

Codifica di Testi - Introduzione Teorica a.a. 2023-2024

Angelo Mario Del Grosso

`angelo.delgrosso@ilc.cnr.it`

CNR-ILC

Istituto di Linguistica Computazionale “A. Zampolli”,
1st March 2024

Contenuto della lezione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

- 1 Elementi Teorici e Metodologici della codifica
- 2 Codifica del Testo
- 3 Modello del Testo
- 4 Importanza della Codifica
- 5 Codifica dei caratteri
- 6 Linguaggi di marcatura
- 7 Approfondiamo e concludiamo

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo, oggetto di studio

Le discipline umanistiche hanno come *oggetto di studio* privilegiato: **i testi**.

L'avvento delle tecnologie informatiche ha *cambiato la produzione, la trasmissione e lo studio del sapere* e della conoscenza (quindi anche il testo, in senso lato).

Il testo in ambiente digitale

Il nuovo formato digitale ha determinato **profonde trasformazioni nelle pratiche scientifiche** che hanno per oggetto o per fine lo studio, la produzione, la pubblicazione di testi.

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Vero e proprio cambio di paradigma

Gli studiosi si sono interrogati sul *nuovo rapporto che le discipline umanistiche devono avere con il testo* che diventa elettronico o digitale.

Nuove sfide

La natura del formato digitale ha favorito *molte ricerche* volte alla **risoluzione di nuovi problemi** connessi alla **rappresentazione formale dei testi**.

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Discussione nella comunità

L'informatica allo stato attuale è in grado di fornire allo studioso umanista **un ampio ecosistema di tecnologie** finalizzato all'implementazione di *efficaci sistemi di rappresentazione formale del testo* (anche conseguenti sistemi di elaborazione).

Nuove sfide

La discussione sulla rappresentazione digitale dei testi ha rivelato un intreccio di complesse **questioni teoriche, metodologiche e pratiche.**

Perché è importante la codifica dei testi

Motivazioni teoriche

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Scienze umane vs Scienze informatiche

Il rapporto tra sapere umanistico e informatica non è solo una questione meramente strumentale:

l'informatica non è solo un utensile dalle notevoli capacità.

Salto di paradigma sia da un punto di vista teorico sia metodologico.

Perché è importante la codifica dei testi

Motivazioni teoriche

La codifica come metodologia

L'attività di codifica diviene **metodo di analisi della conoscenza** nell'ambito delle discipline che si occupano della cura scientifica del testo.

La codifica come analisi teorica

Il linguaggio di codifica adottato può essere considerato come un **linguaggio teorico**:

Esplicitare e formalizzare le ipotesi interpretative su un certo oggetto di studio

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Definizione di codifica informatica dei testi

Fabio Ciotti 2007

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Definizione

Per **codifica** informatica dei testi intendiamo la *rappresentazione formale* di un **testo** ad un qualche livello descrittivo, su di un **supporto digitale**, in un formato utilizzabile da un elaboratore (*Machine Readable Form*) mediante un opportuno **linguaggio informatico** (*formalismo*).

Concetti alla base della codifica di testi

Gli elementi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

- Cosa significa Codifica
- Cosa significa Testo
- Cosa significa Codifica del Testo
- Cosa significa Linguaggio Informatico

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Codifica del testo

Elementi del modello della comunicazione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica e Decodifica

- Il termine codifica (in inglese *encoding*) può essere associato alla creazione di un testo
- Il termine decodifica (in inglese *decoding*) può essere associato alla comprensione e interpretazione di un testo
- Il messaggio può essere visto come contenitore di significato (\sim *testo*)

Codifica del testo

definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codice

Sistema di regole e di segni (*convenzione*) per convertire informazioni da una forma (anche astratta) in un'altra forma o rappresentazione **per la comunicazione tramite un canale.**

La comunicazione Umana

Modello della Comunicazione - C. Shannon e W. Weaver (1949)

The Mathematical Theory of Communication

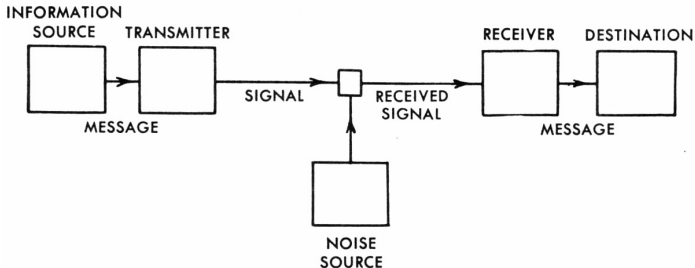


Fig. 1. — Schematic diagram of a general communication system.

La comunicazione Umana

Diasistema

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Messaggio

$M_e \neq M_d$ **Il Messaggio codificato è diverso dal messaggio decodificato**

La comunicazione Umana

Diasistema

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Messaggio

Il sistema di codifica e il sistema di decodifica sono due sottosistemi di uno stesso **diasistema**.

La comunicazione Umana

Diasistema

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Comunicazione come diasistema

Comunicazione = Sistema dell'emittente (S1) ~ Sistema del
destinatario (S2).

