

Mining Twitter for real-time trend and information discovery

VI Jornadas MAVIR

Arkaitz Zubiaga

NLP & IR Group @ UNED

November 16th, 2011

Index

1 Motivation

2 Our Work (I)

3 Our Work (II)

4 Outlook

Twitter



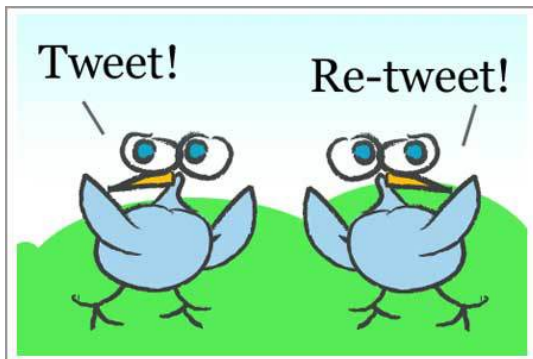
- Twitter is a microblogging service with over 200 million users.
- Users share short messages of up to 140 characters (*tweets*).

Twitter: following users



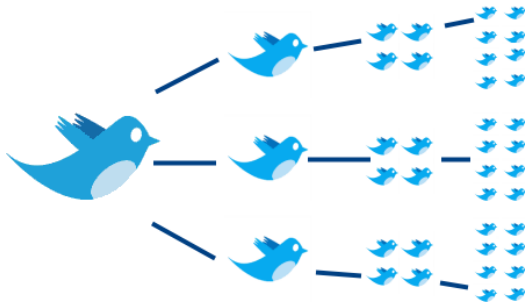
- Different from Facebook, following is not reciprocal.

Twitter: retweeting



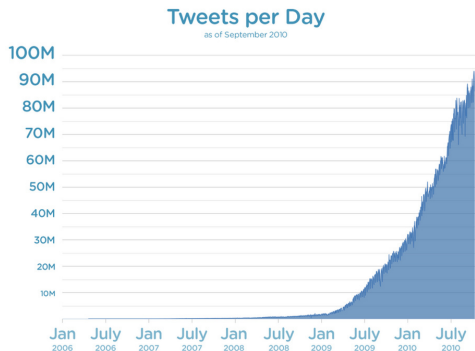
- Retweet: users can help spread tweets by others.

Twitter



- Retweeting enables fast spread of messages.

Increase of activity on Twitter



- As of October 2011, Twitter received 250 million tweets per day.

Variety of Twitter accounts



MundoDeportivo.com

@mundodeportivo Barcelona

Seleccionamos las noticias más relevantes del Barça y del mundo del deporte para ti. Para seguir todas las noticias

@md_total y durante los partidos

<http://www.mundodeportivo.com>



Madrid

@MADRID Madrid

Twitter oficial del Ayuntamiento de Madrid.

<http://www.madrid.es>



Lady Gaga

@ladygaga New York, NY

mother moe

<http://www.ladygaga.com>



Publico.es

@publico_es Madrid, Spain

Twitter oficial del Diario Público

<http://www.publico.es>



Hypertext 2012

@hypertext2012

Twitter account for the Hypertext 2012 conference

<http://www.ht2012.org>



Jimmy Wales

@jimmy_wales Florida

You know, the Wikipedia and Wikia guy.

<http://www.facebook.com/pages/Jimmy-Wales/1065515679>



The New York Times

@nytimes New York, NY

Where the Conversation Begins. Follow breaking news,

NYTimes.com home page articles, special features and more.

<http://www.nytimes.com/>



Real Sociedad info

@Real_Sociedad Donostia - San Sebastian

Información y opiniones sobre la Real Sociedad de San

Sebastián de fútbol - Gora Errealta - fans site -



CIKM2011 - Glasgow

@CIKM2011 Glasgow, UK

Official updates for the 20th ACM Conference on Information

and Knowledge Management, October 24-28, 2011

<http://www.cikm2011.org>



A Googler

@google Mountain View, CA

News and updates from Google

<http://www.google.com/support/>

Usefulness of Twitter

- Twitter provides...
 - ① ...large amounts of data in real-time,
 - ② from a wide variety of sources,
 - ③ with the ability to spread rapidly.

Twitter's popularity

Twitter has gained widespread popularity as a tool for...

Using Twitter for... following events

How to follow an event on twitter

CentrevillePatch

Using Twitter to Follow Live Events and Favorite Topics

KnowtheNetwork.com

How to follow the #Egypt events on Twitter

- (1) Live-tweeting about and following events.

Using Twitter for... helping others



Using Twitter to help Haiti



Twitter Helping Haiti Earthquake Victims

Twitter updates valuable during quake crisis



College student uses twitter
to help natural disaster
victims



Japan earthquake: how Twitter and Facebook helped

- (2) Helping others, as in natural disasters.

Using Twitter for... finding out about news



Twitter As Breaking News Source



How The Attack On Osama Bin Laden Was Live-Tweeted



This just in...News no longer breaks, it Tweets



New York plane crash: Twitter breaks the news, again



Man live-tweets bin Laden operation

- and (3) Finding out about breaking news.

