

# Ho sconfitto gli emendamenti di Calderoli con un software

Jacopo Notarstefano

29 maggio 2016

Mi chiamo Jacopo Notarstefano e sono Junior Fellow al CERN.  
In particolare lavoro a INSPIRE, un motore di ricerca per articoli, preprint e dati nell'ambito della fisica delle alte energie.

Il mio Twitter è **@Jaconotar** e i miei DM sono aperti per domande, commenti e chiarimenti. Ask Me Anything (AMA)!

# Un chiarimento importante

Il CERN non ha niente a che vedere con ciò di cui parlerò oggi.

Il codice di cui parlerò è stato sviluppato nel mio tempo libero, le opinioni che esprimerò sono personali e non riflettono necessariamente quelle dei miei colleghi, superiori o del CERN.

# Article spinning

Aziende e privati competono in ogni modo per raggiungere le prime posizioni nei risultati di Google per certe keyword.

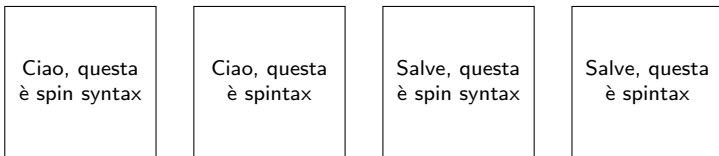
L'**article spinning** è una tecnica ormai in disuso per ottenere questo obiettivo. A partire da una pagina base ne vengono generate un numero elevato di variazioni: ciascuna di queste linka una pagina di cui si desidera migliorare il posizionamento.

# Spintax

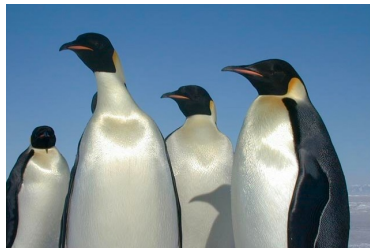
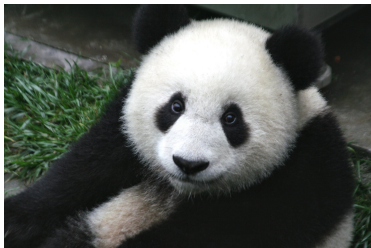
La generazione di queste pagine avviene usando un'opportuna sintassi, talvolta detta **spintax**. Ad esempio

`{Ciao|Salve}`, questa è `{spin syntax|spintax}`.

genera



# La “fine” dell’article spinning



A partire dal 2010 una serie di aggiornamenti di Google chiamati “Panda” e “Penguin” ha cominciato a penalizzare fortemente i siti contenenti testi duplicati.

L'ESCAMOTAGE

## **Article spinning: l'arma tecnologica di Calderoli contro il governo**

Il senatore si affida a un programma per realizzare migliaia di emendamenti

di Martina Pennisi

I senatori Calderoli e Crosio trovano una nuova applicazione dell'article spinning: la generazione di grandi quantità di emendamenti a scopo ostruzionistico.

RIFORME

## **Nuovo Senato: il Pd ritrova l'unità, ma la Lega ha 85 milioni di emendamenti**

La maggioranza ha raggiunto un accordo nella notte e ha depositato tre emendamenti al ddl Boschi su funzioni, articolo 2 e Consulta firmati da Anna Finocchiaro

di Redazione Online

La prima applicazione di questa tecnica ostruzionistica ottiene un successo parziale: la legge passerà, ma verranno rinegoziati alcuni articoli e slitteranno alcuni tempi tecnici.



CIRINNÀ

## **La Lega ritira 4.500 emendamenti e il M5S si sfila sul voto al «canguro»**

Seduta sospesa, il voto sul testo Marcucci slitta a mercoledì mattina. I grillini: non si può comprimere il confronto per soli 500 proposte di modifica. Ira Pd: decisione grave

di Alessandra Arachi

È più complicato valutare il successo della seconda applicazione, visto il peculiare clima politico che si è creato intorno all'emendamento Marcucci.

# Il caso del DDL Cirinnà, 1/3

Andiamo a vedere che aspetto hanno due emendamenti generati in questo modo.

1.450

Sostituire il comma 1 con il seguente:

"1. In osservanza del principio costituzionale di cui agli articoli 29 e 31 della Costituzione, ai sensi della lettera m) secondo comma ex art. 117 della Costituzione, la presente legge tutela e garantisce il ruolo sociale dell'educazione dei figli attraverso il riconoscimento delle figure genitoriali quali madre e padre."