*Il risultato di un compromesso tra il sistema S1 e il sistema S2**

* Segre, 1985

Codifica del testo

Sinteticamente

La codifica digitale di un testo

la codifica testuale è la rappresentazione formale di un testo e delle sue caratteristiche mediante un linguaggio artificiale. (Ciotti)

Grammatica formale

Questo linguaggio consente di **descrivere esplicitamente in modo non ambiguo** le caratteristiche di un testo applicando **etichette meta-testuali** la cui funzione e struttura è descritta da **regole definite in una grammatica formale** (formalismo e meta-linguaggio).

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo

Linguaggio formale che diventa linguaggio teorico

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatatura

Approfondiamo

La codifica del testo come modello formale

La codifica digitale del testo diviene a questo punto un **linguaggio teorico formale** attraverso il quale lo studioso **costruisce modelli formali** del testo.

Linguaggi di Markup

Ai fini della modellizzazione di testi, tra i linguaggi formali si sono imposti quei sistemi basati sui cosiddetti **linguaggi di marcatura descrittivi** (*markup language*).

Codifica del testo

Sistema linguistico formalizzato

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Secondo Lou Burnard 1995

La codifica informatica di un testo impone l'**adozione di un sistema linguistico formalizzato** che permetta a uno studioso di **rendere esplicita una interpretazione di un testo** e le varie operazioni che la hanno prodotta.

I sistemi dichiarativi

I sistemi dichiarativi/descrittivi forniscono un potente dispositivo metalinguistico.

Capacità rappresentazionale

**Il principale requisito di uno schema di codifica, pertanto,
è la capacità rappresentazionale che esso offre allo
studioso**

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Modellare il testo

Il testo come oggetto del dominio di studio

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Informatica nelle scienze umane

L'informatica umanistica si occupa, tra le altre cose, **della rappresentazione e dell'elaborazione degli oggetti** d'interesse e di studio per le discipline umanistiche.

Testo come oggetto di studio

Il testo è, tra questi, l'oggetto più ricorrente e privilegiato

Modellare il testo

qual è la natura del testo?

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modellare il testo

Per trattare e rappresentare il testo in ambiente digitale,
bisogna **formulare un suo modello**

Modello del testo

Un efficace modello del testo non può prescindere da una
determinazione di **cosa si intenda per testo e quale sia la
sua natura.**

Modellare il testo

Testo: oggetto complesso

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

Il testo è un oggetto complesso in quanto è in grado di veicolare **significato su più livelli strutturali** (logico, ontologico, linguistico, autoriale, editoriale, fisico, ecc.), anche attraverso *l'instaurazione di molteplici relazioni tra più dimensioni.*

Modellare il testo

Testo: oggetto complesso

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ma..

**Non abbiamo a disposizione nessuna teoria completa e
condivisa del testo**

Modellare il testo

Tentativo..

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

Quale concezione o modello ontologico del testo è implicata nella rappresentazione informatica di esso?

A quali teorie possiamo rivolgerci?

Molteplici teorie del testo, quasi tutte sbilanciate sul livello verbale-semantico (*linguistica testuale*).

Modellare il testo

Tentativo..

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

Il termine “testo” si riferisce a un oggetto plurale, in cui co-esistono:

- **Un livello astratto:** la *sequenza verbale*, la quale a sua volta genera una serie di *livelli di contenuti semantici*.
- **Un livello materiale:** il supporto e le tracce d'inchiostro.
- **Un livello dinamico:** il testo viene **creato** da un autore e **ricreato** dal lettore.

Modellare il testo

Il testo: Cercando una definizione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo

- Un documento materiale composto da fogli che contengono inchiostro variamente disposti?
- Un discorso linguistico fissato tramite la scrittura su un documento materiale?
- Un'opera dell'ingegno che viene costituita e fissata tramite un insieme di simboli/segni?
- Il contenuto di un messaggio?

Modellare il testo

Il testo: Cercando una definizione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è il testo - uno sguardo più specialistico

- Lo stato linguistico di un singolo testimone materiale di un'opera?
- Lo stato linguistico di un medesimo testimone di un'opera che presenta diverse lezioni identificabili?
- Una versione edita di un'opera?
- Una sequenza coerente di enunciati in una lingua naturale?
- Uno scritto che può essere trasmesso anche oralmente?

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo

Dal latino *textum*, intreccio di unità di significato (**monemi**)
veicolato da simboli (**grafemi**).

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo secondo Segre 1985

Il testo è dunque una successione fissa di significati grafici.
Questi significati sono poi portatori di significati semantici.

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo secondo Tito Orlandi 2010

La produzione di un'attività linguistica intesa in senso stretto,
cioè in una delle lingue storicamente date

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ancora Tito Orlandi

Il Testo si presenta al lettore sotto la forma di
rappresentazione di un'espressione linguistica, e poiché
l'attività linguistica è multiforme, potrà essere una
rappresentazione puramente mentale, *astratta*, ovvero *fonica*
(verbale), ovvero *grafica* (scrittura).

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

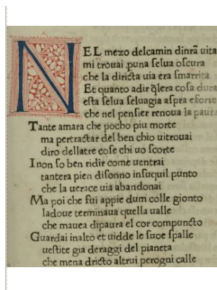
Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



Nel mezzo del cammin di nostra vita,
Mi ritrovai per una selva oscura,
Che la diritta via era smarrita.
E quanto a dir qual era è cosa dura
Questa selva selvaggia, ed aspra e forte,
Che nel pensier rinnova la paura!
Tanto è amara che poco è più morte;
Ma, per trattar del ben ch'i' vi trovai,
Dirò dell'altre cose ch'i' v'ho scorte.
I' non so ben ridir com'i' v'entrai;
Tant'era pien di sonno in su quel punto,
Che la verace via abbandonai.
Ma po' ch'i' fui appiè d'un colle giunto,
Là ove terminava quella valle,
Che m'avea di paura il cor compunto;
Guardai in alto, e vidi le sue spalle
Vestite già de' raggi del pianeta
Che mena dritto altrui per ogni calle.

Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura,
ché la diritta via era smarrita.
Ah! quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinnova la paura!
Tant'è amara che poco è più morte;
ma per trattar del ben ch'i' vi trovai,
dirò de l'altre cose ch'i' v'ho scorte.
Io non so ben ridir com'i' v'entrai,
tant' era pien di sonno a quel punto
che la verace via abbandonai.
Ma poi ch'i' fui al piè d'un colle giunto,
là dove terminava quella valle
che m'avea di paura il cor compunto,
guardai in alto e vidi le sue spalle
vestite già de' raggi del pianeta
che mena dritto altrui per ogni calle.

Possiamo concludere che

Il testo è l'invariante, la successione di valori invariabili
rispetto alle variabili materiali dei caratteri, della scrittura
(*principio di identità*).

Modellare il testo

Complessità del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo oggetto complesso multidimensionale/pluralista

- analisi grafematica (caratteri, ideogrammi)
- analisi strutturale (struttura editoriale, interna)
- analisi metrica (piedi, versi, stanze, strofe)
- analisi stilistica e retorica
- analisi tematica
- analisi semantica
- ecc..

Modellare il testo

Il documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il Testo non è il Documento

La parola **documento** deriva dalla parola latina "*documentum*" che ha la stessa radice di "*doceo*" che significa "*insegnare*" / "*dimostrare*" con l'aggiunta del suffisso "*-umentum*" che intende **lo strumento per fare**.

Il Documento

La parola "*documento*" designa quindi uno **strumento** per **insegnare** o uno strumento atto a **dimostrare qualcosa**.

Modellare il testo

Il documento: Cercando una definizione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Cos'è un documento

- Una prova o una evidenza?
- Un atto cartaceo originale e/o ufficiale attraverso il quale dimostrare o provare qualcosa?
- Qualsiasi cosa materiale che serva come evidenza o prova?
- Un qualcosa di scritto contenente informazioni?
- Un supporto materiale sul quale sono stati rappresentati fatti o pensieri attraverso un qualche tipo di convenzione fatta di simboli e segni?

Modellare il testo

Definizioni

Testo

Il testo è una entità astratta invariante, che in ogni operazione di realizzazione materiale della sequenza di simboli grafici determina la struttura fisica di un oggetto sensibilmente concreto (ovvero capace di attivare uno dei canali recettivi dell'uomo verso stimoli esterni).

Documento

L'oggetto materiale, sensibile, concreto che **costituisce il supporto**, stabile e riproducibile dell'informazione testuale che determina una modalità d'interazione (lettura, decifrazione).

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modellare il testo

Codifica del documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Il testo non è il documento

La codifica di un documento è diversa dalla codifica di un testo
(**testo != documento**).

Documento digitalizzato

Corrispondenza perfetta con l'originale: **documento
digitalizzato**

Modellare il testo

Struttura del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Elementi di un testo

Oltre alla sequenza di grafemi, **un testo si presenta strutturato in segmentazioni logiche** e partizioni interne di interi blocchi

Struttura di un testo

Il testo ha una certa struttura, i cui **elementi sono determinati dalla struttura logico-semantic del discorso** con *finalità* e *funzioni* ben distinte (titolo di un capitolo, corpo di un paragrafo)

Modellare il testo

Struttura del documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Struttura di un documento

Un documento, ugualmente, è costituito da **elementi significativi, non solo verbali, da esplicitare formalmente** (layout, segmentazione, foliazione, aspetti paratestuali, aspetti tipografici, aspetti topografici).

Modellare il testo

Ordered Hierarchy of Content Objects (OHCO)

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La natura del testo come OHCO

Gli oggetti di contenuto testuale sono perlopiù le **strutture editoriali astratte** di cui si compone un testo (divisioni logiche, segmentazioni e partizioni di un libro).

La natura del testo come OHCO

Questi oggetti sono **gerarchici** e di **contenuto** in quanto tali elementi ne possono contenere altri, ma sempre entità che veicolano contenuto. Sono **ordinati** in quanto esiste una *relazione lineare tra due oggetti posti sul medesimo livello gerarchico* (pensiamo alle parole di una frase all'interno di un paragrafo).

Modellare il testo

OHCO come primo modello concettuale TEI

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

OHCO e TEI

Il primo modello implementato dalla *Text Encoding Initiative* si è basato sull'**impianto teorico OHCO**.

OHCO e TEI

Sono però stati riscontrati una serie di *problemi di rappresentazione ed espressività*: esistono diversi insiemi di **elementi di contenuto che non possono essere ricondotti a una struttura gerarchica unitaria** (pensiamo alla rappresentazione di parole e di righe).

Modellare il testo

Prospettive analitiche

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Strutture sovrapposte

Gli elementi di contenuto che si sovrappongono si comportano come se appartenessero a **diverse gerarchie di oggetti testuali**.

