

# Data Mining The Social Web

By Gary Short

Developer Evangelist

Developer Express

# Introduction

- Gary Short
- Developer Evangelist
- [gary@garyshort.org](mailto:gary@garyshort.org)
- [garys@devexpress.com](mailto:garys@devexpress.com)
- [www.garyshort.org](http://www.garyshort.org)
- [@twitter](#)

# What About You Guys?



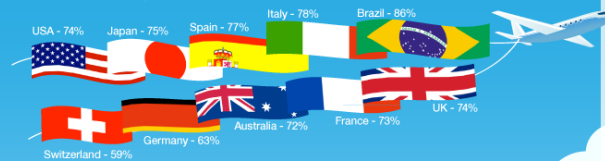
# Why Bother Data Mining The Social Web?

## Social Media Facts & Figures for B2B Sales

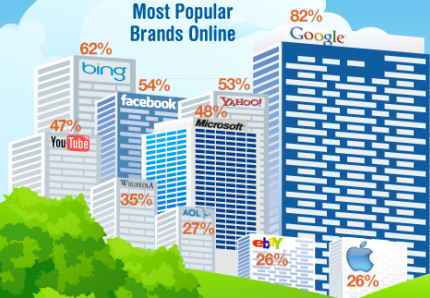
An estimated \$716 million will be spent on social media this year

### Social Media Reach and Usage by Country

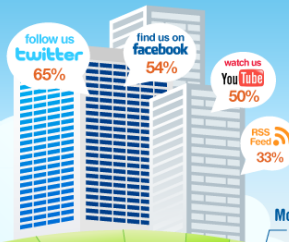
April 2010



### World's Most Popular Brands Online



### Global Fortune 100 Companies With Social Profiles



### Future Growth

Forrester estimates that \$716 million will be spent on social media this year, growing to \$3.1 billion in 2014. At that point, social media will be a bigger marketing channel than both email and mobile, but still just a fraction of the size of search or display advertising (\$31.6B and \$16B, respectively).

**Volatility:**  
A negative customer review on YouTube, Twitter or Facebook can cost a company about 30 customers.

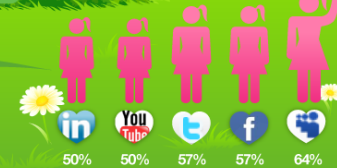
### Most Popular Social Media Sites for Generating B2B Traffic



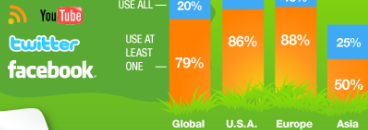
The world now spends over **110 billion minutes** on social networks and blog sites

This equates to **22% of all time** online or 1 in every 4:30 minutes

### Women on Social Media Sites



### Companies Using:



### Average Social Media Activity of Fortune 100 companies:



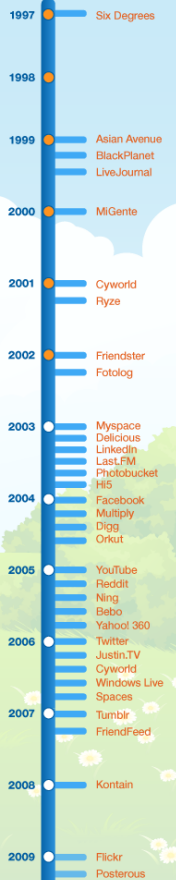
The average visitor spends **66% more time** on social sites than a year ago

