

# Creazione e Analisi dei Testi Digitali

Angelo Mario Del Grosso

*CNR-ILC*

<http://ilc.cnr.it/>

[angelo.delgrosso@ilc.cnr.it](mailto:angelo.delgrosso@ilc.cnr.it)

**INTRODUZIONE ALL'INFORMATICA UMANISTICA**

*AIUCD-AICA*

Istituto di Linguistica Computazionale "A. Zampolli",  
24th November 2023

# Argomenti trattati

- 1 Introduzione personale
- 2 Panoramica Rappresentazione del Testo
- 3 XML, TEI guidelines e Domain Specific Languages
- 4 Modelli di rappresentazione e Vector Semantics
- 5 Esempi
- 6 Voyant-Tools

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Progress status

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

**1** Introduzione personale

2 Panoramica Rappresentazione del Testo

3 XML, TEI guidelines e Domain Specific Languages

4 Modelli di rappresentazione e Vector Semantics

5 Esempi

6 Voyant-Tools

# Di cosa mi occupo

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Filologia Digitale e Computazionale

Attività di ricerca per lo sviluppo di sistemi di linguistica e filologia digitale/computazionale volti alla produzione, rappresentazione, analisi, fruizione e interrogazione di testi di interesse umanistico.



## Profilo professionale e di ricerca

## Creazione e Analisi dei Testi Digitali

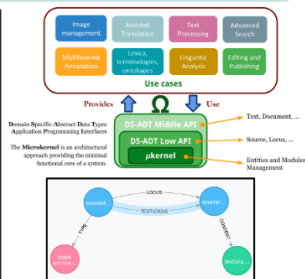
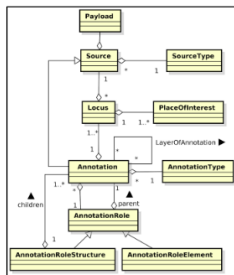
A.M. Del  
Grosso

## Introduzione personale

## In sintesi

Ingegnere Informatico prestatò alla filologia computazionale

```
text = Text.of("Literary Text to process", URI.create("//source/text/000"));
annotation = AnnotationText.of("Annotation on the text",
URI.create("//annotation/text/123"));
annotation.addLocus(text, 13, 18);
annotation.save();
```



# Temi del seminario

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

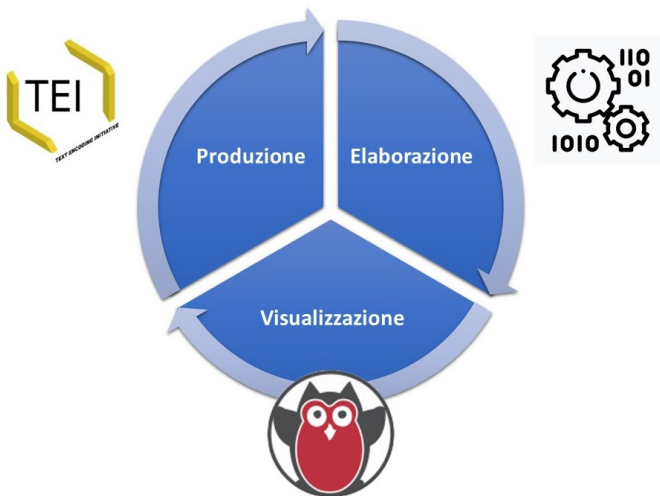
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Obiettivo del seminario

## Argomenti del seminario

Introduzione alle tecnologie per la rappresentazione digitale del testo, per il suo trattamento automatico e la visualizzazione.

## Argomenti del seminario

- **Prima parte:** Concetti di base: rappresentazione ed elaborazione del testo (*XML/TEI, DSL*)
- **Seconda parte:** Panoramica *Semantic Vectors* e *Embeddings*
- **Terza parte:** Esempi e Visualizzazione del testo tramite *Voyant-tools*

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Progress status

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

**Panoramica  
Rapp-  
resentazione  
del Testo**

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

1 Introduzione personale

**2 Panoramica Rappresentazione del Testo**

3 XML, TEI guidelines e Domain Specific Languages

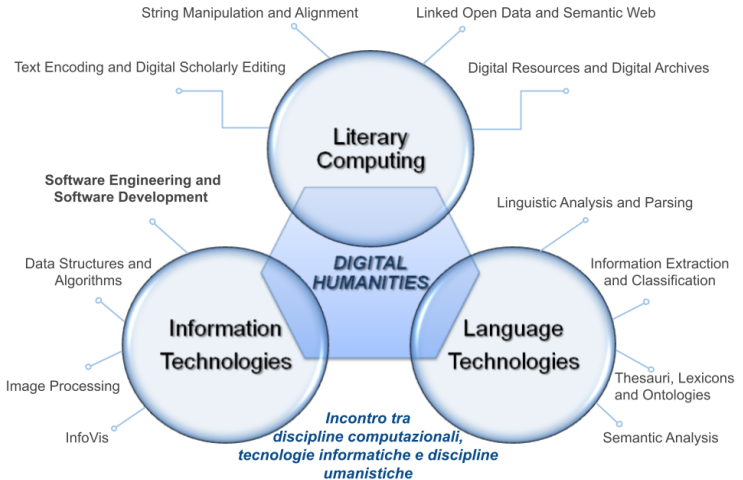
4 Modelli di rappresentazione e Vector Semantics

5 Esempi

6 Voyant-Tools

# Informatica Umanistica

Ambito disciplinare



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Informatica Umanistica

## Tanti tipi di documenti

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

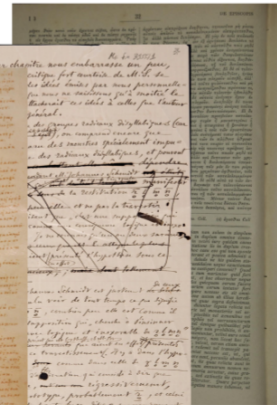
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Informatica Umanistica

Prospettive di un'unica entità?

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

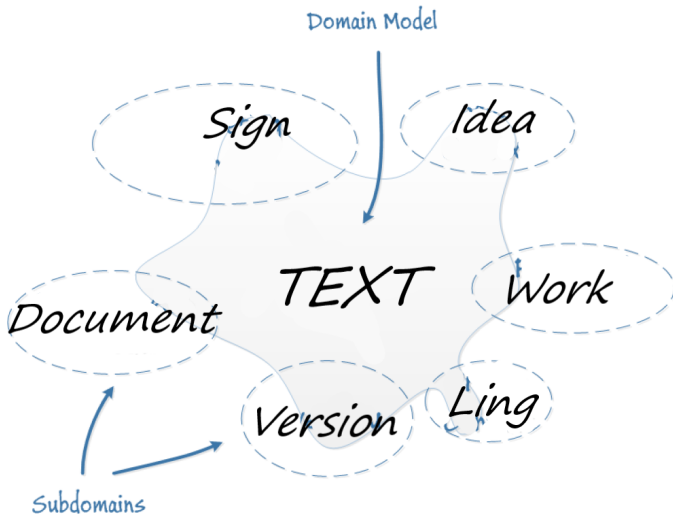
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Informatica Umanistica

Modelli diversi coerenti con le diverse prospettive

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

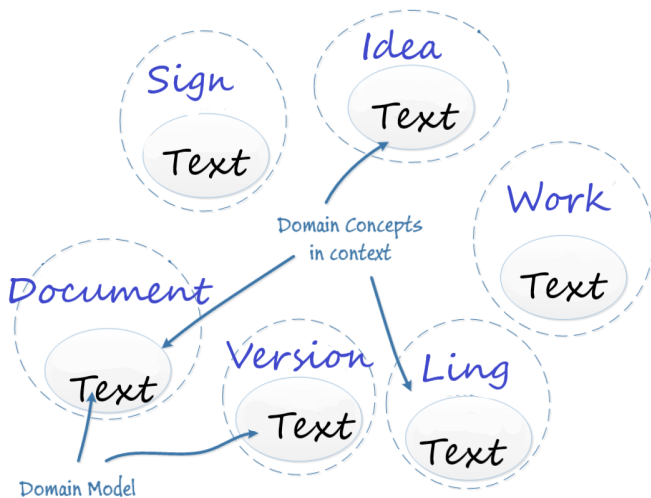
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools





# Text Encoding

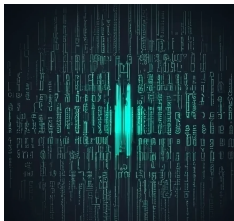
## Plain Text

### Il testo digitale

Di cosa stiamo parlando?

I dispositivi elettronici gestiscono i dati come **simboli** rappresentati mediante **codici numerici binari**.

**Il testo non fa eccezione:** sequenza di simboli codificati tramite un sistema numerico binario.



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Text Encoding

## Plain Text

### Il testo digitale

#### bit byte e sistema numerico binario

*bit*

|   |
|---|
| 0 |
|---|

*Byte*

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
|---|---|---|---|---|---|---|---|

*16 bit-  
Word*

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

*32 bit-  
Word*

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |

Binario → Esadecimale

binario      esadecimale

0000 0  
0001 1  
0010 2  
0011 3  
0100 4  
0101 5  
0110 6  
0111 7  
1000 8  
1001 9  
1010 A  
1011 B  
1100 C  
1101 D  
1110 E  
1111 F

$(10001111)_2$

$(F)_{16}$

# Text Encoding

## Plain Text

### Il testo digitale

*Il testo è una sequenza di simboli che chiamiamo **caratteri** codificati in una struttura dati che chiamiamo **stringa** (sequenza di caratteri)*



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Elementi di Codifica dei Caratteri

## American Standard Code for Information Interchange

### Tabella Code Page ASCII 7 bit - Anni '60

|    | 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

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Elementi di Codifica dei Caratteri

## Esempio codifica binaria

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

codifica *ciao mondo!* 7 bit ASCII

6369 616F 206D 6F6E 646F 210A

codifica *ciao è mondo!* 8 bit ASCII

6369 616F 20**e8** 206D 6F6E 646F 210A

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

6369 616F 20**c3 a8**20 6D6F 6E64 6F21 0A

# 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).

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# 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.

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Progress status

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

1 Introduzione personale

2 Panoramica Rappresentazione del Testo

3 XML, TEI guidelines e Domain Specific Languages

4 Modelli di rappresentazione e Vector Semantics

5 Esempi

6 Voyant-Tools



# Ricordiamo che

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Codifica del Testo

- Rappresentazione dei caratteri (codifica di livello 0)
- Rappresentazione della struttura e dei fenomeni testuali (codifica di alto livello)

# I linguaggi di codifica

## introduzione

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### Definizione di codifica digitale del testo

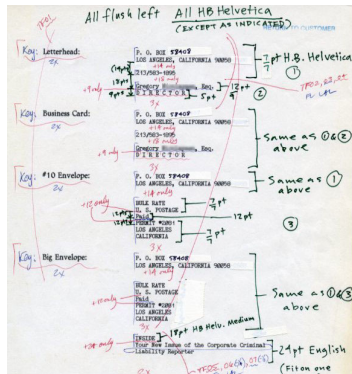
Per **codifica** digitale dei testi intendiamo la *rappresentazione formale* di un **testo** ad un qualche livello descrittivo, su di un supporto digitale, in un formato utilizzabile da un elaboratore (*Machine Readable Form*) mediante un opportuno **linguaggio informatico** (F. Ciotti).

