

[article](#)[discussion](#)[edit this page](#)[history](#)

Editing Wikitext

From Wikipedia, the free encyclopedia.

Preview

Wikitext or *wiki markup* is a [markup language](#) that offers a simplified alternative to [HTML](#), and is used to write pages in [wiki](#) websites.

There is no commonly accepted [standard](#) [wikitext language](#).

The grammar, structure, features, keywords and so on are dependent on the particular [wiki software](#) used on the particular website. For example, all wikitext markup languages have a simple way of [hyperlinking](#) to other pages within the site, but there are several different syntax conventions for these links.



"Wikitext" or "wiki markup" is a [\[\[markup language\]\]](#) that offers a simplified alternative to [\[\[HTML\]\]](#) and is used to write pages in [\[\[wiki\]\]](#) websites.

There is no commonly accepted [\http://meta.wikimedia.org/wiki/Wikitext_standard standard [\[\[wikitext\]\]](#) language.

The grammar, structure, features, keywords and so on are dependent on the particular [\[\[wiki software\]\]](#) used on the particular website. For example, all wikitext markup languages have a simple way of [\[\[hyperlink\]\]ing](#) to other pages within the site, but there are several different syntax conventions for these links.

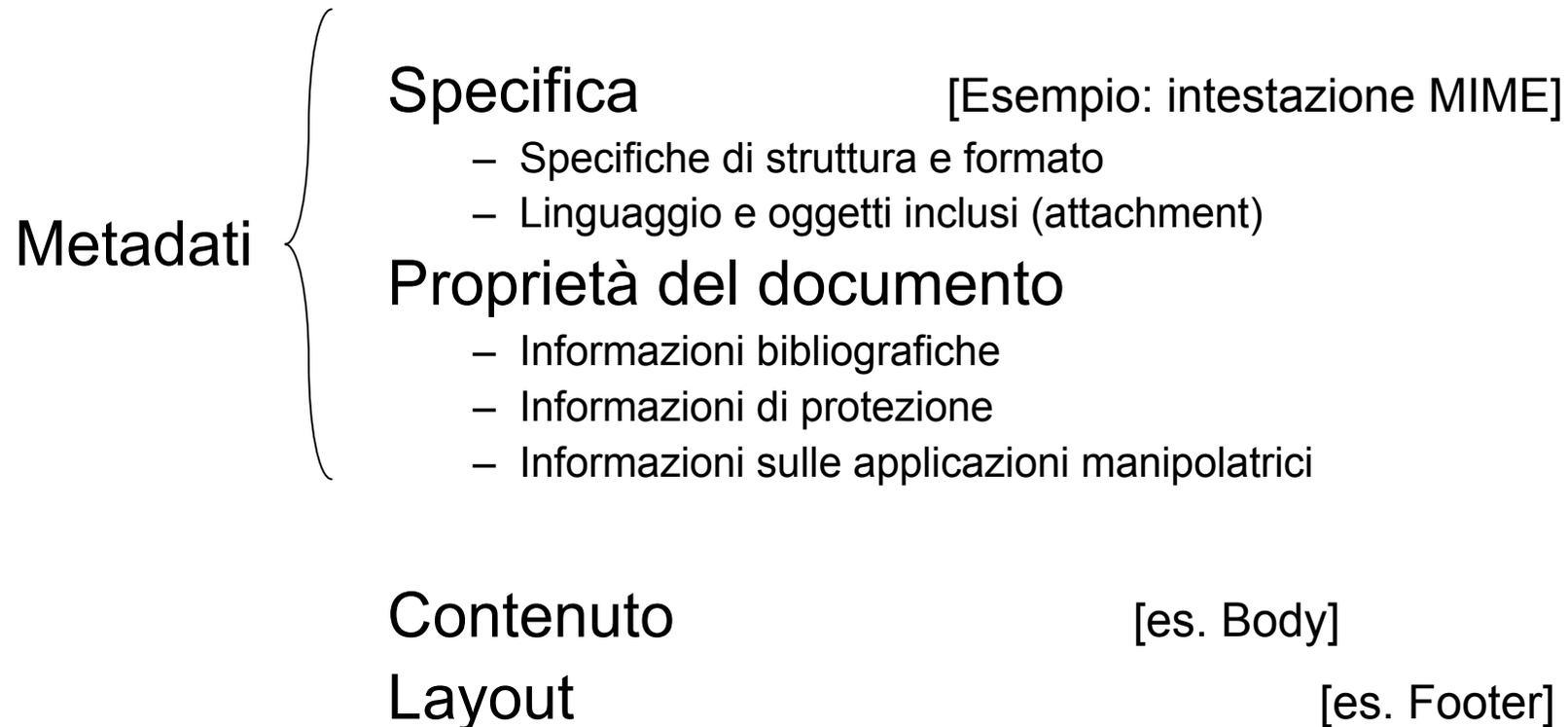
I linguaggi di markup e le loro applicazioni ai documenti digitali

Obiettivo della lezione

- Cos'è un linguaggio di markup?
- XML
- A che serve XML?
- I linguaggi di markup basati su XML
- XML nel mondo editoriale

La descrizione dei documenti digitali

I documenti digitali hanno le seguenti componenti:



Esempio: Lemma di dizionario

Markup: XML

Il file XML specificato apparentemente non ha un foglio di stile associato. L'albero del documento è mostrato di seguito.

```
- <DOC>
  <Tipo>N</Tipo>
  <Cod>0000000013</Cod>
  - <Intestazione>
    <Titolo>abbandonato</Titolo>
    <Numero></Numero>
    - <Silla>
      ab<span class="fontSpeciale">.</span>ban<span
      class="fontSpeciale">.</span>do<span
      class="fontSpeciale">.</span>nà<span class="fontSpeciale">.</span>to
    </Silla>
    <Var></Var>
    <Pron></Pron>
    <Lingua></Lingua>
    <Gramlingua></Gramlingua>
    <Gramm>agg.</Gramm>
    <Partic></Partic>
  </Intestazione>
  - <Trattazione>
    - <definizione tipo="num">
      <testo>Participio pass. di <i>abbandonare</i>.</testo>
    </definizione>
    - <definizione tipo="num">
      - <testo>
        Trascurato, privo di cura o di assistenza: <i>un campo abbandonato</i>;
        <i>il problema dei bambini abbandonati</i>.
      </testo>
    </definizione>
    - <definizione tipo="num">
      <testo>Deserto: <i>un paese abbandonato</i>.</testo>
    </definizione>
  </Trattazione>
  <Etimologia></Etimologia>
</DOC>
```

Struttura e formato

- La **struttura** è una proprietà intrinseca del documento, legata al suo tipo
- **Esempi:**
 - La struttura di un **libro** è una **sequenza di capitoli**
 - La struttura di un **articolo di giornale** è:
 - **titoletto, titolo, sottotitolo, corpo**
- Il **formato** è un attributo secondario del documento, dipendente da come viene creato o visualizzato o salvato o stampato
- **Esempi:**
 - il formato di fruizione di un documento è **PDF**
 - il formato della pagina fisica in stampa è **A4**

Struttura e layout

- La struttura è una proprietà **univoca** del tipo del documento, il formato dipende da aspetti esterni, come il supporto di fruizione (es. dimensione video)
- Uno degli aspetti del formato è il **layout**, che è un *comportamento* del documento, dipendente dallo **stile di presentazione**

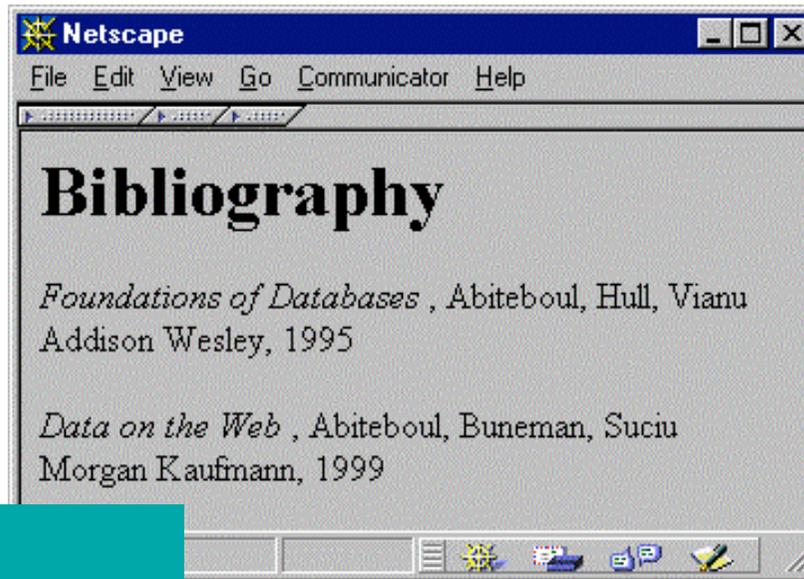
Esempio: un libro viene stampato in un layout su carta A4 due pagine alla volta

- Lo stesso documento può essere visualizzato secondo diverse modalità di layout
- I comportamenti di layout sono molto vari e possono dipendere dall'applicazione visualizzatrice; per un esempio di layout "liquido" di HTML vedere

www.maxdesign.com.au/presentation/liquid/example13.htm

Descrivere struttura e layout

- HTML e LaTeX sono linguaggi che descrivono sia struttura che layout di un documento
- I due concetti (struttura e layout) sono però distinti; XML è un linguaggio per descrivere contenuto



```
<h1> Bibliography </h1>
```

```
<p> <i> Foundations of Databases </i>
```

```
Abiteboul, Hull, Vianu
```

```
<br> Addison Wesley, 1995
```

```
<p> <i> Data on the Web </i>
```

```
Abiteoul, Buneman, Suci
```

```
<br> Morgan Kaufmann, 1999
```

HTML describe il layout

```
<bibliography>
```

```
<book> <title> Foundations... </title>
```

```
<author> Abiteboul </author>
```

```
<author> Hull </author>
```

```
<author> Vianu </author>
```

```
<publisher> Addison Wesley </publisher>
```

```
<year> 1995 </year>
```

```
</book>
```

```
...
```

```
</bibliography>
```

XML describe il contenuto

Formati dei documenti digitali

- Testuale
 - in chiaro
 - con markup di formattazione
 - con markup semantico
- Codificato
 - Binario, proprietario
 - Binario, cifrato
- Compresso

Il formato testuale

La codifica del testo in chiaro (file .txt o .asc) è soggetta a diversi aspetti tecnici:

- ASCII vs. EBCDIC vs. Unicode
- PC vs. UNIX vs. Mac
- Standard di traslitterazione
- Memorizzazione su dispositivi

Formati proprietari

I codici di formato dipendono dal software che crea il documento; di solito sono in binario e inaccessibili all'utente

- RTF (Rich-Text Format); testuale
- I formati di Microsoft Word ed Excel, WordPerfect, WordStar, ecc.
- Postscript: formato per stampanti
- PDF (Portable Document Format)

Cos'è un *Markup Language*

- Un linguaggio capace di descrivere singoli aspetti di un documento
- Specifica informazioni di struttura e/o layout
- Collega informazioni entro il documento
- Collega insieme diversi documenti
- Può incorporare componenti multimediali
- Può esprimere informazioni utili per l'indicizzazione del documento

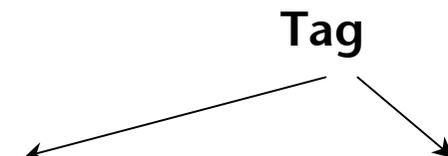
Markup

- Manoscritto (a macchina)

Questa è una nozione importante.