Prospettiva analitica

Ogni prospettiva analitica su un testo determina una struttura gerarchica di oggetti di contenuto. (F. Ciotti)

Modellare il testo

Sottoprospective

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Sottoprospective analitiche

In questa ottica **il testo diventa un sistema multi-dimensionale, ciascuna dimensione corrisponde a specifici punti di vista** dell'osservatore. (F. Ciotti)

Modellare il testo

La struttura gerarchica del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Le gerarchie del testo

Le strutture del testo sono strutture gerarchiche: le sottoprospective, comunque esse siano definite, danno luogo a gerarchie.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Digitalizzare e codificare un testo

Digitalizzare e codificare per favorire la **conservazione**,
l'accesso, elaborazione e il trattamento automatico dei
testi

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Trattamento dei testi

- analisi di tipo linguistico (linguistica computazionale, database testuali, corpora linguistics)
- analisi di altro tipo (metrica, stilistica, ecc.)
- ricerca testuale avanzata
- pubblicazione in vari formati (sul Web, come ebook, a stampa)
- didattica

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Digitalizzare un testo

Per facilitare e garantire una **universalità di accesso** al loro contenuto

Vantaggi della digitalizzazione

- Edizioni elettroniche garantiscono **diffusione** capillare (via Web) e nuove **funzionalità** (ipertesti, ricerca, ecc.)
- Permettono anche di **preservare** i documenti più **antichi** (e fragili) riducendone la *consultazione diretta*

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Superare i problemi dei documenti digitali

- disponibilità di hardware e software
- sistemi proprietari chiusi
- elevata obsolescenza e limitata manutenibilità
- difficile portabilità su piattaforme diverse

Importanza della codifica digitale

Perché effettuare la codifica

Massimizzare le seguenti proprietà: Portabilità

- **indipendenza dall'hardware:** processore, supporto, output
- **indipendenza dal software:** sistemi operativi, applicazioni di authoring, applicazioni di visualizzazione
- **indipendenza dai sistemi di **codifica dei caratteri****
- **indipendenza logica:** da un particolare processo applicativo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Portabilità

Un documento digitale (come qualsiasi risorsa informativa digitale) è portabile se è **interscambiabile tra sistemi diversi**, **riutilizzabile** in molteplici processi computazionali anche a distanza di tempo, e **integrabile** da ulteriori risorse informative omogenee

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Standard

Le modalità di codifica devono divenire uno standard per **favorire la portabilità** del documento digitale.

Standard

Insieme di norme relative a una particolare tecnologia emesse da un **ente istituzionale** nazionale o internazionale deputato a tale scopo (*ISO - International Standardization Organization*).

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Standardizzare: vantaggi

- produrre **risultati confrontabili e condivisibili** all'interno di una comunità che adotta lo stesso tipo di standard
- lavorare in modo tale che i **prodotti** di oggi siano **usfruibili** anche in **futuro**
- codificare testi che possano essere usufruiti anche da chi non ha il tipo di **piattaforma hardware e software** che ha prodotto la codifica

Importanza della codifica digitale

Perché effettuare la codifica

Linguaggio di codifica del contenuto

Il sistema di codifica deve assistere nel processo di **rappresentazione del testo** focalizzando l'attenzione sul **contenuto** (o sulla struttura del contenuto) piuttosto che sulla sua *forma grafica*.

Linguaggio di codifica dichiarativo

I sistemi di codifica dichiarativa si sono rivelati estremamente efficaci nel rappresentare strutture complesse come **referimenti incrociati** e **collegamenti tra elementi** all'interno di un testo, ma anche tra più testi.

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio di codifica dichiarativo - Markup

I sistemi di markup dichiarativo introducono consistenti vantaggi poiché un medesimo **schema di codifica** dichiarativo che corrisponde ad un **modello di codifica** può essere utilizzato in *molteplici forme di trattamento informatico*.

markup dichiarativo: annotano la struttura e il significato degli elementi costitutivi del testo, tralasciandone l'aspetto.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Descrizione formale del testo

Così è possibile descrivere formalmente le caratteristiche di un testo in modo **indipendente** da particolari **finalità** di trattamento da contingenti **forme di presentazione** grafica su un qualsivoglia **supporto fisico**.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio di codifica come linguaggio teorico

In questo modo la codifica permette allo studioso di **esplicitare le sue ipotesi interpretative.**

Linguaggio di codifica dichiarativo flessibile

I linguaggi di markup dichiarativi permettono di **predicare** l'appartenenza di un dato segmento testuale a una **classe di strutture testuali** definita dall'utente.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Gerarchia ordinata di oggetti di contenuto (OHCO)

I linguaggi di markup dichiarativi sono particolarmente **efficienti** dal punto di vista computazionale se modellano una **struttura gerarchica**.

