# Generazione automatica di word cloud dinamiche

Ingegneria Informatica e dell'Automazione

Università degli Studi di Perugia Facoltà di Ingegneria

> Candidato: Enrico Spataro Relatori: Prof.ssa Carla Binucci Prof. Walter Didimo

> > 19.02.2016



## Contenuti

- 1 Introduzione
- Creazione word cloud Word cloud statica Word cloud dinamica
- 3 Risultati sperimentali
- 4 Conclusioni



## Obiettivi

#### Mostrare l'evoluzione di un testo tramite:

- creazione word cloud semantiche ad intervalli regolari;
- animazioni tramite morphing.



## Cos'è una word cloud?

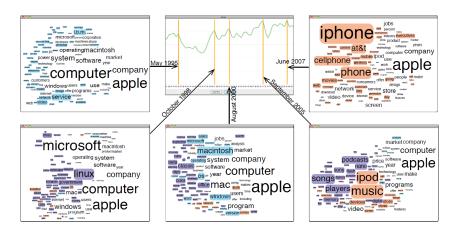




#### Word cloud semantiche

interrogator ESP kind mistakes identification guess Christmas Lovelace Babbage pupil fault clairvoyance telepathy criterion satisfy moment circumstances
Turing solipsist analogy infinite discrete-state **skin** really viva voce think inclined surprise fallacies suppose Yes conjectures storage random machine wheel imperatives and the surprise fallacies suppose Yes conjectures storage capacity wheel imperatives and the surprise fall surpr pile neutron disabilities imitation think inclined game behaviour man argument quite clicks homework answer wrong try instance convenient packets store instruction obey computer





# Morphing

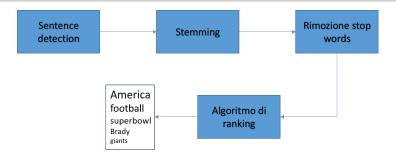




- Estrazione parole chiave
- Calcolo similarità
- Clustering
- Disegno word cloud

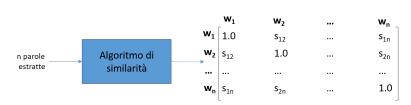


- Estrazione parole chiave
- Calcolo similarità
- Clustering
- Disegno word cloud





- Estrazione parole chiave
- 2 Calcolo similarità
- Clustering
- 4 Disegno word cloud



$$s_{ij}$$
 =  $s_{ji}$  ,  $s_{ij}$   $\in$   $[0, 1]$ 

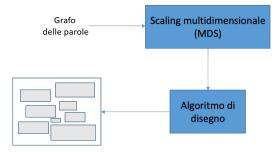


- 1 Estrazione parole chiave
  2 Calcolo similarità
  3 Clustering
  4 Disegno word cloud
  - Algoritmo di clustering



Word cloud statica

- Estrazione parole chiave
- Calcolo similarità
- Clustering
- 4 Disegno word cloud





00000

Obiettivo: create K word cloud dal testo, si vuole passare da una word cloud alla successiva in modo graduale tramite tecniche di morphing.



000000

Word cloud dinamica

# Morphing

#### Caratteristiche

- Numero di frame
- Gestione stato delle parole
- Gestione del colore delle parole



#### Le parole, tra una word cloud e la successiva, possono:

- scomparire
- apparire
- rimanere nel layout (variando posizione)



Le parole, tra una word cloud e la successiva, possono:

- scomparire







Le parole, tra una word cloud e la successiva, possono:

- apparire

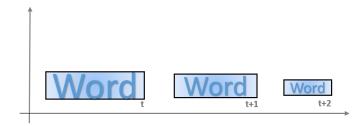






Le parole, tra una word cloud e la successiva, possono:

- scomparire
- apparire
- rimanere nel layout (variando posizione)

