

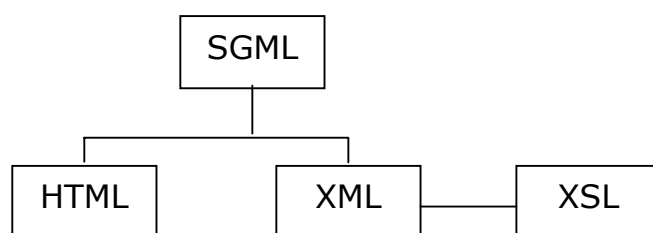
XML – Extensible Markup Language

XML è un linguaggio di marcatura estensibile (insieme di regole per creare altri linguaggi di marcatura) sviluppato e controllato dal World Wide Web Consortium (W3C).

Il suo sviluppo è iniziato nel 1996 ed è stato adottato come standard da W3C nel 1998.

L'attuale versione di questo standard è la 1.0, ma è in progressiva evoluzione¹.

XML è un sottoinsieme di SGML (Standard Generalized Markup Language) che definisce regole per scrivere linguaggi di marcatura, e non comprende tutte le funzionalità di SGML.



2

La marcatura consiste in un insieme di istruzioni, in genere rappresentate da tag, che si aggiungono ad un file. I tag sono utilizzati da un'applicazione in grado di interpretarli, per controllare la struttura del file o la presentazione delle informazioni contenute nel file marcato.

XML, contrariamente a HTML, è estensibile (metalinguaggio), non avendo tag predefiniti, e consente di scrivere altri linguaggi. Diverse organizzazioni hanno creato differenti “vocabolari” di marcatura specializzati in determinati settori.

La principale differenza tra HTML e XML risiede nel fatto che il primo, descrivendo le caratteristiche di presentazione delle informazioni, è uno standard per la presentazione, mentre il secondo, descrivendo il contenuto delle informazioni, è uno standard per l'elaborazione dei dati. XML, quindi, è usato per descrivere la struttura di un documento e non il modo con il quale deve essere presentato³.

In altri termini, HTML è un metodo universale per visualizzare i dati, XML è un metodo universale per “lavorare direttamente” con i dati, per descrivere e scambiare informazioni strutturate.

I linguaggi espressi tramite la grammatica XML, inoltre, sono conformi ad uno standard internazionale, possono essere “compresi” da diverse applicazioni e, al tempo stesso, sono calibrati per soddisfare specifiche esigenze proprio perché creati dall'utente.

L'interoperabilità tra applicazioni può essere facilitata dall'uso massivo dello standard XML sfruttando le caratteristiche di essere indipendente da sistemi e piattaforme. Tuttavia, non contenendo tag predefiniti ed essendo estensibile, sorge la necessità di condividere con altri attori (uomini o sistemi) il linguaggio creato, dichiarando il significato delle estensioni usate. La

¹ Per informazioni aggiornate consultare il sito <http://www.w3c.org>.

² XSL sta per Extensible Stylesheet Language e indica i fogli di stile creati specificatamente per documenti XML

³ Sintetizzando, un documento facilmente interpretabile deve contenere almeno 3 parti fondamentali rappresentate in modo distinto: contenuto, struttura, stile (rappresentazione “tipografica”).

descrizione della struttura di un file XML è fatta tramite DTD (document type definition, definizione del tipo di documento).

Alcune esigenze che spingono verso una tecnologia più evoluta rispetto ad HTML possono essere sintetizzate:

- internazionalizzazione del publishing elettronico che richiede indipendenza dalle piattaforme, dai media, dai linguaggi usati: il medesimo contenuto deve essere utilizzabile “ovunque”.
- nuove applicazioni web based che richiedono scambio di informazioni tra database, manipolazione dei dati da parte delle postazioni di lavoro (selezione, ordinamento,...), personalizzazione delle informazioni con agenti “intelligenti”, ecc.

La tecnica degli script, già usata per realizzare alcune delle funzionalità indicate, comporta difficoltà nella gestione (creazione, manutenzione, aggiornamento), e non consente di governare agevolmente la crescita delle informazioni da trattare.

Alcuni obiettivi ufficiali dello standard XML, definiti dal W3C, sono:

- facilità e usabilità su Internet
- supporto di ampie gamme di applicazioni
- facilità nella scrittura di programmi per l'elaborazione di documenti XML
- leggibilità da parte delle persone e ragionevole chiarezza
- rapidità della fase di disegno
- disegno formale e conciso

Usi di XML

XML si presta a poter essere utilizzato in un'ampia varietà di settori, particolarmente ove risulta critica l'identificazione del contenuto informativo.

L'uso più comune di XML è dato dall'elaborazione di testi, essenzialmente in quanto XML fornisce un insieme di regole specifiche per la creazione e la strutturazione di documenti. In quanto tale, XML può definirsi complementare piuttosto che concorrente rispetto ad HTML poiché quest'ultimo fornisce istruzioni di formattazione e visualizzazione unitamente al contenuto informativo XML.

Un'altra area è costituita dai database, in particolare per il publishing su Web. Si adatta bene per applicazioni che mostrano lo stesso documento in modi diversi con diversi dispositivi di visualizzazione.

Una terza area è data dai motori di ricerca. I motori di ricerca che fanno uso di HTML possono fornire risultati poco accurati, anche se eseguite con parametri di ricerca adeguati. Dal momento che XML tratta il significato e la struttura dei dati è in grado di migliorare in maniera significativa i risultati delle ricerche.

L'area di utilizzo che presenta maggiori prospettive di crescita è lo sviluppo di applicazioni. A causa delle caratteristiche mirate all'interscambio di dati, XML è una tecnologia adatta per

applicazioni di tipo distribuito, in particolar modo nei casi in cui i dati sono soggetti a frequenti modifiche.

Fra le prime organizzazioni che hanno adottato XML figurano gli operatori di e-commerce. In particolare, le aree applicative che possono trarre beneficio dall'uso di XML sono Business to Business (B2B) e Application to Application (A2A).

Caratteristiche di XML

Struttura: possono essere rappresentati dati gerarchici complessi

Validazione: i dati possono essere controllati nella loro "correttezza strutturale"

Indipendenza dai media: lo stesso contenuto può essere pubblicato su più media

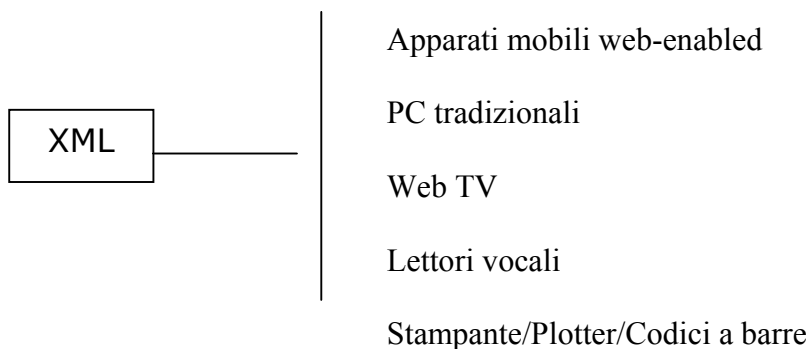
Svantaggi

- 1) performance
- 2) nuovo e in continua evoluzione
- 3) sicurezza

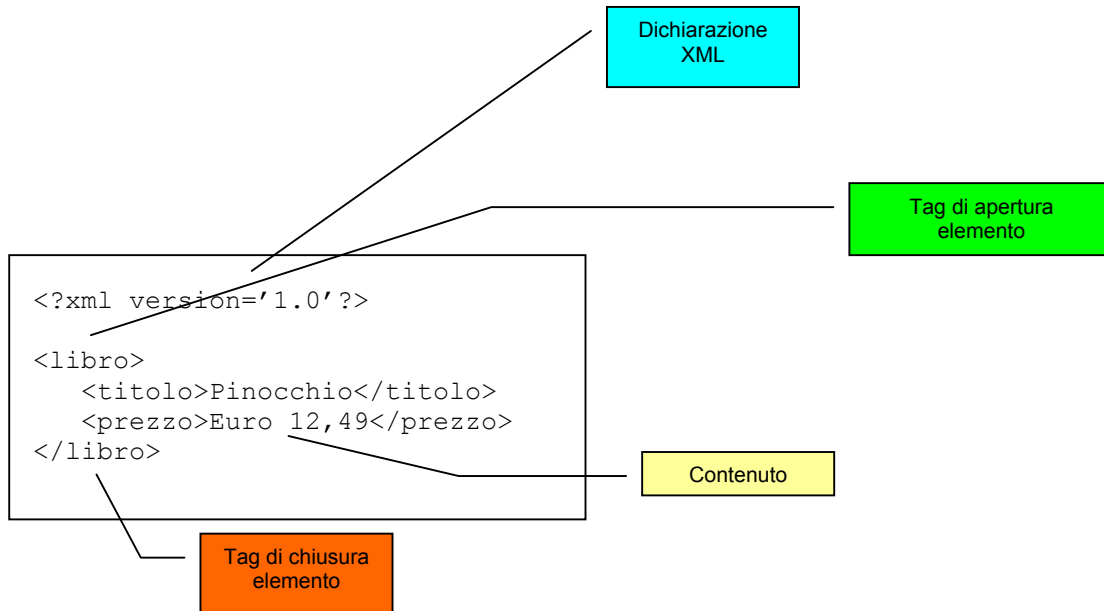
Tecnologia post-html

Gli script sono usati, in genere, per scopi specifici, la loro gestione può risultare difficile, non facilitano la gestione e "crescita" dei dati.

