



UNIVERSITÀ DEGLI STUDI DI PERUGIA  
FACOLTÀ DI INGEGNERIA

Tesi di Laurea in  
INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE

## Generazione automatica di Word Cloud dinamiche

Relatore

*Prof. Carla Binucci*

Candidato

*Enrico Spataro*

Co-relatore

*Prof. Walter Didimo*

Anno Accademico 2014/2015

*EVENTUALE DEDICA*

# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
<b>2</b>	<b>Architettura del sistema</b>	<b>6</b>
<b>3</b>	<b>Creazione word cloud</b>	<b>7</b>
3.1	Estrazione keywords . . . . .	7

## Elenco delle figure

## Elenco delle tabelle

## Capitolo 1

# Introduzione

## Capitolo 2

# Architettura del sistema

Questo capitolo illustra in maniera piuttosto generica l'architettura del sistema, la quale consiste di 3 passi principali:

1. Elaborazione dei dati.
2. Generazione delle word clouds.
3. Morphing tra successive word clouds.

Ogni fase si compone di diversi passi, che verranno approfonditi nei prossimi capitoli.

### **Elaborazione dei dati**

## Capitolo 3

# Creazione word cloud

La creazione di una word cloud consiste di diverse fasi, a partire dall'estrazione delle parole chiave (*keywords*), alle quali viene assegnato un punteggio relativo all'importanza delle parole all'interno del testo, per poi calcolare la correlazione semantica tra le parole, fino ad arrivare agli algoritmi di visualizzazione delle word clouds, i cui input sono proprio le parole estratte. Questo capitolo analizza ognuna di queste fasi, esponendo quelle che, allo stato dell'arte, sono le tecniche utilizzate in questa tesi per l'estrazione delle keywords e la loro visualizzazione.

### 3.1 Estrazione keywords

L'estrazione delle keywords avviene da un input testuale, che nel nostro caso è costituito dalle trascrizioni dei discorsi che si tengono nella conferenza *TED(Technology Entertainment Design)* [?]. Per l'estrazione dei sottotitoli è stato utilizzato il tool online *ted2srt* [?].