Linguaggi di codifica gerarchici

La **sintassi** del linguaggio di codifica gerarchico può essere usata per **rappresentare le relazioni tra gli elementi** strutturali di un testo, a un determinato livello di descrizione.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Gerarchia ordinata di oggetti di contenuto (OHCO)

I linguaggi di markup dichiarativi gerarchici, in particolare **SGML prima e XML poi**, si sono rivelati ottimi strumenti di supporto alla **codifica e all'analisi computazionale dei testi**.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Elementi di Codifica dei Caratteri

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Rappresentare il testo in formato digitale

L'adozione di metodologie informatiche per il trattamento dei testi richiede in primo luogo la disponibilità di un'adeguata **rappresentazione dei dati testuali in formato digitale.**

Elementi di Codifica dei Caratteri

Problemi di rappresentazione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

A.M. Del Grosso

Codifica di Testi - Introduzione Teorica a.a. 2023-2024

64/ 113



lasciato bianca 42r e ha scritto ancora su 43r
capovolgendo

¶ Nous commençons par dire que τ était la "forme faible" de τ

- 1° À quoi se reconnaît un sonantiste?
- 2° nous pouvons examiner les objections contre la conception courante
3. or τ \rightarrow τ Schmidt p. 50.
4. C'est d'abord contre et séparément
l'existence de preuves historiques d'un τ ,
puis d'un n qui se tourne.
- 5° Seule indication p. 3: "Es ist nun keineswegs gleichgültig [sic], ob man etc. ..."

Si quelque'un sent après cela quel principe décisif est opposé à un autre, quelle vue typique on trouve chez les sonantistes ou chez les adversaires, je consens à ce que ma remarque soit fautive. Et qu'on ne s'y trompe pas c'est M. Schmidt ici qui a raison, (du moins ce n'est pas lui qui a le premier tort.)

Si quelque part les sonantistes, tout en menant grand bruit de leur théorie, n'avaient songé à nous la formuler et à nous éclairer par là en quoi elle constitue une théorie on ne verrait pas réduit à formuler ce qu'il combat d'une

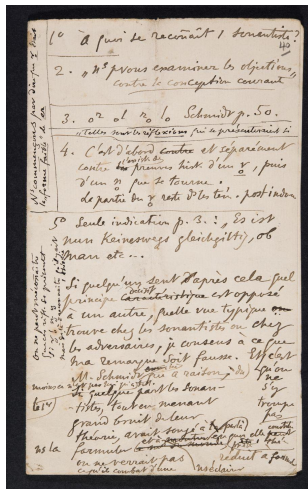
¶ On ne peut méconnaître que ces réflexions se présentent si τ τ n n τ [...] mais de là à reconnaître qu'il y ait une théorie

¶ Si ceux-ci avaient réellement médité la nature du τ en phonologie, il se serait, selon nous, non-seulement abstenus [...] mais positivement prononcés en

¶ Sur la marge droite, écrit traverso: Nous commençons... de τ

- contre
- τ \rightarrow l'existence de τ
- caractéristique \rightarrow décisif
- qu'on ne s'y trompe pas
- ici
- du moins ce n'est pas lui qui a le premier tort
- qu'on ne se trompe pas
- (montrer) \rightarrow nous éclairer par là

¶ Sur la marge gauche, écrit traverso: On ne peut... qu'il y ait une théorie



Elementi di Codifica dei Caratteri

Definizioni

Perché è importante la codifica dei caratteri

La codifica dei caratteri costituisce il **grado zero** (basso livello) della rappresentazione di testi su supporto digitale.

Le codifiche dei caratteri sono la base di qualsiasi schema di codifica testuale.

Rappresentazione digitale dei caratteri

I caratteri vengono rappresentati all'interno di un elaboratore mediante una sequenza di codici binari costituiti da opportune disposizioni di cifre composte da 0 e 1: 01100001 *lettera a*

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Elementi di Codifica dei Caratteri

American Standard Code for Information Interchange

Tabella Code Page ASCII 7 bit

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

7 bit = 128 possibili caratteri; 32 caratteri di controllo; 96 caratteri effettivi

Elementi di Codifica dei Caratteri

Esempio codifica binaria

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

codifica *ciao mondo!* 7 bit ASCII

6369 616F 206D 6F6E 646F 210A

codifica *ciao è mondo!* 8 bit ASCII

6369 616f 20e8 206d 6f6e 646f 210a

codifica **ciao è mondo!** UNICODE UTF-8

6369 616f 20c3 a820 6d6f 6e64 6f21 0a

Elementi di Codifica dei Caratteri

Definizioni

Character Set Per le discipline che studiano i sistemi di scrittura e l'analisi del linguaggio naturale, un **insieme di caratteri astratti** è detto Character set (unità alfabetiche). Astratto perché non riguarda la rappresentazione materiale della forma sul supporto, ma è relativo alla **forma mentale**, fatta di simboli di codifica (referenti).

Coded Char Set Per poter trattare un insieme di unità alfabetiche in formato digitale bisogna assegnare a ciascun carattere **un numero intero non negativo detto code point**.

Elementi di Codifica dei Caratteri

Definizioni

Character Encoding Il fine ultimo della codifica è quello di rappresentare una sequenza di caratteri in una **sequenza di byte**. La codifica di un carattere utilizza uno “encoding schema” che a sua volta mappa o *trasforma ciascun code point in una sequenza di byte* e quindi in ultima istanza in una sequenza di bit.

Tabella del code page Generalmente i code points sono espressi attraverso un sistema numerico esadecimale e disposti in una tabella di associazione.

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Elementi di Codifica dei Caratteri

In sintesi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica dei caratteri

Quindi trasformare una sequenza di caratteri appartenenti ad un char set in una sequenza di byte (bit) significa prima di tutto trasformare/mappare ciascun carattere nel proprio corrispettivo code point e successivamente codificare/serializzare questo code point nella relativa sequenza di byte (bit).