# Markup language e XML

soluzione corrente per la codifica dei testi

## XML per la descrizione e la codifica

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



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Fondamenti XML

## eXtensible Markup Language

### eXtensible Markup Language

*L'XML è un meta-linguaggio mantenuto dal W3c, usato per creare linguaggi di marcatura (detti **vocabolari**).*



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Fondamenti XML

## eXtensible Markup Language: Esempio TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/css" href="customStyle.css"?>
<!--The following document is made online by the Perseus Project -->
<!--Added the TEI-lite DTD and a processing instruction -->
<!DOCTYPE TEI.2 SYSTEM "teixbaby.dtd">

<TEI.2>
  <text lang="en">
    <body>
      <div1 type="book" n="1" org="uniform" sample="complete">
        <div2 type="section" n="327A" org="uniform" sample="complete">
          <p>
            327A - 328B Socrates describes how he visited the Piraeus in company with Glauco, and
            was induced by Polemarchus and others to defer his return to Athens.
          </p>
          <p>
            <lemma lang="greek" targOrder="U" from="ROOT" to="DITTO">κατέβην κτλ.</lemma>
            Dionys. Hal.
            <title lang="la">de comp. verb.</title>
            p. 208 (Reiske)
            <foreign lang="greek">
              ὁ δὲ Πλάτων, τοὺς
              ἑαυτοῦ διαλόγους κτενίζων καὶ βοστρυχίζων, καὶ πάντα τρόπον ἀναπλέκων, οὐ
              διέλιπεν ὀγδοήκοντα γεγονόσ' ἔτη. πᾶσι γὰρ δὴ πού τοις φιλολόγοις γνῶριμα
              τὰ περὶ τῆς φιλοπονίας τάνδρὸς ἱστορούμενα, τὰ τ' ἄλλα, καὶ δὴ καὶ τὰ
              περὶ τὴν δέλτον ἦν τελευτήσαντος αὐτοῦ λέγουσιν εὐρεθῆναι ποικίλως
              μετακειμένην τὴν ἀρχὴν τῆς πολιτείας ἔχουσαν τήνδε "κατέβην χθές
              εἰς Πειραιᾶ μετὰ Γλαῦκωνος τοῦ Ἀριστῶνος
            </foreign>
            ."
          </p>
        </div2>
      </div1>
    </body>
  </text>
</TEI.2>

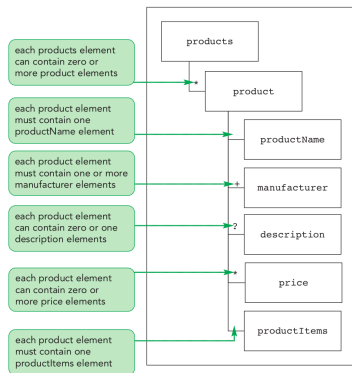
<!-- This document is not completed and was cut without a special meaning -->
```

# Fondamenti XML

eXtensible Markup Language

## XML vista ad albero

XML ha un **modello dei dati gerarchico** e può quindi essere rappresentato come un **albero etichettato e ordinato**. XML ottimizza la rappresentazione di strutture padre-figlio, contenuto-contenitore a diversi livelli di profondità e molteplicità con ordinamento predefinito.



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Fondamenti XML

## eXtensible Markup Language

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

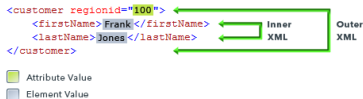
Esempi

Voyant-Tools

## XML per rappresentare il testo

Un documento XML è composto da **elementi** e **attributi**.

Gli elementi sono la base, le unità fondamentali di qualsiasi documento XML.



- Una sintassi e una grammatica regolano l'applicabilità del linguaggio di marcatura
- **Sintassi:** documento **well formed** (ben formato)
- **Grammatica:** documento **valido**

# Fondamenti XML

## eXtensible Markup Language: Esempio TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<div type="narrative" n="6">
  <head>Sixth Narrative</head>
  <head>contributed by Sergeant Cuff</head>
  <div type="fragment" n="6.1">
    <opener>
      <dateline>
        <name type="place">Dorking, Surrey,</name>
        <date>July 30th, 1849</date>
      </dateline>
      <salute>To <name>Franklin Blake, Esq.</name> Sir, </salute>
    </opener>
    <p>I beg to apologize for the delay that has occurred in the
      production of the Report, with which I engaged to furnish you.
      I have waited to make it a complete Report ...</p>
    <closer>
      <salute>I have the honour to remain, dear sir, your
        obedient servant </salute>
      <signed>
        <name>RICHARD CUFF</name> (late sergeant in the
          Detective Force, Scotland Yard, London). </signed>
      </closer>
    </div>
  </div>
```

*immagine dal sito TEI Guide Lines*



# Markup language e XML

soluzione corrente per la codifica dei testi

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## TEI-XML

Le specifiche messe a punto dalla Text Encoding Initiative (TEI-XML) sono considerate ad oggi lo **standard de facto** per una corretta **rappresentazione digitale dei testi** per gli scopi d'indagine umanistica.

## Testo di riferimento

Guidelines for Electronic Text Encoding and Interchange  
( <http://www.tei-c.org/Guidelines/> )

# Intro Text Encoding Initiative

## TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### TEI Guidelines: Obiettivi

- better **interchange** and **integration** of scholarly data
- support for **all texts**, in **all languages**, from **all periods**
- guidance for the *perplexed*: **what to encode** - hence, a user-driven codification of existing best practice
- assistance for the *specialist*: **how to encode** - hence, a loose framework into which unpredictable extensions can be fitted

# Intro Text Encoding Initiative

## TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### Che cosa offre la TEI

- un ricco (e complesso) manuale di codifica, (le Guidelines for Electronic Text Encoding and Interchange)
- un numero elevato di elementi (sia strutturale sia semantico)
- schemi di codifica
- infrastruttura modulare e personalizzabile

# Intro Text Encoding Initiative

## TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
rappre-  
sentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rappre-  
sentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



The screenshot shows the TEI website header with the logo and the title "<Text Encoding Initiative>". Below the header is a navigation bar with links: Home, Guidelines, Activities, Tools, Membership, Support, About, News. A search bar is present with a dropdown menu showing "P5 Guidelines — English" and a "Search" button. The main heading is "P5: Guidelines for Electronic Text Encoding and Interchange", followed by the version information: "Version 4.5.0. Last updated on 25th October 2022, revision 3e98e619e".

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



### Front Matter

#### [Title](#)

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

### Back Matter

- xi. [Appendix A: Model Classes](#)

### Text Body

- 1 [The TEI Infrastructure](#)
- 2 [The TEI Header](#)
- 3 [Elements Available in All TEI Documents](#)
- 4 [Default Text Structure](#)
- 5 [Characters, Glyphs, and Writing Modes](#)
- 6 [Verse](#)
- 7 [Performance Texts](#)
- 8 [Transcriptions of Speech](#)
- 9 [Dictionaries](#)
- 10 [Manuscript Description](#)

### TEI sourcecode

- [Getting and Using the TEI Sources](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

# Intro Text Encoding Initiative

## TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

 < Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — English

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.5.0. Last updated on 25th October 2022, revision 3e98e619e

### Table of contents

Appendix C.1 About the Elements  
Appendix

« Appendix B Attribute Classes

» Appendix D Attributes

Home

## Appendix C Elements

### Appendix C.1 About the Elements Appendix

This appendix gives you links to reference pages for all elements in the TEI Guidelines. There are 585 TEI elements in revision [3e98e619e](#) of TEI P5 [Version 4.5.0](#) of the TEI Guidelines.

The elements listed here are in the TEI Namespace: <http://www.tei-c.org/ns/1.0> unless otherwise noted on that element's reference page.

Sorted alphabetically

a b c d e f g h i j k l m n o p q r s t u v w x z

Show all

Show by module

a

[ab](#) [abbr](#) [abstract](#) [accMat](#) [acquisition](#) [activity](#) [actor](#) [add](#) [additional](#) [additions](#) [addName](#) [address](#) [addLine](#) [addSpan](#) [adminInfo](#) [affiliation](#) [age](#) [alt](#) [alternate](#) [altGrp](#) [altIdent](#) [altIdentifier](#) [am](#) [analytic](#) [anchor](#) [annotation](#) [annotationBlock](#) [anyElement](#) [app](#) [appInfo](#) [application](#) [arc](#) [argument](#) [att](#) [attDef](#) [attList](#) [attRef](#) [author](#) [authority](#) [availability](#)

↳

# Intro Text Encoding Initiative

## TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale


Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

 < Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — English

### P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.6.0. Last updated on 4th April 2023, revision f18deffba

|  |
|--|
| <b>Table of contents</b>                   |
| Appendix D.1 About the Attributes Appendix |
| ◀ Appendix C Elements                      |
| ▶ Appendix E Datatypes and Other Macros    |
| Home                                       |

## Appendix D Attributes

### ↓ Appendix D.1 About the Attributes Appendix

This app. [dix](#) gives you a list of attributes and links to the reference pages for the elements on which they appear. There are 270 distinctly-named attributes in revision [f18deffba](#) of TEI P5 [Version 4.6.0](#) of the TEI Guidelines.

|                 |  |
|-----------------|--|
| <b>absolute</b> | <a href="#">when</a>   |
| <b>active</b>   | <a href="#">interaction</a> <a href="#">relation</a>                     |
| <b>adj</b>      | <a href="#">node</a>   |
| <b>adjFrom</b>  | <a href="#">node</a>   |
| <b>adjTo</b>    | <a href="#">node</a>   |
| <b>age</b>      | <a href="#">person</a> <a href="#">personGrp</a> <a href="#">persona</a> |
| <b>agent</b>    | <a href="#">att.damaged</a> <a href="#">gao</a> <a href="#">unclear</a>  |
| <b>aloud</b>    | <a href="#">said</a>   |
| <b>ana</b>      | <a href="#">att.global.analytic</a>                                      |

# Intro Text Encoding Initiative

## TEI

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### TEI: Struttura modulare - Moduli essenziali

- **tei**: definisce le classi di elementi, le macro e i datatype che verranno usati per tutti i moduli
- **header**: l'intestazione contenente i metadati relativi al documento TEI XML
- **textstructure**: elementi strutturali per qualsiasi tipo di testo
- **core**: elementi utili in qualsiasi tipo di documento

# Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Struttura di un documento TEI

- *struttura fondamentale all'interno della radice (<TEI>)*
- una intestazione TEI (<teiHeader>)
- un testo: <text> (o più testi, cfr. infra)



# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-model href="tei-lite.rng"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- metadati -->
  </teiHeader>
  <text>
    <div> <p> <!-- contenuto testuale --> </p> </div>
  </text>
</TEI>
```

# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### documento TEI - schema di intestazione TEI minima

Metadati essenziali riguardano

- il titolo
- la modalità di diffusione
- la fonte originaria di un testo codificato

*Permettono classificazione, archiviazione ed elaborazione bibliografica*

# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<teiHeader>
  ...
  <fileDesc>
    <titleStmt>
      <title>La Divina Commedia: versione elettronica</title>
      <respStmt>
        <resp>Conversione TEI P5 a cura di</resp>
        <name>M. Rossi</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <publisher>Università di Pisa</publisher>
      <date>2002-11-07</date>
      <availability status="restricted"><p></p></availability>
    </publicationStmt>
    <sourceDesc><bibl>
      <title>La Divina Commedia</title>
      <author>Dante Alighieri </author>
      <publisher>Mondadori</publisher>
      <date>1988</date>
    </bibl></sourceDesc>
  </fileDesc>
</teiHeader>
```

# Intro Text Encoding Initiative

Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Le altre componenti dell'intestazione TEI

- `<encodingDesc>` informazioni riguardo lo schema (e il modello di codifica) utilizzato
- `<profileDesc>` descrizione del testo: quando è stato creato, da chi, usando quali lingue etc.
- `<revisionDesc>` informazioni sulle versioni del file

*I metadati sono una componente essenziale di qualsiasi progetto di digitalizzazione*

# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<TEI>
...
  <teiHeader>
    <!-- informazioni del TEI Header -->
  </teiHeader>
  <text>
    <front>
      <!-- premessa, dedica ... -->
    </front>