Interoperabilità generalizzata



Dal punto di vista sintattico l'elemento è composto dal tag di apertura, dal tag di chiusura e dal contenuto:



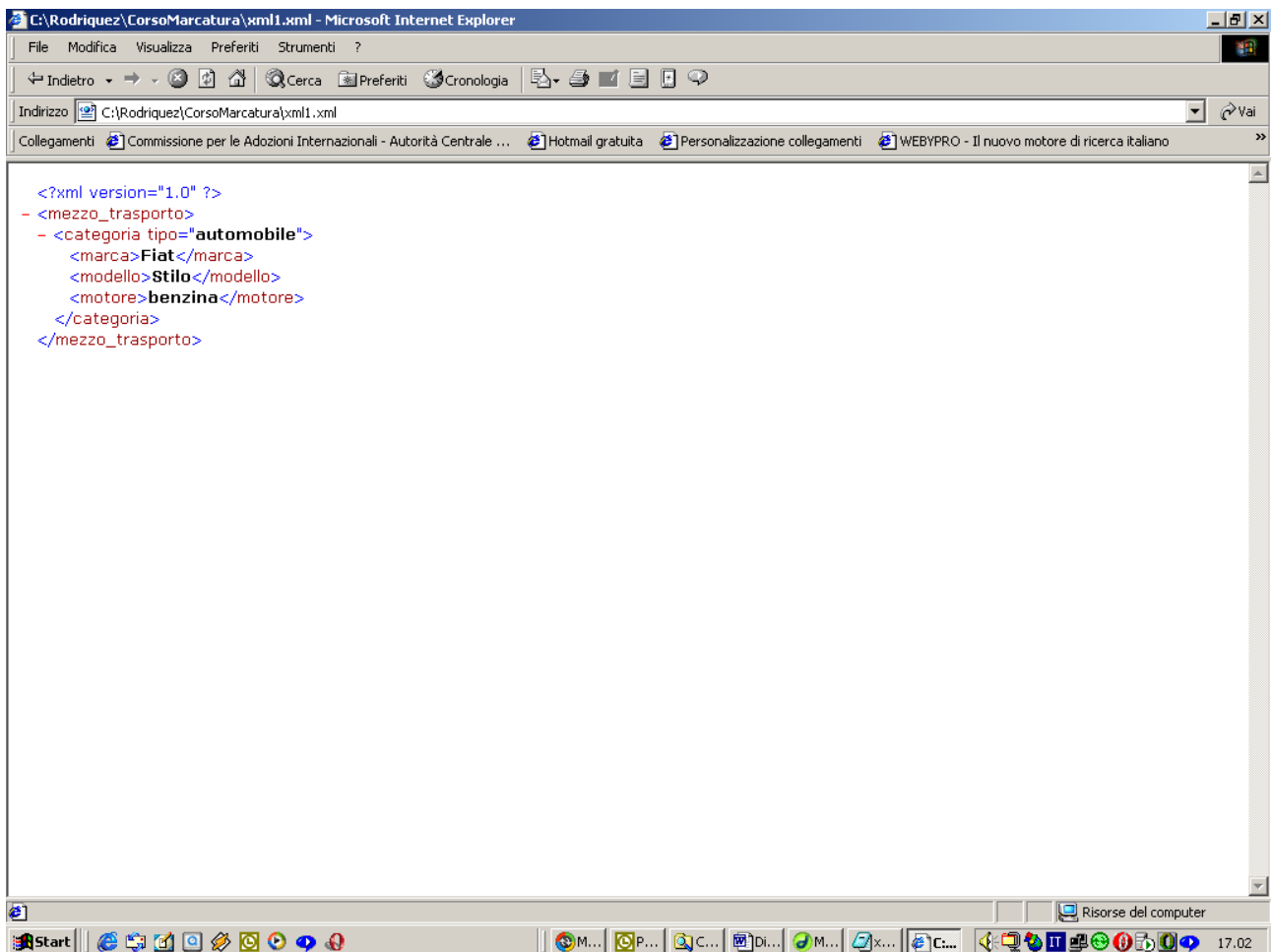
Esempio

```
<?xml version="1.0"?>
<mezzo_trasporto>
  <categoria tipo="automobile">
    <marca>Fiat</marca>
    <modello>Stilo</modello>
    <motore>benzina</motore>
  </categoria>
</mezzo_trasporto>
```

Si noti che:

- i tag sono autoesplicativi e indicano la natura del dato
- non sono presenti tag di gestione della formattazione

Aprendo il file con un browser che include un parser XML si ottiene l'immagine seguente:



Un documento XML può essere:

- ben formato
- valido

Un documento XML ben formato significa sintatticamente corretto (ciò può essere verificato dal parser).

E' valido se, *oltre* ad essere ben formato, è conforme alla DTD interna o esterna utilizzata (esistono anche parser che eseguono il controllo di validità).

Si tenga presente che una DTD non è un documento XML.

Esempio documento ben formato: `<saluto>Hello,world</saluto>`

Esempio documento non ben formato: `<saluto>Hello,world</Saluto]`

(il parser XML genera un errore perché il tag iniziale non corrisponde esattamente al tag finale: la grammatica XML è sensibile ai caratteri maiuscoli/miniscoli)

Es. documento valido

```
<!DOCTYPE saluto SYSTEM "saluto.dtd">
<!ELEMENT saluto (#PCDATA)>
<!ATTLIST saluto type (informale|formale) "formale">
```

il documento è valido perché è conforme a questa DTD.⁴

Elementi, informazioni di controllo, entità

Un documento XML è composto da:

- elementi
- informazioni di controllo
- entità

Elementi

Le unità di informazioni marcate con specifici tag sono dette elementi.

Il documento deve avere *un solo* elemento radice e un numero variabile di sezioni. Anche le sezioni sono elementi e sono delimitate da tag.

I tag dei vari elementi devono essere nidificati⁵ e sono obbligatori i tag di chiusura.

Inoltre la grammatica di XML è case-sensitive.

Attributi

Ogni elemento può essere integrato con ulteriori informazioni. Tali informazioni associate agli elementi sono dette attributi, e sono usati per contenere valori che descrivono “meglio” gli elementi ai quali si riferiscono.

E' da sottolineare che la definizione degli elementi e dei loro attributi descrittivi è un passo importante.

Decidere se un attributo deve essere portato a livello di elemento così come selezionare i possibili attributi che meglio descrivono gli elementi è un compito che rientra nella fase di analisi e modellazione dei dati.

In maniera indicativa si potrebbe affermare che i dati essenziali del modello (ovvero quei dati da sottoporre ad elaborazione diretta) dovrebbero essere dichiarati come elementi. Gli attributi dovrebbero invece essere quei dati che rivestono significato solo nel contesto di riferimento o che non variano (significativamente) con il trascorrere del tempo.

Ad es. in un inventario di magazzino si suppone che il nome di un bene non vari nel tempo, mentre la quantità ed il valore possono variare. Di conseguenza il nome potrebbe essere un attributo e la quantità ed il valore dovrebbero essere codificati come elementi.

⁴ Successivamente verranno esaminate le caratteristiche della DTD.

⁵ Significa che i tag di chiusura devono apparire in ordine inverso rispetto all'ordine con il quale appaiono i rispettivi tag di apertura.

Un altro esempio può essere costituito dalla necessità di trattare informazioni di minore importanza che si desidera memorizzare ma non necessariamente visualizzare e, pertanto, tali informazioni possono essere assegnate ad attributi invece di collocarle nel contenuto di un documento (la visualizzazione di un documento XML con fogli di stile CSS non mostra gli attributi).

