

XML (eXtensible Markup Language)

The **Extensible Markup Language** (XML) is a **W3C**-recommended **general-purpose markup language** for creating **special-purpose markup languages**, capable of describing many different kinds of data.

In other words:

XML is a way of describing data and an XML file can contain the data too, as in a database.

It is a simplified **subset** of **Standard Generalized Markup Language (SGML)**.

Its primary purpose is to facilitate the sharing of data across different systems, particularly systems connected via the Internet.

Languages based on XML (for example, Geography Markup Language (GML), RDF/XML, RSS, MathML, XHTML, SVG, and MusicXML) are defined in a formal way, allowing programs to modify and validate documents in these languages without prior knowledge of their form.

([**http://en.wikipedia.org/XML**](http://en.wikipedia.org/XML))

XML

is a W3C-recommended

Cosa è il W3C?

<http://www.w3c.it>

<http://www.w3.org>

general-purpose markup language

Cosa significa markup?

Cosa è un linguaggio di programmazione?

**for creating special-purpose markup languages,
capable of describing many different kinds of data.**

In other words:

**XML is a way of describing data and an XML file can contain
the data too, as in a database.**

Database e XML, descrizione del contenitore e del contenuto.

Markup – Marcare

Il termine deriva dall'ambiente tipografico dove veniva usato per definire le annotazioni fatte su una bozza, allo scopo di segnalare al compositore o al dattilografo il modo con cui alcune parti del testo andavano evidenziate o corrette.

(http://it.wikipedia.org/wiki/Linguaggio_di_markup)

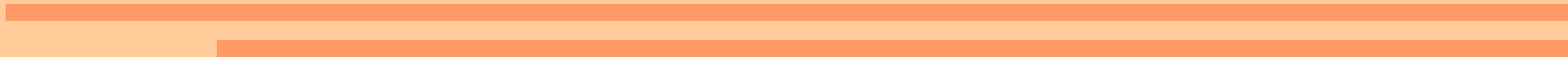
Ma l'idea stessa di corredare i testi con particolari note o accorgimenti grafici, che servissero a rendere un particolare significato, o una determinata enfasi, è ben più antica della stampa.

Programmare

I computer elaborano ogni operazione in linguaggio binario, ovvero in sequenze di 0 e 1.

Per poter costruire un programma o un'applicazione che interagisca con la macchina, si utilizzano delle "astrazioni intellegibili" di istruzioni che si possono comporre in varia maniera.

Un insieme di istruzioni del genere può vagamente dare l'idea di ciò che è un linguaggio di programmazione.



It is a **simplified subset** of: **Standard Generalized Markup Language (SGML)**.

Its primary purpose is **to facilitate the sharing of data across different systems**, particularly systems connected via the Internet.

1960-69

Charles **G**oldfarb, Edward **M**osher e Raymond **L**orie

In 1966... I knew nothing about computers, but I knew there had to be a better way to produce documents than dictating them, reviewing a draft, marking up the draft with corrections, reviewing the retyped draft, and then, in frustration, seeing that the typist had introduced more errors while making the corrections...

- Charles Goldfarb, The Roots of SGML -- A Personal Recollection, 1996.

Sviluppano: **GML (Generalized Markup Language)**

Per conto di: **IBM**

Questo dato è molto importante, evidenzia quanto il mondo della finanza internazionale sia a suo modo stato decisivo per l'informatica

Nel **1980**

American National Standards Institute (**ANSI**)
lavorando su GML

rilascia uno Standard denominato: **SGML**.

Questo viene adottato dalle massime istituzioni Statunitensi
come il Ministero degli interni o il Dipartimento della difesa

Una bozza viene recepita anche dalla CEE in 1985.

Lo Standard definitivo venne pubblicato nella norma:
ISO 8879:1986.

Pochi anni dopo uno scienziato del CERN, Tim Berners-Lee,
costruisce **un subset di SGML** che denomina: **HyperText
Markup Language (HTML)** al fine di definire una
struttura per documenti da pubblicare su Internet.

Ma HTML mostra, nel tempo e con l'espansione di Internet, i suoi limiti:

- 1) Si limita a descrivere la forma dei testi
- 2) Accetta molte deroghe alle regole base del linguaggio affidandosi alla capacità interpretativa dei browser.

Ad esempio:
per definire un grassetto si fa:

```
<b>prova</b>
```

o per un paragrafo:

```
<p>bla bla bla bla bla
```

e non è obbligatorio chiudere il tag `</p>`

Così a metà anni novanta (1995-1996) un gruppo di 11 sviluppatori, con a capo James Clark, si propone di sviluppare un nuovo subset di SGML, che permetta di identificare tanto porzioni formali del testo quanto di poter astrarne la semantica.