    <body>
      <!-- corpo del testo ... -->
    </body>

    <back> <!-- postfazione, appendice ... --></back>
  </text>
</TEI>
```

# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<TEI>
...
<teiHeader> <!-- intestazione --> </teiHeader>
<text>
  <front> <!-- frontespizio testo composito --> </front>
  <group>
    <text>
      <front> <!-- frontespizio del primo testo --> </front>
      <body> <!-- body del primo testo --> </body>
      <back> <!-- appendice del primo testo --> </back>
    </text>
    <text>
      <front> <!-- frontespizio del secondo testo --> </front>
      <body> <!-- body del secondo testo --> </body>
      <back> <!-- appendice del secondo testo --> </back>
    </text> <!-- altri testi o gruppi di testi -->
  </group>
  <back> <!-- appendice del testo composito -->
</text>
</TEI>
```

# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<teiCorpus>
  <teiHeader> <!-- metadati per il corpus --></teiHeader>
  <TEI>
    <teiHeader> <!-- metadati relativi al I testo--></teiHeader>
    <text> <!-- primo testo del corpus--> </text>
  </TEI>
  <TEI>
    <teiHeader> <!-- metadati relativi al II testo--></teiHeader>
    <text> <!-- secondo testo del corpus--> </text>
  </TEI>
</teiCorpus>
```

# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<text>
  <body>
    <div type="book" n="1">
      <p>Dell'Iliade d'Omero<lb/> tradotta in veneziano </p>
      <p>Canto primo</p>
      <p>Preci, peste<lb/> ingiustizie e regii sdegni </p>
      <p>Sentenza del canto<lb/>
        <hi>A ministri di dei si dee rispetto</hi>
      </p>
      <lg n="1.0">
        <head>ARGOMENTO</head>
        <l n="1.0.1">Per Criseida da peste il greco è infetto.</l>
        <l n="1.0.2">La rende Atride; ma Briseida al figlio</l>
        <l n="1.0.3">Di Teti toglie. Achille di dispetto</l>
        <l n="1.0.4">Arde; e la Madre a greci aspro periglio</l>
        <l n="1.0.5">Da Giove ottien. Giuno a sviar l'effetto</l>
        <l n="1.0.6">Il Dio rimbrota nel divin consiglio;</l>
        <l n="1.0.7">Giove si sdegna. È da Vulcan placata.</l>
        <l n="1.0.8">Ebria va a letto l'immortal brigata</l>
      </lg>
    </div>
  </body>
</text>
```



# Intro Text Encoding Initiative

## Schemi di codifica TEI – Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<text>
  <body>
    <div n="ch1" type="chapter">
      <pb n="1" />
      <p>[...] risulta chiaro se avete letto
        <title rend="underline" xml:lang="fra">
          Lesfleurs du mal
        </title> [...]
      </p>
      <p>[...] un grande esempio di
        <foreign xml:lang="fra">savoir faire</foreign> [...]
      </p>
    </div>
    <!-- altri elementi <div> -->
  </body>
</text>
```

# Intro Text Encoding Initiative

## Schemi di codifica TEI: Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<text>
...
  <body>
    <div n="ch1" type="chapter">
      <pb n="1" />
      <p n="1">[...] descritto altrove (si veda ad esempio
        | <ref target="#Rossi94">Rossi 1994</ref>) [...</p>
    </div>
    <div n="bib" type="bibliography"> [...]
      <bibl xml:id="Rossi94">
        | <author>Rossi, M.</author>[...]
      </bibl>
    </div>
  </body>
</text>
```

# Intro Text Encoding Initiative

## Schemi di codifica TEI: Moduli base

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

<del>

Home  
C Elements

<del> (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, or a previous annotator or corrector: [3.5.3 Additions, Deletions, and Omissions]

| Module       | core — Elements Available in All TEI Documents   |
|--------------|--|
| Attributes   | <ul style="list-style-type: none"><li>• <b>att.global</b>: @xml:id, @n, @xml:lang, @xml:base, @xml:space<ul style="list-style-type: none"><li>◦ <b>att.global.rendition</b>: @rend, @style, @rendition</li><li>◦ <b>att.global.linking</b>: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select</li><li>◦ <b>att.global.analytic</b>: @ana</li><li>◦ <b>att.global.facs</b>: @facs</li><li>◦ <b>att.global.change</b>: @change</li><li>◦ <b>att.global.responsibility</b>: @cert, @resp</li><li>◦ <b>att.global.source</b>: @source</li></ul></li><li>• <b>att.transcriptional</b>: @status, @cause, @seq<ul style="list-style-type: none"><li>◦ <b>att.editLike</b>: @evidence, @instant</li><li>◦ <b>att.written</b>: @hand</li></ul></li><li>• <b>att.typed</b>: @type, @subtype</li><li>• <b>att.dimensions</b>: @unit, @quantity, @extent, @precision, @scope<ul style="list-style-type: none"><li>◦ <b>att.ranging</b>: @atLeast, @atMost, @min, @max, @confidence</li></ul></li></ul> |
| Member of    | <a href="#">model.linePart</a> <a href="#">model.pPart.transcriptional</a>   |
| Contained by | <a href="#">analysis</a> : <a href="#">cl</a> <a href="#">pc</a> <a href="#">pnr</a> <a href="#">s</a> <a href="#">w</a>   |

# Intro Text Encoding Initiative

## Pubblicazione - TEIPublisher

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

TEI  
Publisher

Start

Documentazione

Notizie

DTS

Scarica

Informazioni

Help

Accedi

Bibliografia Staropobla - Simple

...

Aktem

Aktem komiznym bodlaye Gymnazijum Leczyskiok. Leczne imprm. Mich. Buk. 1639. folio, s. n. 5.  
Krasin.

Aktem

Aktem obacz Zhylnowski Andrzei (Historia 1588).

Aktorowie

Aktorowie ob. Majestas (1696) — Vismirits (1686).

Aktów.

Aktivni spisanie ob. Filipowski M. Spizarnia (1640. 1655).

Aktów

Sta aktivni bodlaye, w 4co, str. XII. 102. Jocher 6740.  
Ousl.

Aktów

Soek aktivni, to jost: Akt I. Uslugi obratu jnigego. Akt II. Przygotowania publicznego przyspogujacych do i. wieczery Pankiszy. Akt III. Uhugowania i. wieczery Pankiszy. Akt IV. Uhugowania i. wieczery Pankiszy przy obserch. Akt V. Dawaia slobu w stan ioc. matizelski wstugujacy. Akt VI. Nawiedzenia dwoych. Dla przelozego i oguladniznego nczynania i Agendy Zboru ewangielickich komuny i W X. Lz. wytych. W Krolewiec drukował Jan Henryk Hottung, Roku 1742 [nie: B. m. b. r. (Krolewiec Hartung 1742)] w sce k. i str. 64. — Akt polwiczienia domu prywatnego str. 12. — Procnocia o przycipiu na ten swiat prawdziwego Mesyazsa Jensa str. 14 (got.). ob. Agenda mniejsza.  
Jagoll. — Begel — Curion. — Zbor wteliski.

Akty

Akty i affekty cwiit, uctnik o miluisk ob. Rho Jan Igo. (1646).

c. XII. s.

...

Aktem

Aktem komiznym bodlaye Gymnazijum Leczyskiok. Leczne imprm. Mich. Buk. 1639. folio, s. n. 5.  
Krasin.

Aktem

Aktem obacz Zhylnowski Andrzei (Historia 1588).

Aktorowie

Aktorowie ob. Majestas (1696) — Vismirits (1686).

Aktów.

Aktivni spisanie ob. Filipowski M. Spizarnia (1640. 1655).

Aktów

Sta aktivni bodlaye, w 4co, str. XII. 102. Jocher 6740.  
Ousl.

Aktów

Soek aktivni, to jost: Akt I. Uslugi obratu jnigego. Akt II. Przygotowania publicznego przyspogujacych do i. wieczery Pankiszy. Akt III. Uhugowania i. wieczery Pankiszy. Akt IV. Uhugowania i. wieczery Pankiszy przy obserch. Akt V. Dawaia slobu w stan ioc. matizelski wstugujacy. Akt VI. Nawiedzenia dwoych. Dla przelozego i oguladniznego nczynania i Agendy Zboru ewangielickich komuny i W X. Lz. wytych. W Krolewiec drukował Jan Henryk Hottung, Roku 1742 [nie: B. m. b. r. (Krolewiec Hartung 1742)] w sce k. i str. 64. — Akt polwiczienia domu prywatnego str. 12. — Procnocia o przycipiu na ten swiat prawdziwego Mesyazsa Jensa str. 14 (got.). ob. Agenda mniejsza.  
Jagoll. — Begel — Curion. — Zbor wteliski.

Akty

Akty i affekty cwiit, uctnik o miluisk ob. Rho Jan Igo. (1646).

...

Aktem

Aktem komiznym bodlaye Gymnazijum Leczyskiok. Leczne imprm. Mich. Buk. 1639. folio, s. n. 5.  
Krasin.

Aktem

Aktem obacz Zhylnowski Andrzei (Historia 1588).

Aktorowie

Aktorowie ob. Majestas (1696) — Vismirits (1686).

Aktów.

Aktivni spisanie ob. Filipowski M. Spizarnia (1640. 1655).

Aktów

Sta aktivni bodlaye, w 4co, str. XII. 102. Jocher 6740.  
Ousl.

Aktów

Soek aktivni, to jost: Akt I. Uslugi obratu jnigego. Akt II. Przygotowania publicznego przyspogujacych do i. wieczery Pankiszy. Akt III. Uhugowania i. wieczery Pankiszy. Akt IV. Uhugowania i. wieczery Pankiszy przy obserch. Akt V. Dawaia slobu w stan ioc. matizelski wstugujacy. Akt VI. Nawiedzenia dwoych. Dla przelozego i oguladniznego nczynania i Agendy Zboru ewangielickich komuny i W X. Lz. wytych. W Krolewiec drukował Jan Henryk Hottung, Roku 1742 [nie: B. m. b. r. (Krolewiec Hartung 1742)] w sce k. i str. 64. — Akt polwiczienia domu prywatnego str. 12. — Procnocia o przycipiu na ten swiat prawdziwego Mesyazsa Jensa str. 14 (got.). ob. Agenda mniejsza.  
Jagoll. — Begel — Curion. — Zbor wteliski.

Akty

Akty i affekty cwiit, uctnik o miluisk ob. Rho Jan Igo. (1646).

# Intro Text Encoding Initiative

## Pubblicazione - Edition Visualization Technology

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

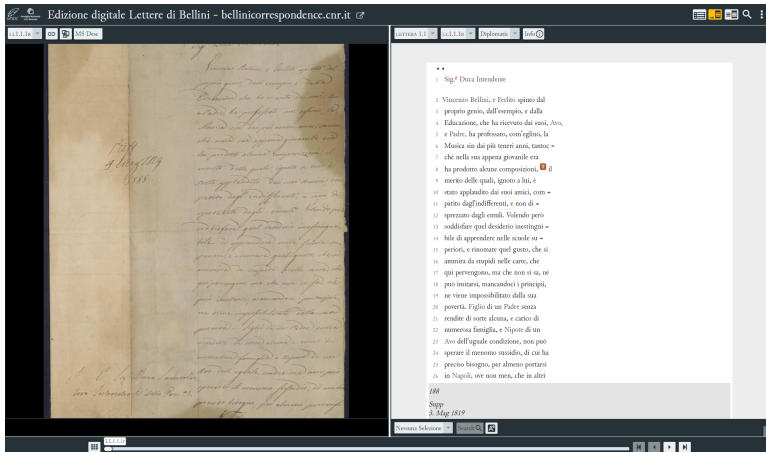
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Domain Specific Language DSL

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## XML-TEI

Verboso. Lontano dalle pratiche usuali. Difficile alla lettura.  
Sforzo Cognitivo

# Domain Specific Language

## DSL

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## DSL

Un linguaggio specifico di dominio (in inglese: Domain-Specific Language, DSL) è un linguaggio formale ottimizzato per un compito specifico

## DSL vs XML

- è conciso
- è familiare al dominio d'interesse
- è semplice da comprendere e modificare

# Domain Specific Language

## Context Free Grammar

σκευαζειντοπρος[...]



