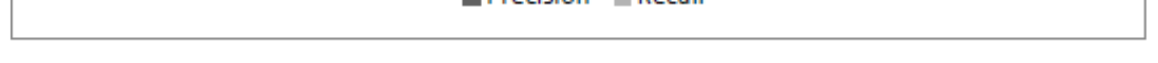
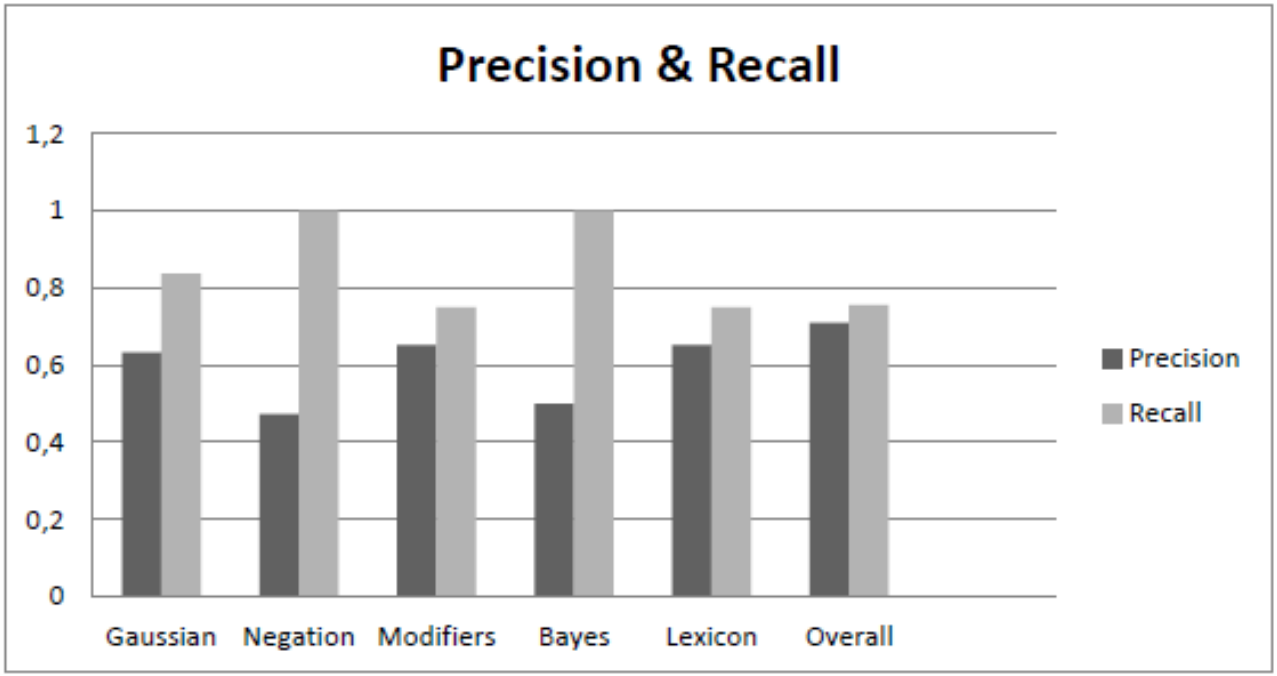


Results for unfiltered data

Results for data annot



Results for data without repeated characters

Conclusions about the p

- A significant improveme negations and the modi
- The performance of pre
- The Bayes classifier doe
- The overall score has ar with the unfiltered data

