

Codifica di Testi - Introduzione Teorica a.a. 2018-2019

Angelo Mario Del Grosso

`angelo.delgrosso@ilc.cnr.it`

CNR-ILC-LicoLab

Istituto di Linguistica Computazionale “A. Zampolli”,
27th September 2018

Contenuto della lezione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Linguaggi di marcatura

6 Approfondiamo e concludiamo

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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).

Il testo in ambiente digitale

Il nuovo formato digitale ha determinato profonde trasformazioni nelle pratiche scientifiche che hanno per oggetto o fine i testi.

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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 aperto nuove sfide al fine di determinare e trovare soluzioni ai problemi connessi alla rappresentazione formale dei testi.

Codifica e Modellizzazione

Problemi e opportunità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Discussione nella comunità

L'informatica allo stato attuale è in grado di fornire allo studioso di scienze umane una serie di tecnologie per implementare dei sistemi di rappresentazione testuale.

Nuove sfide

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

Definizione di codifica informatica dei testi

Fabio Ciotti 2007

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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**

Concetti alla base della codifica di testi

Gli elementi

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcaturo

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Linguaggi di marcatura

6 Approfondiamo e concludiamo

La comunicazione Umana

per comprendere la codifica bisogna partire da lontano

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

In principio c'è la comunicazione

Comunicare è la facoltà di trasmettere qualcosa a qualcuno.
Per comunicare bisogna condividere un sistema di riferimento.
(Comunicare deriva dal latino communis che indica proprio la
condivisione)

Primo schema di base di un sistema di comunicazione

Qualcuno/Qualcosa che Produce; Qualcuno/Qualcosa che
Trasmette; Qualcuno/Qualcosa che Consuma

La comunicazione Umana

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

The Mathematical Theory of Communication

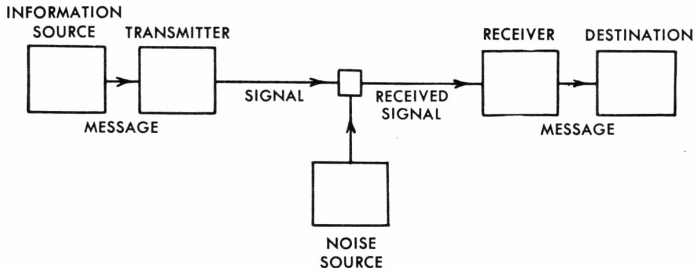
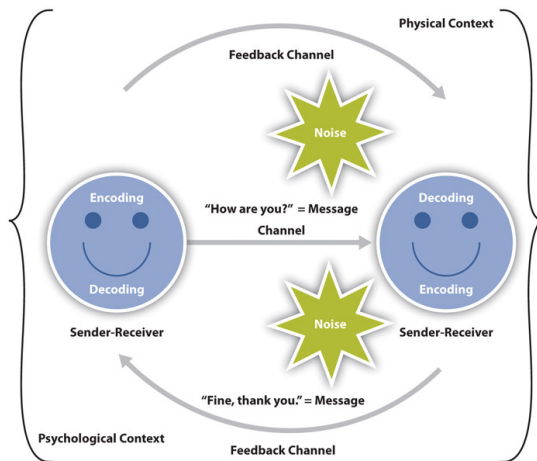


Fig. 1. — Schematic diagram of a general communication system.

La comunicazione Umana

Modello della Comunicazione -Roman Jakobson (1966)



<http://open.lib.umn.edu/>

Estratto Linguistics and Poetics

"The addresser sends a message to the addressee. To be operative the message requires a context referred to ('referent' in another, somewhat ambivalent, nomenclature), seizable by the addressee, and either verbal or capable of being verbalized, a code fully, or at least partially, common to the addresser and addressee (or in other words, to the encoder and decoder of the message); and finally, a contact, a physical channel and psychological connection between the addresser and the addressee, enabling both of them to stay in communication."
(R. Jakobson, 'Closing Statement: Linguistics and Poetics', in Sebeok, Thomas A (Ed.) (1960): *Style in Language*. Cambridge, MA: MIT Press, p. 353)

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Modello Comunicazione: Funzione e Fattore

Funzione Fattore

Referenziale Contesto

Espressiva Mittente

Conativa Destinatario

Fatica Contatto

Metalinguistica Codice

Poetica Messaggio

Messaggio

Nel sistema della comunicazione umana non c'è una assoluta corrispondenza tra il sistema di codifica e il sistema di decodifica.

Codice dell'agente emittente è diverso dal codice dell'agente destinatario

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Diasistema - Weinreich, Uriel. - Linguista (Vilna, Polonia, 1926
- New York 1967)

Sistema (linguistico) di livello superiore, che riunisce due o più sistemi omogenei tra i quali ci siano somiglianze parziali (sul piano fonemico, morfologico, lessicale).