Nomi oltre XML in gara per il nuovo linguaggio:
"MAGMA" (Minimal Architecture for Generalized Markup Applications),
"SLIM" (Structured Language for Internet Markup)
"MGML" (Minimal Generalized Markup Language)

XML 1.0 divenne una "W3C Recommendation" nel Febbraio del 1998

Qualità di XML

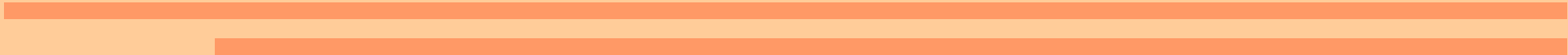
- 1) E' contemporaneamente intellegibile dall'uomo e dalla macchina
- 2) E' basato su Unicode, quindi supporta tutti i caratteri e gli alfabeti esistenti
- 3) Ha una sintassi molto rigida e non ammette equivoci.
- 4) La sua struttura gerarchica permette di rappresentare gran parte dei possibili documenti.
- 5) Funziona su file di testo semplici ed è "platform independent", questo dovrebbe assicurare stabilità e durata

Regole di Base di XML

1) Tutti i tag devono avere una apertura ed una chiusura

SBAGLIATO: `<p>paragrafo1`

GIUSTO: `<p>paragrafo1</p>`



2) I tag devono essere correttamente innestati

SBAGLIATO:

```
<p>Normal  
<em>emphasized  
<strong>strong emphasized</em>  
strong</strong>  
</p>
```

GIUSTO:

```
<p>  
Normal  
<em>  
emphasized  
<strong>strong emphasized</strong>  
</em>  
</p>
```

3) Ogni documento XML deve avere ***uno ed un solo*** elemento radice. Gli attributi devono sempre essere inclusi tra apici **doppi**.

```
<?xml version="1.0" encoding="UTF-8"?>
<recipe name="bread" prep_time="5 mins" cook_time="3 hours">
  <title>Basic bread</title>
  <ingredient amount="3" unit="cups">Flour</ingredient>
  <ingredient amount="0.25" unit="ounce">Yeast</ingredient>
  <ingredient amount="1.5" unit="cups"
state="warm">Water</ingredient>
  <ingredient amount="1" unit="teaspoon">Salt</ingredient>
  <instructions>
    <step>Mix all ingredients together, and knead thoroughly.</step>
    <step>Cover with a cloth, and leave for one hour in warm
room.</step>
    <step>Knead again, place in a tin, and then bake in the oven.</step>
  </instructions>
</recipe>
```

```
<?xml version="1.0"?>
<biblioteca>
  <libro codice="R414">
    <titolo>2001: Odissea nello spazio</titolo>
    <autore>
      <cognome>Clarke</cognome>
      <nome>Arthur Charles</nome>
    </autore>
    <editore>Rizzoli</editore>
    <parola_chiave>romanzo</parola_chiave>
    <parola_chiave>fantascienza
    </parola_chiave>
  </libro>
</biblioteca>
```

Strumenti in XML

XML è in realtà una famiglia di tecnologie

alla base

XML

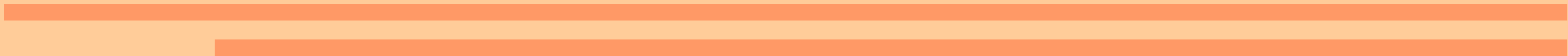
-- il documento

DTD / XSCHEMA

– le leggi che lo governano

XSL e XSLT

– le leggi che lo rendono a vista e lo trasformano



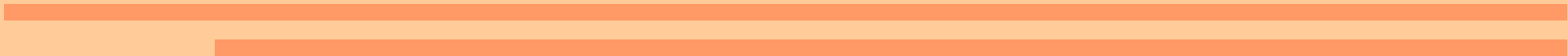
```
<xsl:template match="paragrafo">
  <h3>
    <xsl:choose>
      <xsl:when test="@tipo='abstract'">
Abstract dell'articolo
      </xsl:when>
      <xsl:when test="@tipo='bibliografia'">
        Bibliografia
      </xsl:when>
      <xsl:when test="@tipo='note'">
        Note finali
      </xsl:when>
      <xsl:otherwise>
<xsl:value-of select="@titolo" />
      </xsl:otherwise>
    </xsl:choose>
  </h3>
  <xsl:apply-templates/>
</xsl:template>
```

XPath Meccanismo di puntamento ai singoli nodi di un documento

XQuery Linguaggio di query in stile XML.

XML namespaces Sistema per accettare "tag" da differenti Schemi o DTD

XPointer Meccanismo di linking per XML.



Che ci faccio con XML?

Il **Resource Description Framework (RDF)** del W3C è un XML formato testo che supporta la descrizione di risorse e applicazioni metadata come liste musicali, album di foto e bibliografie.

RSS (acronimo di **RDF Site Summary** ed anche di **Really Simple Syndication**) è uno dei più popolari formati per la distribuzione di contenuti Web, in particolare di notizie.

VoiceXML/ CCXML

Voice markup and telephony call control; per abilitare l'accesso al web tramite voce

Geography Markup Language (GML)

Scalable Vector Graphics (SVG)

<http://en.wikipedia.org/wiki/Image:Orc.svg>