

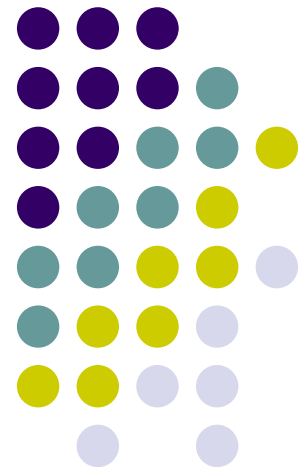
BAYES-SWARM



www.bayes-swarm.com

www.bayesfor.eu

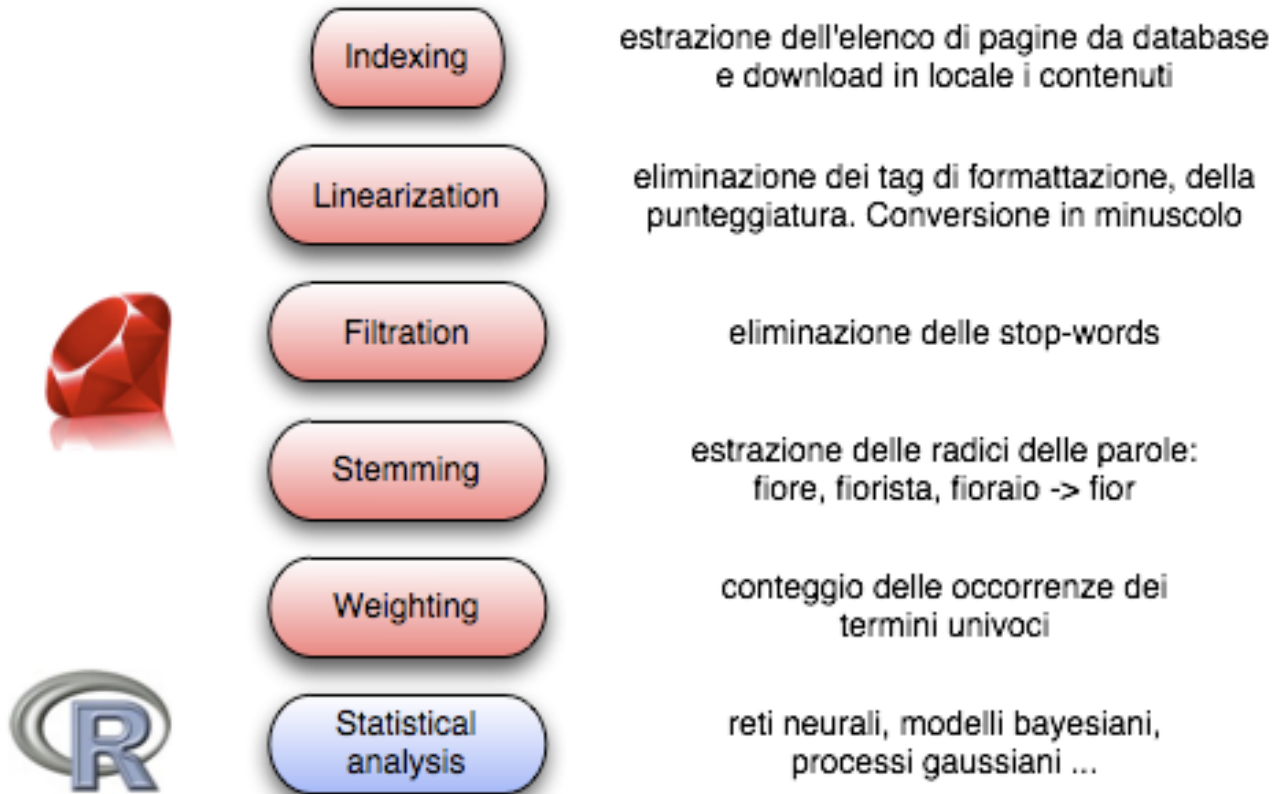
info@bayesfor.eu





- Progetto di ricerca che mira a fare **spidering** delle fonti di informazione sul **web**, trovando:
 - Correlazione tra parole nel tempo.
 - Associazioni tra parole nelle fonti.
 - Correlazione tra uso di parole e notizie.