- HTML

Questa è una nozione `importante`.



```
graph TD; Tag --> em["<em>"]; Tag --> em_slash["</em>"];
```

- VoiceXML

```
<block>
  Questa è una nozione
  <emp>importante</emp>.
</block>
```

Markup stilistico (o presentazionale)

- TeX e LaTeX
- HTML (HyperText Markup Language)
- Dynamic HTML
 - CSS (Cascading Style Sheets)
- XSL (per XML)
- Soluzioni proprietarie per E-Publishing
- Wiki

Wiki: la minimizzazione del markup presentazionale

The screenshot shows the Wikipedia editing interface for the article "Editing Wikipedia:Introduction". At the top right, there are links for "project page", "discussion", "edit this page", and "history". A "Create account / log in" link is also present. The article title is "Editing Wikipedia:Introduction" with the subtitle "From Wikipedia, the free encyclopedia." Below the title is a "Preview" section with a red note: "Note: Remember that this is only a preview, and has not yet been saved!". A navigation bar shows three tabs: "1. Introduction" (highlighted), "2. Learn more about editing", and "3. Explore Wikipedia". The main content area starts with a link to "Wikipedia:Manual of Style#Introduction" and a paragraph: "See the 'edit this page' tab at the top of each article? On Wikipedia, you can edit pages whenever you want, logged in or not." Below this is the section "What is Wikipedia?" with a paragraph: "Wikipedia is an [encyclopedia](#) written [collaboratively](#) by many of its readers. Lots of people are constantly improving Wikipedia, making thousands of changes an hour, all of which are recorded on the [page history](#) and the [Recent Changes](#) page. [Nonsense](#) and [vandalism](#) are usually removed quickly." The next section is "How can I help?" with a paragraph: "Don't be afraid to edit pages on Wikipedia—*anyone* can edit, and we encourage users to **be bold!** Find something that can be improved, either in content, grammar or formatting, then fix it. Worried about breaking Wikipedia? Don't be: it can always be fixed or improved later. So go ahead, edit an article and help make Wikipedia the best source for information on the Internet!". On the left side, there is a sidebar with the Wikipedia logo, a search box, a toolbox with links like "What links here", "Related changes", and "Special pages", and a section for "in other languages" with links for Arabic, Bosanski, Català, Česky, and Cymraeg. A smaller inset image in the bottom right shows a close-up of the "edit this page" button, which is circled in red.

project page discussion **edit this page** history

Create account / log in

Editing Wikipedia:Introduction

From Wikipedia, the free encyclopedia.

Preview

Note: Remember that this is only a preview, and has not yet been saved!

- 1. Introduction**
2. [Learn more about editing](#)
3. [Explore Wikipedia](#)

For article introduction guidelines, see: [Wikipedia:Manual of Style#Introduction](#)

See the "edit this page" tab at the top of each article? On Wikipedia, you can edit pages whenever you want, logged in or not.

What is Wikipedia?

Wikipedia is an [encyclopedia](#) written [collaboratively](#) by many of its readers. Lots of people are constantly improving Wikipedia, making thousands of changes an hour, all of which are recorded on the [page history](#) and the [Recent Changes](#) page. [Nonsense](#) and [vandalism](#) are usually removed quickly.

How can I help?

Don't be afraid to edit pages on Wikipedia—*anyone* can edit, and we encourage users to **be bold!** Find something that can be improved, either in content, grammar or formatting, then fix it. Worried about breaking Wikipedia? Don't be: it can always be fixed or improved later. So go ahead, edit an article and help make Wikipedia the best source for information on the Internet!

WIKIPEDIA
The Free Encyclopedia

article discussion **edit this page**

Wikipedia

From Wikipedia, the free encyclopedia.

Wikipedia is a multilingual "copyleft" [encyclopedia](#) to be read and edited by anyone. It is collaboratively maintained by thousands of users via [wiki](#) software, hosted and supported by the [non-profit Wikimedia Foundation](#). In addition to typical encyclopedia entries, Wikipedia also contains information more often associated with glossaries.

Click **edit this page** to change an article

Go Search

What links here
Related changes
Special pages

in other languages

- العربية
- Bosanski
- Català
- Česky
- Cymraeg

Wiki markup

[article](#)[discussion](#)[edit this page](#)[history](#)

Editing Wikitext

From Wikipedia, the free encyclopedia.

Preview

Wikitext or *wiki markup* is a [markup language](#) that offers a simplified alternative to [HTML](#) and is used to write pages in [wiki](#) websites.

There is no commonly accepted [standard wikitext](#) language.

The grammar, structure, features, keywords and so on are dependent on the particular [wiki software](#) used on the particular website. For example, all wikitext markup languages have a simple way of [hyperlinking](#) to other pages within the site, but there are several different syntax conventions for these links.



""Wikitext"" or "wiki markup" is a [\[\[markup language\]\]](#) that offers a simplified alternative to [\[\[HTML\]\]](#) and is used to write pages in [\[\[wiki\]\]](#) websites.

There is no commonly accepted [\http://meta.wikimedia.org/wiki/Wikitext_standard standard wikitext] language.

The grammar, structure, features, keywords and so on are dependent on the particular [\[\[wiki software\]\]](#) used on the particular website. For example, all wikitext markup languages have a simple way of [\[\[hyperlink\]\]](#)ing to other pages within the site, but there are several different syntax conventions for these links.

WikiTeX

TexIdioms Go back to [TexIdioms](#) without saving.

Can be edited by owner and .

You may not be able edit this: try [logging in](#) as *sergey* or *winny* if you can.

Click on "Edit" to see the page source with LaTeX snippets.

Superscripts and subscripts: $\$x_1\$, \$\alpha^2\$, \$\epsilon^{10}\$, \$z^{a+b}\$$. Simple expressions with sub/superscripts that do not require special symbols are converted to HTML.

Notice typical errors: $_y^{10}\$, and _z^{(a+b)\$$. If your subscript or superscript is longer than one character, use $\{\}$ for grouping.

Greek letters (can't live without them!): $_alpha, _beta, _gamma, _delta, _epsilon, _zeta, _eta, _theta, _iota, _kappa, _lambda, _mu, _nu, _xi, _o, _pi, _rho, _sigma, _tau, _upsilon, _phi, _chi, _psi, _omega\$$.

I cant type tabs. Please [ConvertSpacesToTabs](#) for me when I save.

Process as: Wiki HTML (process only math and Wiki pagenames) Script

Header and footer: built-in custom:

Show *Edit* links to all users only those who can edit

Default TeX style and size

Edit box rows: columns: [Enable cookies to set this for every page]

Markup per indicizzazione

- L'uso di tag specifici facilita le ricerche
- I tag specifici vanno definiti da **ontologie**
- L'indicizzazione dipende dal Search Engine

Nota: l'uso di trucchi di posizionamento ha portato ad una certa alterazione dei meccanismi di rilevanza, e gli utenti al momento preferiscono ricerca mediante chiavi libere rispetto all'uso di ontologie

Esempio (da Docbook)

```
<articleinfo>
<!-- Use "HOWTO", "mini HOWTO", "FAQ" in title, if appropriate -->
<title>Sample HOWTO</title>
<author>
  <firstname>Your Firstname</firstname>
  <surname>Your Surname</surname>
  <affiliation>
    <address><email>your email</email></address>
  </affiliation>
</author>
<editor>
  <firstname>Paolo</firstname>
  <surname>Ciancarini</surname>
  <contrib>Linguaggi di markup</contrib>
</editor>
<othercredit role='converter'>
  <firstname>Filippo</firstname>
  <surname>Rossi</surname>
  <contrib>Conversione da HTML a DocBook v3.1.</contrib>
</othercredit>
```

Metadati:

Specifiche e standard

- Che tipo di documento? [Formato]
- In che lingua è il testo? [Linguaggio]
- Che formato hanno i componenti speciali? [Media]
- Ci sono aspetti grafici speciali ? [Layout]
- Come si trova il documento? [Indicizzazione]
- Questo documento va collegato ad altri? [Correlazione]

Markup strutturale

- `<libro>`
 `<capitolo>`
 `<paragrafo>`
 `</paragrafo>`
 `</capitolo>`
 `<capitolo>`
 `</capitolo>`
 `</libro>`

Struttura corretta

- `<libro>`
 `<paragrafo>`
 `<capitolo>`
 `</paragrafo>`
 `</capitolo>`
 `<capitolo>`
 `</capitolo>`
 `</libro>`

Struttura scorretta

Sintassi

Definizione: La *sintassi* di un linguaggio è l'insieme delle regole che definisce l'insieme delle “frasi ben formate” in quel linguaggio

Definizione: Una *grammatica* (libera da contesto) ha la forma:

$G = (\text{Vocabolario}, \text{Categorie}, \text{Struttura-di-Frase}, \text{RegoleSintattiche})$

Esempio:

Vocabolario = {il_cane, il_gatto, morde, mangia, scappa}

Categorie = {<frase>, <soggetto>, <pred_t>, <pred_i>, <compl_o>}

Struttura-di-Frase = <frase>

RegoleSintattiche:

<frase> ::= <soggetto> <pred_t> <compl_o> | <soggetto> <pred_i>

<soggetto> ::= il_cane | il_gatto

<compl_o> ::= il_cane | il_gatto

<pred_t> ::= morde | mangia

<pred_i> ::= scappa

Esempio

Fraasi ben formate secondo la grammatica dell'esempio:

- `il_gatto morde il_cane`
- `il_cane mangia il_cane`
- `il_gatto scappa`

Fraasi NON ben formate secondo la grammatica dell'esempio:

- `morde il_cane il_gatto`
- `il_cane mangia`
- `il_gatto mangia il_topo`
- `il cane morde il gatto`

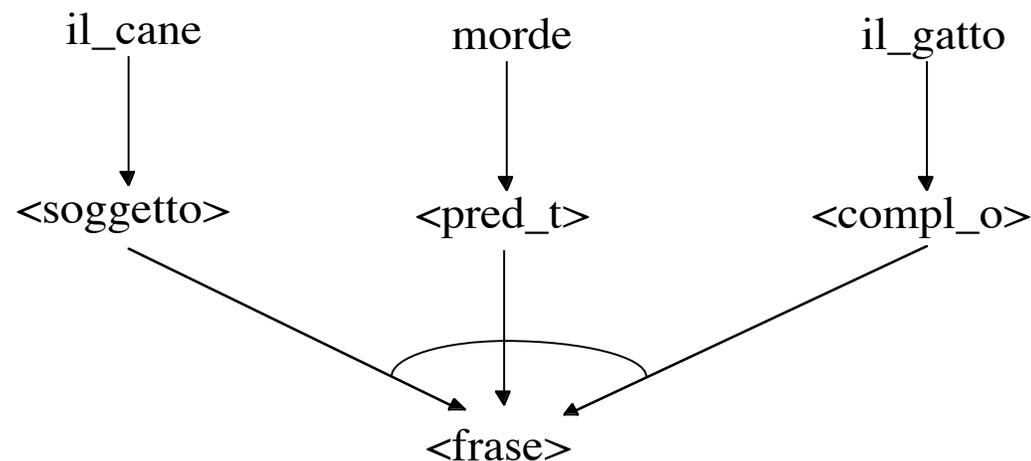
Analisi sintattica

- È possibile costruire programmi (detti “parser”, o “analizzatori sintattici”) che controllano se una frase è sintatticamente corretta rispetto ad una grammatica

Esempio:

La frase `il_cane morde il_gatto` è sintatticamente corretta rispetto alla grammatica del lucido precedente?