### [parser]

|        |                          |   |
|--------|--------------------------|---|
| lacuna | → L_BRA (u   opt)+ R_BRA | // textual lacuna                           |
| opt    | → L_SML_PAR u R_SML_PAR  | // optional uncertain or missing character  |
| u      | → DOT   GS_DOT           | // uncertain character or missing character |
| grcSeq | → GRC_CHAR+              | // sequence of greek characters             |
| Text   | → (grcSeq   lacuna)+;    | // text definition                          |

### [lexer]

|           |                                      |   |
|-----------|--------------------------------------|---|
| L_BRA     | → '['                                | // open in lit. ed. lacuna integrated by editors  |
| R_BRA     | → ']'                                | // close in lit. ed. lacuna integrated by editors |
| L_SML_PAR | → '{'                                | // open optional char                             |
| R_SML_PAR | → '}'                                | // close optional char                            |
| DOT       | → ('.'   '·'   '·')                  | // unreadable or uncertain char                   |
| GS_DOT    | → '\ue5ce'                           | // dot rendered by the specific font              |
| GRC_CHAR  | → [\u0370-\u03ff\u1f00-\u1fff\u2019] | // greek characters                               |



# Domain Specific Language DSL

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

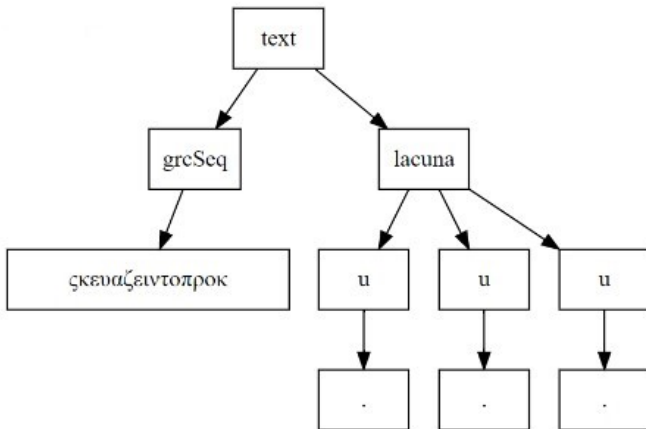
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Domain Specific Language

## Abstract Syntax Tree

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

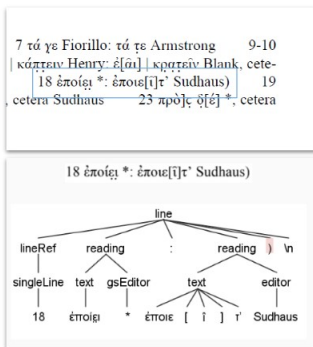
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



Philological Apparatus X

Col. 64

3-8 Sudhaus

6 [ὑπόκωφον ἦν οὐτ' η] del. librarius

7 τὰ γε Fiorillo: τὰ τε Armstrong

9-10 [μη]δέ Sudhaus: [οὐ]δέ Blank

10-11 ἐ[πι]κ[α]ρ[ω]τεῖν \*: ἐ[αἰ]κάπτειν Henry: ἐ[αἰ]κρατεῖν I

12-16 Cirillo

17-18 λε[ῖ]π[ων] \*: λέ[γ]ων Sudhaus

18 ἐποίει \*: ἐποιε[ῖ]τ' Sudhaus)

19 Cirillo\_Cpe

20 Sudhaus

21 πέφ[υ]κε Sudhaus: πέφ[η]νε Janko, cetera Sudhaus

23 πρό]ς ὁ[έ] \*, cetera Sudhaus

# Progress status

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

1 Introduzione personale

2 Panoramica Rappresentazione del Testo

3 XML, TEI guidelines e Domain Specific Languages

**4 Modelli di rappresentazione e Vector Semantics**

5 Esempi

6 Voyant-Tools

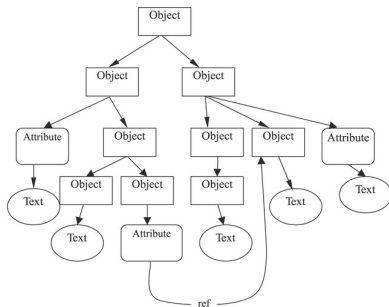
# Alcuni modelli del testo

OHCO, bag of words, set of sequence

## Modelli del testo

### Ordered Hierarchy of Content Objects (OHCO)

Un testo è formato da entità di contenuto di diverso tipo (content objects) contenute in entità più ampie di livello superiore (hierarchies) in cui la relazione d'ordine tra gli elementi è significativa (ordered) "*That is what text really is*" (Allen Renear's 1990).



# Alcuni modelli del testo

OHCO, bag of words, set of sequence

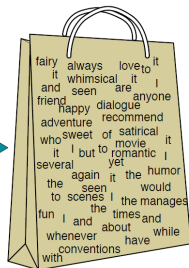
## Modelli del testo

### Bag of Words

Un testo è rappresentato dalle parole con la corrispondente frequenza di apparizione.

Conditional independence assumption (parole indipendenti tra loro)

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



|           |   |
|-----------|---|
| it        | 6 |
| I         | 5 |
| the       | 4 |
| to        | 3 |
| and       | 3 |
| seen      | 2 |
| yet       | 1 |
| would     | 1 |
| whimsical | 1 |
| times     | 1 |
| sweet     | 1 |
| satirical | 1 |
| adventure | 1 |
| genre     | 1 |
| fairy     | 1 |
| humor     | 1 |
| have      | 1 |
| great     | 1 |

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Alcuni modelli del testo

## OHCO, bag of words, set of sequence

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

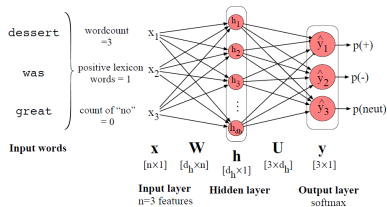
Esempi

Voyant-Tools

Generative Supervised Ap-  
proach

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} \underbrace{P(d|c)}_{\text{likelihood}} \underbrace{P(c)}_{\text{prior}}$$

Discriminative Supervised  
Approach



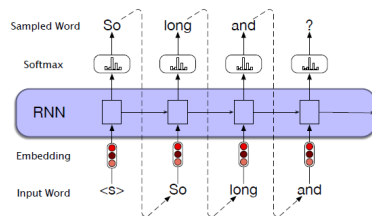
# Alcuni modelli del testo

OHCO, bag of words, set of sequence

## Modelli del testo

### Set of Sequence

Un testo è rappresentato da un insieme di sequenze di informazioni indipendenti (e.g. frasi, paragrafi). L'ordine di apparizione dei dati è significativo. Il modello è usato in applicazioni che richiedono maggiore rappresentazione semantica.

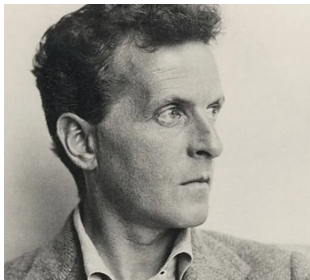


# Embeddings

## Vector Semantics

### Il significato di una parola

"Il significato di una parola è il suo uso nel linguaggio".  
"Anziché utilizzare un linguaggio logico per definire ogni parola".  
(Wittgenstein, 1953)



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Vector Semantics

Una parola può essere modellata come un vettore (lista di valori numerici)

Il significato di una parola è in relazione con la distribuzione delle proprie occorrenze nei contesti d'uso (Joos 1950, Harris 1954)

Una parola modellata come un vettore è rappresentata da un punto in uno spazio euclideo multidimensionale (Salton 1971)



# Embeddings

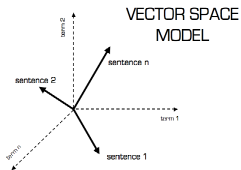
## Vector Semantics

### Vector Semantics

Il significato di una parola può essere rappresentato in uno spazio vettoriale

Parole "simili" (in relazione sintagmatica/paradigmatica) condividono contesti "simili" (co-occorrenze)

La similarità tra parole può essere misurata in termini di distanza tra punti in uno spazio vettoriale (Semantic Space Model)



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Embeddings

## Ipotesi distribuzionale

### Ipotesi Semantica Distribuzionale

Due parole che co-occorrono con parole molto simili hanno probabilmente lo stesso significato

|            | Valence | Arousal | Dominance |
|------------|---------|---------|-----------|
| courageous | 8.05    | 5.5     | 7.38      |
| music      | 7.67    | 5.57    | 6.5       |
| heartbreak | 2.45    | 5.65    | 3.58      |
| cub        | 6.71    | 3.95    | 4.24      |



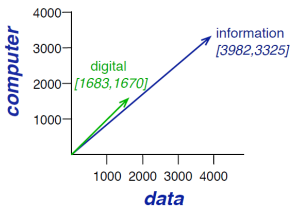
|             | computer | data | result | pie | sugar |
|-------------|----------|------|--------|-----|-------|
| cherry      | 2        | 8    | 9      | 442 | 25    |
| strawberry  | 0        | 0    | 1      | 60  | 19    |
| digital     | 1670     | 1683 | 85     | 5   | 4     |
| information | 3325     | 3982 | 378    | 5   | 13    |

# Embeddings

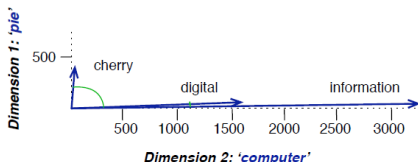
## Semantic Space Model

### Vector Semantics

I vettori che rappresentano il significato di una parola sono chiamati embeddings



$$\text{cosine}(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}}$$



# Embeddings

## Positive Pointwise Mutual Information

### PPMI

PMI è una metrica utile per identificare parole che sono fortemente associate.

Misura quanto spesso si verificano assieme due eventi rispetto a quanto ci aspetteremmo se fossero indipendenti (favorisce eventi rari).

$$\text{PPMI}_\alpha(w, c) = \max(\log_2 \frac{P(w, c)}{P(w)P_\alpha(c)}, 0)$$

|             | computer | data | result | pie  | sugar |
|-------------|----------|------|--------|------|-------|
| cherry      | 0        | 0    | 0      | 4.38 | 3.30  |
| strawberry  | 0        | 0    | 0      | 4.10 | 5.51  |
| digital     | 0.18     | 0.01 | 0      | 0    | 0     |
| information | 0.02     | 0.09 | 0.28   | 0    | 0     |

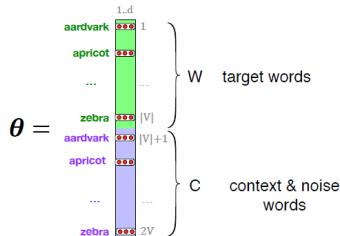
# Embeddings

## Word2vec

### Matrici/Vettori "Short" and "Dense"

Vettori con numero di dimensioni troppo elevato e con valori sparsi (molti valori a zero) non sono efficaci

*word2vec* (Mikolov et al. 2013) è un algoritmo veloce ed efficiente per generare "**word embeddings**" molto efficaci.



