



Informatica Umanistica A.A. 2005/2006

LEZIONE 5

eXtensible Markup Language
XML - DTD



Lezione 5 SOMMARIO

1. Ripasso XML e concetti DTD
2. Dichiarazione DOCTYPE
3. Primo esempio (rubrica.xml)
4. Elementi (esempi)
5. Attributi (esempi)
6. Entità
7. Validare per imparare

Cosa è XML?

- È un "**markup language**" molto simile all'HTML
- È pensato per **descrivere dati e la loro struttura** (non la loro presentazione)
- Non esistono tag predefiniti: **lo sviluppatore definisce i suoi tag personali**
 - Per questo motivo l'XML si può considerare un **meta-linguaggio di markup** = linguaggio che permette di definire altri linguaggi di markup
- Utilizza un documento di definizione dei tipi (**Document Type Definition** - DTD) per descrivere i dati.
 - Documento che definisce i tag utilizzabili in un documento XML, la loro struttura e altre info sugli attributi dei tag

Ma qualcuno lo usa questo XML? i feed RSS

- <http://www.repubblica.it/servizi/rss/index.html>
- Repubblica.it offre i suoi contenuti con il sistema RSS (Really Simple Syndication). E i blog? Anche.
- Un'applicazione in grado di interpretare un documento RSS ne effettua il parsing (=scansione), ovvero analizza il documento per individuare i tag e isolare gli elementi, per poi convertire i contenuti decodificati nel formato utile all'obiettivo specifico: per esempio un feed reader può estrarre i titoli di tutti gli elementi item allo scopo di visualizzare la lista degli articoli di un giornale online; un aggregatore Web può estrarre i contenuti del feed per convertirli in HTML e incorporarli all'interno delle proprie pagine.



Un documento XML è composto da tre parti

- **XML** (eXtensible Markup Language)
 - i dati (struttura fisica - *document instance*)
 - *Quale informazione e' contenuta nel file?*
- **DTD** (Document Type Definition)
 - lo schema dei dati (struttura logica)
 - *Quali tag posso usare? Dove? Come?*
- **XSL** (eXtensible Stylesheet Language)
 - la presentazione dei dati (*stylesheet*)
 - *Come visualizzo questi dati?*



Documenti ben formati

- I nomi degli elementi sono **case-sensitive** (devono coincidere in apertura e chiusura).
- **Ogni elemento aperto deve essere chiuso** entro la fine del documento.
- Gli elementi possono essere nidificati, e in tal caso **vanno chiusi esattamente nell'ordine inverso a quello di apertura**.
- Un documento XML deve avere **un unico elemento "radice"**, in cui tutti gli altri sono nidificati
- I valori si racchiudono tra **singoli (') o doppi (" ") apici**
- Gli elementi non devono avere due attributi con lo stesso nome
- I nomi degli elementi e degli attributi non devono contenere caratteri speciali (&,<,> ...)
- I nomi degli elementi non devono contenere spazi

Ma che succede se il mio documento XML non è ben formato?

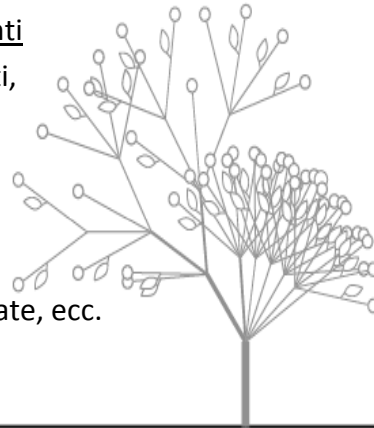
- Quando cercherò di aprirlo **il parser individuerà l'errore!**
- Significa che...
- In generale la **pagina non verrà visualizzata**; verrà invece indicato il **numero di riga** in cui è stato trovato l'errore.

Documenti ben formati e anche validi

- Abbiamo visto che con XML possiamo **definire i tag a piacimento**
- Ma è necessario **definire una grammatica**: un insieme di regole che indicano quali *vocaboli* (elementi) possiamo usare per formare le *frasi* (documenti) → questa grammatica si definisce con il **DTD – Document Type Definition**
- **Un documento XML è *valido* se è ben formato e rispetta le regole del DTD associato.**
 - Cosa è un DTD? Lo vediamo tra poco...
 - Come si validano i documenti? Lo vediamo tra poco...
- **Un documento ben formato può non essere valido?**
- **Un documento valido può non essere ben formato?**

Struttura gerarchica ad albero di un XML

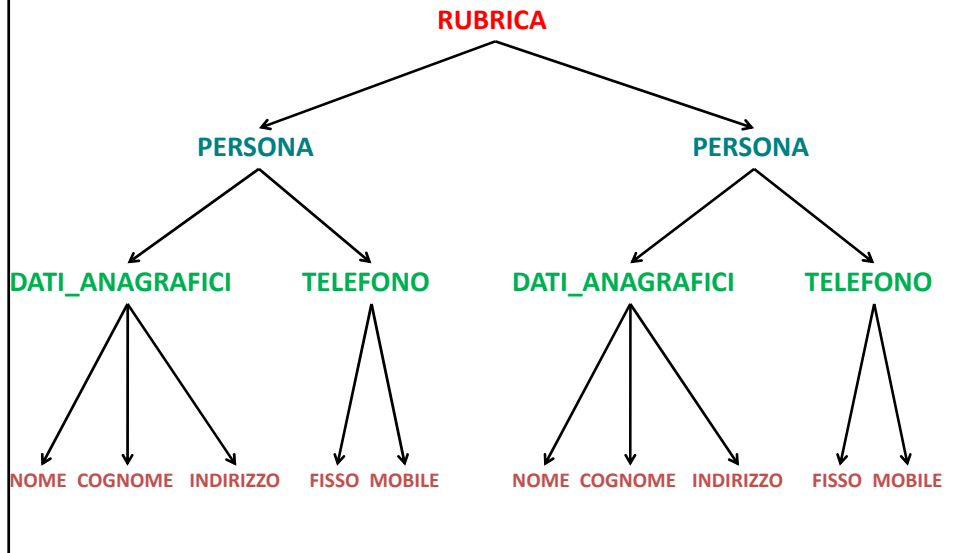
- **Radice (root):** l'elemento principale
- **Nodi intermedi (figli):** elementi
 - con sottoalberi di elementi, attributi e valori
- **Nodi foglia:** attributi e valori
 - attributi: coppie (nome, valore)
 - valori: stringhe, numeri, date, ecc.
- Albero XML (radice, foglie, ...)