Inoltre si consideri che:

- gli attributi non sono facilmente espandibili per apportare successivamente eventuali modifiche
- è più complesso manipolare gli attributi tramite il codice del programma
- è più complesso testare i valori degli attributi con una DTD

Informazioni di controllo

Le informazioni di controllo sono:

- commenti
- istruzioni di elaborazione
- Document Type Definition

Commenti

Lo scopo dei commenti è quello di facilitare le operazioni di sviluppo e manutenzione, anche a distanza di tempo, fornendo chiarimenti, esempi, presupposti, regole, ecc., soprattutto quando il lavoro è svolto da più persone che lavorano in gruppo o quando si tratta di DTD *pubbliche*.

La sintassi per l'inserimento di commenti è la stessa sia per i documenti XML sia per i DTD.

```
<!--Questo è un esempio di commento XML -->
```

I commenti possono essere inseriti in qualsiasi punto del documento XML purché esterno ad altri tag e non possono contenere i caratteri “--“ (doppio trattino).

Ad es. il seguente commento non è valido perché collocato all'interno di un tag:

```
<elemento <!-- -Questo è un esempio di commento XML non valido --> >
    Documento XML </elemento>
```

I commenti possono essere inseriti anche all'interno della DTD.

Istruzioni di elaborazione

Lo scopo delle istruzioni di elaborazione è quello di “comunicare” con le applicazioni, fornendo informazioni che il processore XML passa all'applicazione.

Il processore XML è un software che:

- legge e memorizza il contenuto del documento XML
- consente di accedere al contenuto del documento

L'applicazione è un altro software che:

- accede al contenuto del documento tramite il processore
- manipola e visualizza il contenuto del documento

Ad es. il browser Explorer (v. 5) fornisce il processore XML e una parte dell'applicazione. Se il documento XML è visualizzato tramite codice HTML e script per la manipolazione /visualizzazione si è fornita parte dell'applicazione.

Esempi istruzioni

```
<?xml version='1.0'?>
```

Questa istruzione comunica all'applicazione che il contenuto del file è conforme agli standard XML versione 1.0. Deve necessariamente essere la prima riga del file XML. Sebbene lo standard preveda l'obbligatorietà di questa istruzione alcuni parser (es.: Explorer), in mancanza, la considerano implicitamente presente.

```
<?xml-stylesheet type="text/css" href="Biblioteca01.css"?>
```

Questa istruzione comunica all'applicazione di utilizzare il foglio di stile disponibile nel file Biblioteca01.css.

Analogamente a quanto previsto per i commenti, le istruzioni di elaborazione possono essere inserite in qualsiasi punto del documento XML purchè esterno ad altri tag.

Esempio

```
<?xml version="1.0"?>

<!-- Nome file: Biblioteca.xml -->

<?xml-stylesheet type="text/css" href="Biblioteca01.css"?>

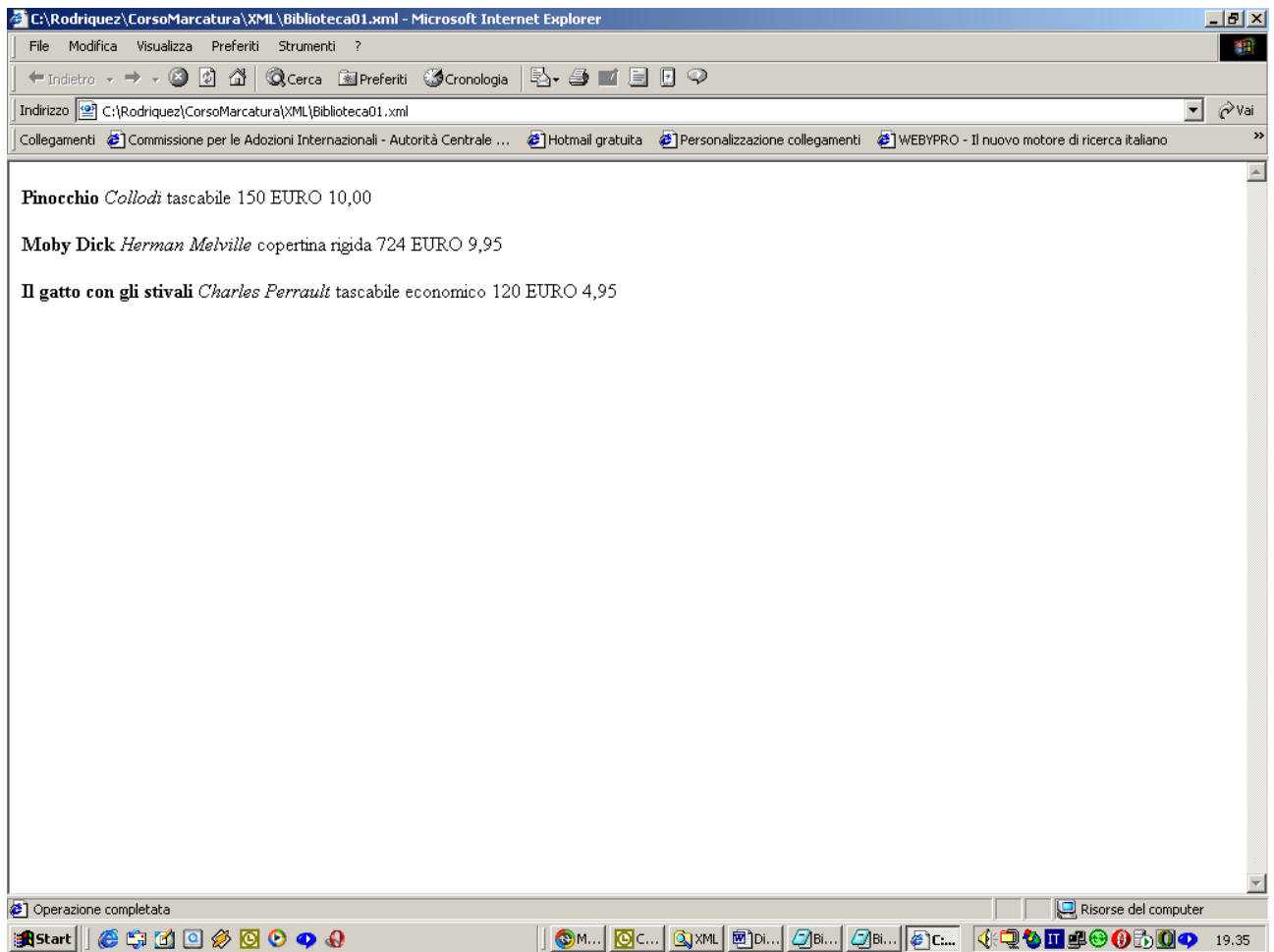
<BIBLIOTECA>
  <BOOK>
    <TITLE>Pinocchio</TITLE>
    <AUTHOR>Collodi</AUTHOR>
    <BINDING>tascabile</BINDING>
    <PAGES>150</PAGES>
    <PRICE>EURO 10,00</PRICE>
  </BOOK>
  <BOOK>
    <TITLE>Moby Dick</TITLE>
    <AUTHOR>Herman Melville</AUTHOR>
    <BINDING>copertina rigida</BINDING>
    <PAGES>724</PAGES>
    <PRICE>EURO 9,95</PRICE>
  </BOOK>
  <BOOK>
    <TITLE>Il gatto con gli stivali</TITLE>
    <AUTHOR>Charles Perrault</AUTHOR>
    <BINDING>tascabile economico</BINDING>
    <PAGES>120</PAGES>
    <PRICE>EURO 4,95</PRICE>
  </BOOK>
</BIBLIOTECA>
```

```
/* Nome file: Biblioteca01.css */

BOOK
{display:block;
 margin-top:14pt;
 font-size:12pt}

TITLE
{font-weight:bold}

AUTHOR
{font-style:italic}
```



Document Type Definition (DTD)

L'utilizzo delle DTD è opzionale in quanto in determinati casi non sono rilevanti mentre in altre situazioni rivestono un aspetto determinante. Non sono rilevanti, ad es., per informazioni scambiate all'interno di applicazioni, come avviene in un unico database, o di una singola organizzazione, ma sono critiche per scambi informativi tra applicazioni e tra organizzazioni diverse.

Ciascun documento XML può essere associato ad una DTD che può essere contenuta nello stesso file XML che la utilizza o può essere specificata in un file esterno. In quest'ultimo caso la DTD potrà essere utilizzata da più file XML.

