Follow the Flow.

Terms based Timeseries.

7 mag 2014
A **time series** is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals.

**Examples:** of time series are the daily closing value of the Dow Jones Industrial Average.

Time series are very frequently **plotted** via line charts.

Time series are used in **statistics**, **signal processing**, **pattern recognition**, **econometrics**, and **mathematical finance**.
Parts

- Time Series are composed:
  - Time
  - Value

- Two values array (one for time aspect and one for data aspect) is the simplest representation.
Term Time Series

- Term Time Series are time series that keep track of the number of occurrences of single term over time.

- Normally are fixed interval Time Series (minutes, hour, day interval).
Term Time Example

Monday, Apr 15, 2013
- Boston: 16,611
- Victim: 1,120
- Police: 1,054
- Bomb: 6,025
How To

- If we have a temporal aspect in a document (for example Tweets timestamp) we should count all the documents that contain that term $w_j$ in a specific time interval $t_i$.

- That produce the value of $w_j$ in $t_i$. 
Class Work (1)

- Using the streaming api (Twitter4J) collect the stream that keep track of the following “target” terms:
  - Renzi
  - Grillo
  - Berlusconi
  - Di Battista
  - Meloni

- For each “target” term create (at realtime) the relative time series. Using 2 minutes time fixed interval.

- Design your own implementation for a time series representation.
Let’s Try?!?!
Class Work (2)

- Using the streaming api (Twitter4J) collect the stream that keep track of the following “target” terms:
  - Renzi
  - Grillo
  - Berlusconi
  - Di Battista
  - Meloni

- Build a Lucene index for all tweets.

- For each “target” extract from the Lucene index the relative time series. Fixed time interval should be chosen later.