Esempio: una rubrica in XML

```
<?xml version="1.0" standalone="yes" ?>
<RUBRICA>
  <PERSONA>
    <DATI_ANAGRAFICI>
      <NOME>Giulio</NOME>
      <COGNOME>Cesare</COGNOME>
      <INDIRIZZO>
        <PIAZZA>Bruto</PIAZZA><N_CIVICO>10</N_CIVICO><CITTA>Roma</CITTA><CAP>10100</CAP>
      </INDIRIZZO>
    </DATI_ANAGRAFICI>
    <TELEFONO>
      <FISSO><NUMERO prefisso="06">321654</NUMERO></FISSO>
      <MOBILE><NUMERO compagnia="blu">32557684785</NUMERO></MOBILE>
    </TELEFONO>
  </PERSONA>
  <PERSONA>
    <DATI_ANAGRAFICI>
      <NOME>Giovanni</NOME>
      <COGNOME>Pompeo</COGNOME>
      <INDIRIZZO>
        <PIAZZA>Catilina</PIAZZA><N_CIVICO>12</N_CIVICO><CITTA>Milano</CITTA><CAP>12100</CAP>
      </INDIRIZZO>
    </DATI_ANAGRAFICI>
    <TELEFONO>
      <FISSO><NUMERO prefisso="02">96877564</NUMERO></FISSO>
      <MOBILE><NUMERO compagnia="tim">3356545342</NUMERO></MOBILE>
    </TELEFONO>
  </PERSONA>
</RUBRICA>
```

La rubrica, vista come albero



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

```
<articolo titolo="Titolo dell'articolo">
```

```
<paragrafo titolo="Titolo del primo paragrafo">
```

```
<testo> Blocco di testo del primo paragrafo
```

```
</testo>
```

```
<immagine file="immagine1.jpg">
```

```
</immagine>
```

```
</paragrafo>
```

```
<paragrafo titolo="Titolo del secondo paragrafo">
```

```
<testo> Blocco di testo del secondo paragrafo
```

```
</testo>
```

```
<codice>
```

Esempio di codice

```
</codice>
```

```
<testo>Altro blocco di testo</testo>
```

```
</paragrafo>
```

```
<paragrafo tipo="bibliografia">
```

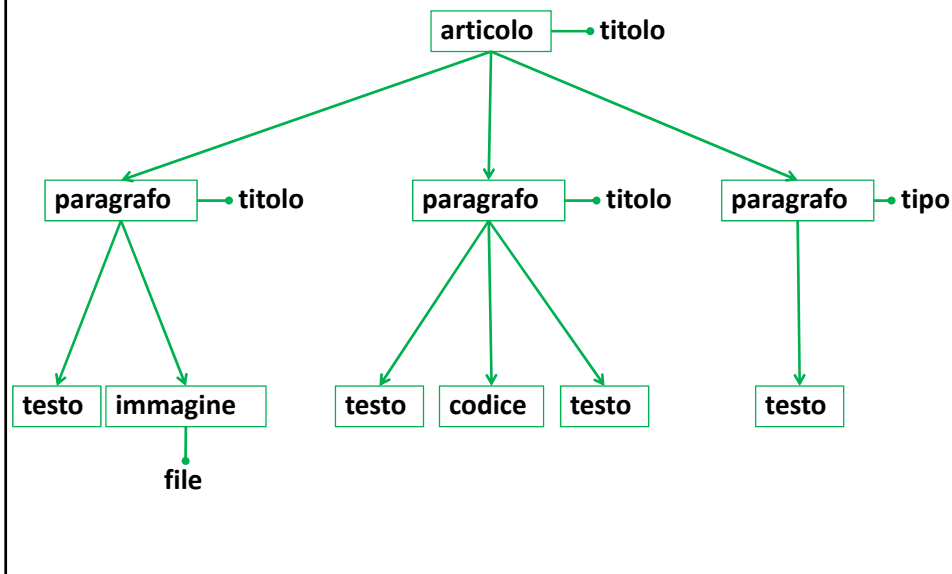
```
<testo> Riferimento ad un articolo </testo>
```

```
</paragrafo>
```

```
</articolo>
```

Altro esempio:
articolo

L'articolo, visto come albero



DTD – Document Type Definition

- Un DTD è un documento che descrive i **tag utilizzabili** in un documento XML, la loro **struttura** nel documento e **altre informazioni sugli attributi** dei tag (ed eventualmente sulle entità, se ce ne sono).
- Quali tag posso usare? Quante volte? Un tag può contenere un altro tag? Quale?
- Queste e altre cose si specificano nel **DTD (Document Type Definition)**.



Importanza DTD

- Il DTD (Document Type Definition) consente di dichiarare in maniera **univoca e formale** la **struttura di markup** mediante la definizione dello schema dei tag.
 - *Ovvero Quali tag posso usare? Quante volte? Un tag puo' contenere un altro tag? Quale?*
- Questo tipo di procedura consente una rapida interpretazione del materiale pubblicato sia dalla macchina che dall'utente umano.
- Un file XML e' **valido se è ben formato e rispetta il DTD**
- *Chi scrive il DTD? Chi definisce cosa è valido e cosa no per un certo documento XML?*



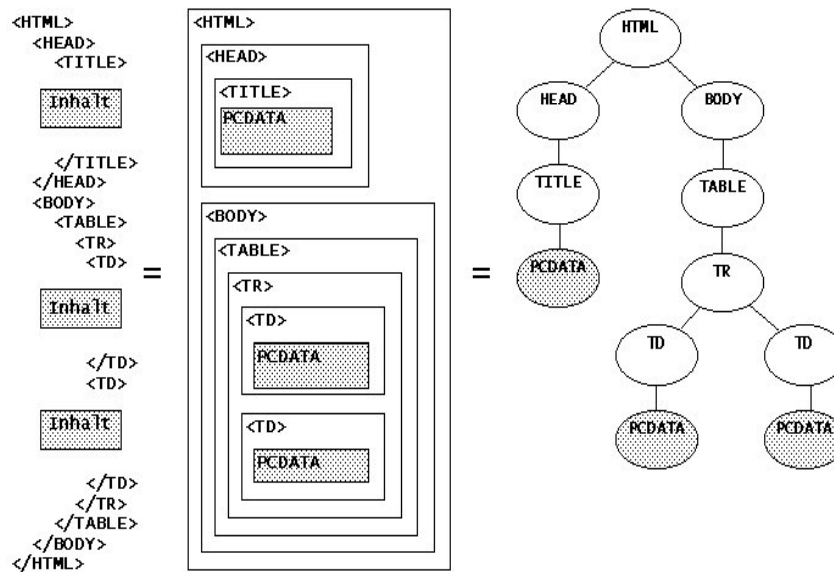
L'HTML e' XML?