Conseguentemente, sopprimere gli articoli da 2 a 10 e all'articolo 11 sopprimere le parole: "o da un'unione civile".

1.451

Sostituire il comma 1 con il seguente:

"1. In osservanza del principio costituzionale di cui agli articoli 29 e 31 della Costituzione, ai sensi della lettera m) secondo comma ex art. 117 della Costituzione, la presente legge tutela e garantisce il ruolo sociale dell'educazione dei figli attraverso il riconoscimento delle figure genitoriali quali madre e padre."

Conseguentemente, sopprimere gli articoli da 2 a 10.

# Il caso del DDL Cirinnà, 2/3

I due precedenti emendamenti differiscono soltanto per l'aggiunta di una frase finale. Il successivo aggiunge un ulteriore paragrafo.

1.452

Sostituire il comma 1 con il seguente:

"1. In osservanza del principio costituzionale di cui agli articoli 29 e 31 della Costituzione, ai sensi della lettera m) secondo comma ex art. 117 della Costituzione, la presente legge tutela e garantisce il ruolo sociale dell'educazione dei figli attraverso il riconoscimento delle figure genitoriali quali madre e padre."

Conseguentemente, sopprimere gli articoli da 2 a 10 e all'articolo 11 sopprimere le parole: "o da un'unione civile".

Conseguentemente, dopo l'articolo, aggiungere il seguente:

"Art. 23-bis. (Disciplina transitoria)

1. La presente legge entra in vigore dopo l'approvazione di una legge costituzionale, adottata ai sensi dell'articolo 138 della Costituzione, di modifica degli articoli 29, 30 e 31 della Costituzione."

# Il caso del DDL Cirinnà, 3/3

Non è difficile ricostruire la spintax che potrebbe aver generato i precedenti tre emendamenti.

Sostituire il comma 1 con il seguente:

```
"1. In osservanza del principio costituzionale di cui agli articoli 29 e 31 della Costituzione, ai sensi della lettera m) secondo comma ex art. 117 della Costituzione, la presente legge tutela e garantisce il ruolo sociale dell'educazione dei figli attraverso il riconoscimento delle figure genitoriali quali madre e padre."
```

Conseguentemente, sopprimere gli articoli da 2 a 10{ e all'articolo 11 sopprimere le parole "o da un'unione civile"}.

```
{|
```

Conseguentemente, dopo l'articolo, aggiungere il seguente:

```
"Art. 23-bis. (Disciplina transitoria)
```

```
1. La presente legge entra in vigore dopo l'approvazione di una legge costituzionale, adottata ai sensi dell'articolo 138 della Costituzione, di modifica degli articoli 29, 30 e 31 della Costituzione."}
```

# Gli Open Data del Senato

Il sito del Senato offre una grande varietà di dati su leggi, iter legislativi e attività dei Senatori in formati liberi (CSV, XML, JSON, ...) e in modo facile da interrogare (SPARQL).

La qualità di dati e metadati è ottima, la documentazione è scarna ma chiara.