- Associazione **BayesFor**



- I contenuti delle pagine web sono organizzati in un database.
- Il set di dati è studiato usando standard dell'analisi statistica, tecniche di data mining e concetti di information-retrieval.



Dati:

- **92** fonti di informazioni online monitorate.
- **188** pagine da cui sono estratti i dati.
- **Parole chiave** in italiano ed inglese.

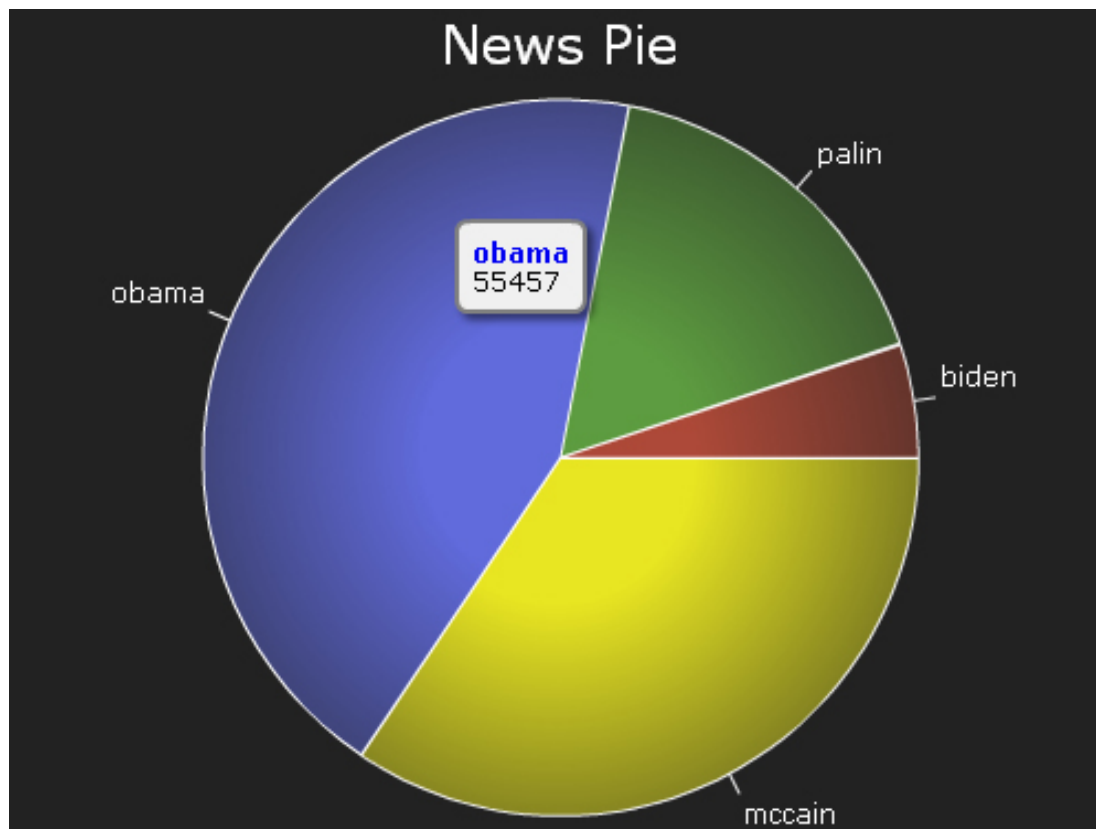
Servizi:

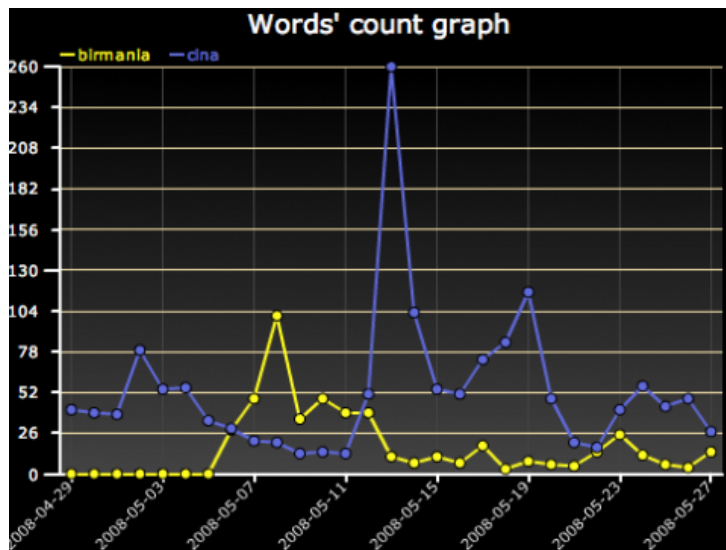
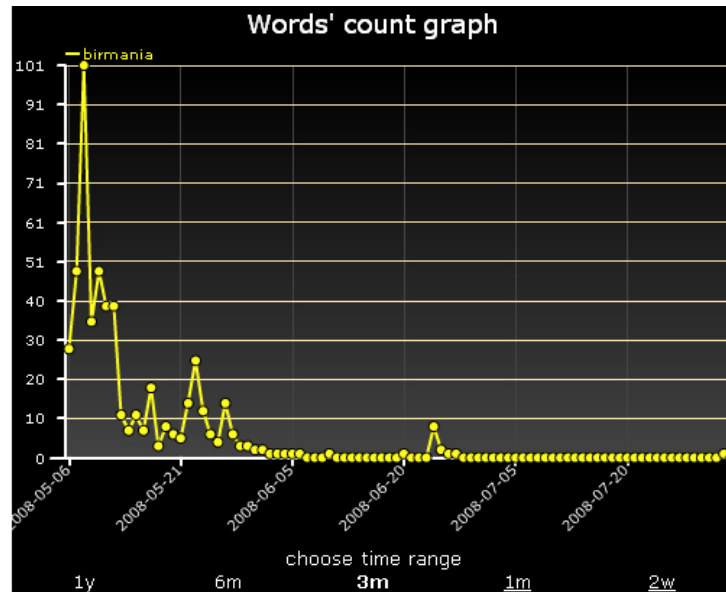
- **Ricerca per parola.**
- **Grafico** delle occorrenze di ogni parola nell'arco di tempo desiderato.
- **Macchina del tempo:** screenshot della pagina da cui ogni occorrenza è stata estratta.



Presidenziali USA 2008:

- visibilità dei candidati e dei vice.