- Un modo di analizzare sintatticamente una frase consiste nel costruire un albero sintattico che abbia per radice la Struttura-di-Frase della grammatica



Markup sintattico

- Un altro modo (testuale) di descrivere la struttura di un albero sintattico è mediante mark-up sintattico

<frase>

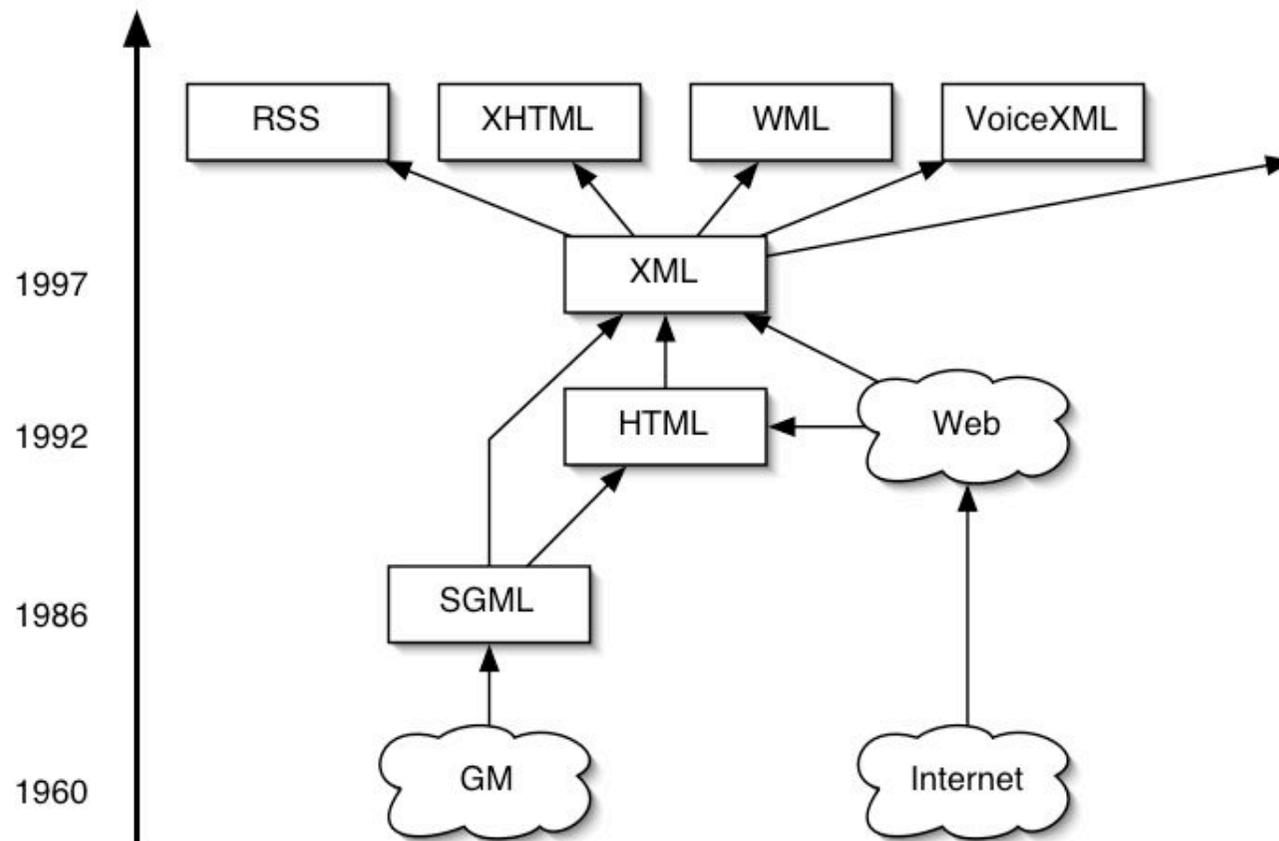
<soggetto> il_cane </soggetto>

<pred_t> morde </pred_t>

<compl_o> il_gatto </compl_o>

</frase>

Storia dei Linguaggi di Markup



Esempi di linguaggi di markup

- RUNOFF, troff, nroff
- SGML Standard Generalized Markup Language
- HTML HyperText Markup Language
- GML Geography Markup Language
- TeX, LaTeX
- LMNL Layered Markup Annotation Language
- RTF Rich Text Format
- WML Wireless Markup Language
- VRML Virtual Reality Markup Language
- XML eXtensible Markup Language

HTML

- Quando nel 1989, al CERN, venne inventato il WWW, si usava un browser non-grafico che gestiva una versione detta HTML 0
- HTML 1 (1993) era quello supportato dal browser Mosaic; la principale innovazione furono le immagini e la gestione grafica della navigazione ipertestuale
- HTML 2 (1994) venne definito da IETF usando SGML; introdusse la gestione delle tabelle e degli script
- HTML 3.0/3.2 (1996) aumentò le funzionalità ipertestuali; nel marzo del 1995 il W3C iniziò a lavorare su una versione 3.0 mentre Microsoft e Netscape rilasciavano i browser contenenti nuovi tags tra loro diversi: il W3C abbandonò la versione 3.0 per passare direttamente alla 3.2, in cui le maggiori novità furono le tabelle, i font colorati, gli applet Java, i tag superscript e subscript
- HTML 4 (1998) è la versione più recente dello standard; introdusse i fogli di stile

Cascading StyleSheets (CSS)

- Uno **stylesheet** è un insieme di regole che istruiscono un browser su come presentare (a video, a stampa) un documento
- Ci sono vari modi di collegare regole di stile a documenti HTML
- Un modo semplice consiste nell'uso dell'elemento **STYLE** di HTML, che viene messo nell'intestazione (**HEAD**) del documento e contiene le regole di stile della pagina
- Ogni regola si compone di due parti: *selector { property: value }*
 - un selettore (*selector*): di solito un elemento HTML come **BODY**, **P**, o **EM**;
 - lo stile da applicare al selettore, definito da coppie *property/value*
- Ogni elemento può avere parecchie proprietà (*properties*)
- Ogni proprietà ha un valore che descrive come va presentato il selettore

Esempio: stylesheet CSS

- Definiamo le proprietà *color* e *font-size* per gli elementi H1 e H2:

```
<HEAD>
```

```
<TITLE>CSS Example</TITLE>
```

```
<STYLE TYPE="text/css">
```

```
  H1 { font-size: x-large; color: red }
```

```
  H2 { font-size: large; color: blue }
```

```
</STYLE>
```

```
</HEAD>
```

- Questo stylesheet dice al browser di mostrare i titoli di livello 1 con fonte rossa *extra-large*, mentre per i titoli di livello 2 deve usare una fonte blu *large*

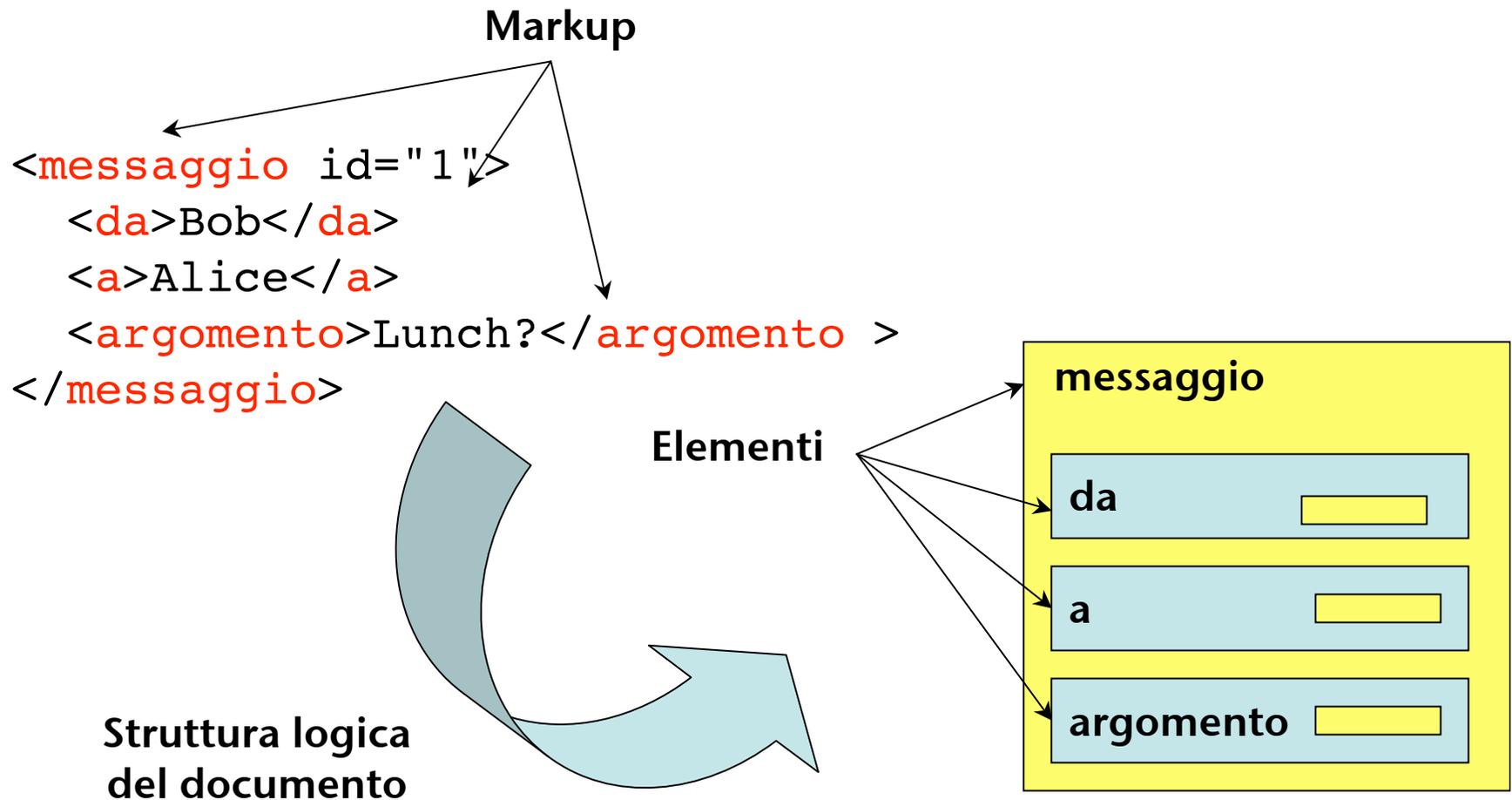
SGML

- SGML venne inventato da IBM e nel 1986 fu standardizzato da ISO
- SGML inizialmente doveva servire a definire documenti strutturati nel campo legale
- E' un metalinguaggio complesso
- Sia HTML che XML derivano da SGML

XML

- XML venne introdotto nel 1996 e standardizzato nel 1998 dal W3C come metalinguaggio di markup ("SGML per il Web")
- E' derivato da SGML, che semplifica
- Si basa sulle stesse tecnologie di base di HTML
- Elimina la debolezza di HTML (i cui tag mescolano struttura e layout), separando la manipolazione della struttura dei documenti dalle problematiche di presentazione del layout

Esempio di frammento XML



SGML vs HTML vs XML

- HTML è un linguaggio con uno specifico vocabolario (insieme di tag predefiniti) per definire struttura e presentazione di pagine ipertestuali su Web
- SGML venne inventato per costruire database di documenti strutturati
- XML è un metalinguaggio libero, cioè senza tag predefiniti, applicabile in molte e diverse situazioni: viene usato soprattutto per definire linguaggi di markup specializzati per qualche applicazione specifica

XML: terminologia

- Un documento XML è fatto di **elementi**, che sono delimitati da tag
 - Tag iniziale: es. `<titolo>`
 - Tag finale: es. `</titolo >`
- Elemento: sequenza di tag ben annidati
- Attributo: proprietà di un elemento
- Documento: frammento XML che combina elementi e contenuto

Esempio: `<titolo> Professore </titolo>`

Terminologia

- Annidamento: elemento entro un altro elemento
- Struttura: annidamento ordinato di elementi

Esempio

```
<Nome>  
  <primo>Paolo</primo>  
  <secondo></secondo>  
  <cognome>Ciancarini</cognome>  
</Nome>
```

- Modello: ordine, frequenza e regole di annidamento degli elementi di un documento definito dal DTD
 - **Esempio:** L'elemento Nome deve contenere primo, seguito da secondo (che può essere vuoto) e da cognome

Terminologia

- **DTD**: Document Type Definition; documento che descrive la grammatica ed il vocabolario di un tipo di documento
- **Parser**: programma che legge un documento e lo convalida se è conforme al suo DTD
- **Documento valido**: documento elaborato da un parser che assicura che struttura e contenuto di un documento siano corretti rispetto al DTD
- **Ben formato**: annidamento degli elementi corretto
 - `<P>Questo markup è ben formato!</P>`
- **Mal formato**: annidamento scorretto
 - `<P>questo markup è mal formato</P>!`
- Un documento valido è sempre ben formato
- Un documento ben formato non sempre è valido

Perché XML è estendibile?