April 2009: 3 hrs. 31 mts.  
April 2010: 6 hours

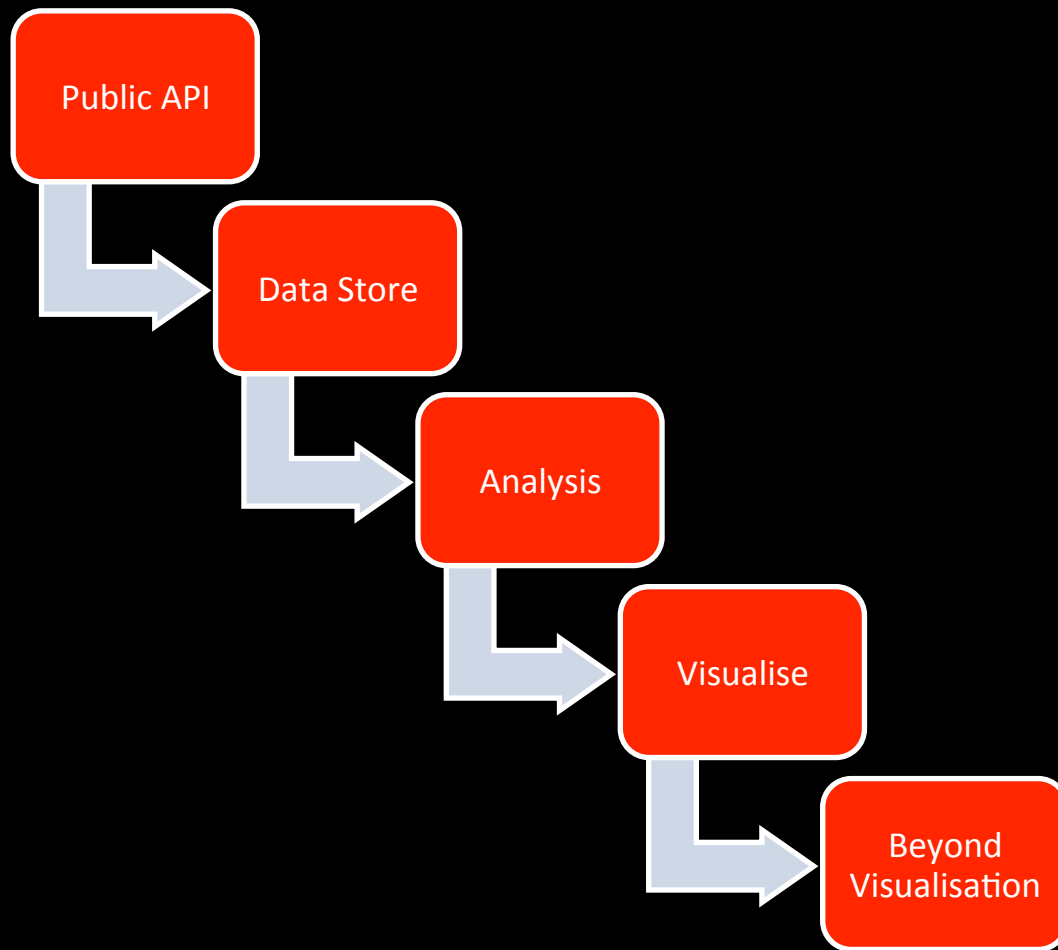
The number of people visiting these sites increased by **24%** over last year