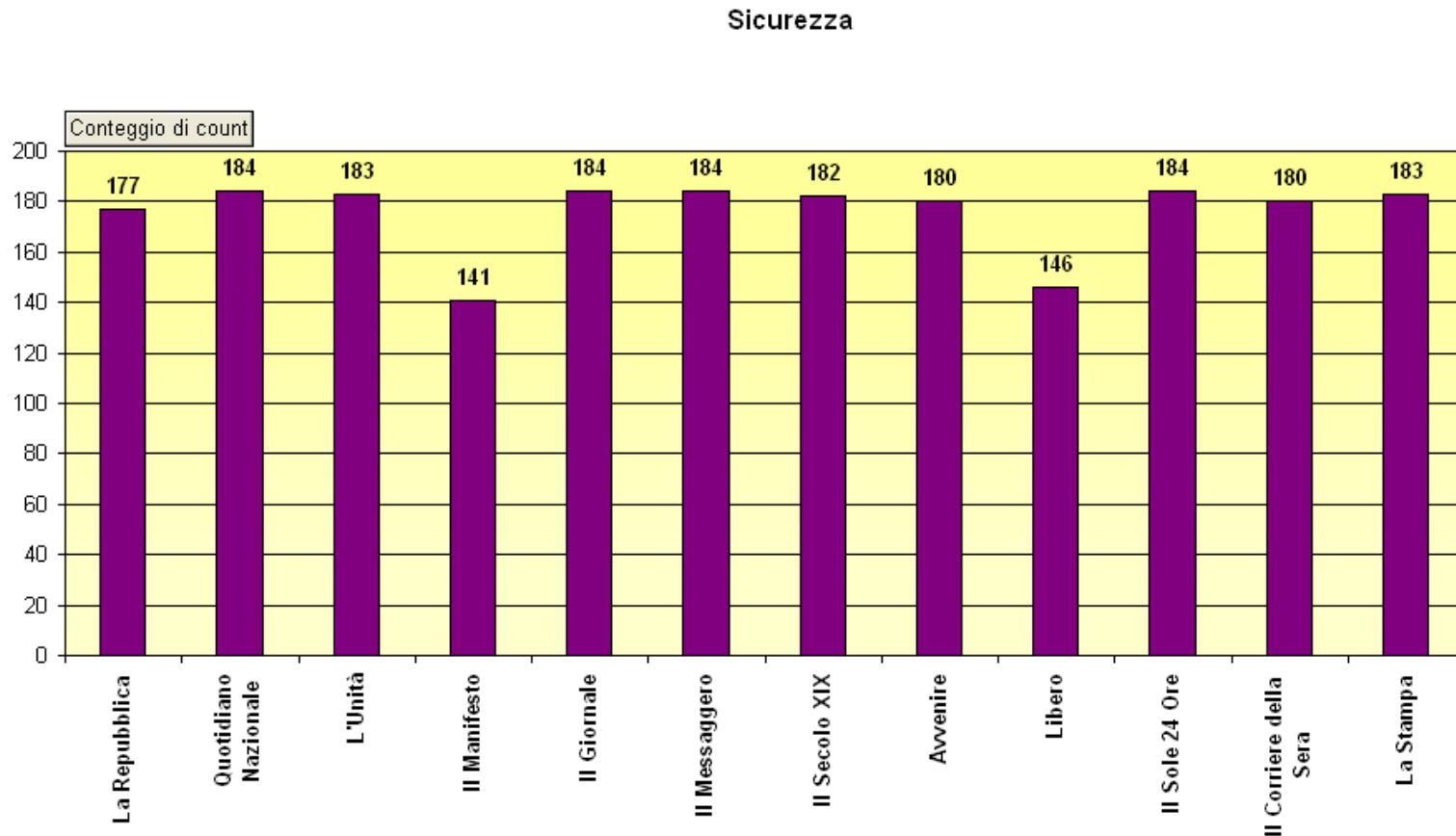
Birmania:

- Rapido affievolirsi dell'interesse dei media.
- Confronto con la visibilità della **Cina** in occasione delle catastrofi naturali che le hanno toccate contemporaneamente nel Maggio scorso.



Sicurezza:

- Presenza del tema sulle home page dei maggiori quotidiani italiani negli ultimi 8 mesi.





100 parole più frequenti nei siti
di informazione in lingua **italiana**

Politici:

Berlusconi Gelmini
Veltroni Colaninno
Sacconi Epifani
Maroni Alemanno
Brunetta La Russa
Fini Tremonti
Bossi Letta
Alfano Carfagna
Di Pietro Matteoli
Napolitano
Obama Bush
McCain Sarkozy
Putin

Politichese:

Politica
Stato
Governo
Presidente
Cgil Pd
Parlamento
Elezioni
Pdl Diritto
Destra
Sinistra
Cisl
Democrazia

**Concetti
astratti:**

Pace
Libertà
Valori
Diritti

Crisi
Sicurezza
Scontro
Allarme
Sconfitta

Geografia:

Italia
Milano
Firenze
Europa
Sardegna
Perugia
India Europa
Brasile
Cina
Israele
Afghanistan
Taiwan
Iraq

Energia:

Petrolio
Nucleare
Energia
Acqua
Gas



100 parole più frequenti nei siti di informazione in lingua **inglese:**

Politici:

Obama McCain
Kucinich
Clinton Bush
Huckabee
Biden Rodham
Edwards Giuliani
Thompson
Bayh Kaine
Blair Hagel
Sarkozy Putin
Chavez Gore
Rendell

Economia:

Max
Lehman
Bond Bank
Credit
Labour
Euro
Finance
Loan
Fund Ecb
Subprim

Terrorismo:

Nuclear
Attack
Taliban
Arab
Risk
Muslim
Qaeda
Bomb
Virus
Islam

Geografia:

World
Iraq China
Asia India
Russia Korea
Pakistan
Japan Iran
Taiwan Israel
Foreign
East Gulf
England
Syria Europe
Spain italy

Ambiente:

Climate
Earth
Warming

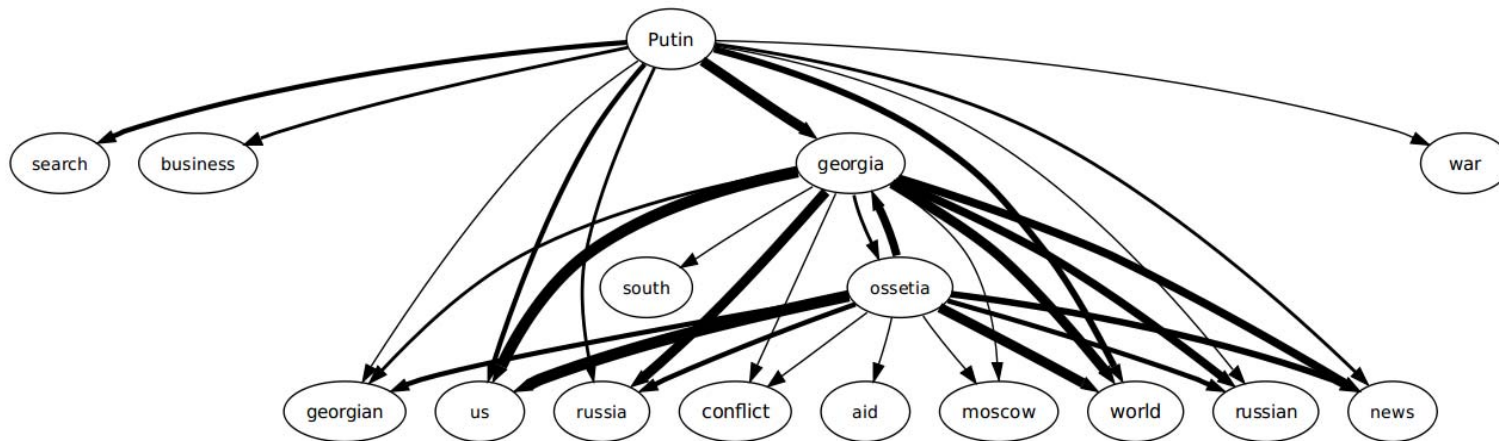


- Parole che compaiono in entrambe le liste:
italia/italy – india – europa - cina/china - israele/israel - taiwan - iraq
obama – bush - mccain - sarkozy - putin - papa/pope
right/diritto-destra - elezioni/elect - bce/ebc - nucleare/nuclear - sesso/sex.
- Religione:
IT *papa, benedetto, vaticano*
EN *atheists, god, pope.*
- Internet e computer:
EN *facebook, podcast, ibm.*
- **Gossip:** unica parola inglese presente nella lista di fonti italiane.
Mafia: unica parola italiana presente nella lista di fonti inglesi.



Limiti – obiettivi per il futuro del progetto:

- Normalizzazione dei dati.
- **Mean Machine**: collegamento tra parole chiave.



Per informazioni tecniche riguardo al funzionamento del database ed il processo di estrazione:

e-mail info@bayesfor.eu

BAYES-SWARM



www.bayes-swarm.com

www.bayesfor.eu

info@bayesfor.eu

Grazie.