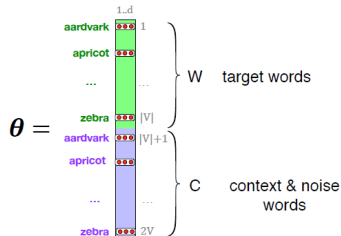
# Embeddings

## Word2vec

### word2vec

Fa uso di una rete neurale molto semplice che in fase di addestramento genera una rappresentazione embedding di ciascuna parola

$$\begin{aligned} L_{CE} &= -\log \left[ P(+|w, c_{pos}) \prod_{i=1}^k P(-|w, c_{neg_i}) \right] \\ &= -\left[ \log P(+|w, c_{pos}) + \sum_{i=1}^k \log P(-|w, c_{neg_i}) \right] \\ &= -\left[ \log P(+|w, c_{pos}) + \sum_{i=1}^k \log (1 - P(+|w, c_{neg_i})) \right] \\ &= -\left[ \log \sigma(c_{pos} \cdot w) + \sum_{i=1}^k \log \sigma(-c_{neg_i} \cdot w) \right] \end{aligned}$$



# Embeddings

## Visualizzazione

### Come Visualizzare le informazioni codificate

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

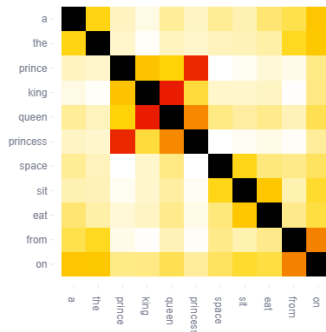
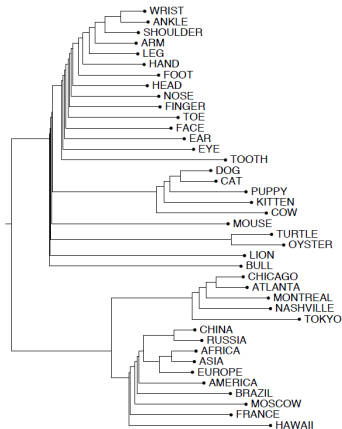
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools





# Embeddings

## Visualizzazione

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

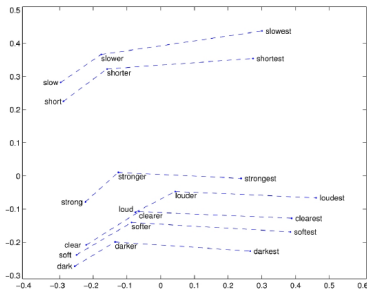
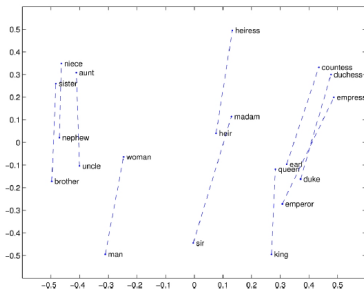
Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Analizzare la rappresentazione

### Analogie e Morfologia (Pennington et al., 2014)

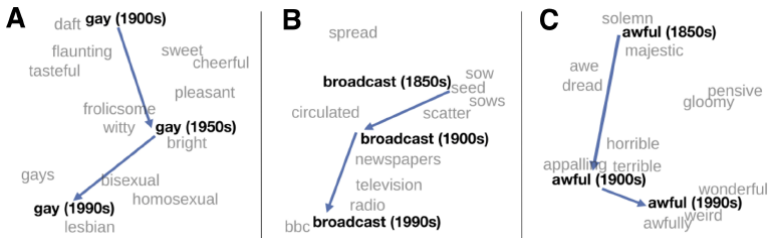


# Embeddings

## Visualizzazione

### Usare i word embeddings

Studio diacronico dello shift semantico (Hamilton et al., 2016)

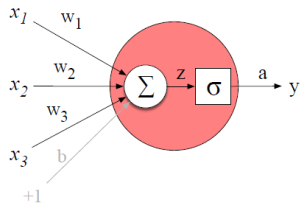
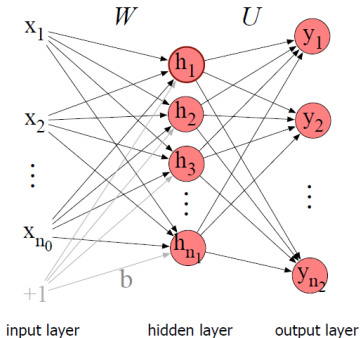


# Embeddings

## Contextual embeddings

### Reti Neurali Artificiali (ANN)

Ispirati dal funzionamento dei neuroni umani, sono una astrazione di dispositivi computazionali



$$\mathbf{h} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

$$\mathbf{z} = \mathbf{U}\mathbf{h}$$

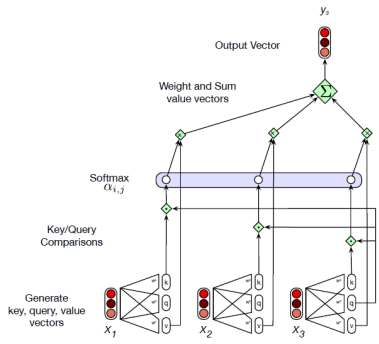
$$\mathbf{y} = \text{softmax}(\mathbf{z})$$

# Embeddings

## Contextual embeddings

### Transformer

ANN basate sul meccanismo del "self-attention": codificano informazioni rilevanti dalle parole in contesto sinistro (tempo)



$$\text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

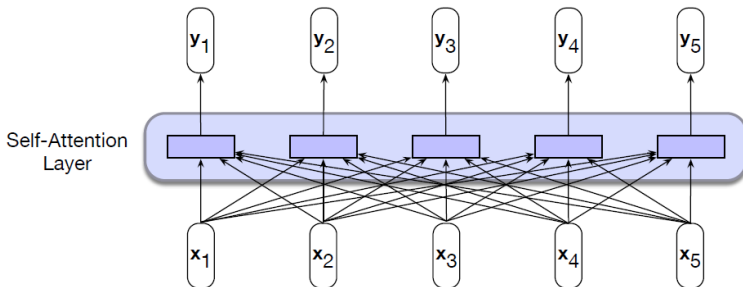
Voyant-Tools

# Embeddings

## Contextual embeddings

### BERT (Devlin et al., 2019)

Bidirectional Transformer Encoders: Generano rappresentazioni contestualizzate (sinistro-destro) degli embeddings in input



Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

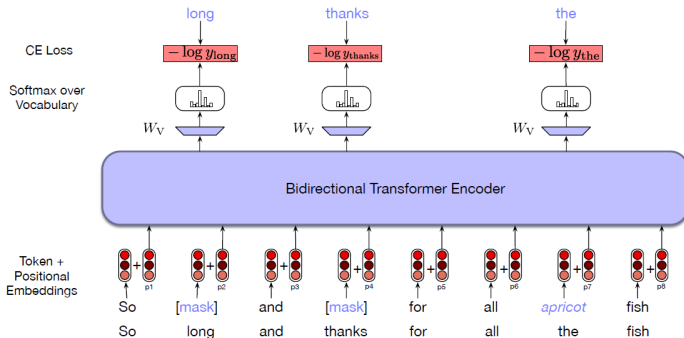
Voyant-Tools

# Embeddings

## Contextual embeddings

### Masked Language Model

Reti che vengono addestrate (pre-training) con enormi quantità di dati al fine di identificare informazioni mancanti/incoerenti dalla stringa di input



# Embeddings

## Static vs Contextual embeddings

### Word Embeddings vs Contextual Embeddings

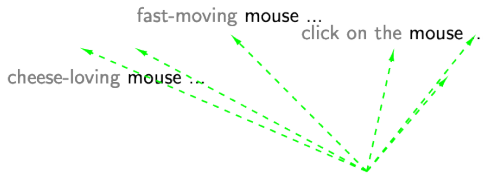
Static Embeddings: significato di un "parola tipo", Contextual Embeddings: significato di una "parola token".

Una singola rappresentazione vettoriale per ogni parola (unica) nel vocabolario (forma, parola tipo)

*Rappresentazione Statica*

Ogni parola è rappresentata da un vettore diverso ogni volta che appare in un contesto differente.

*Rappresentazione Dinamica*



## Pregiudizi e Potenziali Danni

## Gli embeddings codificano stereotipi e pregiudizi!!





# Progress status

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

**Esempi**

Voyant-Tools

1 Introduzione personale

2 Panoramica Rappresentazione del Testo

3 XML, TEI guidelines e Domain Specific Languages

4 Modelli di rappresentazione e Vector Semantics

**5 Esempi**

6 Voyant-Tools

# Embeddings

## Esempi

### Dynamic Contextual Embedding

I Contextual embeddings possono essere utilizzati come rappresentazione del significato di ogni parola in input per qualsiasi applicazione che elabora il significato delle parole.

Fill-Mask

Mask token: [MASK]

Roma è la [MASK] d'Italia.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.034 s

capitale

Capitale

città

regina

porta

JSON Output

Examples

Source Sentence

Questa è una persona felice

Sentences to compare to

Questo è un cane felice

Questa è una persona molto felice

Oggi è una giornata di sole

Add Sentence

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.320 s

Questo è un cane felice

Questa è una persona molto felice

Oggi è una giornata di sole

Maximize

0.932  
0.042  
0.004  
0.001  
0.001  
0.072  
0.976  
0.376

# Embeddings e Generative AI

## Esempi

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Codifica XML/TEI di Entità Nominate

Certo, ecco la tabella aggiornata con le parole latine specifiche che supportano ogni informazione.

### Tabelle

#### Entità Nominate

| Entità Nominate   | Tipo    | Descrizione  | Dettagli  | Parole Latine di Supporto  |
|-------------------|---------|--|---|--|
| Sebastianus Člani | Persona | Duca di Venezia, Dalmazia, e Croazia per grazia di Dio   | Fa una dichiarazione o garanzia a Romano Mayrano. Si riferisce a lui e ai suoi eredi.         | "Sebastianus Člani Dei gratia Venecie, Dalmatie atque Croatiae dux"                    |
| Romano Mayrano    | Persona | Individuo nel confinio di Sancte Trinitatis e successivamente in quello di Sancti Iohannis Evangeliste | Riceve una dichiarazione o garanzia da Sebastianus Člani. Si riferisce a lui e ai suoi eredi. | "Romano Mayrano de confinio Sancte Trinitatis in confinio Sancti Iohannis Evangeliste" |
| Dominicus Barbo   | Persona | Testimone  | Ha sottoscritto il documento.   | "Ego Dominicus Barbo testis subscripsi"  |
| Vitalis Citinus   | Persona | Testimone  | Ha sottoscritto il documento.   | "Ego Vitalis Citinus testis subscripsi"  |
| Marcus Paulinus   | Persona | Diacono e Notaio   | Ha completato e rafforzato il documento.  | "Ego Marcus Paulinus diaconus et notarius"   |

Il formato TEI XML è un modo standardizzato per rappresentare informazioni testuali in un formato strutturato. Ecco come potrebbe apparire il documento che hai fornito, formattato in XML-TEI, con le informazioni integrate utilizzando il tag `<crs>` e altri tag XML appropriati:

```
xml
Copy code
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Documento 252</title>
      </titleStmt>
      <publicationStmt>
        <p>Documento pubblicato con l'integrazione di informazioni estratte</p>
      </publicationStmt>
      <sourceDesc>
        <p>Fonte originale in latino</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <div n="252">
        <head>In nomine Domini Dei et Salvatoris nostri Iesu Christi.</head>
        <p n="1.1">Anno Domini <date when="1174">millesimo centesimo septuag</p>
        <p n="1.2">Plenan et irrevocabilem securitatem facimus nos quidem </p>
        <p n="1.3">Munc autem quia tu de quocumque continetur in supra sc</p>
        <p n="1.4">Quia nichil inde remansit unde te amplius requirere valea</p>
        <p n="1.5">Predictam namque cautionis cartulae tibi reddidimus, si e</p>
        <p n="1.6">Quod si quocumque tempore de suprascriptis omnibus capit</p>
        <p n="2.1"><signed><rs type="person" ref="#SebastianusČlani">Ego Seb</p>
        <p n="3.1"><signed><rs type="person" ref="#DominicusBarbo">Ego Domin</p>
        <p n="4.1"><signed><rs type="person" ref="#VitalisCitinus">Ego Vital</p>
        <p n="5.1"><signed>S.T. <rs type="person" ref="#MarcusPaulinus">Ego</p>
      </div>
    </body>
  </text>
</TEI>
```

# Trattamento automatico del testo

## Tokenizzazione

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### Token

Con **tokenizzazione** ci riferiamo alla segmentazione del testo e più in particolare alla definizione delle unità di testo da analizzare come parole

## Vocabolario

Con **vocabolario** ci riferiamo alla lista dei termini presenti nel testo analizzato. Si differenzia da risorse quali il **lemmario** che si riferisce ad una lista di parole ricondotte alla forma canonica. L'insieme di parole non lemmatizzate prendono il nome anche di **formario**.