Elementi di Codifica dei Caratteri

Complessità e rappresentazione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

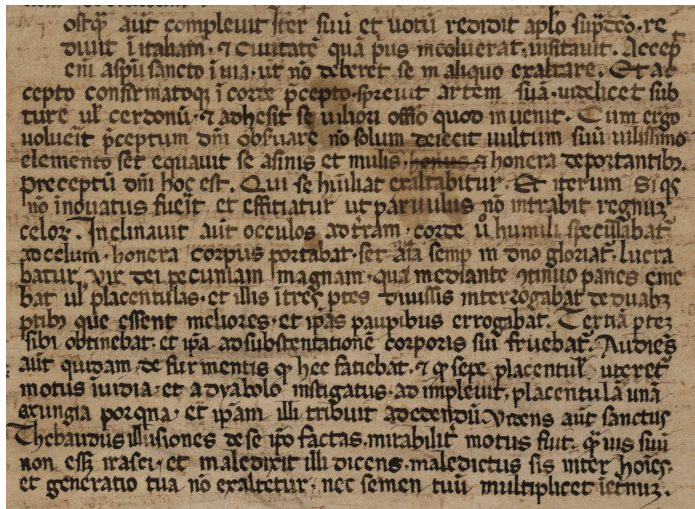
Approfondiamo

Complessità di rappresentazione universale dei caratteri

Se si considerano tutti i possibili alfabeti del mondo e le molteplici esigenze poste dalla scrittura delle fonti manoscritte antiche e medievali, ci si accorge che la realizzazione di un sistema universale per la codifica dei caratteri è un progetto molto complesso con svariate sfide da affrontare.

Complessità della Codifica dei Caratteri

Un Esempio



Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

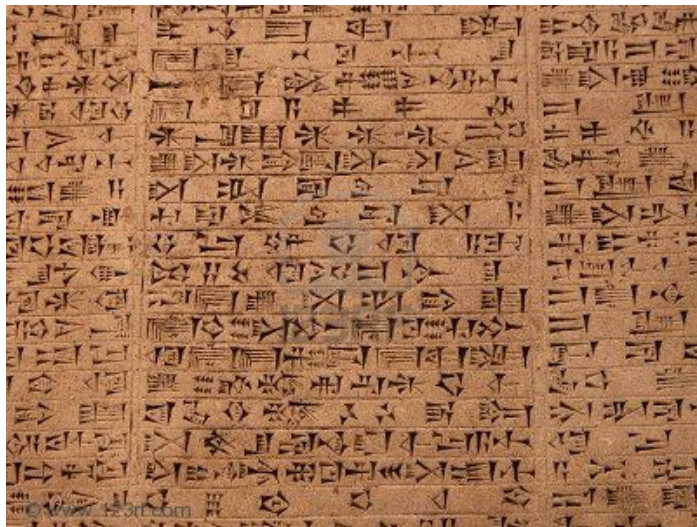
Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Complessità della Codifica dei Caratteri

Un Esempio



Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Elementi di Codifica dei Caratteri

Unicode

Complessità di rappresentazione universale

Ad oggi, lo standard de facto per la codifica dei caratteri è lo **UNICODE**. Esso è in grado di codificare **più di un milione di differenti unità alfabetiche**, segni di interpunzione e diacritici, appartenenti a centinaia di diverse lingue.

Complessità di rappresentazione universale

Unicode assegna i propri code point in un range che va da $0x0$ a $0x10FFFF$. In Unicode il code point viene indicato con una "U" seguita da un segno "+" seguito a sua volta dall'esadecimale con padding del codice (es: U+0041 lettera A).

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Elementi di Codifica dei Caratteri

Unicode

Unicode Transformation Format

Lo Unicode è un Coded Char Set e per essere concretamente serializzato su un supporto elettronico deve essere trasformato attraverso qualche tipo di schema di codifica. L'UTF (Unicode Transformation Format) mappa i code point Unicode in sequenze di byte (bit).

UTF standards

Esistono tre tipi di schemi di codifica che vanno sotto il nome di UTF, ciascuno è identificato dal minimo numero di bit necessario a codificare ciascun code point: UTF-8; UTF-16; UTF-32.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Markup language e XML

soluzione corrente per la codifica dei testi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

XML per la descrizione e la codifica

Ad oggi la soluzione considerata ottimale per una corretta rappresentazione del testo è l'adozione dei markup language descrittivi basati su XML.

TEI-XML

Standard de facto per la codifica dei testi è considerato lo schema XML messo a punto dalla Text Encoding Initiative (TEI-XML).

Lo schema TEI-XML

Estratto di documento TEI-XML

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

```
1 <?xml version="1.0" encoding="utf-8" standalone="yes"?>
2 <!DOCTYPE div SYSTEM "divTEIsnippet.dtd">
3 <divTEIsnippet>
4   <div type="destination">
5     <ab>
6       <stamp type="postmark">
7         <placeName>El Paso</placeName>
8         - TX 799 -
9         <date notBefore="1980-07-26">
10          <unclear>PM JUL</unclear>
11        </date>
12      </stamp>
13      <stamp type="postage">
14        Profil masculin, avec un avion et un radar au second
15        plan:
16        <mentioned>US Airmail 21 c.</mentioned>
17      </stamp>
18    </ab>
19    <ab>
20      <address>
21        <addrLine>
22          Madame
23          <name>Lefrère</name>
24        </addrLine>
25        <addrLine>4, allée George Rouault</addrLine>
26        <addrLine>75020 Paris</addrLine>
27        <addrLine>France</addrLine>
28      </address>
29    </ab>
30  </div>
31 </divTEIsnippet>
```

Perché TEI

La Text Encoding Initiative (TEI) è un autorevole progetto internazionale, a cui afferiscono varie organizzazioni e università, il cui scopo è fornire agli studiosi di informatica umanistica uno strumento il più espressivo e flessibile possibile per rappresentare qualsiasi aspetto di interesse relativo alla risorsa testuale da rappresentare digitalmente.