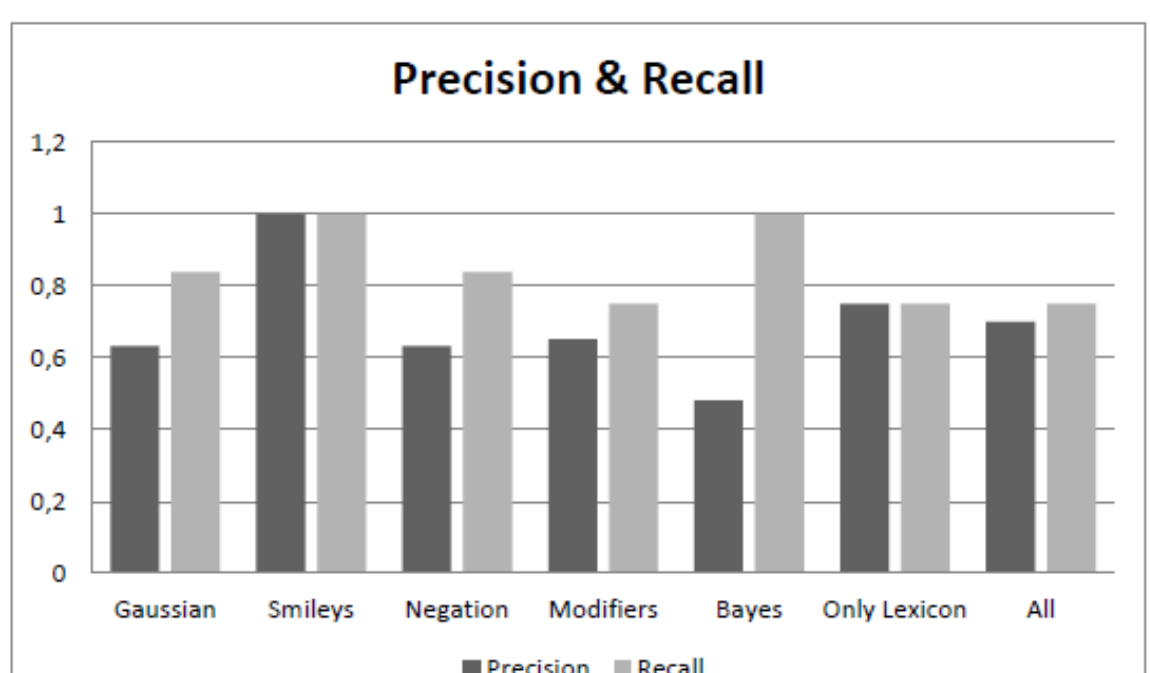
negative, neutral or positive and measures the reputation of a brand

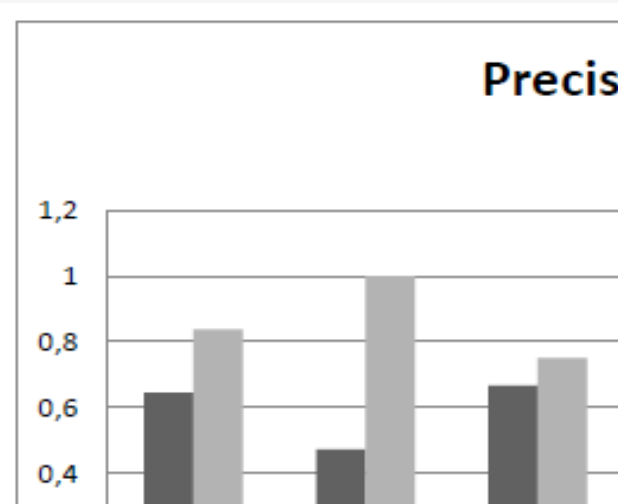
Results

Experimental Set up:

- Use an existing twitter corpus, provided by Sanders Analytics
- Split this data set for training and testing
- Measure precision and recall
- Use the query “Microsoft” for testing

Four different ways are use
❖The Gaussian distance.
❖The detection of negatio
❖The Bayes classifier and
positive scores.
❖The effect of the pres
score.





Reputat

Problem

Before :

I like this company, and you ?

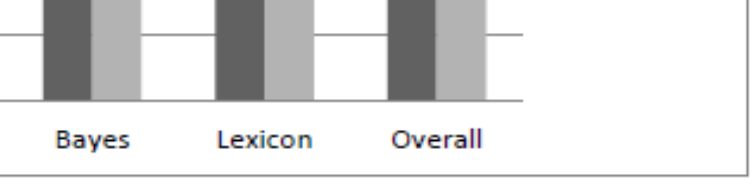
No I don't, I think it's not a good one.

It was difficult to track brand reputation.

Nov

It is possible to kr monitori

We worked on an online reputation management system that cat



without URL and
ations

preprocessing step :

ent occurs at recall for the
fiers.

ecision has been decreased.

s not indicate any difference.

n improvement in comparison
.

Tokenization : Use StandardTokenizer from Luce

Send to this@mail.com. This is perfect.

Tokenizer

- Send
- to
- this@mail.c
- This
- is
- perfect

Sentiment analysis

Lexicon : find the score of a tweet to label it as neg
neutral or positive.