- Dipende, l'xHTML si'.
- Chi definisce il DTD per l'xHTML?

Chi definisce il DTD dell'xHTML?

- Per il linguaggio HTML, il W3C definisce il DTD, ad esempio dice che
 - L'element root e' <html> (in realta' <html> e' il tag)
 - Che deve contenere uno e un solo <head>
 - Che deve contenere uno e un solo <body>
 - Che dentro a <body> ci possono essere un numero qualunque di <p> (da 0 a n)
 -

L'xHTML visto come albero



Chi definisce il DTD per la rubrica?

- NOI
- Come?

Un documento XML è composto da tre parti

- **XML** (eXtensible Markup Language)
 - i dati (struttura fisica - *document instance*)
 - *Quale informazione e' contenuta nel file?*
- **DTD** (Document Type Definition)
 - lo schema dei dati (struttura logica)
 - *Quali tag posso usare? Dove? Come?*
- **XSL** (eXtensible Stylesheet Language)
 - la presentazione dei dati (*stylesheet*)
 - *Come visualizzo questi dati?*

Dove sta il DTD?

Due modi per collegare un documento XML con il suo DTD

Un **DTD** definisce la **grammatica** per il markup del nostro documento XML. Dobbiamo quindi **mettere in relazione il documento XML con il suo DTD** (in modo che il parser possa verificare la validità del documento XML)

Esistono due modi per fare questo (noi usiamo il primo):

1. **Scriviamo il DTD all'interno del documento XML**
2. **Il DTD è in un file esterno**, e nel documento XML è presente un **riferimento** a tale file

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

```
<!DOCTYPE ARTICOLO SYSTEM "articolo.dtd">
```

Nome del documento

URL del DTD definito in articolo.dtd

La dichiarazione XML

- Abbiamo già visto che la **dichiarazione XML** è **obbligatoria** e deve essere posta *all'immediato inizio del documento (ovvero come primo carattere! Niente spazi prima!)*:

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
```

- Gli attributi sono:
 - **version**: (*obbligatorio*) la **versione di XML** usata.
 - **encoding**: (*opzionale*) nome della **codifica dei caratteri** usata nel documento (default: Unicode UTF-8 o 16)
 - **standalone**: (*opzionale*) (default: no)
 - standalone=yes → il file non fa **riferimento ad altri file esterni**, ovvero il solo file XML contiene dati e DTD
 - **Noi useremo sempre standalone="yes"**
 - standalone=no → e' necessario un file esterno (che contiene il DTD)

Dopo la dichiarazione XML: la dichiarazione DOCTYPE

- Il DTD (Document Type Definition) definisce gli **elementi** e gli **attributi** consentiti, e la **struttura** del documento.
- Un DTD può essere interno o esterno al documento XML: noi useremo la dichiarazione interna (dichiarazione DOCTYPE)
- Se il DTD è dichiarato all'interno del documento XML, viene inserito in una dichiarazione DOCTYPE. Idealmente la sintassi è:


```
<!DOCTYPE ElementoRadice [ ...Definizioni del DTD... ]>
```

- **ElementoRadice** (*obbligatorio*) è il nome dell'elemento radice del documento.
- **Definizioni del DTD** (*opzionale*) è un DTD specificato *inline* al documento.

Ricordate la rubrica in XML?

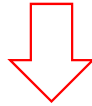
```
<RUBRICA>  
<PERSONA>  
  <DATI_ANAGRAFICI>  
    <NOME>Giulio</NOME>  
    <COGNOME>Cesare</COGNOME>  
    <INDIRIZZO>  
      <PIAZZA>Bruto</PIAZZA><N_CIVICO>10</N_CIVICO><CITTA>Roma</CITTA><CAP>10100</CAP>  
    </INDIRIZZO>  
  </DATI_ANAGRAFICI>  
  <TELEFONO>  
    <FISSO><NUMERO prefisso="06">321654</NUMERO></FISSO>  
    <MOBILE><NUMERO compagnia="blu">32557684785</NUMERO></MOBILE>  
  </TELEFONO>  
</PERSONA>  
<PERSONA>  
  <DATI_ANAGRAFICI>  
    <NOME>Giovanni</NOME>  
    <COGNOME>Pompeo</COGNOME>  
    <INDIRIZZO>  
      <PIAZZA>Catilina</PIAZZA><N_CIVICO>12</N_CIVICO><CITTA>Milano</CITTA><CAP>12100</CAP>  
    </INDIRIZZO>  
  </DATI_ANAGRAFICI>  
  <TELEFONO>  
    <FISSO><NUMERO prefisso="02">96877564</NUMERO></FISSO>  
    <MOBILE><NUMERO compagnia="tim">3356545342</NUMERO></MOBILE>  
  </TELEFONO>  
</PERSONA>  
</RUBRICA>
```

Qual è l'elemento radice?



Esempio di un DTD: rubrica.xml

<!DOCTYPE ElementoRadice [...Definizioni del DTD...]>



DIVENTA:

<!DOCTYPE RUBRICA [

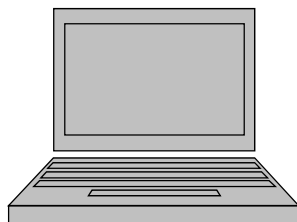
....

qui i vari elementi che possono essere contenuti
nell'ElementoRadice

....

]>

[Come digito le parentesi quadre?]



Ctrl+Alt+[→	[
Ctrl+Alt+]	→]

1. Parentesi quadra aperta [

Tenendo premuti i tasti **Ctrl** e **Alt** digitare [(di solito si trova a sinistra del tasto Enter/Invio)

2. Parentesi quadra chiusa]

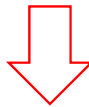
Tenendo premuti i tasti **Ctrl** e **Alt** digitare] (di solito si trova a sinistra del tasto Enter/Invio)

Esempio di un DTD: rubrica.xml

Sappiamo che la **dichiarazione XML** è **obbligatoria** e deve essere posta all'immediato inizio del documento (ovvero come primo carattere! Niente spazi prima!):

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
```