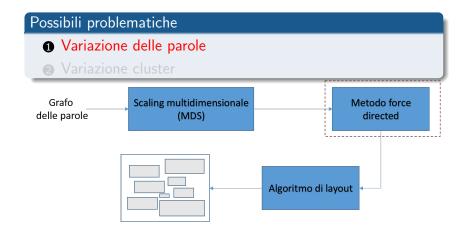




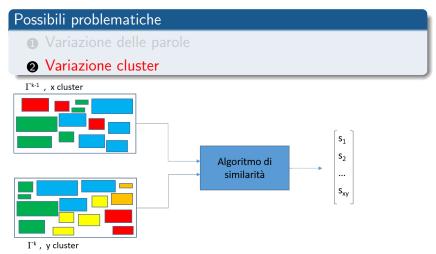
#### Possibili problematiche

- Variazione posizione delle parole
- Variazione dei cluster









Inserire video



## Risultati sperimentali

#### Test suite:

- 200 discorsi estratti (file .txt) dalla conferenza TED.
- Lunghezza media 17/18 minuti.
- 4 campionamenti per ogni testo.
- Parole estratte: 20,40,60



#### Metriche adottate:

• Combination metric  $\Rightarrow \nu = \frac{1}{K} \sum_{k=1}^{K} (\alpha S^k + \beta \vartheta^{k,k-1})$ 

• Distortion metric  $\Rightarrow S^k = \frac{\sum_{ij} c_{ij} s_{ij}}{\sum_{ij} s_{ij}}$ 

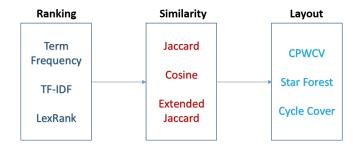
• Coherence metric 
$$\Rightarrow \vartheta^{k,k-1} = 1 - \frac{\sum\limits_{i=1}^p \sigma(w_i^k,w_i^{k-1})}{pD}$$

- Space metric  $\Rightarrow \gamma = 1 \frac{\mu}{\varphi}$
- Tempo d'esecuzione

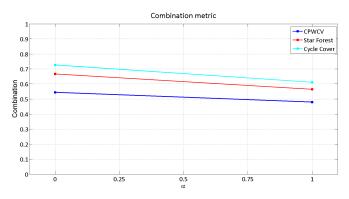


## Risultati sperimentali

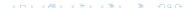
#### Algoritmi utilizzati:



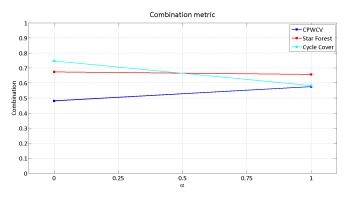
## Combination metric



Parole estratte: 20



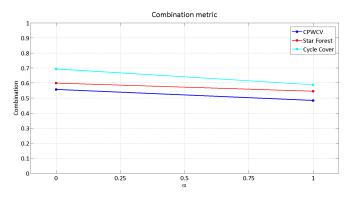
## **Combination metric**



Parole estratte: 40



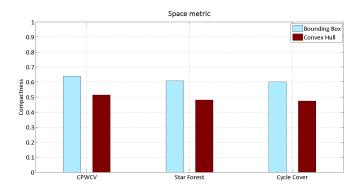
## Combination metric



Parole estratte: 60



# **Space** metric



Parole estratte: 40



# Tempo d'esecuzione

- Configurazione caso peggiore: TFIDF, Jaccard Similarity e Star Forest
- Configurazione caso migliore: Term Frequency, Cosine Similarity e CPWCV

Parole estratte	Caso peggiore	Caso migliore
20	2.65 <i>s</i>	0.77 <i>s</i>
40	3.01 <i>s</i>	0.94 <i>s</i>
60	3.66 <i>s</i>	1.25 <i>s</i>



## Concludendo...

Il lavoro di tesi si è articolato nelle seguenti fasi:

- 1 Lorem ipsum
- 2 Lorem ipsum
- S Lorem ipsum

## Lorem ipsum:

- Lorem ipsum
- Lorem ipsum