Standard lexicon

"It is good" => score = 0+0+1 = 1

Modifiers and negations : add a coefficient for the followir

"It is not good" => score = 0+0+0+(-1)*1 = -1

Distance : the score decreases as the distance between the senti
the name of the company increases

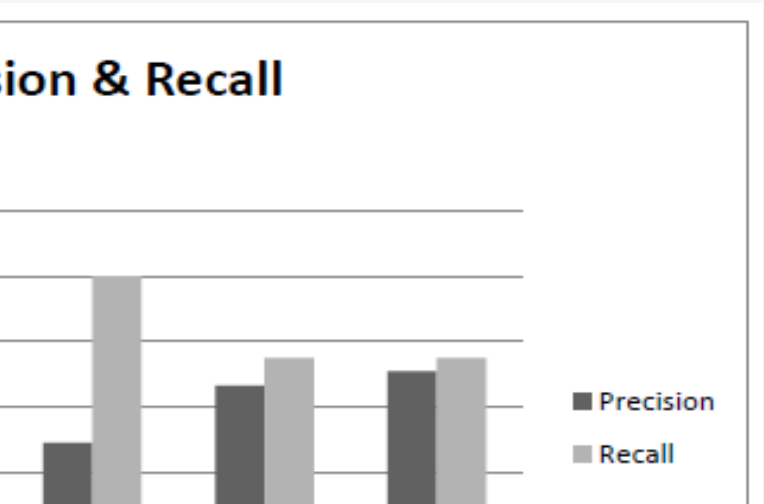
"I like A but I really hate B" => Positive score for A, negative sc

according to that.

ed for the sentiment estimation :

ns and modifiers.
a lexicon with assigned negative and

ence of smileys on the sentiment



Retrieve data about the tweets : Use Twitter RES

Twitter API

- Tweet text
- Author
- Number of fol
- Number of ret
- …

<u>Problem encountered</u> : the rate limit imposed by the API (15 hour, when not authenticated and 350 otherwise).

Solution found : caching.

<u>Preprocessing and Tokenization</u>

Elimination of repeated characters

- Greeeaaaat → Greeaat
- Aweeesome → Aweesome
- Look → Look

Annotations and URL's removal

@Someone : it's awesome. Check www.this.com.

: it's awes
Chec

tion Management System

w :

how brand reputation by
ng social media.

egorizes tweets as

Met

Search

- Search for specific keywords : provided by Twitter

Problem : their in-memory solution is limited to the last 8

- To handle this problem, store the search results in a database.

Query the database as well as the Twitter search.

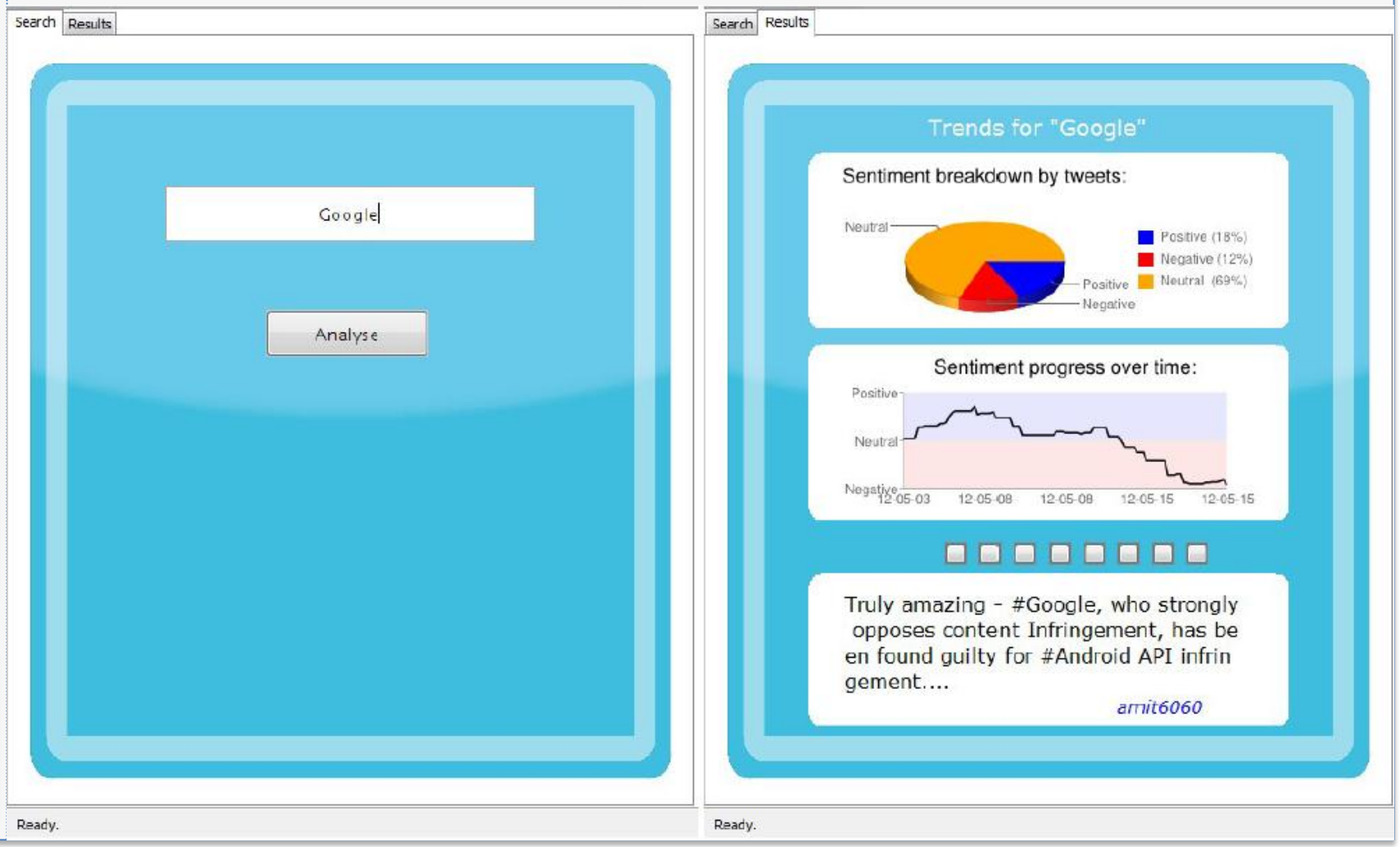
→ Increase recall.