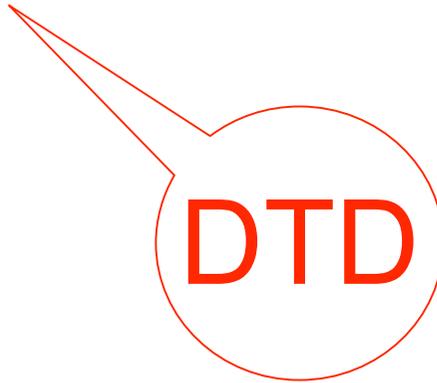
- XML è un **meta-linguaggio**, ovvero è un linguaggio che definisce linguaggi
- Non usa un insieme predefinito di tags, perché permette di definire liberamente i propri tag (HTML ha un insieme prefissato di tag)
- I tag in XML connotano il significato del loro contenuto, quindi il markup XML si può considerare **autodescrivente**
- La struttura "valida" di un documento XML può essere fissata da un DTD (**Document Type Definition**), che può essere annesso al documento

Esempio: documento XML con DTD

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
```

```
<!DOCTYPE BOOK [  
  <!ELEMENT BOOK (TITLE,AUTHOR,ISBN,PRICE,DESCRIPTION)>  
  <ELEMENT TITLE (#PCDATA)>  
  <!ELEMENT AUTHOR (#PCDATA)>  
  <!ELEMENT ISBN (#PCDATA)>  
  <!ELEMENT PRICE (#PCDATA)>  
  <!ELEMENT DESCRIPTION (#PCDATA)>  
>
```

```
<BOOK>  
  <TITLE></TITLE>  
  <AUTHOR></AUTHOR>  
  <ISBN></ISBN>  
  <PRICE></PRICE>  
  <DESCRIPTION></DESCRIPTION>  
</BOOK>
```



DTD

Esempio: DTD

```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE library [
<!ELEMENT library (book*)>
<!ELEMENT book (title, author+, description?,
  (hardcover | softcover))>
<!ELEMENT author (#PCDATA)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT hardcover EMPTY>
<!ELEMENT softcover EMPTY>
<!ATTLIST book isbn CDAT #REQUIRED>
]>
```

Alcune convenzioni nei DTD

- `title` esattamente uno
- `author+` almeno uno
- `description?` Uno o nessuno
- `(hardcover | softcover)` una delle due
- `#PCDATA` caratteri analizzati, possono contenere tag
- `CDATA` caratteri in chiaro, non analizzati per cercare tag
- `EMPTY` solo tag, niente dati
- `#REQUIRED` attributo necessario
- `#IMPLIED` attributo opzionale
- `#FIXED` attributo costante

La struttura di un documento XML

1. Il **prologo**

- Versione di XML `<?xml version="1.0">`
- Codifica Unicode `<encoding=UTF-8>`
- Document Type Declaration (DTD)
 - Esterno o interno

2. L'**elemento documento**, che contiene sottoelementi annidati e riferimenti ad entità esterne (es. abbreviazioni)

Elementi primitivi di XML

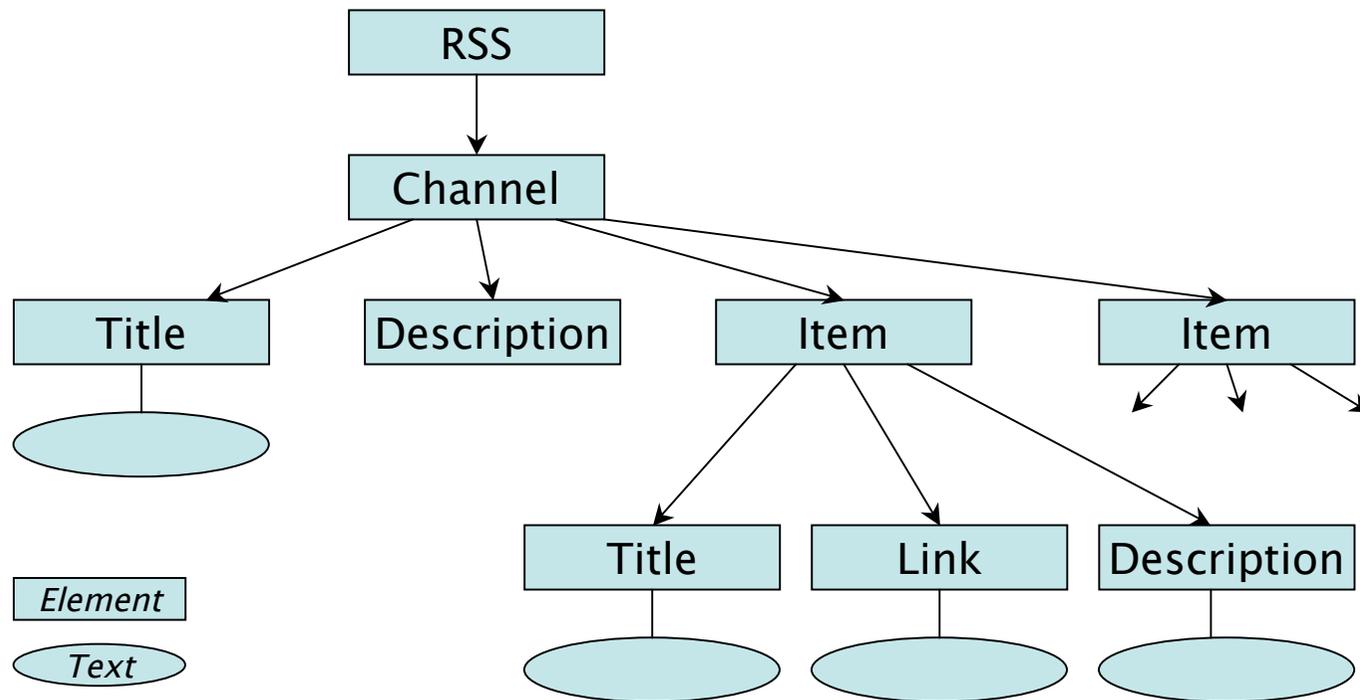
- Un (albero) documento è fatto di nodi
- I tipi di nodo principali sono **elemento**, **attributo** e **testo** (o contenuto)

```
<messaggio id="1">Ciao mondo!</messaggio>
```



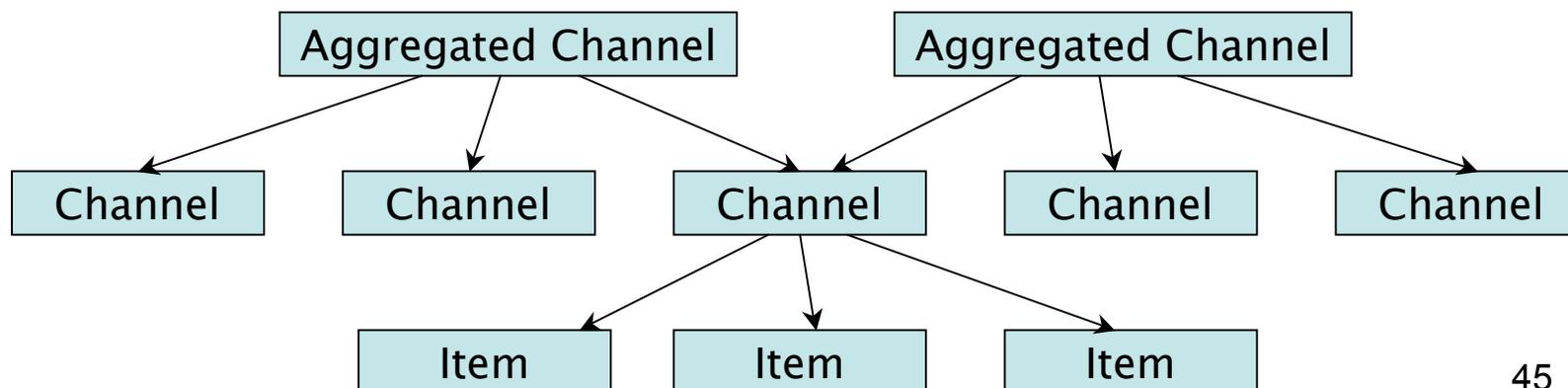
I documenti XML hanno una struttura gerarchica

- La **struttura logica** di un documento XML è un *albero*

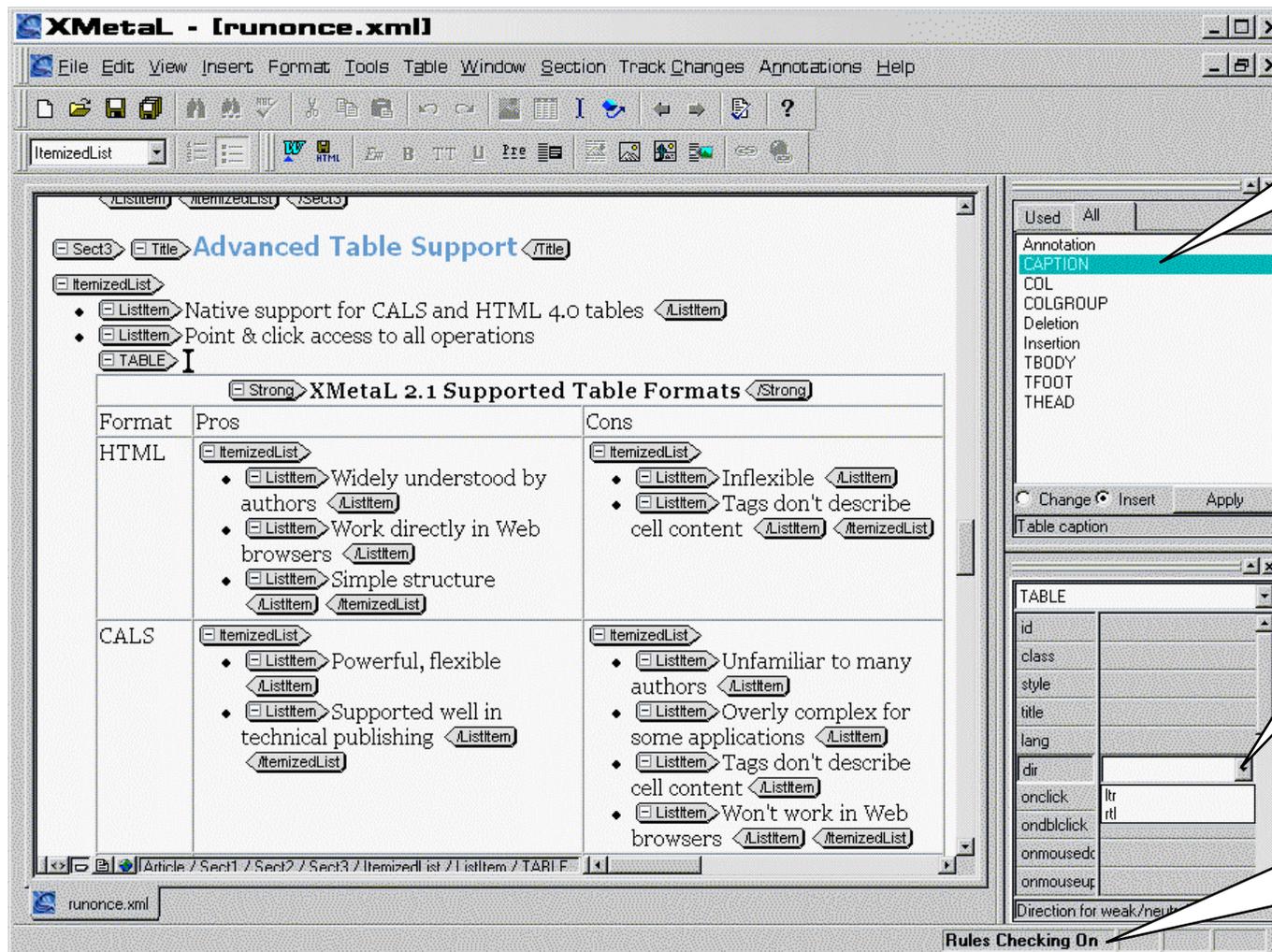


Foreste

- I nodi dell'albero-documento XML possono riferire nodi esterni
- Un insieme di documenti XML con **riferimenti incrociati** è una *foresta*



Un editor XML usa un DTD...



... per dare all'autore la lista completa degli elementi ammessi dal DTD

...e tutti gli attributi degli elementi, inclusi i valori ammessi per gli attributi...

...e forza ad usare il modello durante l'authoring, assicura che gli elementi necessari siano inseriti, e proibisce l'inserimento di elementi fuori posto

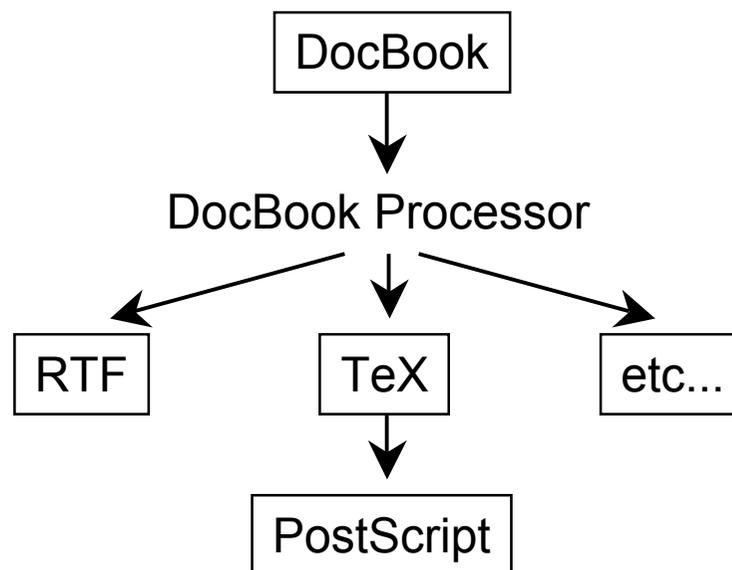
Docbook

- **Docbook** è un DTD per creare libri con modello condiviso internazionalmente
- Creato nel 1991 (in SGML) per la casa editrice O'Reilly
- Si usa prevalentemente per documentazione tecnica, ma si può usare anche in altri campi
- `www.docbook.org`

DocBook

- DocBook venne introdotto nel 1991, quindi prima di XML: usava SGML
- Il DTD originale non era compatibile XML: oggi esiste DTD basato su XML
- OpenJade è un editor di riferimento