Quindi il file rubrica.xml inizierà con:



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

```
<!DOCTYPE RUBRICA [
```

```
....
```

qui i vari elementi che possono essere contenuti nel RootElement

Cosa scrivo qua dentro?

```
....
```

```
]>
```

Come scrivo il DTD per il mio XML: sintassi

La sintassi di un DTD si basa principalmente su **due dichiarazioni**:

- **<!ELEMENT>**

Definisce gli elementi utilizzabili nel documento XML e la **struttura** del documento

- **<!ATTLIST>**

Definisce la lista di attributi per ciascun elemento

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
<!DOCTYPE RUBRICA [
<!ELEMENT RUBRICA (PERSONA)*>
<!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
<!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
<!ELEMENT TELEFONO (FISSO?, MOBILE?)>
<!ELEMENT NOME (#PCDATA)>
<!ELEMENT COGNOME (#PCDATA)>
<!ELEMENT FISSO (NUMERO+)>
<!ELEMENT MOBILE (NUMERO+)>
<!ELEMENT NUMERO (#PCDATA)>
<!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
<!ELEMENT PIAZZA (#PCDATA)>
<!ELEMENT VIA (#PCDATA)>
<!ELEMENT N_CIVICO (#PCDATA)>
<!ELEMENT CITTA (#PCDATA)>
<!ELEMENT CAP (#PCDATA)>
<!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
>

Definizione di ELEMENTO

<!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>

- Un elemento contiene solo altri elementi, senza parti opzionali.
- In questo esempio, dentro all'elemento **PERSONA** ci **deve** essere un elemento **DATI_ANAGRAFICI**, seguito da un elemento **TELEFONO**.

```
<PERSONA>  
  <DATI_ANAGRAFICI> ... </DATI_ANAGRAFICI>  
  <TELEFONO> ... </TELEFONO>  
</PERSONA>
```

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

```
<RUBRICA>
<PERSONA>
  <DATI_ANAGRAFICI>
    <NOME>Giulio</NOME>
    <COGNOME>Cesare</COGNOME>
    <INDIRIZZO>
      <PIAZZA>Bruto</PIAZZA><N_CIVICO>10</N_CIVICO><CITTA>Roma</CITTA><CAP>10100</CAP>
    </INDIRIZZO>
  </DATI_ANAGRAFICI>
  <TELEFONO>
    <FISSO><NUMERO prefisso="06">321654</NUMERO></FISSO>
    <MOBILE><NUMERO compagnia="blu">32557684785</NUMERO></MOBILE>
  </TELEFONO>
</PERSONA>
<PERSONA>
  <DATI_ANAGRAFICI>
    <NOME>Giovanni</NOME>
    <COGNOME>Pompeo</COGNOME>
    <INDIRIZZO>
      <PIAZZA>Catilina</PIAZZA><N_CIVICO>12</N_CIVICO><CITTA>Milano</CITTA><CAP>12100</CAP>
    </INDIRIZZO>
  </DATI_ANAGRAFICI>
  <TELEFONO>
    <FISSO><NUMERO prefisso="02">96877564</NUMERO></FISSO>
    <MOBILE><NUMERO compagnia="tim">3356545342</NUMERO></MOBILE>
  </TELEFONO>
</PERSONA>
</RUBRICA>
```

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

```
<RUBRICA>
  <PERSONA>
    <DATI_ANAGRAFICI>
      <NOME>Giulio</NOME>
      <COGNOME>Cesare</COGNOME>
      <INDIRIZZO>
        <PIAZZA>Bruto</PIAZZA><N_CIVICO>10</N_CIVICO><CITTA>Roma</CITTA><CAP>10100</CAP>
      </INDIRIZZO>
    </DATI_ANAGRAFICI>
    <TELEFONO>
      <FISSO><NUMERO prefisso="06">321654</NUMERO></FISSO>
      <MOBILE><NUMERO compagnia="blu">32557684785</NUMERO></MOBILE>
    </TELEFONO>
  </PERSONA>
  <PERSONA>
    <DATI_ANAGRAFICI>
      <NOME>Giovanni</NOME>
      <COGNOME>Pompeo</COGNOME>
      <INDIRIZZO>
        <PIAZZA>Catilina</PIAZZA><N_CIVICO>12</N_CIVICO><CITTA>Milano</CITTA><CAP>12100</CAP>
      </INDIRIZZO>
    </DATI_ANAGRAFICI>
    <TELEFONO>
      <FISSO><NUMERO prefisso="02">96877564</NUMERO></FISSO>
      <MOBILE><NUMERO compagnia="tim">3356545342</NUMERO></MOBILE>
    </TELEFONO>
  </PERSONA>
</RUBRICA>
```

DTD per la rubrica

```
<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE RUBRICA [
  <ELEMENT RUBRICA (PERSONA)*>
  <ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
  <ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
  <ELEMENT TELEFONO (FISSO?, MOBILE?)>
  <ELEMENT NOME (#PCDATA)>
  <ELEMENT COGNOME (#PCDATA)>
  <ELEMENT FISSO (NUMERO+)>
  <ELEMENT MOBILE (NUMERO+)>
  <ELEMENT NUMERO (#PCDATA)>
  <ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?>
  <ELEMENT PIAZZA (#PCDATA)>
  <ELEMENT VIA (#PCDATA)>
  <ELEMENT N_CIVICO (#PCDATA)>
  <ELEMENT CITTA (#PCDATA)>
  <ELEMENT CAP (#PCDATA)>
  <!ATTLIST NUMERO
    prefisso CDATA #IMPLIED
    compagnia (wind | omnitel | tim | blu ) #IMPLIED>
]>
```

```
<RUBRICA>
  <PERSONA>
    <DATI_ANAGRAFICI>
      <NOME>Giulio</NOME>
      <COGNOME>Cesare</COGNOME>
      <INDIRIZZO>
        <PIAZZA>Bruto</PIAZZA><N_CIVICO>10</N_CIVICO><CITTA>Roma</CITTA><CAP>10100</CAP>
      </INDIRIZZO>
    </DATI_ANAGRAFICI>
    <TELEFONO>
      <FISSO><NUMERO prefisso="06">321654</NUMERO></FISSO>
      <MOBILE><NUMERO compagnia="blu">32557684785</NUMERO></MOBILE>
    </TELEFONO>
  </PERSONA>
  <PERSONA>
    <DATI_ANAGRAFICI>
      <NOME>Giovanni</NOME>
      <COGNOME>Pompeo</COGNOME>
      <INDIRIZZO>
        <PIAZZA>Catilina</PIAZZA><N_CIVICO>12</N_CIVICO><CITTA>Milano</CITTA><CAP>12100</CAP>
      </INDIRIZZO>
    </DATI_ANAGRAFICI>
    <TELEFONO>
      <FISSO><NUMERO prefisso="02">96877564</NUMERO></FISSO>
      <MOBILE><NUMERO compagnia="tim">3356545342</NUMERO></MOBILE>
    </TELEFONO>
  </PERSONA>
</RUBRICA>
```

