

LINGUAGGI DI MARKUP

Documenti su Internet

- Internet permette (tra l'altro) di accedere a documenti remoti
- In generale, i documenti acceduti via Internet sono **multimediali**, cioè che possono essere riprodotti **da diversi più mezzi** e costituiti da oggetti multimediali

Oggetti multimediali

- Un oggetto multimediale può essere
 - **Continuo** (audio, video)
 - **Discreto** (testo, immagini)
 - Hanno caratteristiche diverse e sono presentati con modalità diverse
- QUINDI**
- Necessità di
 - un linguaggio **uniforme** e **trasparente** per descrivere ed elaborare gli oggetti indipendentemente dalla loro natura
 - uniformare le modalità di rappresentazione di **un oggetto**

Architettura di un documento

- L'**architettura** di un documento descrive le connessioni tra i singoli elementi
- Tali relazioni vengono rappresentate tramite modelli
 - Modello di **Presentazione**
 - Modello di **Rappresentazione**
 - Modello di **Manipolazione**

Modelli

- **Modello di Manipolazione**: descrive le operazioni ammissibili per Creare, Modificare e Eliminare un documento
- **Modello di Rappresentazione**: descrive
 - *Protocolli* per lo scambio di informazione
 - *Formati* per la memorizzazione
 - Include le relazioni tra i singoli elementi necessarie durante la presentazione
- **Modello di Presentazione**: descrive la Presentazione del Documento

Struttura di un documento

- Un documento è realizzato mediante un **editor**
- Oltre al suo contenuto il documento può contenere informazioni aggiuntive per la presentazione/formattazione del contenuto
 - SGML
 - HTML
 - XML
 - LaTeX
 - Postscript
 - ...

Iper testi (1)

- Un Documento tradizionale (libro, articolo, ...) è presentato in maniera lineare (sequenziale)
- Un *Iper testo* ha invece la proprietà di poter essere **nonlineare**
- Formato da più testi collegati tra loro tramite **link**
- Può contenere informazioni multimediali di vario tipo
- **L'utente sceglie il modo di navigare nel documento**

Iper testi (2)

- Un **ipertesto** ha la struttura di un grafo
 - Nodi = *unità di informazione*
 - Testi
 - Immagini
 - Audio
 - ...
 - Archi = *collegamenti* (link) tra le varie unità di informazione

Iper testi (3)

- La possibilità di strutturare l'informazione tramite grafi permette di far scegliere all'utente il cammino da seguire
 - Possibilità di scendere nel dettaglio delle informazioni fornite
 - Approfondire solo certi particolari aspetti
 - Ritrovare in maniera più veloce le informazioni desiderate

Sistemi Hypermedia

- Sistemi composti di link di informazione non lineari
- Informazione presente in un nodo rappresentata attraverso *documenti multimediali*
- Generalizza e include Sistemi Iper testuali e Sistemi Multimediali

Markup (1)

- **Markup** è ogni mezzo per rendere esplicita una particolare interpretazione di un testo
- Ad es:
 - L'insieme dei simboli di interpunzione, che permettono di rendere leggibile il testo
 - I simboli di livello (•, -) e i colori usati in questi lucidi
 - Le formattazioni
 - ...

Markup (2)

- Si distingue Markup interno da quello esterno
- Markup **interno**
 - Inserisce **istruzioni di presentazione** all'interno del testo, tra le parole
 - Richiede sintassi particolari per **distinguere** il markup dal contenuto.
 - Tipicamente si adottano segnalatori particolari che cambiano il tipo di interpretazione del documento (*escaping*)

Markup (3)

- Markup **esterno**
 - Prevede due blocchi di informazioni: il contenuto e il markup, separati e collegati da meccanismi di indirazione
 - Richiede un meccanismo di indirazione, basato su indirizzi, *offset* o identificatori, per associare con correttezza il markup al contenuto

Linguaggi di Markup (1)

- Sono sistemi formali che permettono di scambiare e pubblicare informazioni in **formato testo** (documento) in componenti logiche (*elementi*), e di "etichettarle" in modo utile
- Le etichette (i *nomi* degli elementi) specificano il tipo di dato rappresentato da una componente logica
- Le etichette vengono inserite nel documento stesso come speciali sequenze di caratteri, dette **markup tag** o semplicemente **tag**

Linguaggi di Markup (2)

- Le regole del linguaggio stabiliscono
 - quali elementi (e tag) si possono usare in un documento
 - come i tag vengono distinti dal contenuto del documento
 - (eventualmente) il significato dei vari tag
- **NON** esiste **il** linguaggio di markup, esistono **diversi** linguaggi di markup per rappresentare diversi tipi di documenti

Esempio: documento senza markup (1)

Alessandro Bianchi

+39-080 544 2283

Dipartimento di Informatica

Piano 5

alessandro.bianchi@uniba.it

Segreteria Studenti

Dipartimento di Informatica

Piano -1

Esempio: documento senza markup (2)

- Il documento è suddiviso in due parti
 - Una riferita a Alessandro Bianchi
 - Una riferita a Segreteria Studenti
- Ogni parte rappresenta un **contatto**
- Ogni parte può a sua volta essere suddivisa in sottoparti
 - **nominativo** (obbligatorio)
 - **numero di telefono** (opzionale)
 - **edificio** (obbligatorio)
 - **piano** (obbligatorio)
 - **indirizzo email** (opzionale)

Esempio: documento con markup (1)

```
<contact><name>Alessandro Bianchi</name>
<tel>+39-080 544 2283</tel>
<building>Dipartimento di Informatica</building>
<floor>Piano 5</floor>
<email>alessandro.bianchi@uniba.it</email>
</contact>
<contact><name>Segreteria Studenti</name>
<building>Dipartimento di Informatica</building>
<floor>Piano -1</floor>
</contact>
```

Esempio: documento con markup (2)

- Il documento è composto da
 - contenuto (dati), in formato testo
 - markup tag, qui posti prima e dopo il contenuto a cui si riferiscono

Vantaggi dei documenti con markup

- Un programma può facilmente distinguere tra le componenti logiche di un documento con markup e trattarle in modo appropriato
- Lo scopo originale dei markup tag era facilitare la stampa di documenti, indipendentemente dalla stampante o dal programma di trattamento testi utilizzato
- Costituiscono un formato semplice per lo scambio di dati (in formato testo) che può essere compreso sia dal computer che dagli umani

Caratteristiche (1)

- I linguaggi di markup **NON** sono linguaggi di programmazione
 - I documenti scritti in un linguaggio di programmazione ricevono dati in input, li elaborano (secondo le istruzioni fornite), e producono un risultato utile per l'utente
 - I documenti con markup non elaborano informazione e non eseguono alcuna procedura. Sono un modo di rappresentare informazioni che possono essere facilmente utilizzate da un programma per vari scopi

Caratteristiche (2)

- Standard e aperti
 - ogni programmatore può scrivere un *parser*, ovvero un programma che legga e utilizzi il documento con markup
- Basati su standard per la codifica dei caratteri (principalmente UNICODE)
 - i documenti possono essere scambiati tra diversi dispositivi e supportano le diverse lingue