```
<book id="simple_book">  
  <title>Very simple book</title>  
  <chapter id="simplechapter">  
    <title>Chapter 1</title>  
    <para>Hello world!</para>  
  </chapter>  
</book>
```



Docbook: esempio

TOOLS FOR XML CONFERENCE PRESENTATIONS

SchemaSoft is happy to provide the following free downloads for authoring, converting and viewing [SVG Open 2005](#) and [XML 2004](#) Conference Papers and Presentations.



[1. DocBook Authoring Tools](#)

Our *Word to DocBook Converter* is a free online service to convert uploaded Microsoft Word files to the [IDEAlliance](#) subset of DocBook, ready for conference submission. In addition, the service applies our *DocBook Styler*, which uses XSLT to convert your DocBook paper to XHTML + CSS for previewing purposes. The *DocBook Styler* can also be downloaded separately and used standalone, in case you prefer to start by creating your DocBook paper with an XML editor.

The DocBook Converter is powered by SchemaSoft's Document Interpreter™ Technology, including [SchemaSoft's Microsoft Office™ Readers and Writers](#).

[2. gcapaper Web Publisher - version 2.1](#)

This package includes:

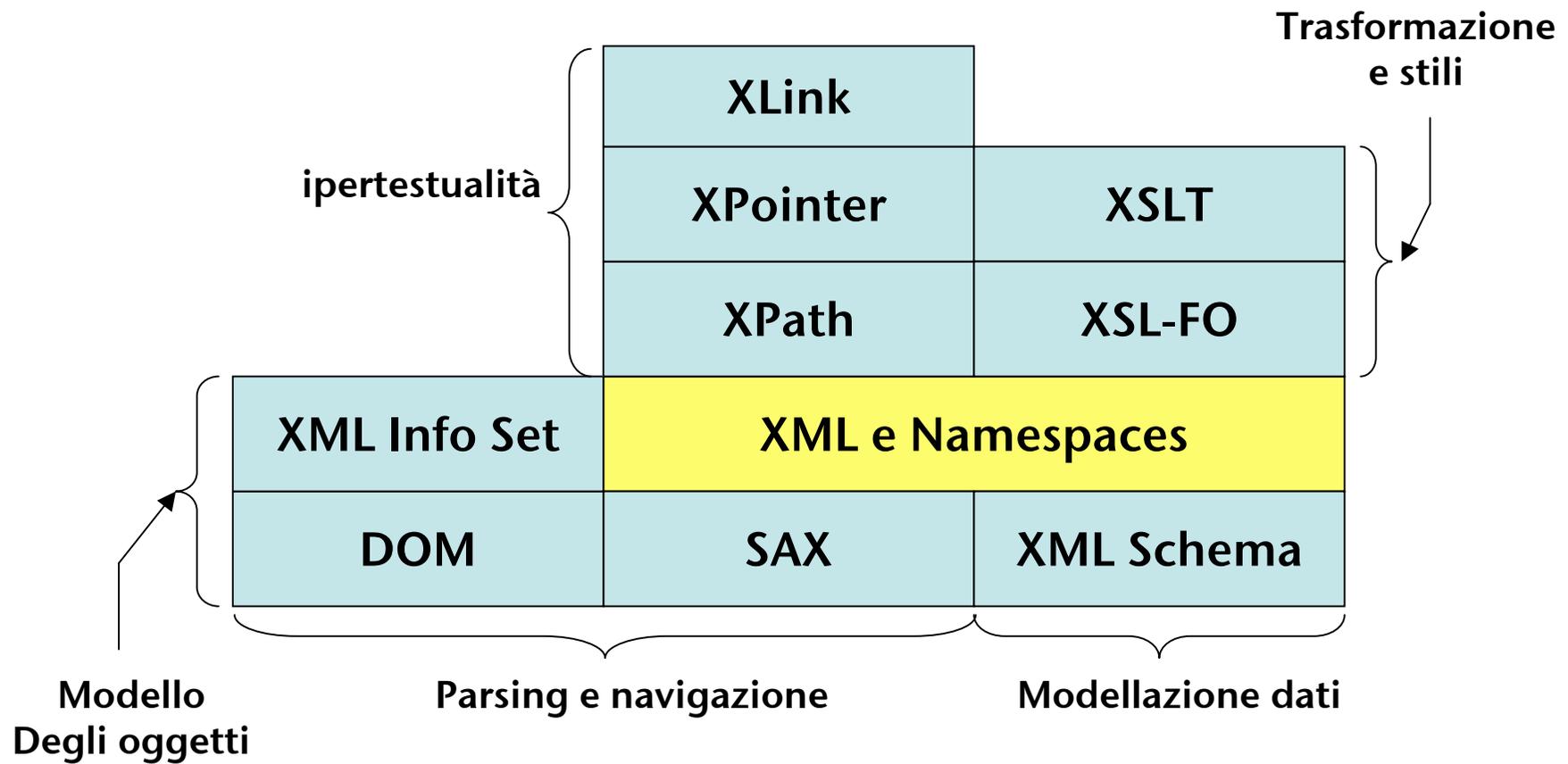
- an XSLT program to convert from the gcapaper XML format to HTML
- a CSS stylesheet to present the output HTML stylishly
- example papers before and after conversion

gcapaper Web Publisher gives comprehensive coverage of all gcapaper features. Version 2 adds support for XHTML1.1 output, SVG images and a number of other enhancements. Here are three SchemaSoft conference papers published with *gcapaper Web Publisher*:

XML è un metalinguaggio

- Un *metalinguaggio* è un linguaggio per descrivere linguaggi
- XML è un formalismo semplice per descrivere la sintassi di qualsiasi linguaggio mediante markup
- Una serie di strumenti di corredo permettono poi di elaborare i linguaggi descritti con XML

La famiglia XML



Esempi: Xlink e Xpointer

XLink linguaggio XML per link ipertestuali sofisticati:

- destinazioni multiple
- collegamenti bidirezionali
- collegamenti con comportamenti speciali:
 - Expand-in-place / Replace / Create new window
 - Link on load / Link on user action
- database di collegamenti
- www.xml.com/pub/a/2000/09/xlink/

• **Xpointer** linguaggio XML per ancore

- referenzia posizioni qualsiasi entro una risorsa XML
- www.w3.org/TR/WD-xptr

Esempio: XHTML

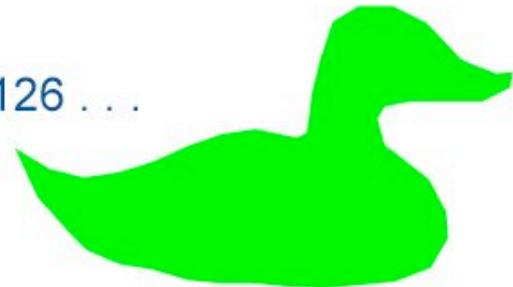
- Linguaggio XML che rappresenta HTML
- Presenta alcune piccole modifiche di HTML:
 - Solo minuscole nei tag `<p>` not `<P>`
 - Attributi tra virgolette ``
 - Tutti gli elementi sono chiusi:
`<p>... </p> `

HTML non evolverà più: il suo successore è
XHTML

Esempio: SVG

- Scalable Vector Graphics
- Linguaggio XML per grafica bidimensionale
- www.w3.org/Graphics/SVG/Overview.htm
- www.w3.org/2004/Talks/1211-Twente-IH/0.svgz

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 20010904//EN"
"http://www.w3.org/TR/2001/REC-SVG-20010904/DTD/svg10.dtd">
<svg width="600" height="300">
  <g transform="translate(10 10)">
    <g stroke="none" fill="lime">
      <path d="M 0.0 112 L 20 124 L 40 129 L 60 126 ...
      ...
    </path>
  </g>
</g>
</svg>
```



Esempio: RSS

- RSS (Rich Site Summary) è un linguaggio di markup per sunti di notiziari

```
<rss version="0.91">
  <channel>
    <title>Acme Product Sale</title>
    <description>Check out our discounts</description>
    <item>
      <title>Sony Playstation Games</title>
      <link>www.acme.com/products?sku=123</link>
      <description>20% off</description>
    </item>
    <item> .. </item>
  </channel>
</rss>
```



Feed Rss - Repubblica.it

Informazioni in tempo reale

Repubblica.it da oggi offre i suoi contenuti con il sistema Rss (Really Simple Syndication), un modo semplice e comodo per essere informati in tempo reale. Grazie ai feed Rss, possibile ricevere sul proprio computer aggiornamenti sulle ultime notizie pubblicate dal sito. Chi ha un blog, inoltre, pu diffondere in maniera semplice e immediata le news di Repubblica.it.



Per accedere ai contenuti Rss con **pochi semplici passi** sufficiente una connessione internet e un apposito programma chiamato "aggregatore". In rete se ne trovano moltissimi (vedi a lato), da scaricare sul proprio pc o da usare attraverso il web. Alcuni aggregatori si integrano perfettamente con i principali browser e con i pi usati programmi di posta elettronica.

Repubblica.it distribuisce tramite Rss il titolo, il sommario e l'indirizzo di tutti gli articoli pubblicati in homepage e nelle varie sezioni. I feed Rss sono dunque il metodo ideale per conoscere le ultime notizie anche mentre si lavora su altri documenti o si visitano altri siti.