Sintassi <!ELEMENT>

<!ELEMENT nome_elemento (#PCDATA)>

- Significa:
è lecito scrivere nel documento XML un tag **<nome_elemento>** che contiene caratteri riconoscibili dal parser XML di tipo **#PCDATA** (ricordate?)

<nome_elemento> qui dentro il testo
</nome_elemento>

PCDATA (Parsed Character DATA)

- È il testo che si trova tra il tag di apertura e il tag di chiusura di un elemento (il contenuto vero e proprio del documento)
- Esso corrisponde alle parole, gli spazi e la punteggiatura che costituiscono il testo.
- Viene anche detto **PCDATA (Parsed Character DATA)** perché a differenza dei CDATA (*character data*, il cui contenuto testuale non viene analizzato) quello degli elementi è soggetto ad azione di parsing (= analisi, per lo più per identificare le entità e sostituirle).

```
<codice> <![CDATA[  
  <libro><capitolo></capitolo>  
  </libro>  
]]> </codice>
```

```
<testo>  
Oggi &egrave; una bella  
giornata.  
</testo>
```

Come si dichiarano gli elementi

- In generale, in un DTD gli elementi XML si dichiarano con una sintassi del tipo:

<!ELEMENT nome_elemento categoria>

Es. EMPTY: l'elemento è vuoto

oppure

<!ELEMENT nome_elemento (contenuto_dell_elemento)>

Es. #PCDATA: testo soggetto ad azione di parsing

Altri tipi di dichiarazioni <!ELEMENT> EMPTY e ANY

Elementi vuoti:

- **<!ELEMENT nome_elemento EMPTY>**

- Es. <!ELEMENT br EMPTY>

In XML diventa:

Elementi che contengono qualsiasi combinazione di dati:

- **<!ELEMENT nome_elemento ANY>**

- Es: <!ELEMENT note ANY>

Altri tipi di dichiarazioni <!ELEMENT> #PCDATA

Elementi che contengono **Parsed Character Data (PCDATA)**:

- <!ELEMENT nome_elemento (#PCDATA)>

Esempio in XML:

```
<testo>
```

```
Oggi &egrave; una bella  
giornata.
```

```
</testo>
```

Altri tipi di dichiarazioni <!ELEMENT> Sequenze di sotto-elementi

Elementi che contengono **solo una sequenza di sotto-elementi** (come nell'esempio della rubrica):

- <!ELEMENT nome_elemento (sotto-elemento1, sotto-elemento2)>

- Es. rubrica:

```
<!ELEMENT PERSONA (DATI_ANAGRAFICI,  
TELEFONO)>
```

Altri tipi di dichiarazioni <!ELEMENT> Sequenze di sotto-elementi

- Es. rubrica:
<!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
- NB: Quando in un DTD i sotto-elementi sono dichiarati in una **sequenza separati da una virgola**, dovranno apparire nello stesso ordine nel documento XML.
- **Anche i sotto-elementi vanno dichiarati nel DTD.**
L'esempio completo del DTD di rubrica.xml è quindi:

```
<!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>  
<!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>  
<!ELEMENT TELEFONO (FISSO?, MOBILE?)>
```

Altri tipi di dichiarazioni <!ELEMENT> solo un sotto-elemento

Elementi che contengono **un unico sotto-elemento**

- <!ELEMENT nome_elemento (sotto-elemento)>

- Es.:
<!ELEMENT NOTE (MESSAGGIO)>

Nell'esempio si dichiara che il sotto-elemento MESSAGGIO (figlio di NOTE) deve apparire solo una volta, e solo all'interno dell'elemento NOTE.

Altri tipi di dichiarazioni <!ELEMENT> una o più occorrenze di un sotto-elemento

Elementi che contengono **minimo una occorrenza di un sotto-elemento**

○ <!ELEMENT nome_elemento (sotto-elemento+)>

- Es. rubrica:

<!ELEMENT FISSO (NUMERO+)>

Il segno **+** dichiara che il sotto-elemento NUMERO può essere presente **una o più volte** all'interno dell'elemento FISSO

Altri tipi di dichiarazioni <!ELEMENT> zero o una occorrenza di un sotto-elemento

Elementi che contengono **zero o una occorrenza di un sotto-elemento**

○ <!ELEMENT nome_elemento (sotto-elemento?)>

- Es. rubrica:

<!ELEMENT TELEFONO (FISSO?, MOBILE?)>

Il segno **?** dichiara che i sotto-elementi FISSO e NUMERO possono essere presente **zero o una volta** all'interno dell'elemento TELEFONO. (Se ci sono, ci sono solo una volta = nella rubrica registriamo, se c'è, solo un numero di telefono fisso e uno mobile)

Altri tipi di dichiarazioni <!ELEMENT> uno solo tra diversi sotto-elementi

Elementi che contengono **uno solo tra una lista di sotto-elementi**

○ <!ELEMENT nome_elemento (sotto-elemento1 | sotto-elemento | sotto-elemento3)>

- Es. rubrica:

<!ELEMENT MESSAGGIO (TO|FROM|HEADER|MESSAGE)>

Il segno | dichiara che l'elemento MESSAGGIO può contenere il sotto-elemento TO, oppure il sotto-elemento FROM, oppure il sotto-elemento HEADER, oppure il sotto-elemento MESSAGE.