La DTD contiene la particolare grammatica specificata per una determinata struttura XML. Contiene quindi l'indicazione degli elementi, il loro ordine di apparizione, i tag utilizzati. Tale grammatica verrà usata dai parser validatori per determinare, oltre alla correttezza sintattica del file (ben formato), anche che è conforme alla grammatica stabilita (valido).

```
<!DOCTYPE Ricetta SYSTEM "Ricetta.dtd">
```

Questa dichiarazione, inserita nel file XML, comunica al parser che il file usa una DTD specificata in un altro file denominato Ricetta.dtd e che tale DTD si trova nella stessa directory del file XML. Per convenzione si usa per la DTD lo stesso nome del file XML (Ricetta.xml - Ricetta.dtd).

La parola chiave SYSTEM è usata per indicare una DTD sviluppata dall'utente per sue proprie esigenze e che ha in genere validità solo all'interno della propria organizzazione. La parola chiave PUBLIC indica invece una DTD sviluppata e riconosciuta pubblicamente, anche a livello internazionale.

Possono essere elaborate sia dichiarazioni interne sia dichiarazioni esterne. Quelle interne *prevalgono* su quelle esterne.

```
<?xml version="1.0"? standalone="yes"?>
```

Il valore "yes" indica che saranno usate solo le dichiarazioni interne.

Come accennato, la DTD contiene le regole di definizione dei tag (markup) e indica gli elementi e il loro ordine nel documento XML e stabilisce, tra l'altro, condizioni obbligatorie e facoltative alle quali i file XML devono conformarsi.

Con la DTD si esprimono quindi i metadati mentre i dati sono contenuti nel file XML.

Si noti che:

- i file di tipo DTD non sono documenti XML
- il solo file XML, senza una DTD che rappresenta la grammatica le cui regole devono essere rispettate, è "privo di significato"
- la DTD è espressa in forma testuale e pertanto può essere compresa da un lettore umano abbastanza facilmente

La strutturazione della DTD riveste un'importanza fondamentale e richiede un'attenta attività di analisi e di modellazione dei dati relativi al contesto di riferimento.

Passi per la definizione di una DTD

- 1) definizione del proprio linguaggio di markup (tag)
- 2) definizione della struttura del documento
- 3) definizione dei vincoli (constraint) e dei valori predefiniti (default) per gli attributi

Il file XML può contenere al suo interno la DTD. In questo caso tutta la DTD è collocata in un unico tag DOCTYPE delimitato da parentesi quadre.

Esempio

```
<?xml version='1.0?'>

<!DOCTYPE recipe [

    <!ELEMENT cookbook (category+)>
    <!--altre istruzioni della DTD →
]>

<cookbook>
    <!--qui si inserisce il resto del file XML →
</cookbook>
```

Elementi (segue)

La struttura di un documento XML è esplicitata nella DTD. Un documento XML è costituito da elementi che possono contenere al loro interno altri elementi. Determinati elementi possono contenere, a loro volta, attributi. Il primo nodo, che contiene tutti gli altri, è detto nodo radice e tutti i documenti XML ne devono contenere obbligatoriamente uno. Se ad es. il nodo radice è denominato *cookbook* e tutti i nodi in esso contenuti sono denominati *category* ciò è così indicato nella DTD:

```
<!ELEMENT cookbook (category+)>
```

Si consideri che gli elementi devono avere un nome univoco nell'ambito di uno stesso documento e che tutti gli elementi sono dichiarati con lo stesso formato.

Il tag element è indicato con un punto esclamativo seguito dalla parola ELEMENT in caratteri maiuscoli, dal nome dato all'elemento e infine da altre informazioni racchiuse tra parentesi. Ogni elemento può contenere altri elementi o può definire dei dati. Se l'elemento contiene altri nodi, nelle parentesi sono indicati i sottoelementi separati da una virgola. Ogni sottoelemento può essere accompagnato da un operatore che indica le regole con le quali il sottoelemento deve essere presente.

Operatori

“*” indica che l'item appare più volte o zero volte (es.: item*)

“,” suddivide l'item in base ad una sequenza di altri item nell'esatto ordine con cui la sequenza si presenta (es.: riassunto, corpo del capitolo)

“()” racchiude un gruppo di item (es.: item1, item2)

“|” indica item alternativi, dei quali se ne può presentare soltanto uno (es.: acceso | spento)

“+” indica che l'item appare almeno una volta (es.: libro+)

“?” indica che l'item appare una sola volta o nessuna (es.: biblioteca?)

Esempio uso operatori

Si consideri una semplice rappresentazione di una biblioteca, costituita solo da libri (senza ulteriori dettagli), un catalogo di ricerca, un chiosco multimediale ed eventuali punti di ristorazione. I dati di questi tre elementi devono essere espressi nello stesso ordine. La biblioteca, oltre ai libri, dispone di un solo catalogo di ricerca e, eventualmente, di un solo chiosco di accesso multimediale. Potrebbe infine disporre di almeno due punti di ristorazione.

La dichiarazione del tipo di elemento *biblioteca* (della specifica biblioteca descritta) è:

```
<!ELEMENT biblioteca (libro+, catalogo, chiosco?, puntoristorazione*)>
```

Con questo formalismo è quindi possibile rappresentare in modo rigoroso realtà nelle quali gli elementi si presentano con una data sequenza, in modo opzionale, con ripetizioni, in alternativa.

Queste possibilità possono essere espresse anche in modo combinato utilizzando gli operatori e le parentesi, al fine di attribuire le corrette priorità all'interno della combinazione.

Si ricorda che ogni documento XML deve contenere un solo elemento radice (elemento di livello più alto all'interno della struttura) che contiene tutte le restanti parti del documento.

Esempio

```
<!ELEMENT recipe (name, ingredient+, cooking+, serves?, instruction*)>
```

Con questa definizione dell'elemento "recipe" si indica che il nome appare solo una volta, è obbligatorio che siano presenti almeno un elemento "ingredient" e "cooking", che è ammesso solo un elemento "serves" e che è previsto un numero non limitato di elementi "instructions".

PCDATA⁶ e CDATA

Gli elementi che contengono dati sono dichiarati con il seguente formato:

```
<!ELEMENT NomeElemento (#PCDATA)>
```

Qui le parentesi contengono il tipo di dati dell'elemento, preceduto dal carattere "#".

I tipi di dato comunemente inclusi negli elementi sono PCDATA e CDATA.

Tutto il testo contenuto all'interno di un documento XML è analizzato dal parser, ad eccezione della parte di testo inclusa in una sezione CDATA.

CDATA consiste in un set di caratteri che non viene esaminato dal parser XML. Questo tipo di dato è utile se, ad esempio, il set contiene caratteri usati per la marcatura come &, <, > e non si vuole che il parser li analizzi per evitare la generazione di falsi errori, oppure se la sezione CDATA contiene parti di script per l'elaborazione dei dati presenti nel documento XML. Un esempio di quest'ultimo caso può essere dato da un documento XML che contiene denominazione, quantità e prezzo unitario di un articolo da acquistare e la sezione CDATA che contiene le istruzioni di elaborazione per il calcolo dell'importo totale dell'articolo acquistato e per la sua visualizzazione.

PCDATA (Parsed character data) è invece un set di caratteri analizzato dal parser XML ed è il tipo di dato di default degli elementi.

Si noti che sebbene un documento XML può contenere diversi tipi di dato è preferibile che quelli maggiormente complessi (es.: gif) siano inclusi in entità.

⁶ PCDATA sta per dati di caratteri analizzati (Parsed Character Data). Il processore XML analizza i caratteri all'interno degli elementi ricercando i tag XML.

Esempio di elemento empty

```
<elementovuoto attributo="abc"/>
```

L'elemento è vuoto perché non contiene dati o altri elementi ma solo un attributo.

Attributi (segue)

Un elemento XML, oltre a contenere dati o altri elementi, può essere corredato da attributi.

Il concetto di attributo è già presente in HTML, come nell'esempio del tag hyperlynk riportato di seguito:

```
<a href="http://www.governo.it">Link al sito del Governo italiano</a>
```

In questo esempio il tag HTML dell'indirizzo è costituito dal testo sul quale cliccare per raggiungere l'indirizzo di destinazione che è invece un attributo del tag.

Gli attributi indicano particolari caratteristiche del contenuto degli elementi. In pratica aggiungono informazioni, e hanno significato solo nell'ambito dell'elemento al quale si accompagnano. Sono composti da una coppia formata dal nome dell'attributo e dal suo valore.

Una volta individuati gli elementi, i loro attributi descrittivi si definiscono nella DTD che rappresenta la struttura del documento XML.