Esistono diversi tipi di aggregatori, a seconda del computer, del sistema operativo e in alcuni casi del browser che si utilizza

Gratis

- Feedreader** Windows
- Sharpreader** Windows
- Sage** plug-ins Firefox/ThunderBird
- Urss** plug-ins Mozilla
- Straw** Linux
- Netnewswire Lite** Mac OS X

A pagamento

- FeedDemon** windows
- Netnewswire** Mac OS X
- Newsgator** Outlook

Istruzioni per l'uso

I passi da compiere per scaricare l'aggregatore e personalizzarlo

La guida



16 subscriptions, 212 headlines unread

Subscriptions

Versio...S X (20)

MacCentral (30)

MacMegasite (15)

MacMerc (15)

MacMinute

A List Apart (19)

Aaron ...rtz (15)

BBC News

Daring ...all (10)

Locker...rce (17)

MobileWhack (7)

ongoing (20)

Ranhero (12)

Wired News (4)

Repubb...age (7)

• Repubblica.it > Homepage headlines

- ... [Ciampi: "Il Parlamento, non la tv è la vera base della democrazia"](#)
- ... [Mussolini, lacrime prima del Tar Prodi: "Violazioni gravissime"](#)
- ... • [Conti pubblici, Eurostat non convalida i dati dell'italia](#)
- ... • [Sciopero generale degli statali "In piazza siamo in duecentomila"](#)
- ... • [Parla il marito della Schiavo "La lascerò morire, glielo devo"](#)
- ... • [Omicidio delle cingalesi, è svolta Fermata la sorella della madre](#)
- ... • [Maxi-rissa al Parlamento somalo](#)
- ... • [Malaysia, per Shumi prove amare nelle libere è solo sedicesimo](#)
- ... • [Il grande fotografo di guerra che amava l'avventura, l'etica e il gin](#)

[Ciampi: "Il Parlamento, non la tv è la vera base della democrazia"](#)

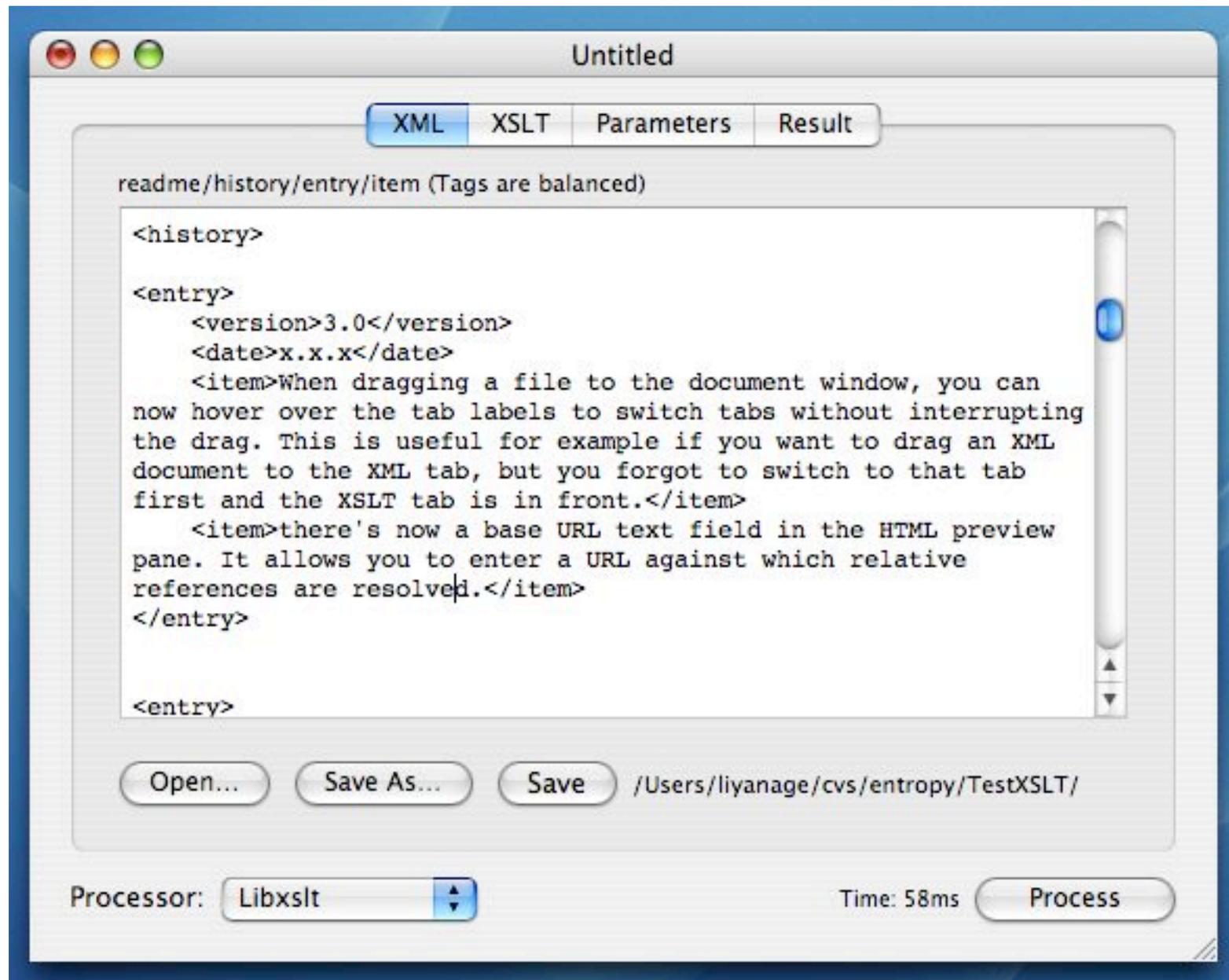
Ultimo giorno di visita a Londra per il Capo dello Stato

"Le decisioni che contano le prendono Camera e Senato"

Ciampi: "Il Parlamento, non la tv è la vera base della democrazia"

(11:22 18/03/2005)

Editor XML



Gli stili di presentazione

- Siccome in XML si definisce solo la struttura di un documento, occorre qualcos'altro per programmare lo stile di presentazione
- XSL (eXtensible Stylesheet Language) è un linguaggio XML per descrivere stili di presentazione - detti **stylesheet**

Stile XSLT

The image shows a window titled "Untitled" with a tabbed interface. The "XSLT" tab is selected, showing the following XML code:

```
xsl:stylesheet/xsl:template/html/head/title (Tags are balanced)
value-of instructions.
-->
<xsl:template match='readme'>

<html>
  <head>
    <title><xsl:value-of select="title"/> v<xsl:value-of
select="$version"/></title>
    <META name="Author" content="{author/text()}" />
  </head>

  <body>
    <xsl:apply-templates/>

    <p>#160;</p>
    <p>#160;</p>
  </body>
```

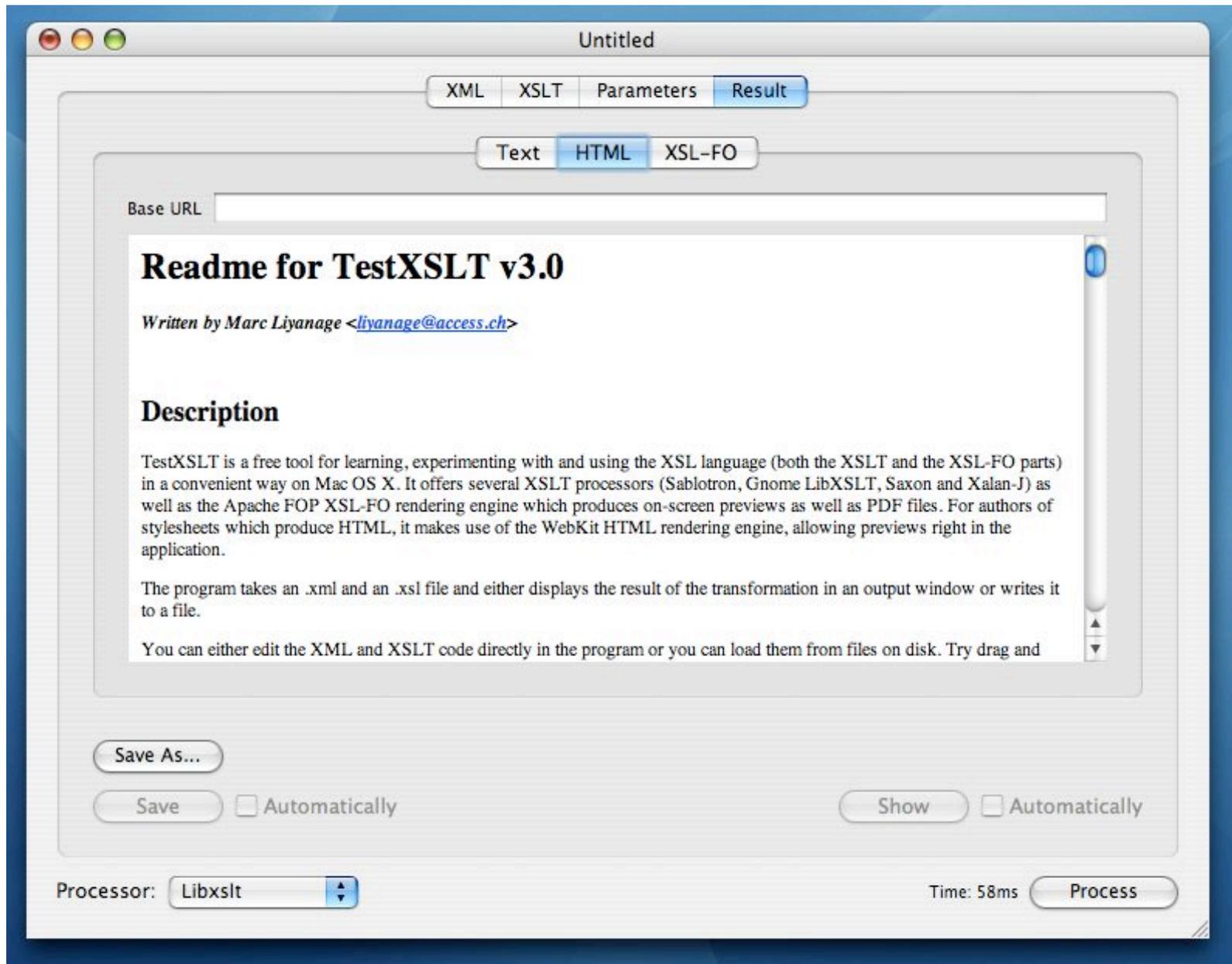
At the bottom of the window, there are several controls:

- Buttons: "Open...", "Save As...", "Save", and "Process".
- File path: "/Users/liyanage/cvs/entropy/TestXSLT/"
- Processor: "Libxslt" (selected in a dropdown menu).
- Time: "Time: 58ms"

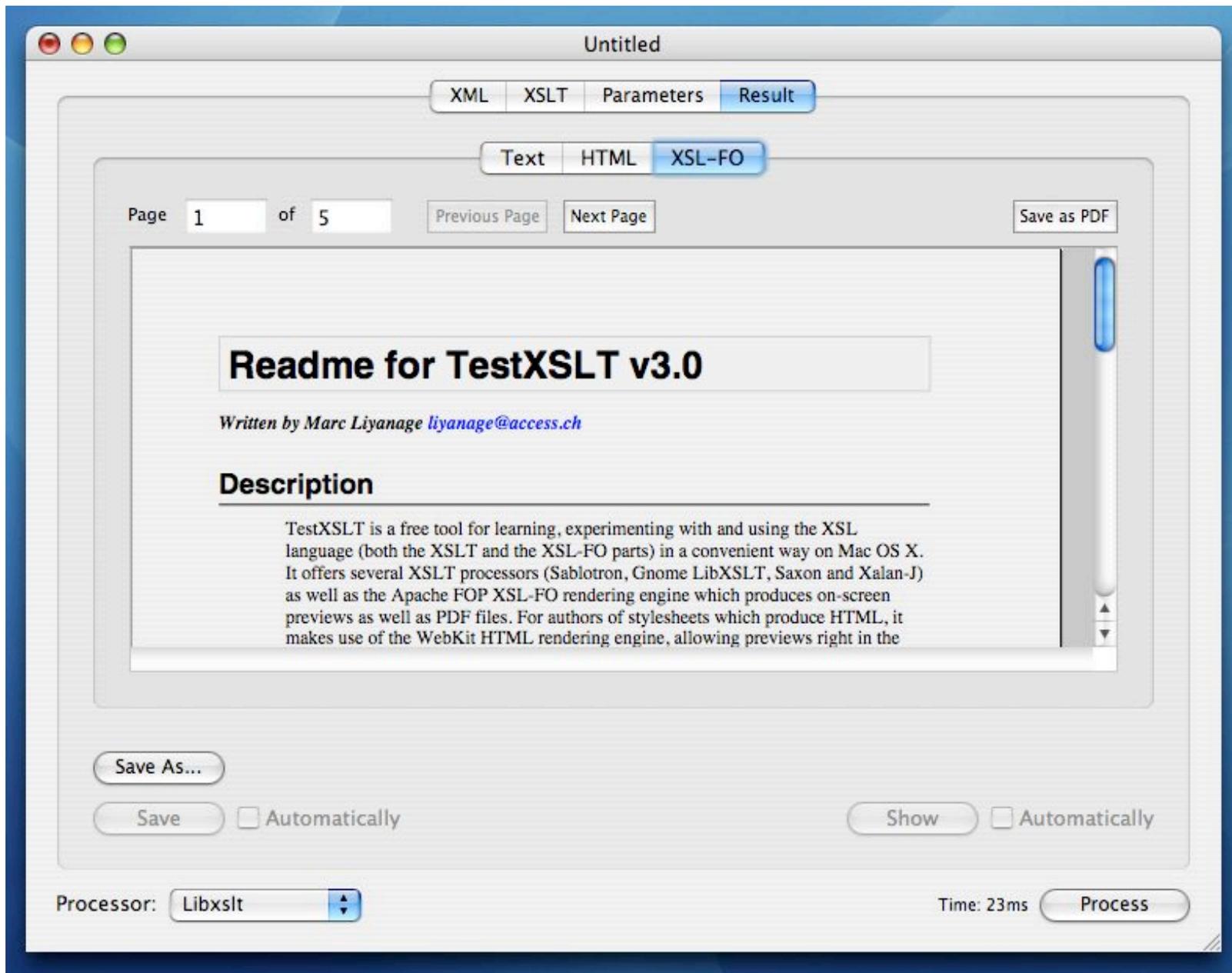
Il rendering

- Il programma che esegue uno stylesheet applicandolo ad un documento di chiama motore di rendering
- HTML su browser è un possibile modo di rendere XML
- Un altro modo è di usare XSL-FO (Formatting Objects), e generare direttamente un PDF