Messaggio

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

Comunicazione come diasistema

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

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

Segre, 1985

Codifica del testo

Elementi del modello della comunicazione

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Codice

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

Codifica del testo

Quando il destinatario è un calcolatore elettronico?

La codifica informatica

Il codice deve essere in comune tra l'autore e la macchina, in un formato comprensibile ad entrambi.

La codifica informatica

- Codice in formato *machine readable* (MRF), cioè un codice in formato comprensibile ad un calcolatore elettronico.
- Codificare per il computer implica una ricodifica del testo affinché vengano esplicitate le caratteristiche e le funzioni del testo identificate.

Codifica del testo

Esplicitare in modo formale

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatatura

Approfondiamo
e concludiamo

La codifica informatica

Codificare in formato digitale un testo significa esplicitare i processi inferenziali effettuati da un interprete nella comprensione del testo stesso.

Codifica del testo

Sinteticamente

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

La codifica digitale di un testo

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

Grammatica formale

Questo linguaggio consente di rappresentare esplicitamente le caratteristiche di un testo applicando etichette meta-testuali la cui funzione e applicazione è descritta da una grammatica formale (per essere “comprese” da una macchina/sistema software).

Codifica del testo

Linguaggio formale che diventa linguaggio teorico

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

La codifica del testo come modello formale

La codifica digitale del testo diviene a questo punto un linguaggio teorico 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 (*markup language*).

Codifica del testo

Sistema linguistico formalizzato

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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 forniscono un potente dispositivo metalinguistico.

Codifica del testo

Finalità

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Prerequisito per una corretta rappresentazione digitale dei testi

Definizione e implementazione di un linguaggio formale che deve essere sia processabile da un computer sia sufficientemente espressivo per rappresentare la complessità del testo.

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcaturo

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Linguaggi di marcatura

6 Approfondiamo e concludiamo

Modellare il testo

Il testo come oggetto del dominio di studio

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Informatica nelle scienze umane

L'informatica umanistica ruota attorno alla rappresentazione e all'elaborazione degli oggetti che costituiscono il dominio delle discipline umanistiche.

Testo come oggetto di studio

Il testo è tra questi, l'oggetto più ricorrente

Modellare il testo

qual è la natura del testo?

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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 la sua natura. Il fatto è che la rappresentazione (e a maggior ragione l'elaborazione) informatica è ontologicamente formale in senso stretto (Ciotti).

Modellare il testo

Testo: oggetto complesso

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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, etc), anche attraverso l'instaurazione di molteplici relazioni tra più livelli.

Ma..

Non possediamo nessuna teoria sufficientemente completa del testo

Modellare il testo

Testo: oggetto complesso

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Cos'è il testo

Si tratta di una entità informativa complessa. in quanto è il prodotto di più agenti e più fattori

Modellare il testo

Tentativo..

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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-semantic (linguistica testuale).

Cos'è il testo

Il termine “testo” si riferisce a un oggetto plurale, in cui esiste

- 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

Cos'è il testo

- Un documento materiale composto da fogli che contengono inchiostro variamente disposte?
- 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?

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Modellare il testo

Il testo: Cercando una definizione

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Il testo

Dal latino *textum*, participio passato di *texere* “tessere” quindi un testo è un “tessuto”, un ordito di unità di significato (**monemi**) veicolato da simboli (**grafemi**)

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Il testo secondo Tito Orlandi 2010

La produzione di un'attività linguistica intesa in senso stretto, cioè in una delle lingue storicamente date, secondo il concetto saussuriano di lingue che produce le parole. Il fatto che essa sia finalizzata a trasmettere un messaggio, pur importante in sé, non interessa per il momento se non marginalmente.

Modellare il testo

Definizioni

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

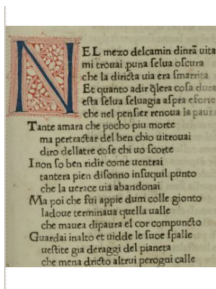
Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo



Nel mezzo del cammin di nostra vita,
Mi ritrovai per una selva oscura,
Che la dritta via era smarrita.
È quanto a dir qual era è cosa dura
Questa selva selvaggia, ed aspra e forte, 5
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; 10
Tant'era pien di sonno in su quel punto,
Che la verace via abbandonai.
Ma po' ch'i' fui appie d'un colle giunto,
Là ove terminava quella valle,
Che m'avea di paura il cor compunto; 15
Guarda' 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 dritta via era smarrita.
Ahi 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'intrai,
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, rispetto alle
variabili dei caratteri, della scrittura.

Modellare il testo

Complessità del testo

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..