# Trattamento automatico del testo

## Concordanza

### Concordanza

Con **concordanza** ci riferiamo ad una struttura derivata dallo **spoglio** di un testo ed è formata dall'insieme di parole contenute nel testo, generalmente ordinate secondo un criterio lessicografico e ciascuna parola è accompagnata da riferimenti e/o dal segmento del testo originale dove la parola occorre.

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Trattamento automatico del testo

KeywordInContext

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

## Key Word In Context

Con **KWIC** ci riferiamo ad una struttura derivata dallo **spoglio** di un testo ed è formata dall'allineamento delle forme centrate sulla stessa colonna. Le parole sono accompagnate da un frammento di contesto destro e sinistro che può essere usato anche come criterio di ordinamento.

# Trattamento automatico del testo

## Esempio MQDQ

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### MQDQ

14 occorrenze per la chiave: **tityre** [prima posizione]

Indici

Esporta

Nuova ricerca

Cercate le alternative grafiche TITYRE, TYTIRE

|                        |  |
|------------------------|--|
| VERG. ecl. 1, 1        | ▶ <b>Tityre</b> , tu patulae recubans sub tegmine fagi           |
| VERG. ecl. 3, 20       | ▶ <b>Tityre</b> , coge pecus", tu <i>post</i> carecta latebas. 📄 |
| VERG. ecl. 3, 96       | ▶ <b>Tityre</b> , pascentis a flumine reice capellas:            |
| VERG. ecl. 9, 23       | ▶ <b>Tityre</b> , dum redeo (breuis est uia) pasce capellas,     |
| VERG. georg. 4, 566    | ▶ <b>Tityre</b> , te patulae cecini sub tegmine fagi.            |
| OBTR. Verg. 1, 1       | ▶ <b>Tityre</b> , si toga calda tibi est, quo tegmine fagi       |
| CALP. SIC. ecl. 3, 19  | ▶ <b>Tityre</b> , quas dixit, salices pete solus et illinc,      |
| CALP. SIC. ecl. 4, 163 | ▶ <b>Tityre</b> , rura prius, sed post cantabimus arma."         |
| TER. MAVR. metr. 1675  | ▶ ' <b>Tityre</b> tu <i>patulae</i> ': concludit syllaba nomen,  |
| DAMAS. carm. 3, 1      | ▶ <b>Tityre</b> , tu fido recubans sub tegmine Christi           |
| ANTH. Lat. 558, 1      | ▶ <b>Tityre</b> , te Latio cecinit mea fistula uersu             |
| ANTH. Lat. 719a, 1     | ▶ <b>Tityre</b> , tu patulae recubans sub tegmine fagi,          |
| ANTH. Lat. 719a, 46    | ▶ <b>Tityre</b> , tunc alius ad caelum hinc ire putandum est     |



# Trattamento automatico del testo

espressioni regolari

## Searching

Tra le operazioni più importanti che posso essere effettuate automaticamente, la ricerca di frammenti di testo è la quella fondamentale

## Espressioni regolari

Accanto alla ricerca di base condotta per forma ortografica letterale, la ricerca con espressioni regolari fornisce un potente strumento di indagine e di editing (**find and replace**)

**Le espressioni regolari definiscono "pattern" utilizzati per descrivere combinazioni di caratteri in stringhe.**

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Trattamento automatico del testo

## espressioni regolari

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

REGULAR EXPRESSION

4 matches (0.3ms)

" ([A-Za-z])+ate+([pa])e "

gm

TEST STRING

Voi+che+vivate+sicuri+  
Nelle+vostre+tiepide+case,+  
Voi+che+trovate+tornando+a+sera+  
Il+cibo+caldo+e+visi+amici:+  
Considerate+se+questo+è+un+uomo+  
Che+lavora+nel+fango+  
Che+non+conosce+pace+  
Che+lotta+per+mezzo+pane+  
Che+muore+per+un+sì+o+per+un+no+[...]

EXPLANATION

▼ " ([A-Za-z])+ate+([pa])e " gm

▼ 1st Capturing Group ([A-Za-z])+ate+([pa])e

▼ 1st Alternative [A-Za-z]+ate+

▼ Match a single character present in the list below [A-Za-z]

matches the previous token between one and unlimited time

matches a single character in the range between A (index 1) and z (index 26)

matches a single character in the range between a (index 27) and z (index 52)

ate+ matches the characters ate+ literally (case sensitive)

▼ 2nd Alternative ([pa])e

▼ pa matches the characters pa literally (case sensitive)

matches any character (except for line terminators)

matches the character e with index 101 (65<sub>10</sub> or 145<sub>8</sub>) literally

▼ Global pattern flags

g modifier: global. All matches (don't return after first match)

m modifier: multi line. Causes ^ and \$ to match the begin/end of each line (not just the whole document's begin/end)

# Trattamento automatico del testo

## XPath

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

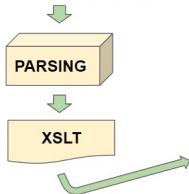
## XPath, XSLT, XQUERY

XPath, XSLT, XQUERY sono linguaggi formali fondamentali per elaborare documenti XML.

# Trattamento automatico del testo

## Esempio DSL

50 alegranza V] #CE alegreza (Mussafia, Casini). 51 saver] salvare V. ciertanza V] #CE cierteza (Mussafia, Casini). 53 qual] quale V. 54 #CE gientilanza (Menichetti)] gientileza V. 55 #CE agrestanza (Menichetti)] agresteza V.



```
<listApp>
  <app loc="50">
    <lem wit="#V">alegranza</lem>
    <rdg type="substantive" resp="#Mussafia #Casini">alegreza</rdg>
  </app>
  <listApp>
    <app loc="51">
      <lem>saver</lem>
      <rdg wit="#V">salvare</rdg>
    </app>
    <app loc="51">
      <lem wit="#V">ciertanza</lem>
      <rdg type="substantive" resp="#Mussafia #Casini"></rdg>
    </app>
  </listApp>
  <app loc="53">
    <lem>qual</lem>
    <rdg wit="#V">quale</rdg>
  </app>
  <app loc="54">
    <lem type="substantive" resp="#Menichetti">gientilanza</lem>
    <rdg wit="#V">gientileza</rdg>
  </app>
  <app loc="55">
    <lem type="substantive" resp="#Menichetti">agrestanza</lem>
    <rdg wit="#V">agresteza</rdg>
  </app>
</listApp>
```

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

# Trattamento automatico del testo

## Esempio DSL

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

```
<body><h3>Congetture editoriali</h3>
  <ul>{
    for $lem in $doc//tei:app/tei:lem[@type="substantive"]
    let $loc := string($lem/parent::node()/@loc)
    return <li>
      <span style="font-weight:bolder">lemma: </span>
      <span><u>{string($lem)}</u></span>
      <span> {{substring-after($lem/@resp,'#')}} </span>
      <span> al verso </span>
      <span><b>{$loc}</b></span>
      <span>: </span>
      <span><i>{string($doc//tei:l[@n=$loc])}</i></span>
    <ul><li>il manoscritto
      <b>{
        substring-after(
          $lem/following-sibling::tei:rdg/@wit,'#')
        }</b>
        riporta <u>{string(
          $lem/following-sibling::tei:rdg)}</u>
      </li></ul></li>
  }</ul></body>
```

# Trattamento automatico del testo

## Esempio DSL

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

trova i casi in cui la mia bozza di edizione prevede una congettura editoriale

### Congetture editoriali

- lemma: gentilanza (Menichetti) al verso 54: *ma piacie tanto a vostra gentilanza*
  - il manoscritto V riporta gentileza
- lemma: agrestanza (Menichetti) al verso 55: *di me dare agrestanza,*
  - il manoscritto V riporta agresteza

trova i casi che nella mia bozza di edizione sono considerati lectiones difficiliores

### Lectiones difficiliores

- lezione gentilanza diff. rispetto a gentilezza
- lezione agrestanza diff. rispetto a agresteza

# Progress status

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

1 Introduzione personale

2 Panoramica Rappresentazione del Testo

3 XML, TEI guidelines e Domain Specific Languages

4 Modelli di rappresentazione e Vector Semantics

5 Esempi

**6 Voyant-Tools**

# Voyant Tools

Text Analysis Web Environment

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



## Voyant Tools

Voyant Tools è un ambiente di lettura e analisi dei testi web-based di facile utilizzo. È un progetto scientifico concepito per facilitare la lettura e le pratiche interpretative per gli studiosi, gli studenti e per il pubblico in generale.

*idea e sviluppo di Stéfán Sinclair e Geoffrey Rockwell*



## Outline

- Informazioni di base su Voyant-Tools
- I pannelli predefiniti e le loro interazioni
- Approfondimento delle funzionalità di Voyant
- Esempio di caricamento di un testo da file
- Direzioni future

# Voyant Tools

## Hermeneutica

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

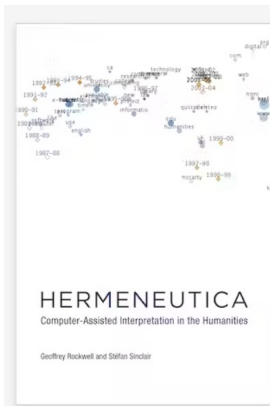
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



## Hermeneutica

Computer-Assisted Interpretation in the Humanities

By Geoffrey Rockwell and Stéfan Sinclair

256 pp., 6 x 9 in, 65 b&w illus., 7 tables

Paperback

ISBN: 9780262545891

Published: June 7, 2022

Publisher: The MIT Press

Penguin Random House

Amazon

Barnes and Noble

# Voyant Tools

Hermeneutica - <https://voyant-tools.org/>

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

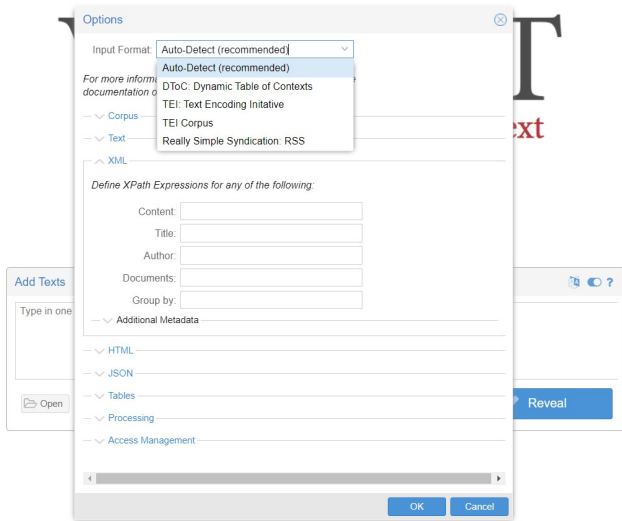
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



# Voyant Tools

## SKIN

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

- La **skin predefinita** include i tools: Cirrus, Reader, Trends, Summary, Contexts;
- **Bubblelines Skin**: lo strumento bubblelines combinato con gli strumenti contexts e reader.
- **Collocates Skin**: termini del corpus, termini dei documenti, collocazioni del corpus, contesti e grafico delle collocazioni
- **Scatter Skin**: grafico di dispersione (PCA e CA) con documenti, tendenze e contesti
- **Subset Skin**: un'interfaccia specializzata per selezionare un sottoinsieme di documenti dal corpus
- **Skin Dynamic TOC**: Il Dynamic Table of Contexts Browser è un ambiente di lettura online che consente di effettuare ricerche complesse combinando vari indici