Resa HTML

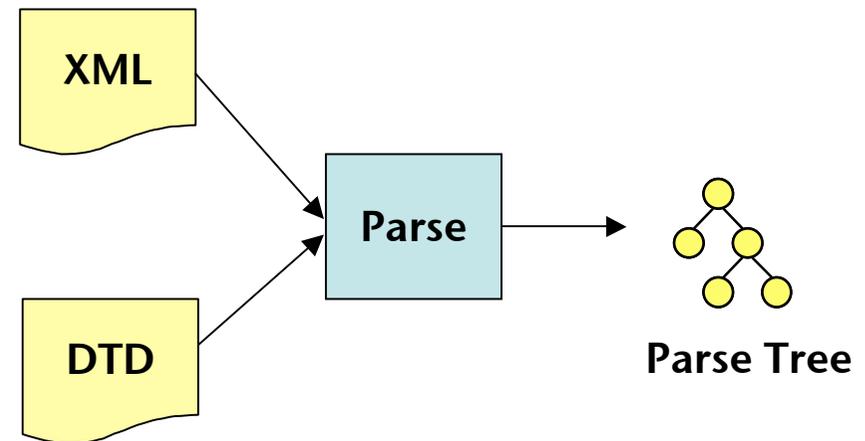


Resa FO



Analisi di un documento XML

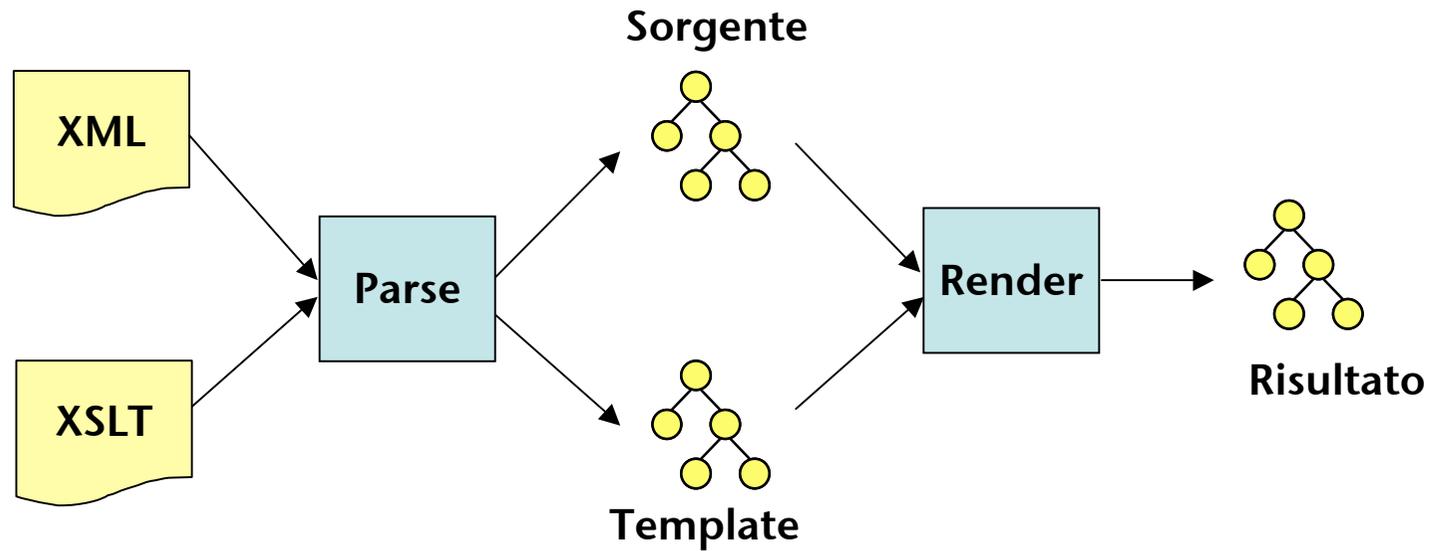
- L'**analisi sintattica** è il processo che trasforma un documento sorgente XML in una rappresentazione interna in forma di albero (detto *parse tree*)
- Se il programma parser dispone di un DTD o uno XML Schema del documento in analisi, può convalidare la conformità del documento ad un formato prestabilito



Trasformazione e stile

- XSL (eXtensible Stylesheet Language) include due tecnologie distinte
 - XSLT (XSL Transformation) definisce i meccanismi di trasformazione da XML in qualsiasi altra struttura linguistica
 - XSL-FO (XSL Formatting Objects) è un vocabolario per specificare markup di presentazione (sovrainsieme di CSS)
- XSLT è largamente usato per manipolare documenti mediante stylesheets

Applicazione di XSLT ad un documento



XSLT Stylesheet per RSS

- Questo stylesheet trasforma RSS in HTML
- Contiene alcune **regole di trasformazione**: il template *channel* crea un contesto cui viene applicato il template *item*

```
<?xml version='1.0'?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="html" encoding="UTF-8"/>

  <xsl:template match="channel"> ← Template
    <html>
      <head>
        <title><xsl:value-of select="title"/></title>
      </head>
      <body>
        <h1><xsl:value-of select="title"/></h1>
```

↑ Inserisce titolo canale

XSLT Stylesheet per RSS

```
<p><xsl:value-of select="description"/></p>
<ul>
  <xsl:apply-templates select="item"/> ← Apply the
</ul>                                     item template against all items in the channel
</body>
</html>
</xsl:template>

<xsl:template match="item">
  <li>
    <a href="{link}"><xsl:value-of select="title"></a>
  </li>
</xsl:template>
↑ Insert item link into attribute

</xsl:stylesheet>
```

Applicazioni di XML

- Portali
- Motori di ricerca
- Publishing
- Commercio elettronico
- Comunicazioni multicanale

Portali

The screenshot shows the My Yahoo! homepage in Microsoft Internet Explorer. The browser title is "Welcome to My Yahoo! - Microsoft Internet Explorer" and the address bar shows "http://my.yahoo.com/". The page features a navigation bar with "My YAHOO!", "Welcome to My Yahoo!", and links for "Yahoo!", "Help", "Sign In", and "INSTANT ROI. Procure with Oracle." Below this is a banner ad for "1-800-MY-YAHOO" promoting services like Mail, Weather, Finance, Sports, and News. The main content area includes a "Sign in to Yahoo!" form, a "Free Personalized Content" section, a "Portfolios" section with stock market data, a "Weather" section for Chicago, IL, and a "Lead Photo" section with a news article about a Chinese scholar.

Layout is simple HTML Table template file

Banner ad selected from database

Yahoo service promotion

Stock Market info from markets around the world

Photo and news copy from wire services

Login form communicates with user authentication application

Weather information syndicated from Weather.com

Indices		
DIA	10353.22	-52.45
NASDAQ	1985.24	+0.92
S&P 500	1189.28	-1.21
Widely Held		
* HWP	23.90	-1.78
* G	28.16	-0.03
* IBM	103.44	-1.45
* INTC	29.19	-0.23
* MSFT	66.19	-1.29
* ORCL	18.78	-0.48
* T	20.35	+0.20

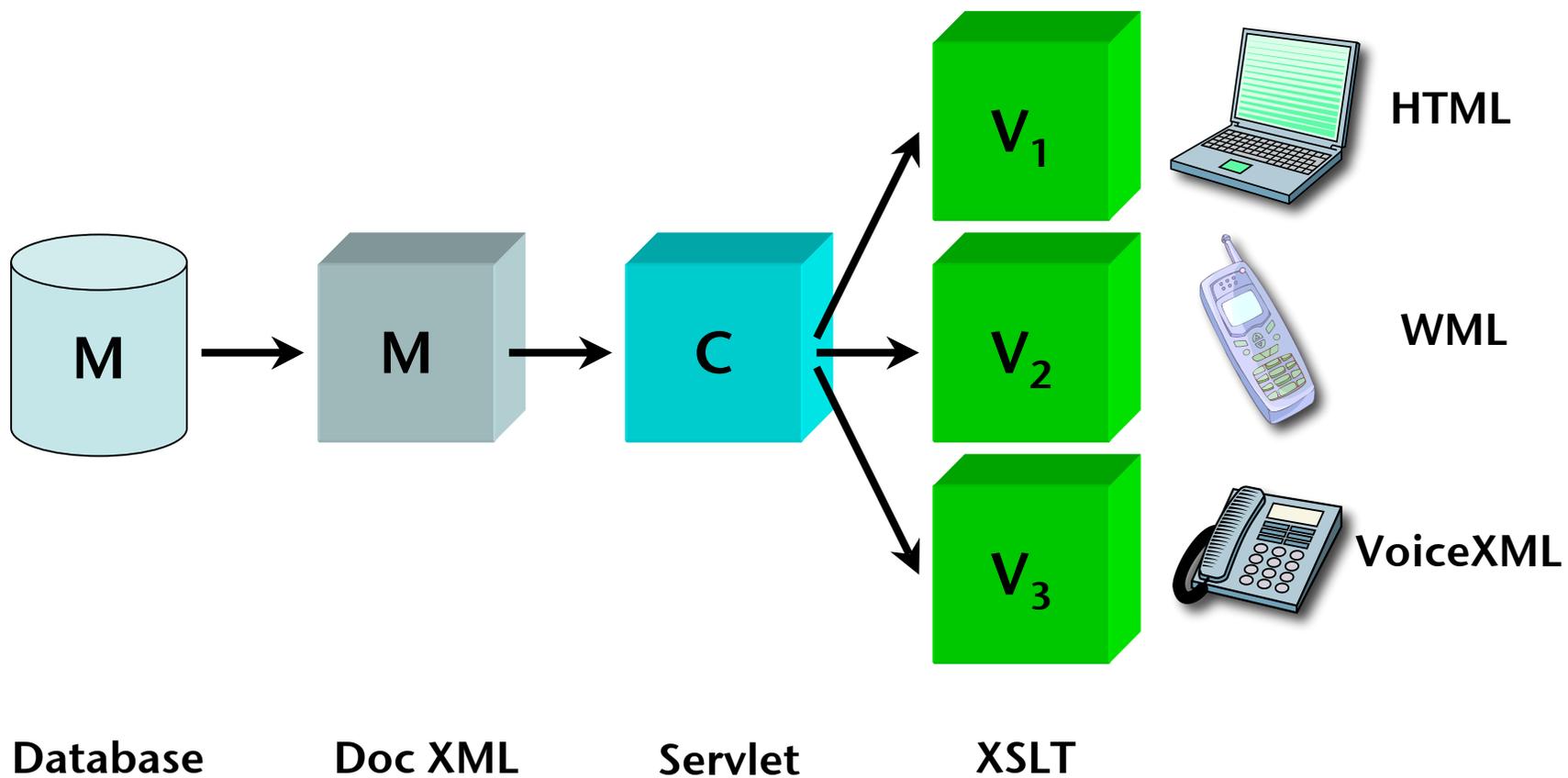
Applicazioni XML

- Motori di ricerca: standard per metadati
 - Ricerca di testo “nel contesto semantico”:
migliora l'accuratezza dei risultati
- Publishing: Open Document
 - Il documento viene presentato in un formato qualsiasi
- E-Commerce: ebXML
 - Integrazione di fonti di dati