La definizione si esegue con la parola chiave ATTLIST, che può includere:

- nome attributo dell'elemento
- tipo di dato di ogni attributo
- eventuale valore di default per ogni attributo
- informazioni di controllo

Gli attributi devono essere inseriti dopo il tag di apertura e il nome dell'elemento al quale si riferiscono. All'interno di uno stesso elemento i nomi degli attributi devono essere univoci. Il valore dell'attributo è racchiuso tra virgolette.

Esempio

```
<!ATTLIST qty  
    amount CDATA #REQUIRED  
    unit CDATA "g">
```

Questa dichiarazione definisce gli attributi, "amount" e "qty", entrambi di tipo CDATA, il che indica che il testo contenuto nei due attributi non verrà preso in considerazione dal parser XML. La parola chiave REQUIRED indica che l'attributo amount è obbligatorio. Ciò significa che se tale attributo non verrà valorizzato quando si usa l'elemento, il parser XML segnalerà l'errore.

Si noti che tale errore è segnalato perché l'attributo è dichiarato come obbligatorio ma che, una volta valorizzato, il parser non ne analizzerà il contenuto in quanto il tipo di dato è dichiarato come CDATA.

Infine, l'attributo "unit" non è obbligatorio e presenta un valore di default "g". Se questo attributo non verrà valorizzato nel file XML, verrà usato il valore "g".

Oltre che REQUIRED un attributo può essere definito FIXED (opzionale ma se presente è valido il solo valore specificato) o IMPLIED (opzionale senza valore di default).

Il tipo di un attributo definisce l'insieme dei possibili valori.
Alcuni tipi di attributo possono essere:

- CDATA in cui è consentito inserire qualsiasi stringa testuale. Si noti che il tipo attributo CDATA è diverso dalla sezione CDATA

Esempio

```
<!ATTLIST libro
    titolo CDATA #REQUIRED ....
```

- ID il cui funzionamento è analogo a quello di una chiave primaria delle tabelle di un database relazionale.

Si consideri che quindi, analogamente a quanto avviene per la chiave primaria di una tabella di database, in un documento non possono esistere valori uguali di ID.

Esempio

```
<!ATTLIST libro
    NumeroInventario ID #REQUIRED ....
```

- IDREF/IDREFS il cui funzionamento è analogo a quello di una chiave esterna delle tabelle di un database relazionale. Analogamente a quanto avviene per la chiave esterna delle tabelle di database, in un documento il valore IDREF può essere ripetuto e fa riferimento all'attributo di tipo ID di un elemento del documento.

- elenco di nomi che indica che il valore di un attributo deve essere uno fra quelli compresi nell'elenco.

Esempio

```
genere (uomo|donna)
```

Documenti XML validi

Un documento XML valido risponde ai seguenti requisiti:

- è un documento XML strutturato correttamente (un documento non strutturato correttamente non può essere considerato un documento XML)
- include una DTD, dichiarazione del tipo di documento, che ne definisce la struttura
- è conforme alla DTD

Attraverso l'inclusione di una DTD un processore XML verifica se il documento XML è conforme a tale DTD. Si consideri che lo standard XML definisce una DTD come la "grammatica di una classe di documenti".

La creazione di documenti validi garantisce l'uniformità tra gruppi di documenti simili. La creazione di una singola DTD e la sua inclusione in tali documenti garantiscono l'uniformità di tutti i documenti a tale struttura.

Esempi di *modello di contenuto*

```
<!ELEMENT film (protagonista* | regista | produttore)>
```

In base a tale modello di contenuto saranno validi ognuno dei tre elementi seguenti:

1)

```
<film >
  <protagonista>Totò</protagonista>
  <protagonista>Massimo Troisi</protagonista>
</film>
```

2)

```
<film >
  <regista>Steno</regista>
</film>
```

3)

```
<film/>      (Elemento vuoto)
```

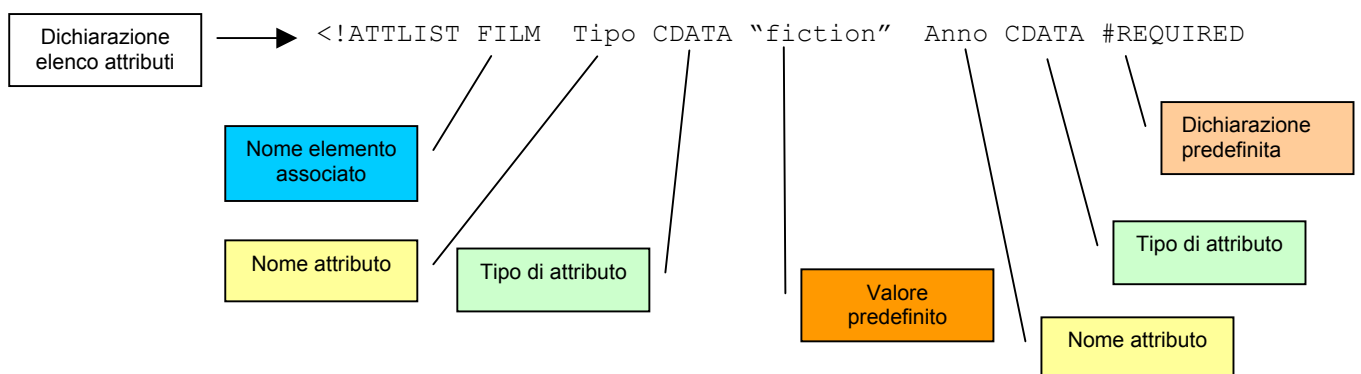
In un documento XML valido vanno dichiarati anche tutti gli attributi associati agli elementi. Il corrispondente tipo di tag è detto *dichiarazione dell'elenco degli attributi*.

Esempio

```
<!ELEMENT FILM (TITOLO, (PROTAGONISTA* | REGISTA | PRODUTTORE))>
```

```
<!ATTLIST FILM  Tipo CDATA "fiction"  Anno CDATA #REQUIRED
```

Con la prima istruzione è stato dichiarato l'elemento denominato film, con la seconda sono stati dichiarati due attributi (tipo e anno) per gli elementi film.



Omettendo l'attributo di un elemento verrà automaticamente assegnato il valore predefinito "fiction" mentre all'attributo anno si può assegnare qualsiasi stringa tra virgolette valida. Anno è un attributo obbligatorio (parola chiave "REQUIRED") e non ha un valore predefinito.⁷

Il tipo di attributo indica il tipo di valore (di tipo stringa, numerico, ecc.) assegnabile all'attributo all'interno del documento.

In particolare va ricordato un tipo di attributo definibile con la parola chiave ID. Tale parola chiave indica che all'attributo così definito deve essere assegnato un valore univoco per ciascun elemento.⁸

Entità

Un documento XML può contenere una o più entità. Le entità sono *oggetti*, non costituiti da elementi semplici, che sono usati di frequente e che possono essere inclusi in più file XML, anziché essere definite di volta in volta per ogni file XML. Si tratta, in altri termini, di nomi simbolici che verranno usati al posto di altre definizioni.

Le entità ospitano un generico contenuto.

Sono dette interne se incluse all'interno del file XML o esterne se ospitate in un altro file.

Possono inoltre essere di tipo parsed o non parsed. E' preferibile definire le entità che ospitano contenuti complessi come di tipo non parsed.

Le entità sono anche un modo per rappresentare caratteri speciali come &,<,>,"

Ad esempio, non è possibile scrivere

```
<test>if prezzo <150 then</test>
```

in quanto il parser interpreta il carattere < come l'inizio di un nuovo elemento, e genererebbe un errore.

Il testo precedente deve essere scritto come segue:

```
<test>if prezzo &lt;150 then</test>
```

⁷ Si tenga presente che se per un elemento si includono più dichiarazioni di attributi e ad un attributo viene dichiarato per uno stesso elemento più di una volta viene utilizzata la prima dichiarazione e la seconda viene ignorata.

Il caso delle dichiarazioni multiple dell'elenco attributi può essere frequente in presenza di DTD interne ed esterne.

⁸ Analogamente a quanto avviene nelle basi di dati nelle quali, tipicamente, le chiavi primarie sono individuate con i nomi Id, Identity, IdRef, ecc.

Per l'utilizzo di caratteri speciali si riporta una tabella riepilogativa

<i>Entità</i>	<i>Descrizione</i>	<i>Valore</i>
<	Parentesi angolare sinistra (minore)	<
>	Parentesi angolare destra (maggiore)	>
&	e commerciale (&)	&
'	Apostrofo	'
"	Virgolette doppie	"

Si tenga comunque presente che i caratteri illegali in XML sono propriamente "<" e "&", mentre è opportuno sostituire gli altri indicati nella tabella, anche se in realtà il loro uso è ammesso.