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Modellare il testo

Il testo letterario

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Particolarità del testo letterario

- L'emittente prepara il messaggio, il destinatario potrebbe riceverlo anche dopo molto tempo
- Il feedback molto spesso non è possibile
- Il codice dell'emittente e quello del destinatario possono risultare diversi (o solo parzialmente compatibili)

Modellare il testo

Il documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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

Cos'è un documento

- Una prova o un 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?

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Modellare il testo

Struttura del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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-semantiche 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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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

Codifica del documento

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Il testo non è il documento

La codifica di un documento è diversa dalla codifica di un testo
(testo != documento). Si può parlare di documenti cartacei
e di documenti digitali, ma non di testi cartacei e testi digitali!

Documento digitale

Corrispondenza perfetta con l'originale: documento digitale

Il testo è strutturato

Il testo è un oggetto reale dotato di una sua struttura che corrisponde alla struttura del linguaggio di rappresentazione;

La natura del testo come OHCO

In un famoso articolo (DeRose et al., 1990) gli autori tentarono di dimostrare che il testo è una “ordered hierarchy of content objects (OHCO)”: una gerarchia ordinata di oggetti di contenuto.

Modellare il testo

ordered hierarchy of content objects (OHCO)

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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 poiché tali elementi ne possono contenere altri, ma sempre 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

OHCO e TEI

Il primo modello implementato dalla Text Encoding Initiative si è basato su questo impianto teorico.

OHCO e TEI

Sono però stati riscontrati una serie di problemi di rappresentazione che determinano diversi insiemi di elementi di contenuto i quali non possono essere ricondotti a una struttura gerarchica unitaria (pensiamo alla rappresentazione di parole e di righe).

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Modellare il testo

Prospettive analitiche

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Sottoprospective analitiche

Se due oggetti testuali evidenziati da una prospettiva teorica si sovrappongono, allora essi appartengono rispettivamente a due sottoprospective diverse della prospettiva teorica principale. (F. Ciotti)

Sottoprospective analitiche

In questa ottica il testo diventa un sistema a più livelli, che corrispondono a diversi punti di vista dell'osservatore (sistema multidimensionale). (F. Ciotti)

Modellare il testo

La struttura gerarchica del testo

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Le gerarchie del testo

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

Modellare il testo

OHCO

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

OHCO: che cosa è un testo?

Un testo è un oggetto linguistico astratto organizzato secondo una struttura gerarchica ordinata di oggetti di contenuto (e varie successive riformulazioni).

OHCO: ricapitolando

L'idea di una preminenza della struttura gerarchica nella testualità ha mantenuto un ruolo descrittivo ed esplicativo essenziale.

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcaturo

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Linguaggi di marcatura

6 Approfondiamo e concludiamo

Importanza della codifica digitale

Perché effettuare la codifica

Digitalizzare un testo

Digitalizzare per poter favorire l'elaborazione e il trattamento automatico dei testi

Trattamento dei testi

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Importanza della codifica digitale

Perché effettuare la codifica

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, etc.)
- Permettono anche di preservare i documenti più antichi (e fragili) riducendone la consultazione diretta

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Sistemi non adatti

- word processing: WYSIWYG (What You See Is What You Get)
- sistemi proprietari chiusi (Word, Adobe, etc)
- elevata obsolescenza e limitata manutenibilità
- difficile portabilità su piattaforme diverse (windows, linux)

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Portabilità

Un documento digitale (come qualsiasi risorsa informativa digitale) è portabile se è intercambiabile 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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

Standard

I vantaggi di uno standard formale o informale, oltre alla portabilità sta anche nella sua apertura, ovvero nella disponibilità pubblica delle sue specifiche.

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Standardizzare

- 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 usufruibili anche in futuro - codificare testi che possano essere usati 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

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Linguaggio di codifica

Se per predisporre una codifica del testo utilizziamo un linguaggio dotato di una sintassi che permetta di specificare le relazioni tra gli elementi, essa può essere usata per rappresentare la struttura e l'organizzazione del testo a un determinato livello di descrizione, o i rapporti tra elementi appartenenti a diversi livelli.

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 prestano ottimamente per rappresentare strutture complesse come riferimenti incrociati e collegamenti tra elementi all'interno di un testo, ma anche tra più testi.

dichiarativa

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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:

Importanza della codifica digitale

Perché effettuare la codifica

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatura

Approfondiamo
e concludiamo

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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

Gerarchia ordinata di oggetti di contenuto (OHCO)

I linguaggi di markup dichiarativi gerarchici, e 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.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Linguaggi di marcatura

6 Approfondiamo e concludiamo

Progress status

Codifica di
Testi -
Introduzione
Teorica
a.a.
2018-2019

A.M. Del
Grosso

Elementi
Teorici e
Metodologici
della codifica

Codifica del
Testo

Modello del
Testo

Importanza
della Codifica

Linguaggi di
marcatore

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Importanza della Codifica

5 Linguaggi di marcatura

6 Approfondiamo e concludiamo