Social network sites are visited by **75%** of global consumers who go online

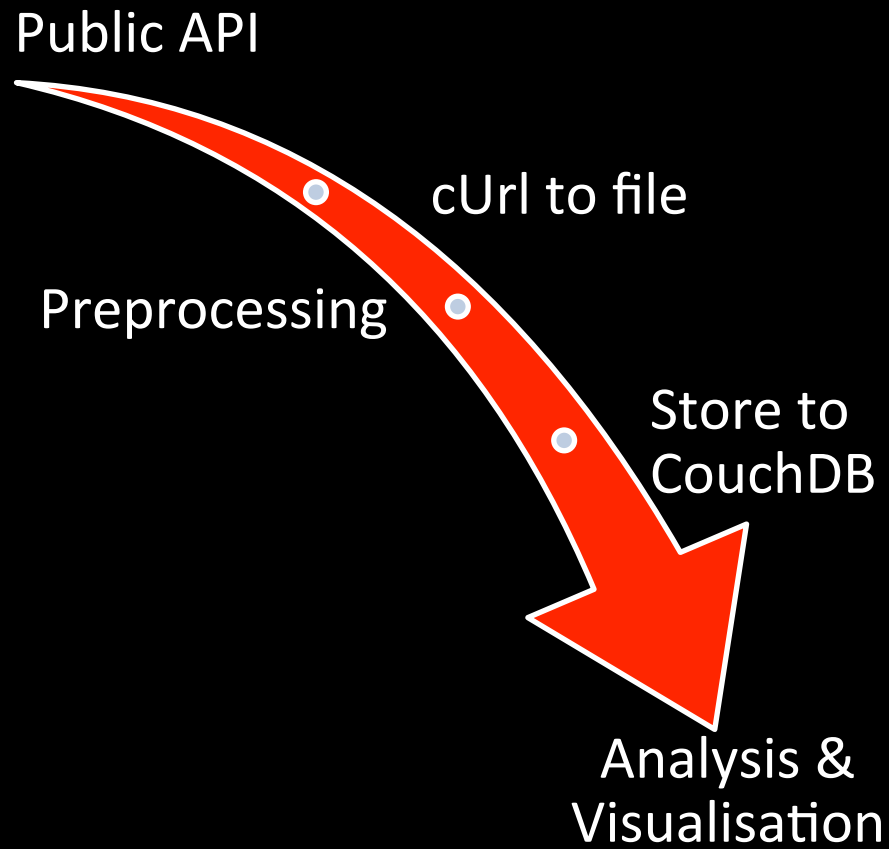
### Early Social Networks



# How do we Data Mine The Social Web?



# My Set UP

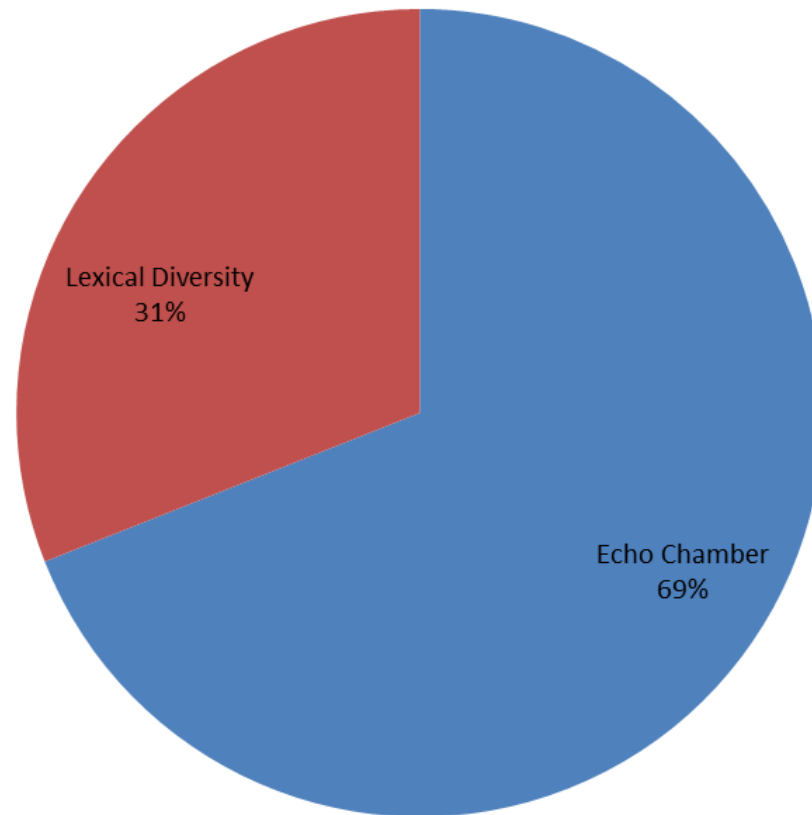


# Today's Sample

- ~6,000 tweets
- Mention
  - Apple
  - Google
  - Microsoft
- Doesn't include last night's announcement 😊