Applicazioni XML

- DocBook: DTD per la creazione di libri tecnici; standard curato dal consorzio OASIS
- XHTML 1.0: DTD per la creazione di documenti Web; standard curato dal W3Consortium
- Text Encoding Initiative: DTD per la creazione di testi su supporto digitale in ambito umanistico; curato dal consorzio TEI
- Encoded Archival Description: DTD per la descrizione di materiali archivistici; realizzato dalla Biblioteca del Congresso USA

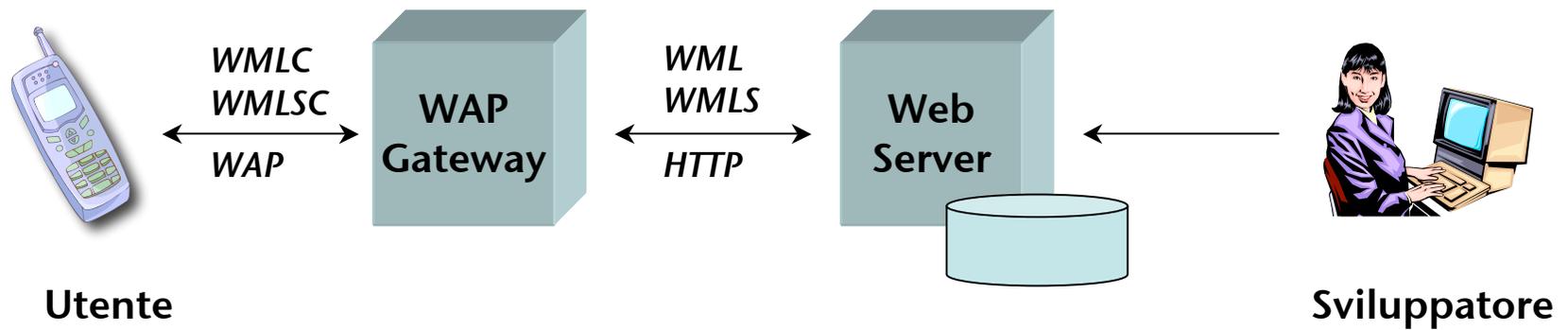
Multicanalità



WAP e WML

- I cellulari sono piccoli computer
- Possono eseguire applicazioni, accedere portali e interrogare database
- WAP (Wireless Application Protocol) è un protocollo usato per accedere da cellulare certi siti di informazione
- WML (Wireless Markup Language) è un linguaggio di markup, basato su XML, adatto ai micro-browser dei cellulari WAP

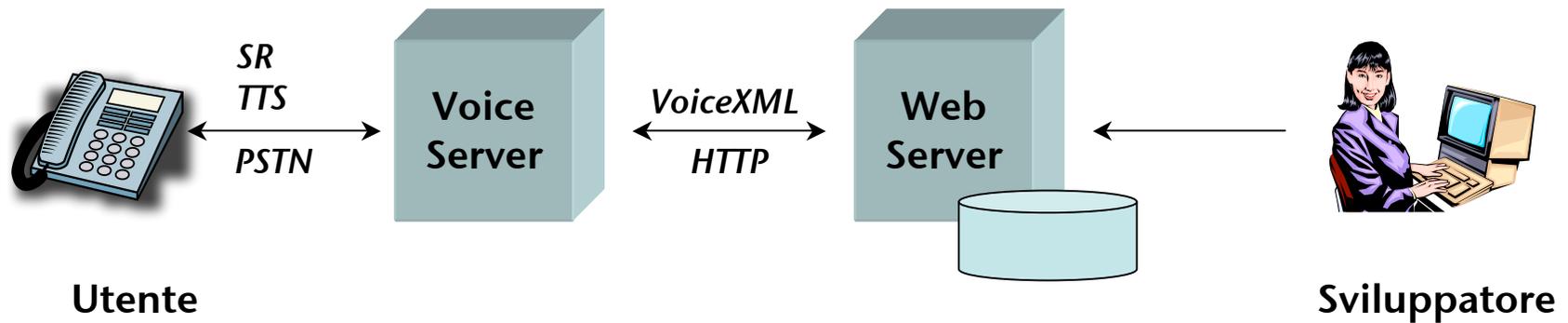
Applicazioni WML



VoiceXML

- Markup per browsers vocali
- Sintesi del parlato (text-to-speech)
- Output di audio registrato
- Speech recognition
- Registrazione del parlato
- Funzioni di telefonia (es. call transfer)
- W3C Speech Interface Framework

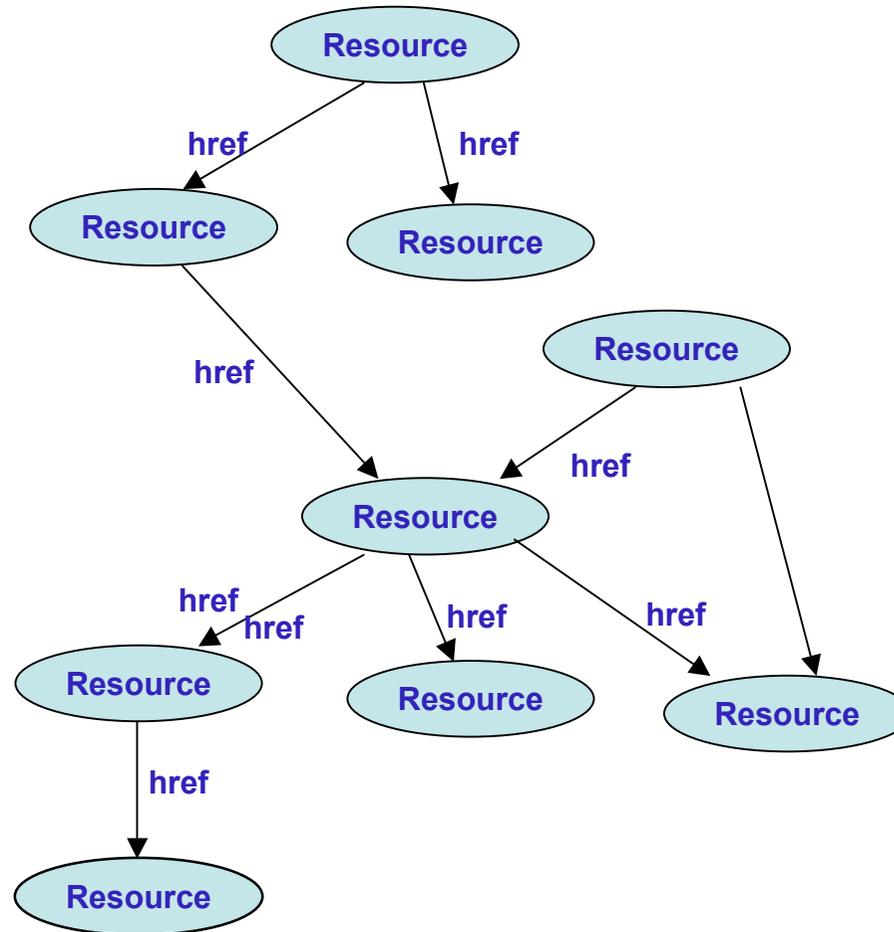
Applicazioni VoiceXML



SR *Speech Recognition*
TTS *Text-to-Speech*

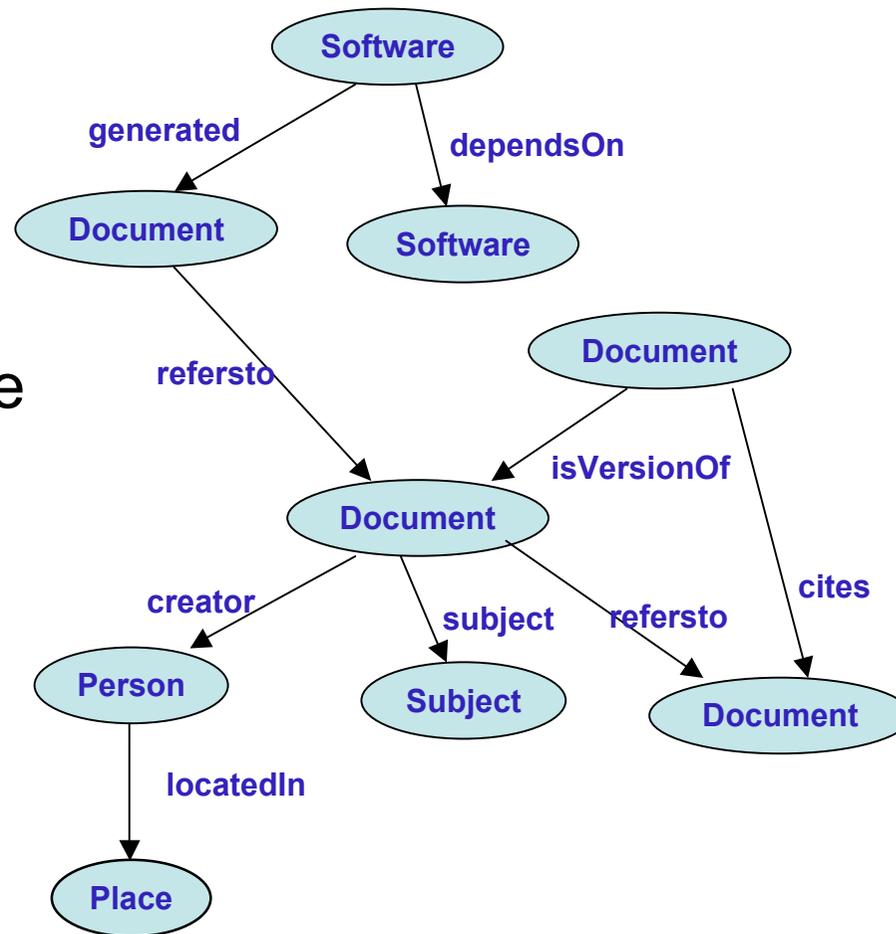
“Vecchio Web”

link ipertestuali
e risorse



"Semantic Web"

relazioni e risorse
con tipo



Obiettivi del Semantic Web

- Usare la potenza di calcolo dei computer per “capire” i contenuti del Web
- Ottenere da una ricerca sul Web più risultati interessanti ed accurati
- Favorire l'integrazione dei dati da sorgenti diverse
- Permettere l'automazione di operazioni

Resource Description Framework (RDF)

- RDF: è un linguaggio di markup, basato su XML, per descrivere e comunicare metadati
 - Una Risorsa (**Resource**) è un entità denotata da una URL
 - Una Proprietà (**Property**) è una risorsa che ha un nome e può essere usata come attributo: esempi sono Autore o Titolo. Di una proprietà conta soprattutto il nome, ma è anche una risorsa, dunque una proprietà può a sua volta avere proprietà
 - Un'asserzione (**Statement**) è una combinazione di una Risorsa (*soggetto* dell'asserzione), una Proprietà (*predicato* dell'asserzione), ed un valore (*oggetto* dell'asserzione)

Esempio RDF

- **Esempio: la frase**

"L'Autore di <http://www.textuality.com/RDF/Why.html> è Tim Bray."

si traduce in RDF così

```
<rdf:Description
  about='http://www.textuality.com/RDF/WhyRDF.html'>
  <Author>Tim Bray</Author>
  <Home-Page rdf:resource='http://www.textuality.com' />
</rdf:Description>
```

Conclusioni

- Cos'è XML?
- Quali problemi risolve?
 - Descrizione di strutture di documenti
 - Content delivery and presentation (portali)
 - Scambio dati (hubs)
- La famiglia XML
- Applicazioni XML
 - Separare il contenuto alla presentazione
 - Metalinguaggio di linguaggi di markup
 - Servizi Web
 - Ontologie (Web semantico)

Lecture

- J. H. Coombs, A. H. Renear, S. J. DeRose, Markup Systems and the future of Scholarly Text Processing, *Communications of the ACM*, 30(11), November 1987.

Riferimenti

- Berners Lee, *L'architettura del nuovo Web*, Feltrinelli, 2001
- Docbook: the complete guide

Siti

- www.w3.org/XML/
- www.tei-c.org/
- www.docbook.org/tdg/en/html/docbook.html
- www.maxdesign.com.au/presentation/liquid/index.htm
- www.griseldaonline.it/informatica/

Domande?