E' anche possibile includere nelle entità i caratteri speciali rappresentandoli mediante il formato &#nnn in cui *nnn* indica il codice decimale di un carattere.

Le entità possono essere anche usate per includere nella struttura il contenuto di file esterni.

Entità interne

Le entità interne sono utili per creare contenuti informativi che devono essere ripetuti frequentemente all'interno della struttura.

Oltre a facilitare la leggibilità e la manutenzione dello schema, le dimensioni del file XML sorgente saranno ridotte.

```
<!ELEMENT TEATRO "Teatro San Carlo">
```

Questa istruzione può essere usata come indicato di seguito, ogni volta che occorre

```
<item> Il tour si conclude con uno spettacolo al &TEATRO;</item>
```

L'uso di un'entità interna, come illustrato nell'esempio, è indicato con il carattere "&" seguito dal carattere ";

Entità esterne

```
<!ELEMENT logoaziendale SYSTEM "logo.png" NDATA PNG>
```

```
<!ELEMENT jingleaziendale SYSTEM "jingle.wav" *>
```

La prima istruzione indica come si crea un'entità esterna che contiene un'immagine png (Portable Network Graphic). La seconda si riferisce ad un file audio (il simbolo "*" indica al parser di non analizzare l'elemento).

La parola chiave SYSTEM indica, come già visto, che si tratta di una DTD creata dall'utente ad uso interno all'organizzazione.

La parola chiave NDATA, alla fine della dichiarazione dell'entità, indica che per questo tipo di dato è stata creata una *notazione* che serve per identificare specifici tipi di dati binari esterni (es. file in formato "gif").

Ciò è richiesto in quanto le applicazioni ed i parser XML sono in grado di elaborare un insieme limitato di tipi di dato. Se un'applicazione usa un tipo di dato non elaborabile dal parser XML (es.: oggetto multimediale) deve essere indicata un'applicazione esterna alla quale trasferire il dato per eseguire l'elaborazione.

```
<!NOTATION PNG SYSTEM "xv">
```

In questo esempio si cede il controllo al programma UNIX standard xv di visualizzazione.

Spazi dei nomi (namespaces)

Poiché i nomi degli elementi in XML non sono fissi possono verificarsi conflitti se più documenti usano gli stessi nomi per descrivere elementi differenti. Può quindi accadere, nel caso di numerosi documenti XML creati all'interno di grandi organizzazioni, che alcuni nomi di elementi possano essere duplicati. Unendo documenti che usano lo stesso nome di elementi ma riferiti a definizioni differenti si generano conflitti. Gli spazi dei nomi forniscono un metodo per evitare tali conflitti.

Gli spazi dei nomi sono, in sintesi, una lista di nomi di tag identificati tramite un riferimento ad una risorsa specifica o ad un URI (Uniform Resource Indicator). Offrono ai tag XML una sorta di polimorfismo grazie al quale un medesimo nome di tag può essere usato per indicare diversi tipi di dati in diversi documenti o applicazioni.

Gli spazi dei nomi sono costituiti da due componenti: un prefisso, che indica dove trovare la risorsa che definisce lo spazio dei nomi, ed una componente locale, che è il nome usato nel documento o nell'applicazione.

Il parser XML usa i nomi degli elementi per creare una rappresentazione degli item al suo interno.

Esempio

```
<persona>
  <nome>Ferdinando Rodriquez</nome>
</persona>

<ufficio>
  <nome>Informatica</nome>
  <ubicazione>Primo piano stanza 9999</ubicazione>
</ufficio>
```

L'unione dei due documenti sopraindicati genera un conflitto tra i nomi di elementi in quanto entrambi usano un elemento <nome> con diverso significato. E' evidente che nel caso di documenti piccoli è possibile ridenominare uno dei due elementi evitando ambiguità, ma il problema può presentarsi nel caso di numerosi schemi creati in grandi organizzazioni. L'uso dei *namespaces* consente di tenere separati i nomi e ne consente l'utilizzo senza errori da parte dell'applicazione.

Facendo precedere il nome dell'elemento da un prefisso seguito da due punti si creano nomi diversi di elementi, a seguito della concatenazione “*prefisso:nome*”.

Usando un prefisso si ottiene

```
<a:persona>
  <a:nome>Ferdinando Rodriquez</a:nome>
</a:persona>

<b:ufficio>
  <b:nome>Informatica</b:nome>
  <b:ubicazione>Primo piano stanza 9999</b:ubicazione>
</b:ufficio>
```

ottenendo quindi due diversi tipi dell'elemento <nome>.

Esempio

```
<?xml version="1.0"?>
<!DOCTYPE Recipes SYSTEM "recipes.dtd">

<!xml:namespace ns=http://URL/namespaces/breads
  prefix="bread">

<!xml:namespace ns=http://URL/namespaces/meats
  prefix="lamb">

<recipes>
<category>
  <bread:name>Basic Loaf</bread:name>
</category>

<category>
  <lamb:name>Roast Lamb</lamb:name>
</category>
</recipes>
```

Invece di fare uso solo dei prefissi nell'esempio precedente è stato aggiunto un *attributo xmlns* per attribuire all'elemento un nome qualificato associato ad un namespace. Quando un namespace è definito nel tag iniziale di un elemento, tutti gli elementi figli con lo stesso prefisso sono associati al medesimo namespace.

Nell'esempio ogni elemento “categories” del nodo radice “recipes” ha un elemento denominato “name”. La dichiarazione dei namespaces consente alle applicazioni di trattare lo schema senza ambiguità.

Visualizzazione di documenti XML

Tramite fogli di stile⁹

Il metodo più semplice di visualizzare documenti XML consiste nell'usare i fogli di stile, la cui sintassi è già stata esaminata in HTML. E' anche il metodo più limitato in quanto, ad es., non consente l'accesso agli attributi né l'elaborazione delle informazioni contenute nel documento.

Contrariamente ad HTML, in XML ciascun utente può liberamente definire i propri elementi ed il browser non può conoscere a priori come visualizzarli in maniera corretta.

Il foglio di stile CSS contiene le istruzioni per la formattazione degli elementi del documento XML e, collegando il foglio di stile al documento XML, il browser sarà in grado di comprendere come visualizzare correttamente gli elementi.

La separazione fra le istruzioni tipografiche e i dati consente l'utilizzo dello stesso testo marcato con XML per differenti scopi, come la visualizzazione su web o la stampa, mediante l'associazione delle opportune caratteristiche tipografiche contenute nei fogli di stile.

Esempio

```
LIBRO
{display:block;
 margin-top:12pt;
 font-size:10pt}
```

Di norma, tranne alcune eccezioni, l'impostazione di una proprietà per un dato elemento è propagata a tutti gli elementi secondari. Se, quindi, l'elemento libro ha elementi secondari (capitoli, pagine, ecc.) questi ereditano le proprietà specificate per l'elemento superiore.

Di conseguenza è opportuno iniziare la definizione del foglio di stile partendo dagli elementi superiori ed apportare impostazioni aggiuntive passando via via agli elementi nidificati.

Se si collegano più fogli di stile le regole (in IE) vengono aggregate tra loro, e nel caso di regole conflittuali prevalgono quelle incluse nei fogli successivi su quelle dei fogli collegati prima.

Quando il foglio di stile non contiene alcuna proprietà per un determinato elemento, il browser usa un valore predefinito incorporato nel browser stesso o un valore che l'utente ha impostato.

Si tenga comunque presente che gli utenti con problemi di vista potrebbero voler visualizzare i caratteri con dimensioni maggiori. In questo caso tali utenti potrebbero decidere di mantenere il controllo sulla formattazione delle pagine, configurando il browser in modo da attribuire priorità alle impostazioni delle proprietà del browser rispetto a quelle presenti nel foglio di stile.¹⁰

⁹ Il documento XML con il foglio di stile allegato può essere aperto direttamente tramite browser che supporta i fogli di stile CSS.

¹⁰ In MS Internet Explorer si sceglie il pulsante Accesso facilitato della scheda Generale della finestra di dialogo Opzioni Internet

Visualizzazione di documenti con il binding dei dati (Internet Explorer v. 5.0 e successive)

Con questa tecnica si visualizza un documento XML attraverso una pagina HTML sfruttando le migliori caratteristiche dei due ambienti. Il binding dei dati lavora con i documenti che sono strutturati come database, ovvero nel quale gli elementi possono essere interpretati come un insieme di record e campi (recordset simmetrici).¹¹

L'oggetto di programmazione a base del binding dei dati è il DSO (Data Source Object).