# Lexical Diversity

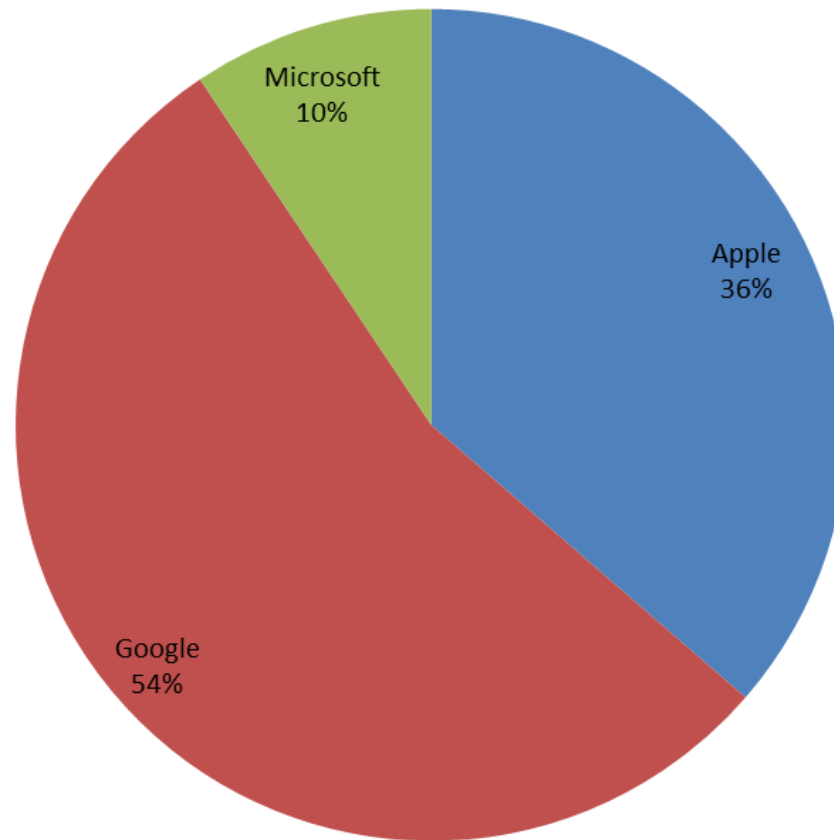
Lexical Diversity of Sample





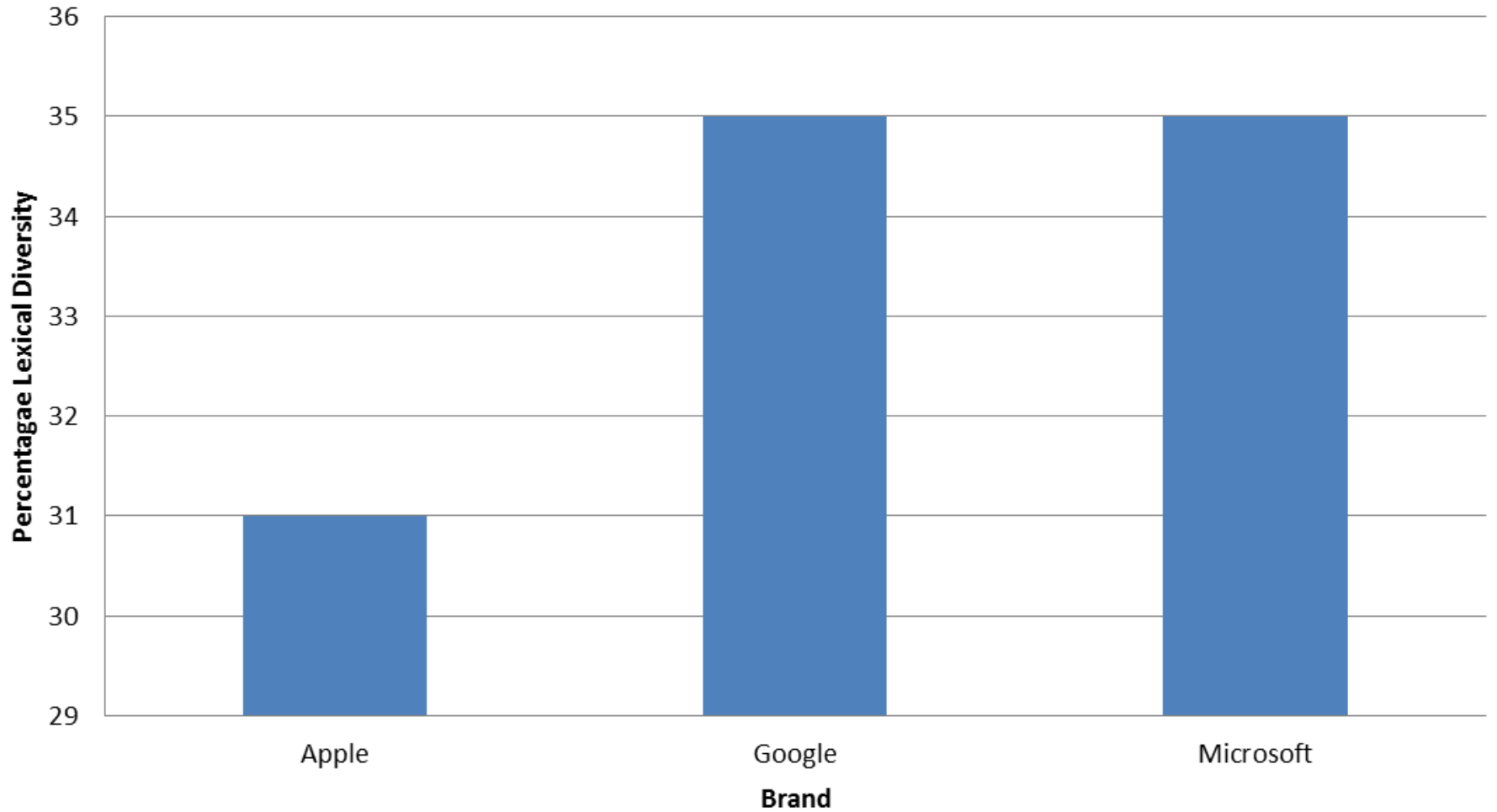
# Competitive Analysis

Share of Buzz



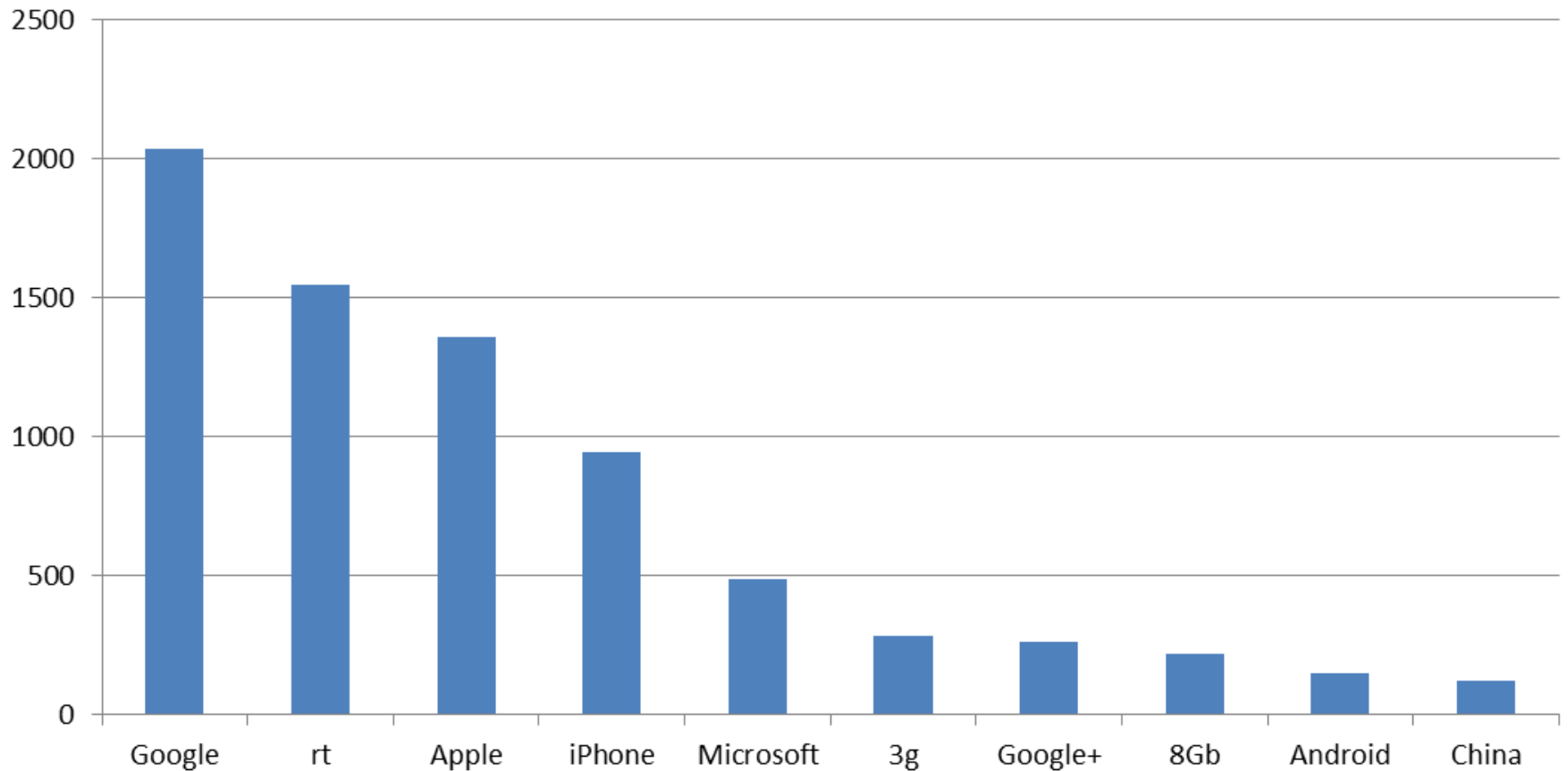
# Lexical Diversity by Brand

Lexical Diversity by Brand