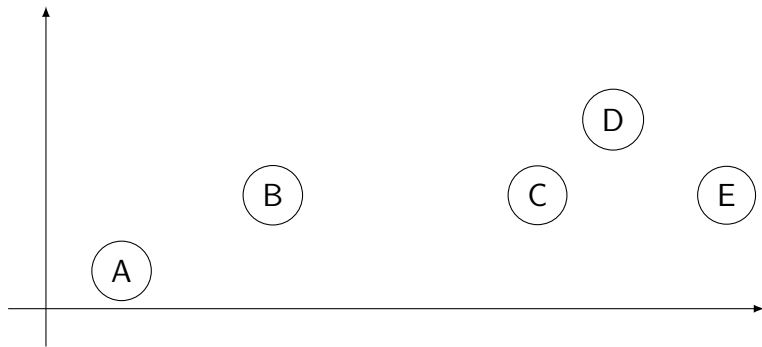
# Un piccolo intoppo

Purtroppo l'endpoint SPARQL non offre i testi degli emendamenti, quindi siamo costretti a scrivere uno spider per scaricarli dal sito del Senato.

```
class SenatoSpider(scrapy.Spider):  
  
    """A spider to download all amendments of a law discussed by the Italian Senate."""  
  
    name = 'senato'  
    allowed_domains = ['senato.it']  
    start_urls = ['http://www.senato.it/leg/17/BGT/Schede/Ddliter/testi/46051_testi.htm']  
  
    def parse(self, response):  
        for href in response.xpath(AMENDMENT_XPATH).extract():  
            match = re.search(HREF_REGEX, href)  
            if match:  
                _id = match.group('_id')  
                parent = match.group('parent')  
                relative_url = AMENDMENT_URL.format(_id=_id, parent=parent)  
                absolute_url = response.urljoin(relative_url)  
                yield scrapy.Request(absolute_url, callback=self.download_amendment)  
  
    def download_amendment(self, response):  
        relative_filename = response.url.split('/')[-1]  
        absolute_filename = os.path.join(os.getcwd(), 'data', relative_filename)  
        with open(absolute_filename, 'wb') as f:  
            f.write(response.body)
```

# Hierarchical Clustering: il problema

Vogliamo raggruppare un certo numero di oggetti in un numero ignoto di partizioni, detti **cluster**, in modo che oggetti vicini appartengano allo stesso cluster.



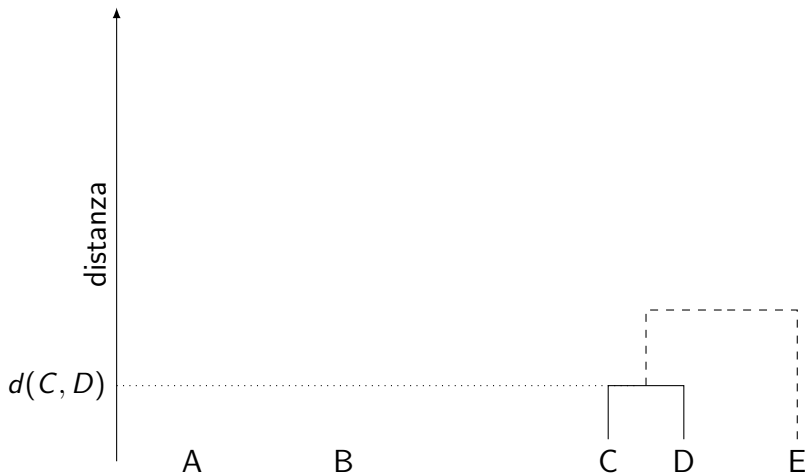
# Hierarchical Clustering: l'idea

L'idea più semplice che ci viene in mente è fondere insieme i due nodi più vicini e ripetere il procedimento, fino a che non arriviamo a un unico cluster.

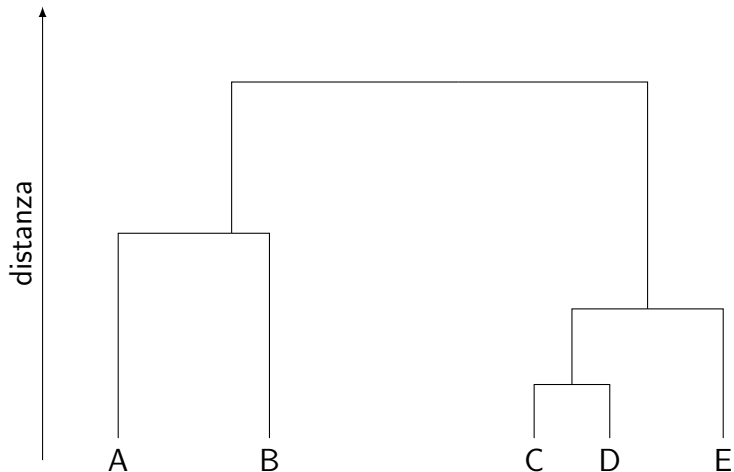




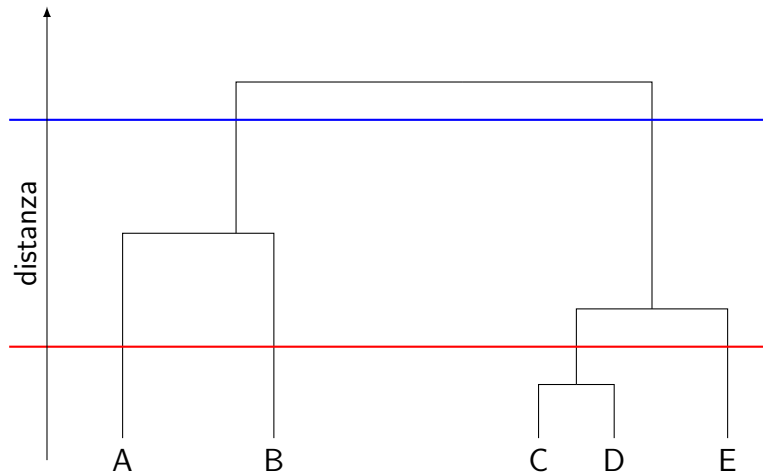
# Hierarchical Clustering: l'albero delle fusioni, 1/3