Perché XML

- separazione dei dati dall'applicativo di authoring/editing
- separazione della rappresentazione dei dati dalla presentazione dei dati
- possibilità di trasformare i dati in qualsiasi altro formato compatibile
- leggibilità dei documenti XML da parte di esseri umani

Perché XML

- standard w3c testuale, aperto, personalizzabile e liberamente utilizzabile
- semplicità di condivisione e scambio dati (interoperabilità e portabilità)
- adatto per codificare dati semistrutturati oltre che dati strutturati
- validazione del documento attraverso grammatiche formali

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ecosistema XML

Le tecnologie XML per la definizione ed elaborazione di documenti XML

- XSD: XML Schema Definition Language
- XPath: XML Path Language
- XSL: eXtensible Stylesheet Language
- XSL-T: XSL – Transformations
- XSL-FO: XSL – Formatting Objects
- XQuery: XML Query Language for XML Databases
- XInclude: XML inclusion Language
- DTD: Document Type Definition Language
- RelaxNG: Regular Expression Language for XML (New Generation)

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatore

Approfondiamo

Linguaggio di marcatura XML

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Perché XML

Adottando la tecnologia e il linguaggio XML abbiamo la possibilità di creare linguaggi di marcatura personalizzati e specifici per ogni esigenza e dominio.

XML come linguaggio per la codifica di testi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Vantaggi

Attraverso XML è possibile strutturare i dati, gestire in modo nativo strutture gerarchiche, elaborare e presentare i dati con strumenti XML nativi, validare i tipi di strutture e i tipi di dati, gestire riferimenti incrociati tramite opportuni meccanismi di dereferenziazione, aggiungere e gestire annotazioni a vari livelli di granularità.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Codifica dei caratteri

6 Linguaggi di marcatura

7 Approfondiamo e concludiamo

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo

L'applicazione di metodologie computazionali nell'ambito della ricerca umanistica comporta due aspetti di formalizzazione

Due elementi

Formalizzazione dei dati e formalizzazione dell'elaborazione

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Formalizzazione dei dati

Definizione e implementazione di strutture dati adeguate alla cattura dei fenomeni di interesse dell'umanista, e in particolare alla rappresentazione formale dei testi;

Formalizzazione dell'elaborazione

Specificazione di algoritmi che, applicati alle strutture dati, siano in grado di simulare i processi di manipolazione dei testi tipici della ricerca umanistica o in generale delle pratiche sociali che hanno a che fare in vario modo con i testi.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo

Il problema della codifica testuale rientra in generale nel primo tipo di formalizzazione (Dati).

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Problemi teorici

La codifica è un processo assai più complesso della semplice e meccanica correlazione biunivoca di strutture rappresentazionali.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Norme TEI

Le linee guida di codifica TEI impongono a chi codifica di effettuare delle scelte teoriche e interpretative che non sono imputabili alla semplice trascrizione.

Processo di codifica

La codifica è un processo interpretativo non solo un processo riproduttivo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica come interpretazione

Conseguentemente ogni processo di codifica (inclusi quelli di codifica informatica del testo) è il risultato di una interpretazione da esplicitare e formalizzare.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Esempio

L'assunzione che una data traccia grafica "A" sia una istanza di una data classe astratta di tracce che identifichiamo come il carattere "a" nella sua forma maiuscola richiede molti sforzi interpretativi e quindi cognitivi.

Rappresentazione e interpretazione

In linea di principio non è sempre possibile dire in modo non ambiguo che una traccia su un supporto fisico appartiene a una certa classe di iscrizioni che chiamiamo carattere.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Certezza e soggettività

Ogni interpretazione può godere di diversi gradi di certezza e di soggettività. In ogni caso non esiste nessun genere di rappresentazione di un testo che si possa definire libera da processi interpretativi.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Linguaggio teorico

Lo schema di codifica è un linguaggio teorico usato per costruire teorie o modelli di fenomeni testuali

Linguaggio teorico

La stessa costruzione di un linguaggio teorico riflette un determinato modello del mondo (soggettivo o condiviso).
Si presuppone una teoria ontologica del testo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Obiettivo

- Sviluppare teorie e modelli formali del testo (o di alcuni suoi livelli descrittivi)
- Individuare formalismi atti a esprimerli in modo computazionalmente accettabile

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

OHCO

Il modello di struttura gerarchica del testo si sposa perfettamente con XML, il quale può essere considerato sia un formalismo sia un modello di dati espresso da quel formalismo, e tale meta-modello è appunto un albero ordinato etichettato.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

OHCO: modello di dati così vincolante

Il modello OHCO presenta dei vincoli, che vanno sotto il nome di “problema delle gerarchie sovrapposte” (overlapping hierarchies).