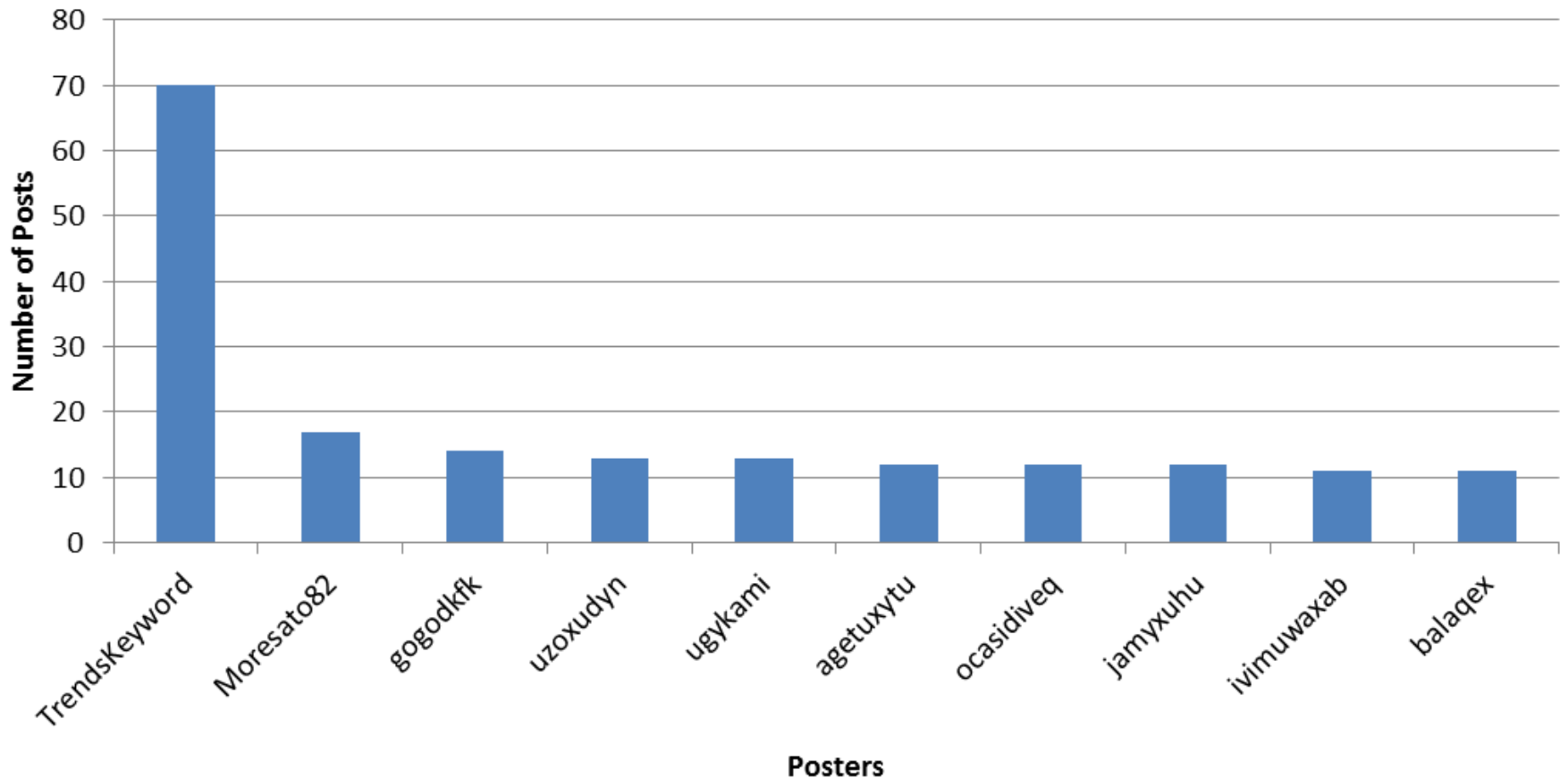
# Top 10 Popular Words

Top Ten Popular Words



# Top 10 Posters by Volume

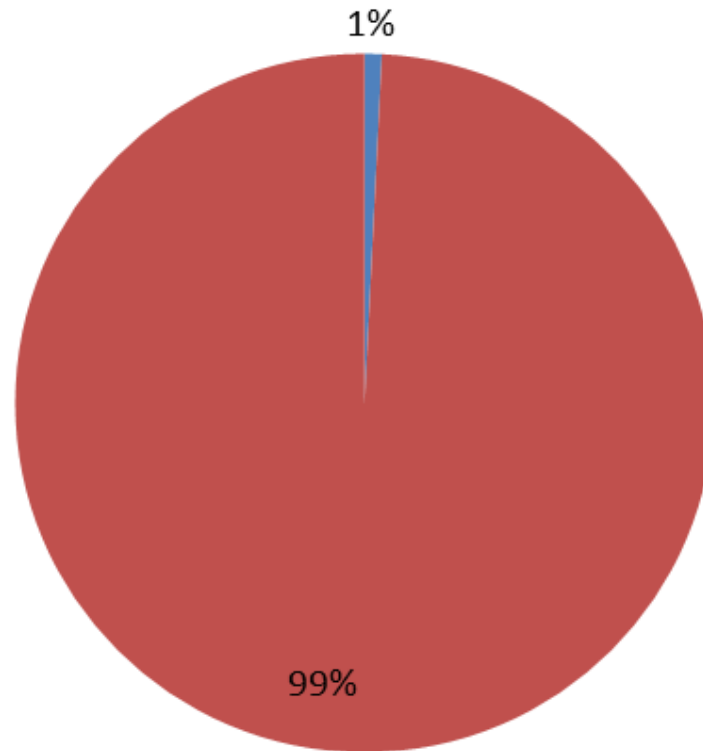
## Top 10 Posters by Volume



# Proportion Of Geo Coded Tweets

## Proportion of Geo Coded Tweets

■ Geo Coded Tweets ■ All





techcentral\_ie

gabfoundnemo

rachiedoos

Luutsiaano

bas\_van\_wijk

HarshBeatz

Goonery

pdbrewer

RHochstenbach

HugoApps

ole4ka

aaliaksandr

Twelvebars

Twelvebars

OZN3R0L

ChamsGunz

ChamsGunz

QuiQue2M

davecampbell

US Dept of State Geographer  
© 2011 Geocentre Consulting  
© 2011 Tele Atlas  
© 2011 Europa Technologies

©2010 Google

48°16'06.11" N 4°35'49.93" E elev 253 m

Eye alt 2641.01 km

# Questions

- [gary@garyshort.org](mailto:gary@garyshort.org)
- [garys@devexpress.com](mailto:garys@devexpress.com)
- [@twitter](#)