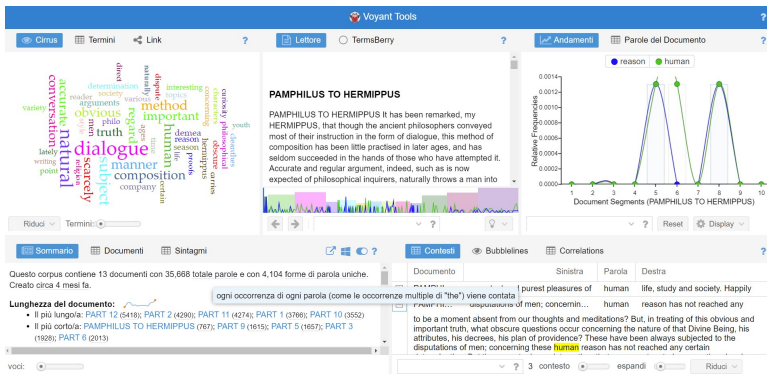
**skin: disposizione e layout degli strumenti**

## Hermeneutica - Skin Predefinita

## Creazione e Analisi dei Testi Digitali

A.M. Del  
Grosso

## Voyant-Tools



# Voyant Tools

## Hermeneutica - Bubblelines Skin

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

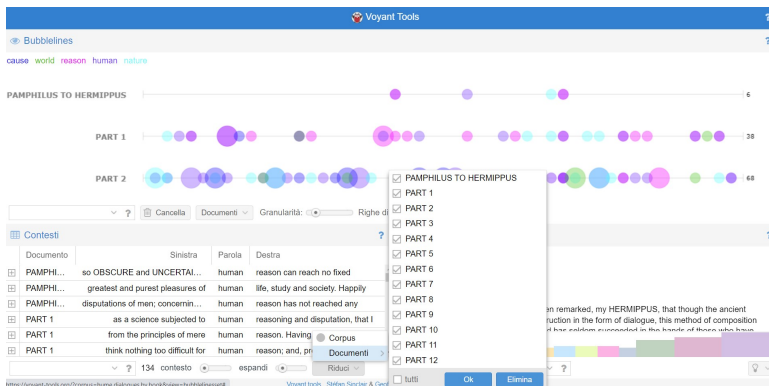
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



`view=bubblelineset`

# Voyant Tools

## Hermeneutica - Collocates Skin:

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

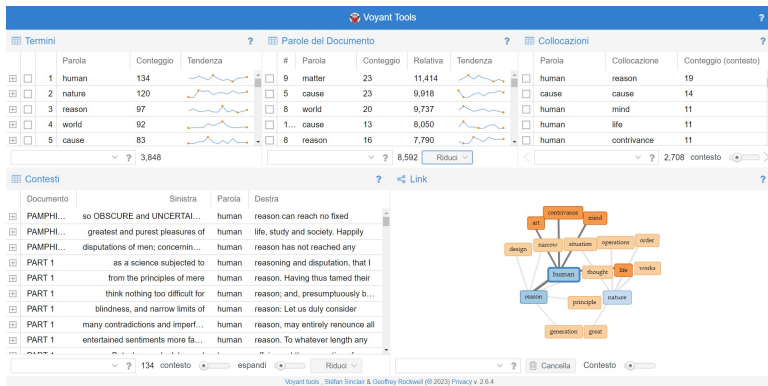
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



view=collocateset





# Voyant Tools

## Hermeneutica - Subset Skin

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

Questo corpus contiene 21 documenti con 8,886,412 totale parole e con 200,310 forme di parola uniche. Creato circa 6 mesi fa.

Titoli:  21 ?

Testo intero:  ?

- txt\_2007 (1)
- txt\_2006 (1)
- txt\_2005 (1)
- txt\_2004 (1)
- txt\_2003 (1)
- txt\_2002 (1)
- txt\_2001 (1)
- txt\_2000 (1)

di Voyant

Scarica il file

Sintassi di ricerca per i documenti (premere invio/return per far partire una ricerca):

- < ul style ="margin-top:3px ; margin-bottom:3px;">
- **cappotto**:trova termine esatto *cappotto*
- **cappotto \***:trova termini che iniziano con *cappotto*
- **cappotto, giacca**:trova ogni termine separato da virgola come termini separati
- **"cappotto invernale"**: *cappotto invernale* come un sintagma
- < b>"guanti cappotto"~ 5: *cappotto* vicino a *guanti* (entro 5 parole)
- < li> + **inverno + cappotto**: trova ogni termine preceduto da un segno più ( + )
- < b> + "cappotto invernale"+ **guanto**": combina le sintassi

view=subset

# Voyant Tools

## Hermeneutica - Dynamic TOC

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

### Dynamic Table of Contents | Regenerations / Régénérations

Canadian Women's Writing / Écriture des femmes au Canada



**Index**   Tags   Stats

Filter    Filter by Chapter ▾

- Connivance (Brossard) (1)
- consumer identity and globalization (3)
- consumer vs. gift-giving culture (1)
- ↳ converts. (1)
- Couser, G. Thomas (1)
- Coward, Rosalind (1)
- Cowley, Malcolm (1)
- ↳ creative process (1)
- ↳ creative writing programs (1)
- ↳ collaborative writing (1)
- history of (4)
- as support networks (11)
- Critically Speaking (Livesay) (1)
- The Crooked Rib (Folksong) (2)
- Cropper (Bergvall) (1)
- ↳ culture (1)
- The Cure for Death By Lightning (Dargatz) (1)
- Curiously "My" Writing (Moure and Averilichian) (2)

**Table of Contents**

Contents / Table des matières

Preface / Préface

Au fil de la narration et des générations

**Women Who Invite Collaboration**

A Modernist Commons in Canada / Une commune moderniste au Canada

L'apport du concept d'ontologie informatique à l'étude de l'histoire littéraire des femmes au XIX<sup>e</sup> siècle

Reclaiming the Past. Nurturing the Future

Du métatféminisme et des histoires au féminin

Gothic Romance in Quebec

Readers Reading Margaret Laurence, 1965-1980

Être écrivain en contexte minoritaire franco-canadien (1970-1989)

Crooked Ribs, Modern Martyrs, and Dull Days

Common Measure

Clear

**Susan Rudy: Women Who Invite Collaboration**

across such a broad span of time" (361). Clearly, the Canadian Writing Research Collaboratory/Collaboratoire scientifique des écrits du Canada (CWRC/CSÉC) follows in this mode.<sup>4</sup>

But my focus is on alternative collaborative modes in innovative women's writing communities. A bridge from Ede and Lunsford's argument to my own may be found in their attention to the work of Kathy Acker, whom they describe as a "self-styled autoplagiarist" (362). In Acker's work they find a practical example of the "theory that language cannot be owned—particularly not by a coherent, autonomous, individual author" (363). Also useful is Ede and Lunsford's discussion of an "invitational rhetoric" that "grows out of collaborative principles," one that offers an alternative to the "agonistic individualism characteristic of so much academic writing" (360). As much as I admire and have learned from their research, I find it a bit unsettling that despite their call for ways of thinking about writing that are genuinely collaborative, what they really want to see is work that is being written by two coherent, autonomous individuals—two authors, rather than one. Their collaborative writing is still, to paraphrase the title of their book, a "singular text", but it is now written by "plural authors".

**Breaking the Authorial Hold: Collaboration as Connivance**

In the work of innovative women's writing in Canada, there is both a critique of the author and an enactment of an alternative: a collaborative space of friendship, networking, and conviviality that Nicole Brossard calls "connivance." In an article on Brossard's *Baroque at Dawn*, Jodi Lundgren

<https://voyant-tools.org/dtoc/>

# Hermeneutica

## Voyant-Tools



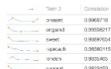
Bubblelines visualizes the frequency and distribution of terms in a corpus.



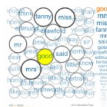
Cirrus is a word cloud that visualizes the top frequency words of a corpus or document.

| Term | Collocate | Count (non) |
|------|-----------|-------------|
| soon | mr        | 62          |
| soon | said      | 57          |
| soon | mrs       | 41          |
| soon | possible  | 36          |
| soon | away      | 32          |
| soon | miss      | 30          |
| soon | began     | 21          |
| soon | came      | 21          |

Corpus Collocates is a table view of which terms appear more frequently in proximity to keywords across the entire corpus.



The Correlations tool enables an exploration of the extent to which term



Bubbles is a playful visualization of term frequencies by document.



Collocates Graph represents keywords and terms that occur in close proximity as a force directed network graph.

| Left            | Term | Right      |
|-----------------|------|------------|
| cession of a    | good | fortune, m |
| and pound...    | good | luck to ca |
| e the general   | good | opinion of |
| into indepe...  | good | livings--a |
| tema, with a    | good | temper, a  |
| erikable, wi... | good | constitut  |
| pact and it...  | good | looks and  |
| vous of ev...   | good | Quality a  |

The Contexts (or Keywords in Context) tool shows each occurrence of a keyword with a bit of surrounding text (the context).

| # | Term      | Count | Relative | Trend |
|---|-----------|-------|----------|-------|
| 6 | ret       | 1153  | 7.502    |       |
| 4 | ret       | 785   | 6.434    |       |
| 7 | catherine | 479   | 5.500    |       |
| 8 | anna      | 447   | 5.340    |       |
| 3 | olivia    | 423   | 5.198    |       |
| 5 | larry     | 410   | 5.060    |       |

Document Terms is a table view of document term frequencies.

# Voyant Tools

## Hermeneutica

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

| Term  | Count | Trend |
|-------|-------|-------|
| me    | 3117  |       |
| me    | 2531  |       |
| said  | 2165  |       |
| miss  | 1942  |       |
| think | 1514  |       |
| know  | 1450  |       |
| good  | 1444  |       |

### Corpus Terms

Corpus Terms is a table view of term frequency in the entire corpus.



### Knots

Knots is a creative visualization that represents terms in a single document as a series of twisted lines.



### MicroSearch

Microsearch visualizes the frequency and distribution of terms in a corpus.

### 1790 Love And Freindshi

LOVE AND FREINDSHIP AND  
(Love And Freindship And Oth  
A Collection of Juvenile Writing  
By Jane Austen  
Transcriber's Note: A few very  
version: Italics have been conv



### Reader

The Reader tool provides a way of reading documents in the corpus, text is fetched on-demand as needed.

| Documents           |         |
|---------------------|---------|
| Title               | Words   |
| 1790 Love And F...  | 33,559  |
| 1805 Lady Susan...  | 23,149  |
| 1811 Sense and ...  | 119,957 |
| 1813 Pride and P... | 122,204 |
| 1814 Mansfield P... | 160,461 |
| 1815 Emma           | 107,096 |

### Documents

The Documents tool shows a table of the documents in the corpus and includes functionality for modifying the corpus.



### Mandala

Mandala is a conceptual visualization that shows the relationships between terms and documents.

| Term | Count | Length | Trend |
|------|-------|--------|-------|
| and  | 14    | 3      |       |
| for  | 14    | 3      |       |
| by   | 14    | 2      |       |
| at   | 13    | 3      |       |
| in   | 12    | 4      |       |
| the  | 12    | 4      |       |
| of   | 12    | 2      |       |
| a    | 10    | 3      |       |
| in   | 10    | 3      |       |
| at   | 10    | 3      |       |
| of   | 10    | 3      |       |
| and  | 10    | 3      |       |
| the  | 10    | 3      |       |

### Phrases

The Phrases tool shows repeating sequences of words organized by frequency of repetition or number of words in each repeated phrase.