L'utilizzo del binding comporta due passaggi: il collegamento XML alla pagina HTML nella quale visualizzare i dati XML e il binding vero e proprio di elementi HTML a elementi XML.

Il collegamento del documento alla pagina può avvenire includendo nella pagina un elemento HTML denominato XML, conosciuto come *isola di dati*.

Nell'esempio seguente l'isola di dati è vuota e include solo l'URL del documento XML.

```
<HTML>
<HEAD>
  <TITLE>Titolo del libro</TITLE>
<HEAD/>

<BODY>

  <XML ID="dsoBook" SRC="Book.xml"></XML> 12

  <!-- - altri elementi HTML - -->

</BODY>
</HTML>
```

Si noti che si è utilizzato un URL parziale, assumendo che Book.xml si trova nella stessa cartella della pagina HTML.

Segue il testo del documento XML:

```
<?xml version="1.0"?>

<!-- Nome file: Book.xml -->

<BOOK>
  <TITLE>Pinocchio</TITLE>
  <AUTHOR>Collodi</AUTHOR>
  <BINDING>tascabile</BINDING>
  <PAGES>150</PAGES>
  <PRICE>EURO 10,00</PRICE>
</BOOK>
```

¹¹ Per i documenti la cui struttura non è assimilabile a quella di un database è possibile utilizzare le tecniche di scripting DOM (Document Object Model), valide per qualsiasi tipo di documento XML.

¹² Si tenga presente che l'elemento denominato XML usato per la creazione dell'isola di dati non è un elemento XML. E' un elemento HTML che contiene elementi XML.

Aperto la pagina HTML con il browser il processore XML crea l'oggetto di programmazione DSO (Data Source Object) che memorizza nella memoria cache i dati XML sotto forma di recordset (insieme di record con i relativi campi). Il DSO consente l'accesso al recordset memorizzato e la sua modifica mediante un insieme di metodi, proprietà ed eventi.

Ad es. tramite un metodo (funzione) è possibile spostarsi da un record ad un altro; tramite le proprietà (impostazioni correnti) si può conoscere se è stato raggiunto l'ultimo record; tramite gli eventi (ad es. un click su un bottone) è possibile la gestione tramite funzioni incluse nello script annesso alla pagina.

Il secondo passaggio consiste nell'effettuare il binding di un elemento HTML ai dati XML. Di seguito è descritto il binding con tabelle.

```
<?xml version="1.0"?>

<!-- Nome file: Biblioteca.xml -->

<BIBLIOTECA>
  <BOOK>
    <TITLE>Pinocchio</TITLE>
    <AUTHOR>Collodi</AUTHOR>
    <BINDING>tascabile</BINDING>
    <PAGES>150</PAGES>
    <PRICE>EURO 10,00</PRICE>
  </BOOK>
  <BOOK>
    <TITLE>Moby Dick</TITLE>
    <AUTHOR>Herman Melville</AUTHOR>
    <BINDING>copertina rigida</BINDING>
    <PAGES>724</PAGES>
    <PRICE>EURO 9,95</PRICE>
  </BOOK>
  <BOOK>
    <TITLE>Il gatto con gli stivali</TITLE>
    <AUTHOR>Charles Perrault</AUTHOR>
    <BINDING>tascabile economico</BINDING>
    <PAGES>120</PAGES>
    <PRICE>EURO 4,95</PRICE>
  </BOOK>
</BIBLIOTECA>
```

```

<!-- Nome file: BibliotecaTable.htm -->

<HTML>

<HEAD>
  <TITLE>Inventario dei libri</TITLE>
</HEAD>

<BODY>

  <XML ID="dsoInventory" SRC="biblioteca.xml"></XML>

  <H2>Inventario dei libri</H2>

  <TABLE DATASRC="#dsoInventory" BORDER="1" CELLPADDING="5">
    <THEAD>
      <TH>Titolo</TH>
      <TH>Autore</TH>
      <TH>Edizione</TH>
      <TH>Pagine</TH>
      <TH>Prezzo</TH>
    </THEAD>
    <TR ALIGN="center">
      <TD><SPAN DATAFLD="TITLE"
        STYLE="font-style:italic"></SPAN></TD>
      <TD><SPAN DATAFLD="AUTHOR"></SPAN></TD>
      <TD><SPAN DATAFLD="BINDING"></SPAN></TD>
      <TD><SPAN DATAFLD="PAGES"></SPAN></TD>
      <TD><SPAN DATAFLD="PRICE"></SPAN></TD>
    </TR>
  </TABLE>

</BODY>

</HTML>

```

collegamento al documento XML (biblioteca.xml) con l'isola di dati che ha ID dsoInventory

elemento table in binding con l'intero documento HTML tramite l'attributo datasrc dell'elemento all'id dell'isola di dati, preceduto dal carattere #.

Aprendo il file BibliotecaTable.htm in Internet Explorer vers. 5.5 o successive si ottiene a video

Microsoft Internet Explorer window titled "Inventario dei libri". The address bar shows the file path: C:\Rodriguez\CorsoMarcatura\XML\BibliotecaTable.htm. The page content displays a table titled "Inventario dei libri" with the following data:

Titolo	Autore	Edizione	Pagine	Prezzo
<i>Pinocchio</i>	Colodi	tascabile	150	EURO 10,00
<i>Moby Dick</i>	Herman Melville	copertina rigida	724	EURO 9,95
<i>Il gatto con gli stivali</i>	Charles Perrault	tascabile economico	120	EURO 4,95

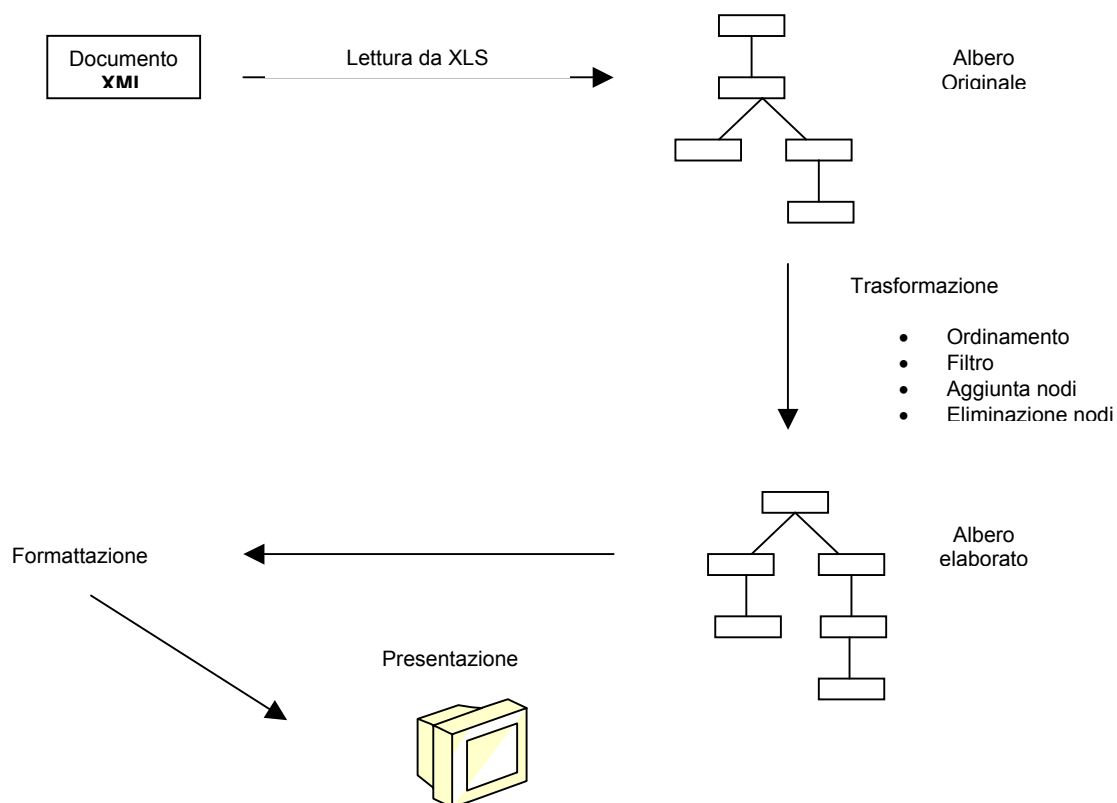
The taskbar at the bottom shows the Start button, several application icons, and the system clock displaying 11.27.