# Hierarchical Clustering: l'albero delle fusioni, 2/3



# Hierarchical Clustering: l'albero delle fusioni, 3/3



# La Distanza di Jaccard: l'idea, 1/2

La vicinanza semantica di due testi può essere dedotta dalla vicinanza sintattica delle *parole* dei due testi. Ad esempio:

- Oggi sono andato al mercato con mia mamma.
- Io e mia mamma oggi ci siamo recati al mercato.
- Meglio viaggiare in treno che viaggiare in macchina.

# La Distanza di Jaccard: l'idea, 2/2

La vicinanza semantica di due testi può essere dedotta dalla vicinanza sintattica delle *parole* dei due testi. Ad esempio:

- **Oggi** sono andato **al mercato** con **mia mamma**.
- Io e **mia mamma oggi** ci siamo recati **al mercato**.
- Meglio viaggiare in treno che viaggiare in macchina.

# La Distanza di Jaccard: la definizione

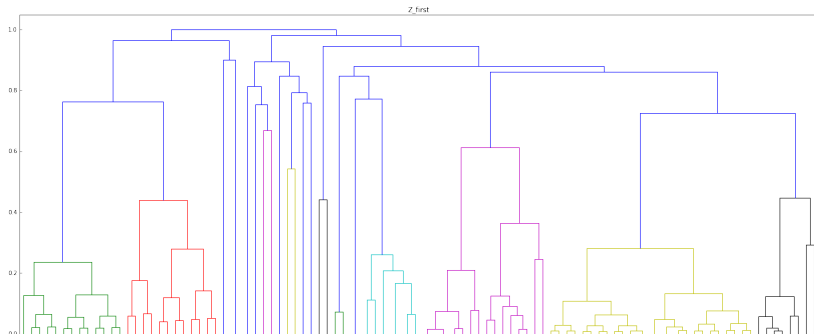
## Definizione (Distanza di Jaccard)

*Siano  $A$  e  $B$  due documenti, e sia  $token(X)$  la funzione che ritorna l'insieme delle parole distinte del documento  $X$ .*

*Allora chiamiamo **Distanza di Jaccard** la seguente funzione:*

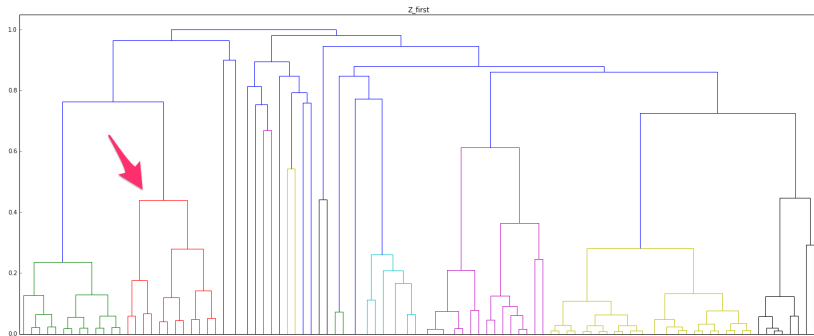
$$d_{Jaccard}(A, B) = 1 - \frac{|token(A) \cap token(B)|}{|token(A) \cup token(B)|}$$

# L'albero delle fusioni dei primi cento emendamenti



Una volta definita una distanza fra emendamenti possiamo applicare HC. Per verificare il buon funzionamento proviamo ad applicarlo ai primi cento emendamenti.