Gerarchie Sovrapposte

Dati due oggetti logici presenti in un testo, le coppie di tag bilanciati che li rappresentano non si annidano propriamente ma si sovrappongono. *Tale situazione è sintatticamente e semanticamente vietata in XML.*

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modelli gerarchici

Vantaggio di poter validare in modo automatico ogni istanza di dati rispetto al modello mediante algoritmi generali ben conosciuti e computazionalmente trattabili, ciò che a sua volta consente di costruire sistemi di elaborazione degli stessi dati consistenti ed efficaci (al netto dei costrutti ID/IDREF).

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Modelli non gerarchici

Negli ultimi anni la ricerca teorica sullo sviluppo di sistemi di markup non gerarchici ha avuto un notevole stimolo

Modelli non gerarchici

Tutte le soluzioni proposte finora hanno grossi limiti per quel che concerne la facilità di gestione e manutenzione

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Ricapitolando

Il trasferimento del testo su supporto informatico pone allo studioso una serie di **quesiti teorici** (oltre che a numerosi problemi pratici) a partire dalla determinazione di quale particolare oggetto del mondo considerare come “**fonte**” della codifica e memorizzazione.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

In conclusione

Secondo la teoria OHCO un testo è una struttura gerarchica di oggetti logici, e la codifica non fa altro che esplicitare questa sua struttura essenziale.

In conclusione

Individuazione di diverse unità o elementi di contenuto testuale a partire da un punto di vista interno alle discipline che studiano ed analizzano il testo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Punti di vista

TEI ha una struttura modulare, in cui ogni modulo corrisponde alla rappresentazione di un determinato punto di vista metodologico sul testo.

TEI e Gerarchie sovrapposte

L'impianto finale dello schema di codifica della Text Encoding Initiative ha accolto in parte la nozione di sottoprospettiva ed ha sviluppato una serie di costrutti sintattici e semantici in grado di rappresentare adeguatamente fenomeni di sovrapposizione e di parallelismo tra elementi testuali in XML.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

TEI e Gerarchie sovrapposte

La TEI ha sviluppato un sistema di codifica sufficientemente potente da permettere di rappresentare le caratteristiche testuali che interessano e di poter esplicitare le sue interpretazioni sul ruolo di tali caratteristiche.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Scopo applicativo

Occorre tenere presente nella rappresentazione del testo anche i possibili processi applicativi a cui esso può essere sottoposto.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

La codifica informatica di un testo

La codifica informatica di un testo può essere vista come il prodotto di un insieme di inferenze che vengono espresse mediante un linguaggio formalizzato (Ciotti).

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo: riassumendo

- elaborazione di un quadro teorico di riferimento entro cui collocare i procedimenti analitici
- definizione di un modello di rappresentazione informatica o codifica del testo e delle strutture rilevanti in relazione al testo di riferimento
- individuazione di metodi e processi di analisi testuale applicabili al modello del testo e loro definizione sottoforma di procedure formali o algoritmi

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo: riassumendo

- implementazione del modello di rappresentazione e dei processi di analisi mediante adeguati linguaggi informatici
- applicazione delle procedure informatiche al testo digitalizzato
- analisi e interpretazione critica dei risultati

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Codifica del testo: riassumendo

Si tratta di costruire un modello informatico del testo nel quadro di un determinato contesto teorico, per poi interrogare opportunamente tale modello e avanzare ipotesi interpretative sul testo.

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Non è codifica di un testo

- fare una scansione di un documento e diffonderne le immagini (digitalizzazione)
- usare un software OCR per ricavarne una versione in formato testo
- creare una pagina html di un testo
- creare un documento con software di word processing
- creare banche dati con testo destrutturato

Approfondimenti e Conclusioni

per comprendere la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo

Fare la codifica di un testo significa

- convertirlo in un formato comprensibile al calcolatore elettronico
- usare, a tal fine, un linguaggio di codifica di tipo formale
- stabilire le caratteristiche del testo che si intendono esplicitare (modello di codifica)
- definire e seguire uno schema di codifica ben preciso

References

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



Ciotti, F. (2007). Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari. Aracne.



Lenci, A., Montemagni S., and Pirrelli V. (2016). Testo e Computer. Elementi Di Linguistica Computazionale. Aulamagna. Carocci.



Pierazzo, E. (2015). Digital Scholarly Editing : Theories, Models and Methods. Farnham Surrey: Ashgate.



Orlandi, T. (2010). Informatica testuale: teoria e prassi. Laterza.



Driscoll, M. J., and Pierazzo, E. (Eds.). (2016). Digital Scholarly Editing: Theories and Practices (Vol. 4). Open Book Publishers.



Segre, C. (1985). Avviamento all'analisi del testo letterario. Einaudi.



Segre, C. (1979). Critica testuale, teoria degli insiemi e diasistema, in Semiotica filologica. Einaudi.

References

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



DeRose, S. J., Durand, D. G., Mylonas, E., and Renear, A. H. (1990). What is text, really? *Journal of Computing in Higher Education*, 1(2), 3–26.



Renear, Allen H.; Mylonas, Elli; Durand, David (1995). Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies.



Ciotti F., e Crupi G, a c. di. (2012). Dall'Informatica umanistica alle culture digitali. ROMA : Casa Editrice Università La Sapienza. open access: <http://www.editricesapienza.it/node/7688>

References

Codifica di
Testi -
Introduzione
Teorica
a.a.
2023-2024

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Codifica dei
caratteri

Linguaggi di
marcatura

Approfondiamo



TEI Guide Reference. <http://www.tei-c.org/>