Altri tipi di dichiarazioni <!ELEMENT> contenuto misto

Elementi che hanno un **contenuto misto**, cioè **dati oppure sottoelementi**:

- Esempio:

<!ELEMENT NOTE
(#PCDATA|TO|FROM|HEADER|MESSAGE)*>

L'asterisco * dichiara che l'elemento in questione può essere presente zero o più volte.

Quindi la dichiarazione dell'esempio ci dice che l'elemento NOTE può contenere zero o più occorrenze (= qualsiasi numero) di Parsed Character Data, oppure dell'elemento TO, oppure FROM, oppure HEADER, oppure MESSAGE.

Riassunto simboli

+	L'elemento cui si applica il simbolo può essere presente una o più volte .
?	L'elemento cui si applica il simbolo può essere presente zero o una sola volta .
*	L'elemento cui si applica il simbolo può essere presente zero o più volte .
 (shift +)	Simbolo "OR": può essere presente solo uno degli elementi in una lista .

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

DTD per la rubrica

```
<!DOCTYPE RUBRICA [
```

```
<!ELEMENT RUBRICA (PERSONA)*>
```

```
<!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
```

```
<!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
```

```
<!ELEMENT TELEFONO (FISSO?, MOBILE?)>
```

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

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

<RUBRICA> è l'elemento radice del documento XML

```
<!ELEMENT FISSO (NUMERO+)>
```

```
<!ELEMENT MOBILE (NUMERO+)>
```

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

```
<!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
```

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

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

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

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

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

```
<!ATTLIST NUMERO
```

```
  prefisso CDATA #IMPLIED
```

```
  compagnia (wind | omnitel | tim | blu) #IMPLIED>
```

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)> **l'elemento <RUBRICA> può**
 <!ELEMENT FISSO (NUMERO+)> **contenere zero o più elementi**
 <!ELEMENT MOBILE (NUMERO+)> **<PERSONA>**
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)> **l'elemento <PERSONA> deve**
 <!ELEMENT COGNOME (#PCDATA)> **contenere gli elementi**
 <!ELEMENT FISSO (NUMERO+)> **<DATI_ANAGRAFICI> e**
 <!ELEMENT MOBILE (NUMERO+)> **<TELEFONO> in sequenza**
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <DATI_ANAGRAFICI>
 deve contenere una sequenza di
 zero o più elementi <NOME>, zero
 o più elementi <COGNOME>, zero
 o un elemento <INDIRIZZO>

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <TELEFONO> deve
 contenere una sequenza di zero o
 uno elementi <FISSO>, e zero o
 uno elementi <MOBILE>

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <NOME> deve contenere Parsed Character Data

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <COGNOME> deve contenere Parsed Character Data

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <FISSO> deve contenere uno o più elementi <NUMERO>

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <MOBILE> deve contenere uno o più elementi <NUMERO>

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <NUMERO> deve contenere Parsed Character Data

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

l'elemento <INDIRIZZO> deve contenere una sequenza di: zero o uno tra gli elementi <PIAZZA> o <VIA>, zero o uno elementi <N_CIVICO>, zero o uno elementi <CITTA>, zero o uno elementi <CAP>

In realtà in questa dichiarazione c'è un errore di sintassi: quale?

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

gli elementi <PIAZZA>, <VIA>, <N_CIVICO>, <CITTA>, <CAP> contengono Parsed Character Data

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**
 <!DOCTYPE RUBRICA [
 <!ELEMENT RUBRICA (PERSONA)*>
 <!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
 <!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
 <!ELEMENT TELEFONO (FISSO?, MOBILE?)>
 <!ELEMENT NOME (#PCDATA)>
 <!ELEMENT COGNOME (#PCDATA)>
 <!ELEMENT FISSO (NUMERO+)>
 <!ELEMENT MOBILE (NUMERO+)>
 <!ELEMENT NUMERO (#PCDATA)>
 <!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
 <!ELEMENT PIAZZA (#PCDATA)>
 <!ELEMENT VIA (#PCDATA)>
 <!ELEMENT N_CIVICO (#PCDATA)>
 <!ELEMENT CITTA (#PCDATA)>
 <!ELEMENT CAP (#PCDATA)>
 <!ATTLIST NUMERO
 prefisso CDATA #IMPLIED
 compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

E la dichiarazione <!ATTLIST> ?
 La vediamo adesso...

Riassumendo - Significato del DTD di rubrica.xml

- <RUBRICA> è l'elemento radice del documento XML
- l'elemento <RUBRICA> può contenere zero o più elementi <PERSONA>
- l'elemento <PERSONA> deve contenere gli elementi <DATI_ANAGRAFICI> e <TELEFONO> in sequenza
- l'elemento <DATI_ANAGRAFICI> deve contenere una sequenza di zero o più elementi <NOME>, zero o più elementi <COGNOME>, zero o un elemento <INDIRIZZO>
- l'elemento <TELEFONO> deve contenere una sequenza di zero o uno elementi <FISSO>, e zero o uno elementi <MOBILE>
- l'elemento <NOME> deve contenere Parsed Character Data
- l'elemento <COGNOME> deve contenere Parsed Character Data
- l'elemento <FISSO> deve contenere uno o più elementi <NUMERO>
- l'elemento <MOBILE> deve contenere uno o più elementi <NUMERO>
- l'elemento <NUMERO> deve contenere Parsed Character Data
- l'elemento <INDIRIZZO> deve contenere una sequenza di: zero o uno tra gli elementi <PIAZZA> o <VIA>, zero o uno elementi <N_CIVICO>, zero o uno elementi <CITTA>, zero o uno elementi <CAP>
- gli elementi <PIAZZA>, <VIA>, <N_CIVICO>, <CITTA>, <CAP> contengono Parsed Character Data

Attributi: dichiarazioni <!ATTLIST>

Abbiamo visto che la sintassi di un DTD si basa principalmente su **due dichiarazioni**:

- **<!ELEMENT>**

Definisce gli elementi utilizzabili nel documento XML e la **struttura** del documento

- **<!ATTLIST>**

Definisce la lista di attributi per ciascun elemento

Sintassi <!ATTLIST>

<!ATTLIST

nome-elemento

Il nome dell'elemento cui
l'attributo si riferisce

nome-attributo

tipo-attributo

Il tipo di valore che può avere

valore-di-default >

valore di default se non è
specificato nulla

- Esempio:

<!ATTLIST pagamento tipo CDATA "assegno">

Significa che l'elemento <pagamento> ha un attributo *tipo* che può avere come valore una qualsiasi combinazione di caratteri, e il cui valore di default è "assegno".

