

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”,
21st 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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Linguaggi di marcatura

5 Importanza della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

il testo come 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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Linguaggi di marcatura

5 Importanza della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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)

34

The Mathematical Theory of Communication

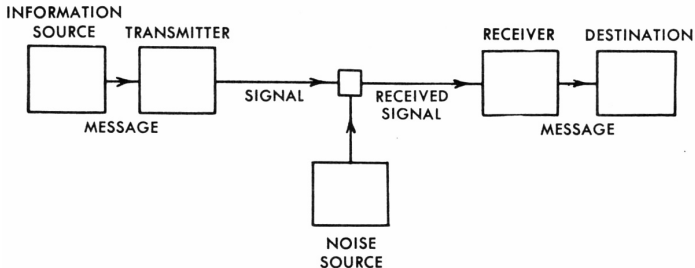
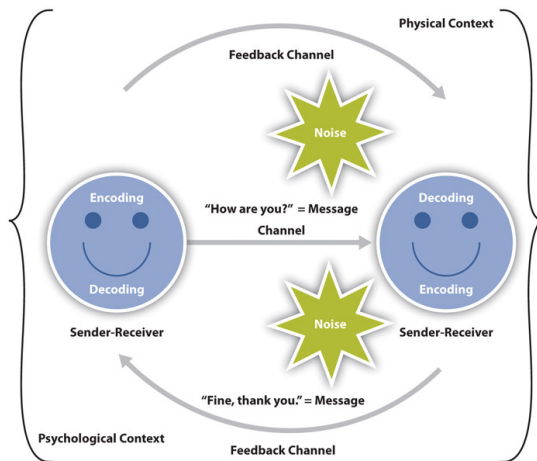


Fig. 1. — Schematic diagram of a general communication system.

La comunicazione Umana

Modello della Comunicazione -Roman Jakobson (1966)



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

La comunicazione Umana

R. Jakobson

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

La comunicazione Umana

R. Jakobson

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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 (~ *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

Linguaggi di
marcatura

Importanza
della Codifica

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 ricodifica il testo affinchè vengano esplicitate caratteristiche e funzioni del testo.

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

Linguaggi di
marcatura

Importanza
della Codifica

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

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

La codifica digitale di un testo

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

grammatica formale

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Linguaggi di marcatura

5 Importanza della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Tentativo..

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

Il testo

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

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

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

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 langue che produce la parole. Il fatto che essa sia finalizzata a trasmettere un messaggio, pur importante in sé, non interessa per il momento se non marginalmente.

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.

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

Linguaggi di
marcatura

Importanza
della Codifica

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 lo chiamiamo 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

Linguaggi di
marcatura

Importanza
della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Linguaggi di marcatura

5 Importanza della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Linguaggi di marcatura

5 Importanza della Codifica

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

Linguaggi di
marcatura

Importanza
della Codifica

Approfondiamo
e concludiamo

1 Elementi Teorici e Metodologici della codifica

2 Codifica del Testo

3 Modello del Testo

4 Linguaggi di marcatura

5 Importanza della Codifica

6 Approfondiamo e concludiamo