Twitter on the media

THE WALL STREET JOURNAL | TECHNOLOGY

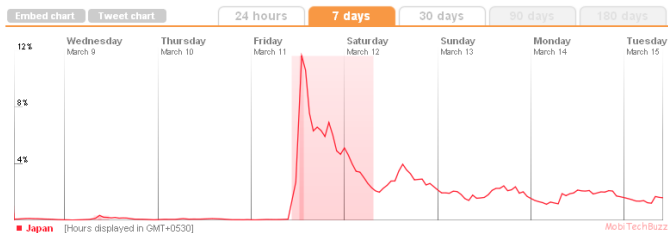
Decoding Our Chatter

Want to monitor an earthquake, track political activity or predict the ups and downs of the stock market? Researchers have found a bonanza of real-time data in the torrential flow of Twitter feeds.

The New York Times
Sunday Review | The Opinion Pages
Twitterology: A New Science?

- Lots of researchers are analyzing tweets.

Trends on Twitter



- The news about the Japan earthquake broke on Twitter.

Research on Twitter

- Most of the research on Twitter focus on the analysis of streams after they happened.
- Very little research deals with the real-time analysis of streams.

Our goal: How can we mine Twitter streams to acquire real-time knowledge about events and trends?

Index

- 1 Motivation
- 2 Our Work (I)**
- 3 Our Work (II)
- 4 Outlook

Trending Topics on Twitter

Trends: Spain · change

#terelu

#hazloFlan

#saltacoOM

Día Mundial de la Diabetes

Deport Sport

TWO DOOR CINEMA CLUB

Buen Lunes

YOU KNOW YOU ARE BORED

Josep Pernaut

Louis Tomlegendson

- Trending topics reflect the top conversations being discussed on Twitter more than usually.

What produces trending topics?

What kinds of events leverage those trending topics?

Typology of Trending Topics

- **News:** Japan earthquake.
- **Current events:** a soccer game.
- **Memes:** funny and viral ideas.
- **Commemoratives:** World AIDS Day.

Main findings

- Trending topics can **accurately** (78.4%) be **categorized** using **social features**:
 - **Outperforming** use of **textual content**.
 - **Without** making use of **external data**.
 - In **real-time** as the trending topic emerges.

Arkaitz Zubiaga, Damiano Spina, Víctor Fresno, and Raquel Martínez.
2011. *Classifying trending topics: a typology of conversation triggers on Twitter*. **CIKM 2011**.

Index

- 1 Motivation
- 2 Our Work (I)
- 3 Our Work (II)**
- 4 Outlook

Events on Twitter

- When **users live-tweet** about **events**:
 - They produce **vast amounts of tweets** about events.
 - Users **want to follow what others say**.
 - Users **cannot follow the overwhelming amounts** of tweets.

Stream summarization

Can we summarize streams of tweets in such a way that:

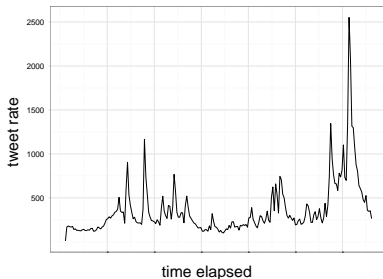
- Users receive a **reduced stream** that they can follow?
- Users **do not miss any key sub-event** occurred during the event?

Study of soccer games

- Copa America 2011 (July 1-26, 2011):
 - **26 soccer games.**
 - **11k-70k tweets per game.**
 - **Tweets are written in 30 languages.**

Summarization of soccer games

- **2-step summarization:**
 - ① **Sub-event detection.**
 - ② **Tweet selection.**
- e.g., Argentina - Uruguay (July 16, 2011): summarization from 70k to 31 tweets.



Main findings

- Use of **state-of-the-art text analysis methods** generates **accurate summaries**:
 - With **precision and recall** values **above 80%** (**100% for key sub-events**).
 - In **real-time** as the game is being played.
 - In **3 different languages** (es, en, pt).
 - **Without** need of **external data**.

Damiano Spina, Arkaitz Zubiaga, Enrique Amigó, Julio Gonzalo. *Towards Real-Time Summarization of Events from Twitter Streams*. **Submitted.**

Index

- 1 Motivation
- 2 Our Work (I)
- 3 Our Work (II)
- 4 Outlook**

Outlook

- **Work 1:**

- Further dig into each type of trending topic, in order to **look for subtypes of trends**.

- **Work 2:**

- Evaluate the performance of the **summarizer on other kinds of scheduled events** (award ceremonies, keynote talks,...)
- Evaluate **novelty of information garnered** from tweets.

Any Questions?



@arkaitz

Arkaitz Zubiaga

#thankyou! Any questions?

Contact: arkaitz@zubiaga.org

6 hours ago via web

[Favorite](#)

[Reply](#)

[Delete](#)