In XML diventa:

<pagamento tipo="assegno" />

Sintassi <!ATTLIST>

<!ATTLIST

nome-elemento

nome-attributo

tipo-attributo → i due più usati sono:

CDATA: il valore dell'attributo può essere una qualsiasi
combinazione di caratteri

(questo | quello): il valore dell'attributo può assumere solo
uno dei valori nella lista fra parentesi (divisi dal simbolo |)
quindi o "questo" o "quello" o "..."

valore-di-default

Sintassi <!ATTLIST>

<!ATTLIST

nome-elemento

nome-attributo

tipo-attributo

valore-di-default → può essere uno dei seguenti:

"valore-di-default" : valore che l'attributo assume se non ne è specificato un altro

#REQUIRED : l'attributo deve essere sempre dichiarato

#IMPLIED : l'attributo può anche non essere dichiarato

#FIXED valore : l'attributo è sempre presente e deve essere dichiarato con il valore fisso: "valore"

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

```
<!DOCTYPE RUBRICA [
```

```
<!ELEMENT RUBRICA (PERSONA)*>
```

```
<!ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
```

```
<!ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
```

```
<!ELEMENT TELEFONO (FISSO?, MOBILE?)>
```

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

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

```
<!ELEMENT FISSO (NUMERO+)>
```

```
<!ELEMENT MOBILE (NUMERO+)>
```

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

```
<!ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
```

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

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

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

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

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

```
<!ATTLIST NUMERO
```

```
  prefisso CDATA #IMPLIED
```

```
  compagnia (wind | omnitel | tim | blu) #IMPLIED>
```

DTD per la rubrica

L'elemento NUMERO ha l'attributo prefisso, che è opzionale e può avere come valore una qualsiasi combinazione di caratteri

<?xml version="1.0" standalone="yes" ?> **DTD per la rubrica**

```

<!DOCTYPE RUBRICA [
<ELEMENT RUBRICA (PERSONA)*>
<ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
<ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
<ELEMENT TELEFONO (FISSO?, MOBILE?)>
<ELEMENT NOME (#PCDATA)>
<ELEMENT COGNOME (#PCDATA)>
<ELEMENT FISSO (NUMERO+)>
<ELEMENT MOBILE (NUMERO+)>
<ELEMENT NUMERO (#PCDATA)>
<ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?)>
<ELEMENT PIAZZA (#PCDATA)>
<ELEMENT VIA (#PCDATA)>
<ELEMENT N_CIVICO (#PCDATA)>
<ELEMENT CITTA (#PCDATA)>
<ELEMENT CAP (#PCDATA)>
<!ATTLIST NUMERO
  prefisso CDATA #IMPLIED
  compagnia (wind | omnitel | tim | blu) #IMPLIED>
]>

```

L'elemento NUMERO ha un altro attributo compagnia, che è opzionale e può avere come valore o wind, oppure omnitel, oppure tim, oppure blu

Altri esempi <!ATTLIST>

```

<!ATTLIST paragrafo
  titolo CDATA #REQUIRED
  tipo (abstract | bibliografia | note ) #IMPLIED>

```

Qui si definiscono due attributi per l'elemento <paragrafo>:

- l'attributo *titolo*, obbligatorio, può contenere una qualsiasi combinazione di caratteri
- l'attributo *tipo*, opzionale, può assumere uno tra i valori *abstract*, *bibliografia*, oppure *note*

Altri esempi <!ATTLIST>

<!ATTLIST mittente azienda CDATA #FIXED "Microsoft">

Significa che l'elemento mittente ha un attributo *azienda*, il cui valore è sempre presente e deve essere dichiarato con la combinazione di caratteri *Microsoft*.

Nel file.xml diventerà:

```
<mittente azienda="Microsoft">
```

Altri esempi <!ATTLIST>

<!ATTLIST pagamento tipo (assegno | contanti) "contanti">

Significa che l'elemento pagamento ha un attributo *tipo*, il cui valore può essere *assegno* o *contanti*: se non è specificato nulla il valore di default sarà *contanti*.

Nel file.xml diventerà:

```
<pagamento tipo="contanti">
```

oppure

```
<pagamento tipo="assegno">
```

DTD per un articolo: articolo.xml

```
<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE articolo [
  <!ELEMENT articolo (paragrafo+)>
  <!ELEMENT paragrafo (immagine*, testo+, codice*)>
  <!ELEMENT immagine EMPTY>
  <!ELEMENT testo (#PCDATA)>
  <!ELEMENT codice (#PCDATA)>
  <!ATTLIST articolo titolo CDATA #REQUIRED>
  <!ATTLIST paragrafo
    titolo CDATA #IMPLIED
    tipo (abstract|bibliografia|note) #IMPLIED>
  <!ATTLIST immagine file CDATA #REQUIRED>
]>
```

Riassumendo - Significato del DTD di articolo.xml

- <articolo> è il tag radice del documento XML
- l'elemento <articolo> può contenere almeno un paragrafo
- l'elemento <paragrafo> contiene la sequenza di elementi <immagine> (0 o più), <testo> (almeno 1), <codice> (0 o più)
- l'elemento <immagine> è vuoto
- gli elementi <testo> e <codice> contengono PCDATA, ossia "parsed character data", cioè contenuto misto (testo o marcatori)
- l'elemento <articolo> contiene un attributo "titolo" (obbligatorio) che può avere come valore una qualsiasi combinazione di caratteri (CDATA - character data)
- l'elemento <paragrafo> contiene gli attributi "titolo" e "tipo": "titolo" è opzionale e contiene una combinazione di caratteri (CDATA - character data); "tipo" è opzionale e assume uno tra i valori "abstract", "bibliografia" o "note" (o "abstract", oppure "bibliografia", oppure "note")
- l'elemento <immagine> contiene un attributo "file" (obbligatorio), che contiene una sequenza qualsiasi di caratteri (CDATA - character data);

Un esempio più completo lettera.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE lettera [
.....
.....
]>

<lettera>
  <to>Pippo</to>
  <from>Nicola</from>
  <intestazione>Domanda: </intestazione>
  <corpo_domanda>Cosa fai stasera? </corpo_domanda>
</lettera>
```

Proviamo insieme a scrivere un DTD per questo XML
Esiste un solo possibile DTD per lettera.xml?

Un esempio più completo lettera.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE lettera [
<!ELEMENT lettera (to, from, intestazione, corpo_domanda)>
.....
.....
]>

<lettera>
  <to>Pippo</to>
  <from>Nicola</from>
  <intestazione>Domanda: </intestazione>
  <corpo_domanda>Cosa fai stasera? </corpo_domanda>
</lettera>
```

Un esempio più completo lettera.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE lettera [
<!ELEMENT lettera (to, from, intestazione, corpo_domanda)>
<!ELEMENT to (#PCDATA)>

]>
<lettera>
  <to>Pippo</to>
  <from>Nicola</from>
  <intestazione>Domanda: </intestazione>
  <corpo_domanda>Cosa fai stasera? </corpo_domanda>
</lettera>
```

Un esempio più completo lettera.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE lettera [
<!ELEMENT lettera (to, from, intestazione, corpo_domanda)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>

]>
<lettera>
  <to>Pippo</to>
  <from>Nicola</from>
  <intestazione>Domanda: </intestazione>
  <corpo_domanda>Cosa fai stasera? </corpo_domanda>
</lettera>
```

Un esempio più completo lettera.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE lettera [
<!ELEMENT lettera (to, from, intestazione, corpo_domanda)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT intestazione (#PCDATA)>
]>
<lettera>
  <to>Pippo</to>
  <from>Nicola</from>
  <intestazione>Domanda: </intestazione>
  <corpo_domanda>Cosa fai stasera? </corpo_domanda>
</lettera>
```

Un esempio più completo lettera.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE lettera [
<!ELEMENT lettera (to, from, intestazione, corpo_domanda)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT intestazione (#PCDATA)>
<!ELEMENT corpo_domanda (#PCDATA)>
]>
<lettera>
  <to>Pippo</to>
  <from>Nicola</from>
  <intestazione>Domanda: </intestazione>
  <corpo_domanda>Cosa fai stasera? </corpo_domanda>
</lettera>
```

Entità

Abbiamo visto che le entità sono sequenze di caratteri speciali dalla forma **&nome;**; XML permette di definire le proprie entità utilizzando la dichiarazione **<!ENTITY>**

○ A cosa servono? Per dichiarare dei **testi o delle iscrizioni fisse che vengono inserite ripetutamente** all'interno dei documenti.

- Esempio. Se devo scrivere "HyperText Markup Language" molte volte all'interno del mio documento XML (lungo!) lo posso sostituire con un'entità (breve):

```
<!ENTITY html "HyperText Markup Language">
```

In questo modo quando nel file XML userò l'entità **&html;** questa verrà sostituita con l'intera stringa

Entità

```
<!ENTITY autore "Dott. Ing. Arch. Prof.  
Leonardo Cosimo Scarpetta ">
```

Si richiama con **&autore;**

Documenti ben formati e anche validi

- Abbiamo visto che con XML possiamo **definire i tag a piacimento**
- Ma è necessario **definire una grammatica**: un insieme di regole che indicano quali *vocaboli* (elementi) possiamo usare per formare le *frasi* (documenti) → questa grammatica si definisce con il **DTD** – Document Type Definition
- **Un documento XML è *valido* se è ben formato e rispetta le regole del DTD associato.**
 - Cosa e' un DTD?
 - Come si validano i documenti? Lo vediamo ora...

Validare per imparare

- Per validare l'HTML
 - http://validator.w3.org/#validate_by_input
- Per validare l'XML
 - Lo apriamo col browser che ci dice se e' valido oppure
http://www.w3schools.com/xml/xml_validator.asp

Rubrica.xml

```
<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE RUBRICA [
<ELEMENT RUBRICA (PERSONA)*>
<ELEMENT PERSONA (DATI_ANAGRAFICI, TELEFONO)>
<ELEMENT DATI_ANAGRAFICI (NOME*, COGNOME*, INDIRIZZO?)>
<ELEMENT TELEFONO (FISSO?, MOBILE?)>
<ELEMENT NOME (#PCDATA)>
<ELEMENT COGNOME (#PCDATA)>
<ELEMENT FISSO (NUMERO+)>
<ELEMENT MOBILE (NUMERO+)>
<ELEMENT NUMERO (#PCDATA)>
<ELEMENT INDIRIZZO (PIAZZA | VIA)?, N_CIVICO?, CITTA?, CAP?>
<ELEMENT PIAZZA (#PCDATA)>
<ELEMENT VIA (#PCDATA)>
<ELEMENT N_CIVICO (#PCDATA)>
<ELEMENT CITTA (#PCDATA)>
<ELEMENT CAP (#PCDATA)>
<!ATTLIST NUMERO
  prefisso CDATA #IMPLIED
  compagnia (wind | omnitel | tim | blu ) #IMPLIED>
]>
```

Ricordate l'errore?
Proviamo a validare!

```
<RUBRICA>
<PERSONA>
<DATI_ANAGRAFICI>
  <NOME>Giulio</NOME>
  <COGNOME>Cesare</COGNOME>
  <INDIRIZZO>
    <PIAZZA>Bruto</PIAZZA><N_CIVICO>10</N_CIVICO><CITTA>Roma</CITTA><CAP>10100</CAP>
  </INDIRIZZO>
</DATI_ANAGRAFICI>
<TELEFONO>
  <FISSO><NUMERO prefisso="06">321654</NUMERO></FISSO>
  <MOBILE><NUMERO compagnia="blu">32557684785</NUMERO></MOBILE>
</TELEFONO>
</PERSONA>
<PERSONA>
<DATI_ANAGRAFICI>
  <NOME>Giovanni</NOME>
  <COGNOME>Pompeo</COGNOME>
  <INDIRIZZO>
    <PIAZZA>Catilina</PIAZZA><N_CIVICO>12</N_CIVICO><CITTA>Milano</CITTA><CAP>12100</CAP>
  </INDIRIZZO>
</DATI_ANAGRAFICI>
<TELEFONO>
  <FISSO><NUMERO prefisso="02">96877564</NUMERO></FISSO>
  <MOBILE><NUMERO compagnia="tim">3356545342</NUMERO></MOBILE>
</TELEFONO>
</PERSONA>
</RUBRICA>
```