Visualizzazione tramite script DOM (Document Object Model)

Il modello DOM è costituito da una serie di oggetti di programmazione che rappresentano i diversi componenti di un documento XML. Questo modello memorizza i dati XML in una struttura gerarchica ad albero che rappresenta la struttura gerarchica del documento XML e consente di accedere a qualsiasi componente del documento.

Visualizzazione tramite fogli di stile XSL

Mentre un foglio di stile CSS fornisce istruzioni per la formattazione di ciascun elemento XML, un foglio di stile XSL è molto più efficace e flessibile, fornendo la possibilità di selezionare i dati XML, di visualizzarli nell'ordine desiderato, di accedere a tutti i componenti XML (elementi, attributi, commenti, istruzioni di elaborazione), di manipolare le informazioni attraverso metodi.



La visualizzazione tramite i fogli di stile XSL è fatta trasformando il documento XML in pagina HTML. Tale trasformazione è eseguibile sia sul client sia sul server. Nel primo caso il server invia al browser il documento XML unitamente al foglio di stile, nel secondo il server stesso esegue la trasformazione e al browser è trasferito il documento XML già convertito in HTML. La seconda ipotesi è particolarmente valida dal momento che la tecnologia XSL, relativamente recente, non è supportata in maniera diffusa dai browser correntemente utilizzati.

Anziché regole, come si verifica per i fogli di stile CSS, XSL include uno o più modelli, ciascuno dei quali contiene informazioni per la visualizzazione di un dato ramo di elementi del documento XML.

Per collegare il foglio di stile XSL al documento XML si aggiunge al documento XML un'istruzione `xml-stylesheet` secondo la seguente sintassi.

```
<?xml-stylesheet type="text/xsl" href=PercorsoFileXSL?>
```

La stringa `PercorsoFileXSL` è l' URL che indica la posizione del foglio di stile, e può essere un URL completo o parziale (che indica la posizione relativa a quella del documento XML).¹³

Esempio

```
<?xml version="1.0"?>

<!-- Nome file: XslDemo01.xsl -->

<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">
  <xsl:template match="/">
    <H3>Descrizione del libro</H3>
    <SPAN STYLE="font-style:italic">Autore: </SPAN>
    <xsl:value-of select="BOOK/AUTHOR"/><BR/>
    <SPAN STYLE="font-style:italic">Titolo: </SPAN>
    <xsl:value-of select="BOOK/TITLE"/><BR/>
    <SPAN STYLE="font-style:italic">Prezzo: </SPAN>
    <xsl:value-of select="BOOK/PRICE"/><BR/>
    <SPAN STYLE="font-style:italic">Tipo di edizione: </SPAN>
    <xsl:value-of select="BOOK/BINDING"/><BR/>
    <SPAN STYLE="font-style:italic">Numero di pagine: </SPAN>
    <xsl:value-of select="BOOK/PAGES"/>

  </xsl:template>
</xsl:stylesheet>
```

```
<?xml version="1.0"?>

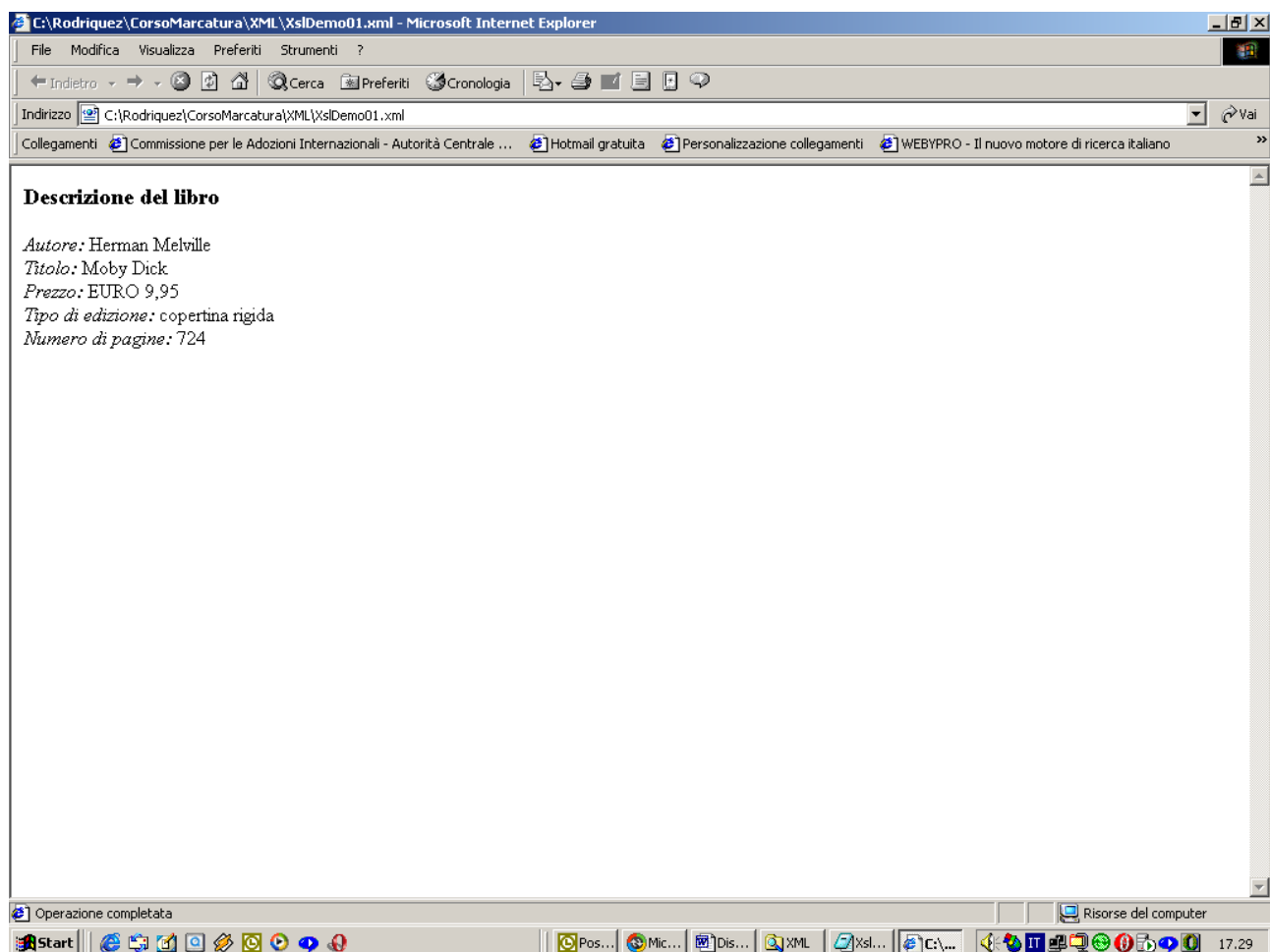
<!-- Nome file: XslDemo01.xml -->

<?xml-stylesheet type="text/xsl" href="XslDemo01.xsl"?>

<BOOK>
  <TITLE>Moby Dick</TITLE>
  <AUTHOR>
```

¹³ Se al documento XML si collegano più fogli di stile XSL il browser utilizzerà solo il primo (contrariamente ai fogli di stile CSS). Se al documento XML si collegano sia un foglio di stile CSS sia un foglio di stile XSL il browser utilizzerà solo il foglio di stile XSL.

```
<FIRSTNAME>Herman</FIRSTNAME>  
<LASTNAME>Melville</LASTNAME>  
</AUTHOR>  
<BINDING>copertina rigida</BINDING>  
<PAGES>724</PAGES>  
<PRICE>EURO 9,95</PRICE>  
</BOOK>
```



Si noti che l'esempio è strutturato in modo da visualizzare l'unico "record" presente nel documento XML..
Per accedere ad eventuali altri record presenti nel file occorrono altre istruzioni.

Di seguito è illustrato il modo in cui il browser genera la prima parte del blocco di tag HTML con il documento XML ed il foglio di stile XSL esemplificati.

Documento XML

```
<AUTHOR>
  <FIRSTNAME>Herman</FIRSTNAME>
  <LASTNAME>Melville</LASTNAME>
</AUTHOR>
```

Foglio stile XSL

```
<H3>Descrizione del libro</H3>
<SPAN STYLE="font-style:italic">Autore: </SPAN>
<xsl:value-of select="BOOK/AUTHOR"/><BR/>
```

HTML

```
<H3>Descrizione del libro</H3>
<SPAN STYLE="font-style:italic">Autore: </SPAN>
Herman Melville
```