Don't use a relevance score based on term frequency and document frequency, since tweets are limited in leng

Get tweets

pie chart; using the Google Charts API.

- To provide time dependent information, a second graph is used to display the cumulative reputation over time.
- Lastly, a list of top-ranked tweets is provided. By clicking on any of the square buttons, a top-sentiment tweet can be viewed.

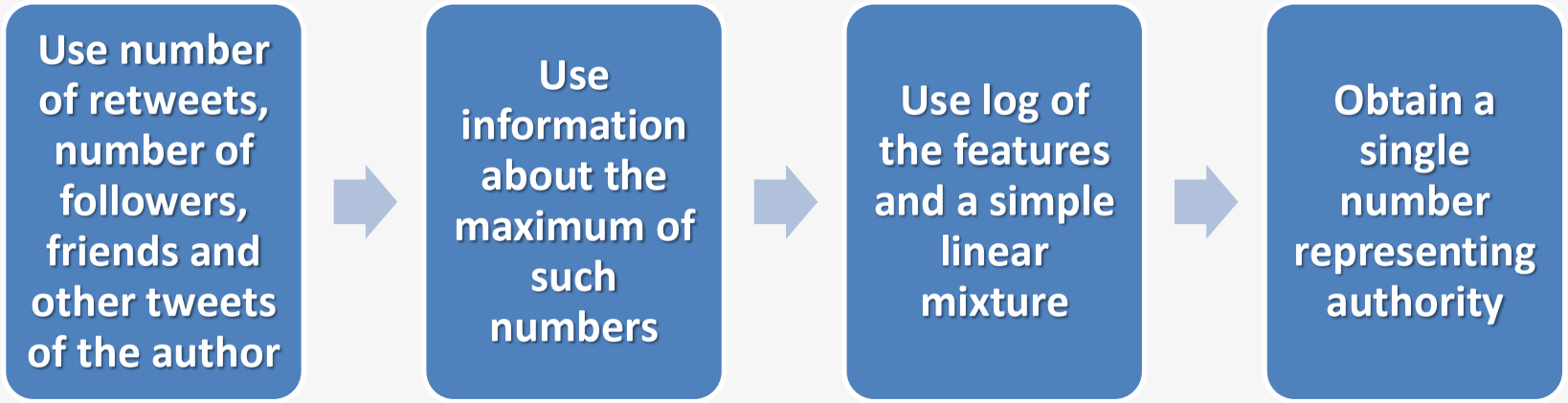


ST API.

llowers
tweets

0 calls per

some.
k.



Problem encountered : the rate limit imposed by the API is a problem for the creation of a user graph. Consequently, it was not possible to apply the original PageRank algorithm.

User interface

Goal : keep the design as simple and self explanatory as possible, while maintaining the essential functionality

Search input

The user is presented with a single text input field and a single button. The only distracting elements are the status bar at the bottom (which provides real time information on the analysis progress) and tab bar, enabling a user to switch to previous search results

Search results

- A break-down of positive, negative and neutral tweets is displayed as

Mihai Damaschin
Matthijs Dorst
Maria Gerontini

Cihat Imamoglu
Caroline Queva

hods

3 days

MySQL

th.

Naïve Bayes

Use the algorithm provided by Weka and the hand-classified data provided by Sanders Analytics

PageRank

Calculate a static quality measure