# Un cluster di emendamenti



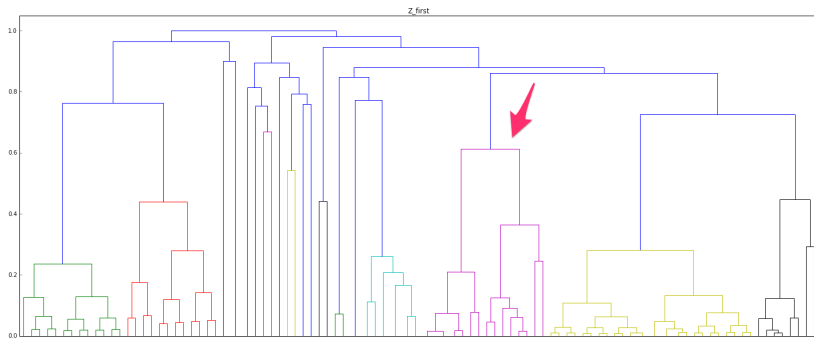
**1.2511:** Sopprimere gli articoli 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11.

**1.2510:** Sopprimere gli articoli 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

**1.2509:** Sopprimere gli articoli 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13.



# Un altro cluster di emendamenti

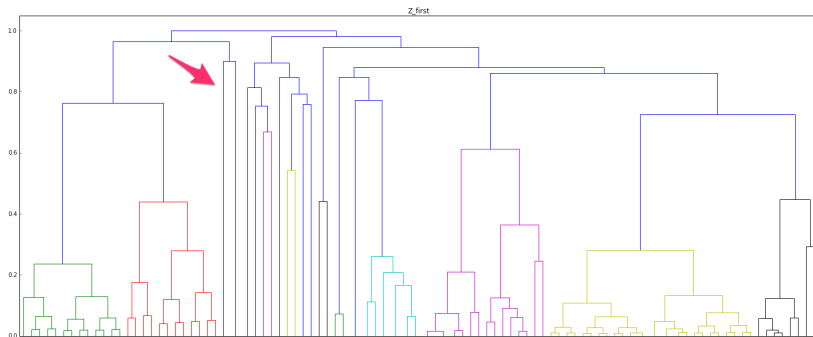


**01.8007:** All'articolo, premettere il seguente: «Art. 01. (Società economiche volte...

**01.8006:** All'articolo, premettere il seguente: «Art. 01. (Esclusività delle prerog...

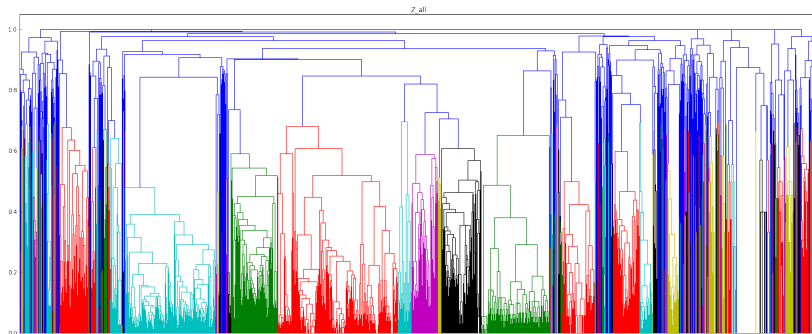
**01.8002:** All'articolo, premettere il seguente: «Art. 01. (Esclusività della famigl...

# Emendamenti che non formano un cluster



È interessante notare che le questioni pregiudiziali, tutte diverse fra loro, non paiono formare alcun cluster.

# L'albero delle fusioni di tutti gli emendamenti



In cinque minuti siamo in grado di far girare l'algoritmo su tutti gli emendamenti del DDL Cirinnà ( $\approx 5000$ ).

Una volta raccolti generati i cluster abbiamo due alternative:

- 1 Chi presiede il Senato può invocare il comma 8 dell'articolo 100 del regolamento del Senato:

*Il Presidente può stabilire, con decisione inappellabile, la inammissibilità di emendamenti privi di ogni reale portata modificativa (...)*

- 2 Un altro parlamentare può depositare un piccolo emendamento “canguro” per ciascun cluster, riducendo drasticamente il numero di votazioni.

# Ho sconfitto gli emendamenti di Calderoli con un software

Jacopo Notarstefano

29 maggio 2016

~~Ho sconfitto~~ È possibile sconfiggere gli  
emendamenti di Calderoli con un software

Jacopo Notarstefano

29 maggio 2016

# Ringraziamenti



Grazie a tutti per l'attenzione e a Wired per avermi invitato.

Il mio Twitter è **@Jaconotar** e i miei DM sono aperti per domande, commenti e chiarimenti. Ask Me Anything (AMA)!