenimento della sessione e quant'altro?

- E' configurabile per adattarsi alle situazioni particolari che non sono nello standard?

Orange-box risponde a tutti questi requisiti essendo basato su tecnologie del mondo Open-Source di provata affidabilità.

Tecnicamente è un filtro di codice HTML in grado di intervenire "chirurgicamente" sul flusso di dati inviato al browser ed apportare le opportune correzioni, l'esistenza del filtro rimane trasparente all'utente che naviga il quale non si accorge neanche della sua esistenza.

Orange-box può sempre rendere accessibile una pagina non accessibile? No, il successo del filtro dipende molto da come sono realizzate le pagine HTML quindi va valutato caso per caso.



### Principio di funzionamento

1. Il browser effettua la richiesta che viene presa da Orange-box
2. Orange-box rigira la richiesta al Server Web contenente il sito
3. Il Server Web invia i dati HTML ad Orange-box
4. Orange-box effettua la sostituzione al volo dei dati in arrivo e restituisce il codice HTML filtrato al browser



# HTML advanced content filter stream analysis system



Gran parte del lavoro di customizzazione del filtro è relativa alla configurazione delle regole che vengono applicate al flusso (*HTML stream*) in transito attraverso di esso.

Per agevolare la configurazione dello strumento è stata realizzata una interfaccia web attraverso la quale è possibile comporre regole di filtro. Vengono fornite una serie di regole prestabilite ma non abilitate di default, attraverso di esse o eventualmente attraverso la configurazione di nuove regole, Orange-box eseguirà una serie di operazioni sul flusso HTML.

Dal pannello principale è possibile modificare le regole esistenti o inserirne di nuove, modificarne l'ordine di esecuzione, attivare o disattivare il filtro. Di seguito illustriamo la tipologia di regole comunemente utilizzate per la configurazione:

## HTML Tags

Consente di selezionare il tag insieme ad eventuali attributi dello stesso sul quale intervenire ed il tipo di operazione da effettuare come ad esempio:

- Sostituzione di uno o più attributi
- Sostituzione del TAG
- Sostituzione del contenuto

## Add Content

Inserisce codice HTML nel flusso in transito

## URL Encode link

Esegue la codifica URL dei link

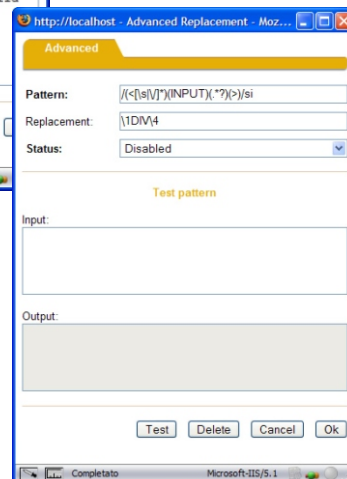
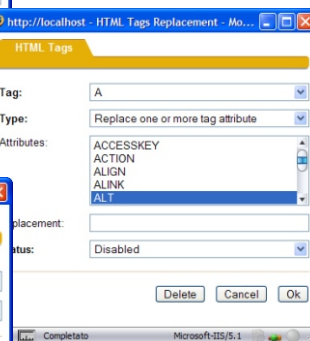
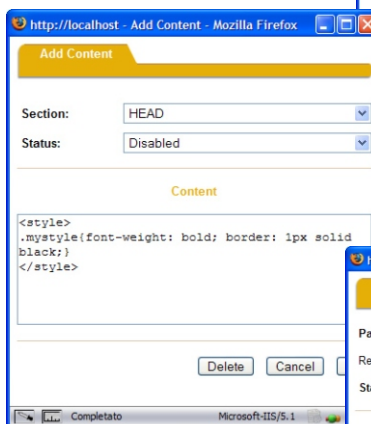
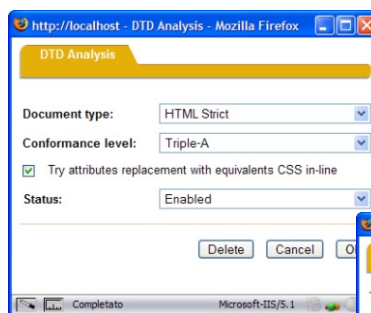
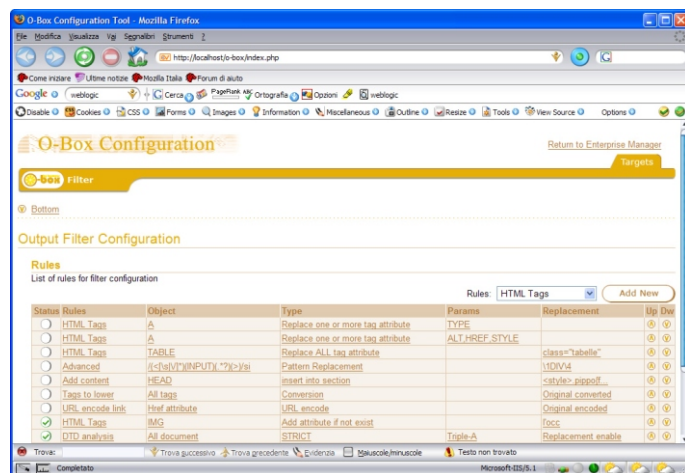
## Tags to lower

Converte in minuscolo tutti i tag

## Advanced

Consente di configurare operazioni avanzate non previste nelle regole standard. Tramite questo pannello di configurazione è possibile eseguire tutte quelle operazioni di sostituzione che richiedono un intervento di massima precisione.

## visto da vicino





# HTML advanced content filter stream analysis system

## caratteristiche

Le caratteristiche tecniche riportate in questa brochure sono relative alla versione 1.2 del prodotto, il produttore si riserva in qualunque momento di apportare modifiche alle caratteristiche tecniche illustrate in base alle esigenze di sviluppo e di mercato.

- Basato su Apache
  - Robusto dal punto di vista della sicurezza
  - Ben documentato, aggiornato e supportato dalla comunità Open Source
  - Gestisce molteplici connessioni non rappresentando un “collo di bottiglia”
- Multiplatforma
- Altamente scalabile
- Non necessita di ricompilazioni né di alcuna modifica dei prodotti originali
- Non presenta problemi di upgrade e porting del prodotto originale
- Facile integrazione negli strumenti di configurazione del prodotto originale  
Dotato di pannello di configurazione per la gestione delle regole di filtro
- Si inserisce perfettamente nel contesto dell'applicazione web che si sta utilizzando essendo la customizzazione grafica velocemente personalizzabile
- Assolutamente trasparente all'utente ed al customizzatore
- Pieno utilizzo di tutte le funzionalità del prodotto  
(*Consente di continuare ad utilizzare gli strumenti originali*)
- Indipendente dal Sistema Operativo
- In qualunque momento è possibile disabilitare temporaneamente il filtro e lavorare direttamente sui sistemi con senza dover effettuare il riavvio

Per qualunque ulteriore informazione, comprese le ultime novità sul prodotto, potete contattarci ai recapiti riportati su questa brochure oppure recarvi sul sito [www.kosmous.com/orange-box](http://www.kosmous.com/orange-box).





**© 2006/2008 Kosmous Srl**

Via G. Scarabelli, 6 - 00157 - Roma

Tel: +39 6 41735400 - Fax: +39 6 41796080