### Scatterplot

ScatterPlot is a graph visualization of how words cluster in a corpus using document similarity, correspondence analysis or principal component

# Voyant Tools

## Hermeneutica

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

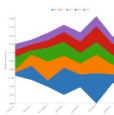
Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

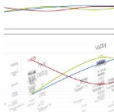
Esempi

Voyant-Tools



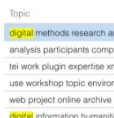
### StreamGraph

StreamGraph is a visualization that depicts the change of the frequency of words in a corpus (or within a single document).



### Terms Radio

TermsRadio is a visualization that depicts the change of the frequency of words in a corpus (or within a single document).



### Topics

he Topics tool provides a rudimentary way of generating term clusters from a document or corpus and then seeing how each topic (term cluster) is distributed across the document or corpus.

### Veliza

Veliza is an experimental tool for having a (limited) natural language exchange (in English) based on your corpus.

- This corpus has 8 documents with 751,746
- **Largest documents (by words)**
  - 1513 Emma (1,122,000)
  - 1514 Elizabethan Prose (945,481)
  - Shortest documents
    - 1503 Lady Susan (25,749)
    - 1753 Love And Friendship (35,555)
- **Highest vocabulary density**
  - 1923 Lady Susan (3,127)
  - 1782 Love And Friendship (3,126)
- **Lexical density**
  - 1513 Emma (0.943)
  - 1514 Elizabethan Prose (0.853)

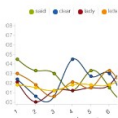
### Summary

The Summary provides a simple, textual overview of the current corpus, including including information about words and documents.



### TextualArc

TextualArc is a visualization of the terms in a document that includes a weighted centroid of terms and an arc that follows the terms in document order.



### Trends

Trends shows a line graph depicting the distribution of a word's occurrence across a corpus or document.

### Word Tree

Word Tree is a tool that allows you to explore how words are used in phrases.



# Voyant Tools

## Hermeneutica

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

Voyant Tools

Sommario

Questo corpus contiene 4 documenti con 35,640 totale parole e con 4,098 forme di parola uniche. Creato circa 4 mesi fa.

**Lunghezza del documento:**

- Il più lungo/a: [Philo](#) (23629); [Cleanthes](#) (6987)
- Il più corto/a: [Pamphilus](#) (1057); [Demea](#) (3967)

**Densità del vocabolario:**

- Maggiore: [Pamphilus](#) (0.425); [Demea](#) (0.302)
- Minore: [Philo](#) (0.140); [Cleanthes](#) (0.226)

**Average Words Per Sentence:**

- Maggiore: [Pamphilus](#) (36.4); [Philo](#) (30.2)
- Minore: [Demea](#) (24.9); [Cleanthes](#) (29.6)

**Readability Index:**

- Maggiore: [Pamphilus](#) (16.745); [Cleanthes](#) (14.903)
- Minore: [Philo](#) (14.290); [Demea](#) (14.417)

**Parole più frequenti nel corpus** [human](#) (134); [nature](#) (120); [reason](#) (97); [world](#) (92); [cause](#) (83); [universe](#) (75); [order](#) (72); [great](#) (68); [man](#) (62); [life](#) (62); [religion](#) (61); [experience](#) (61); [cleanthes](#) (60); [mind](#) (59); [natural](#) (58)

**Parole caratteristiche** (in relazione al resto del corpus)

- [Cleanthes](#): [voice](#) (6), [objections](#) (11), [pretended](#) (5), [cause](#) (23), [staff](#) (4), [speech](#) (4), [species](#) (8), [contrary](#) (8), [analogy](#) (8), [mind](#) (18), [hypothesis](#) (7), [philo](#) (16), [twice](#) (3), [sceptics](#) (6), [maintain](#) (6).
- [Demea](#): [son](#) (3), [bath](#) (3), [silly](#) (13), [succession](#) (10), [mind](#) (10), [cleanthes](#) (10), [united](#) (4), [students](#) (2), [seemingly](#) (2), [turning](#) (2), [reverence](#) (2), [profaneness](#) (2), [pestilence](#) (2), [infidels](#) (2), [lie](#) (4).
- [Pamphilus](#): [dialogue](#) (5), [conversations](#) (5), [philo](#) (10), [book](#) (2), [reader](#) (2), [hermippus](#) (2), [cleanthes](#) (9), [important](#) (3), [demea](#) (7), [worse](#) (1), [vivacity](#) (1), [vehement](#) (1), [unwaried](#) (1), [unities](#) (1), [intenses](#) (1).
- [Philo](#): [world](#) (79), [operations](#) (15), [commonly](#) (14), [analogy](#) (27), [universe](#) (61), [species](#) (24), [position](#) (12), [hypothesis](#) (22), [cause](#) (51), [experience](#) (50), [equal](#) (10), [contrivance](#) (19), [sound](#) (9), [kind](#) (18), [goodness](#) (9).

voci:

# Hermeneutica

## Creazione e Analisi dei Testi Digitali

A.M. Del  
Grosso

## Voyant-Tools

| Contesti  |  |            |   |
|---|--|------------|---|
| Documento   | Sinistra   | Parola     | Destra ↑  |
| PART 10   | a most dangerous issue, and are unaware introducing a total          | scepticism | into the most essential articles of natural and revealed theology |
| PART 1  | elaborate reasoning to prove and establish it. This species of       | scepticism | is fatal to knowledge, not to religion; since we find             |
| PART 1  | We must be far removed from the smallest tendency to                 | scepticism | not to be apprehensive, that we have here got quite               |
| PAMPHI...   | opposed the accurate philosophical turn of CLEANTHES to the careless | scepticism | of PHILO, or compared either of their dispositions with the       |
| PART 1  | use which BAYLE and other libertines made of the philosophical       | scepticism | of the fathers and first reformers, still further propagated the  |
| PART 2  | there is no need of having recourse to that affected                 | scepticism | so displeasing to you, in order to come at this                   |
| PART 3  | that this religious argument, instead of being weakened by that      | scepticism | so much affected by you, rather acquires force from it            |
| PART 1  | they endeavoured, as is pretended, to extend, throughout, the same   | scepticism | which they had learned from the declamations of their schools     |
| PART 1  | to me, from the history of the religious and irreligious             | scepticism | with which you have entertained us. It appears to me              |
| PART 1  | it is impossible for him to persevere in this total                  | scepticism | , or make it appear in his conduct for a few                      |
| PART 1  | mankind? There is indeed a kind of brutish and ignorant              | scepticism | , as you well observed, which gives the vulgar a general          |
| PART 3  | to trace them further. But this is neither dogmatism nor             | scepticism | , but stupidity: a state of mind very different from your         |
| removes set the objects at such a distance, that to him they are lost in darkness and confusion; nor is he actuated by any curiosity to trace them further. But this is neither dogmatism nor <b>scepticism</b> , but stupidity: a state of mind very different from your sifting, inquisitive disposition, my ingenious friend. You can trace causes from effects: You can compare the most distant and remote objects: and your greatest errors |  |            |   |
| PART 11   | never prove these attributes. Such a conclusion cannot result from   | scepticism | , but must arise from the phenomena, and from our confidence      |
| PART 1  | whatever length any one may push his speculative principles of       | scepticism | , he must act, I own, and live, and converse, like                |
| PART 1  | that very suspense or balance, which is the triumph of               | scepticism | . But I observe with regard to you, PHILO, and all                |

scepticism x
21 contesto
espandi
Riduci

# Hermeneutica

A.M. Del  
Grosso

## Voyant-Tools



# Voyant Tools

## Hermeneutica

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools

Voyant Tools

Lettore

in a manner, into company; and unites the two greatest and purest pleasures of human life, study and society. Happily, these circumstances are all to be found in the subject of NATURAL RELIGION. What truth so obvious, so certain, as the being of a God, which the most ignorant ages have acknowledged, for which the most refined geniuses have ambitiously striven to produce new proofs and arguments? What truth so important as this, which is the ground of all our hopes, the surest foundation of morality, the firmest support of society, and the only principle which ought never to be a moment absent from our thoughts and meditations? But, in treating of this obvious and important truth, what obscure questions occur concerning the nature of that Divine Being, his attributes, his decrees, his plan of providence? These have been always subjected to the disputations of men; concerning these human reason has not reached any certain determination. But these are topics so interesting, that we cannot restrain our restless inquiry with regard to them; though nothing but doubt, uncertainty, and contradiction, have as yet been the result of our most accurate researches. This I had lately occasion to observe, while I passed, as usual, part of the summer season with CLEANTHES, and was present at those conversations of his with PHILO and DEMEA, of which I gave you lately some imperfect account. Your curiosity, you then told me, was so excited, that I must, of necessity, enter into a more exact detail of their reasonings, and display those various systems which they advanced with regard to so delicate a subject as that of natural religion. The remarkable contrast in their characters still further raised your expectations; while you opposed the accurate philosophical turn of CLEANTHES to the careless scepticism of PHILO, or compared either of their dispositions with the rigid inflexible orthodoxy of DEMEA. My youth rendered me a mere auditor of their disputes; and that curiosity, natural to the early season of life, has so deeply imprinted in my memory the whole chain and connection of their arguments, that, I hope, I shall not omit or confound any considerable part of them in the recital.

### PART 1

PART 1 After I joined the company, whom I found sitting in CLEANTHES's library, DEMEA paid CLEANTHES some compliments on the great care which he took of my education, and on his unwearied perseverance and constancy in all his friendships. The father of PAMPHILUS was your intimate friend; the son is your pupil; and may indeed be regarded as your adopted son, were we to judge by the pains which you bestow in conveying to him every useful branch of literature and science. You are no more wanting, I am persuaded, in prudence, than in industry. I shall, therefore, communicate to you a maxim, which I have observed with regard to frequency nel documento: 18 how far it agrees with your practice. The method I follow in their education is founded on the saying of an ancient, "That students of philosophy ought first to learn logics, then the nature of the gods." This science of natural theology, according to him, being the most profound and abstruse of any, required the maturest judgement in its students; and none out of a mind ennobled with all the other sciences, can safely be entrusted with it. Are you so late in teaching your children the principles of religion? Is there no danger of their neglecting, or rejecting altogether those opinions of which they have heard so little during the whole course of their education? It is only as a science subjected to human reasoning and disputation, that I postpone the study of Natural Theology. To season their minds with early piety, is my chief care; and by continual precept and instruction, and I hope too by example, I imprint deeply on their tender minds an habitual reverence for all the principles of religion. While they pass through every other science, I still remark the uncertainty of each part; the eternal disputations of men; the obscurity of all philosophy; and the strange, ridiculous conclusions, which some of the greatest geniuses have derived from the principles of mere human reason. Having thus tamed their mind to a proper submission and



# Voyant Tools

## Hermeneutica

Creazione e  
Analisi dei  
Testi Digitali

A.M. Del  
Grosso

Introduzione  
personale

Panoramica  
Rapp-  
resentazione  
del Testo

XML, TEI  
guidelines e  
Domain  
Specific  
Languages

Modelli di  
rapp-  
resentazione e  
Vector  
Semantics

Esempi

Voyant-Tools



### search

- **love**: corrispondenza esatta di parole
- **lov\***: tutte le parole che iniziano con "lov"
- **^lov\***: separa tutte le parole che iniziano con "lov"
- **love|loves**: combina tutte le parole separate da una pipe (OR)
- **love,loves**: combina tutte le parole separate da una virgola (AND)
- **"he loves"**: corrispondenza esatta della frase
- **"she he love\*" 10**: tutte le parole entro una finestra di 10 parole

Grazie

Creazione e Analisi dei Testi Digitali

Angelo Mario Del Grosso

Istituto di Linguistica Computazionale "A. Zampolli", Pisa

24th November 2023



**AICA**  
Associazione Italiana  
per l'Informatica ed  
il Calcolo Automatico

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE

