

Mining Banks Reviews in Google Maps with R

Gabriela A. Werb

Chair of Electronic Commerce

Faculty of Business and Administration

Graduate School of Finance, Economics and Management

*e*finance**lab**

at the HOUSE OF FINANCE

- Background
- Important Concepts in Text Mining
- Preparatory Steps
- Sentiment Analysis
- Data
- Sentiment Analysis in R

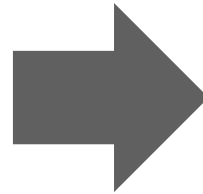


- User Generated Content (UGC): Word of mouth leaves a “digital trace”
 - Large amounts of unstructured data available from social channels, blogs, review sites
 - This data can contain relevant information for the business.
- Main questions
 - **What** are consumers talking about? – Discover underlying **topics** (topic modelling)
 - **How** are consumers talking? – Discover the emotional **tone** (sentiment analysis)



- Bag-of-words model (unigrams)
 - Focus on term frequencies (order does not matter)

“My bank is the best”



Document

Corpus

blog entry, review, tweet

Collection of documents

B Ben H
4 reviews

★★★★★ a year ago
(Translated by Google) Very friendly and helpful. Top Consulting. Always have a Council in stock. My bank of trust!

(Original)
Sehr freundlich und hilfsbereit. Top Beratung. Haben immer einen Rat auf Lager. Meine Bank des Vertrauens!

B Ben H
4 reviews

★★★★★ a ye
(Translated by Google) Very friendly and helpful. Top Consulting. Always have a Council in stock. My bank of trust!

(Original)
Sehr freundlich und hilfsbereit. Top Beratung. Haben immer einen Rat auf Lager. Meine Bank des Vertrauens!

Jette Sievers
6 reviews

★★★★★ 2 weeks
(Translated by Google) Here is an ir not refer to any hotli other stores. Keep it

(Original)
Sehr freundlicher Ku umgehend versucht irgendwelche Hotlin anderen Filialen pas

La Familia
27 reviews

★★★★★ 4 months ago
(Translated by Google) TOP, perfect service, top Internet banking, top prices, think currently the best bank, thank you, keep it up!

(Original)
TOP, perfekter Service, Top Internet Banking, Top Konditionen, denke zur Zeit die beste Bank, herzlichen Dank, weiter so!

Document Term Matrix

		Terms				
		bank	bad	fees	good	service
Documents	Document 1	1	0	1	0	0
	Document 2	3	0	2	1	0
	Document 3	0	0	0	1	1
	Document 4	1	1	0	0	4



Preparatory Steps

"Bin schon 10 Jahre da und sehr zufrieden."

1

Remove unwanted characters

Special characters
Punctuation
Numbers
Words < 2 characters

"Bin schon Jahre und sehr zufrieden"

2

Lower case +
Tokenization

Break down sentences
into words

"bin" "schon" "jahre"
"und" "sehr" "zufrieden"

3

Remove
"stop words"

Non informative
common words

"jahre" "zufrieden"



- Tone = positive – negative
- Term weighting
 - Polarity based: “good” and “excellent” have the same (positive) weight
 - Valence based: “excellent” receives more weight
 - Tf-idf weighting: terms are weighted by their importance in the document and in the whole corpus
- Need a dictionary of positive and negative words
 - We will use SentiWS (University of Leipzig)



- Subsample of Google Maps Reviews
- 41 branches in Frankfurt and surroundings (20 banks)
- 282 reviews in German
- Timespan: 2011 to 2018
- Source: Google Places API

All reviews

WRITE A REVIEW

5 ★
4 ★
3 ★
2 ★
1 ★

4.1
★★★★★
7 reviews

Sort by: Most helpful

J Jena Jena
1 review

★★★★★ 2 months ago
(Translated by Google) My bank branch in Jena is the 2nd week closed for health reasons. Deposits are only possible in this branch and only on Mondays and Wednesdays. The resulting non-deductible rates were dismissed in Frankfurt with a ... [More](#)

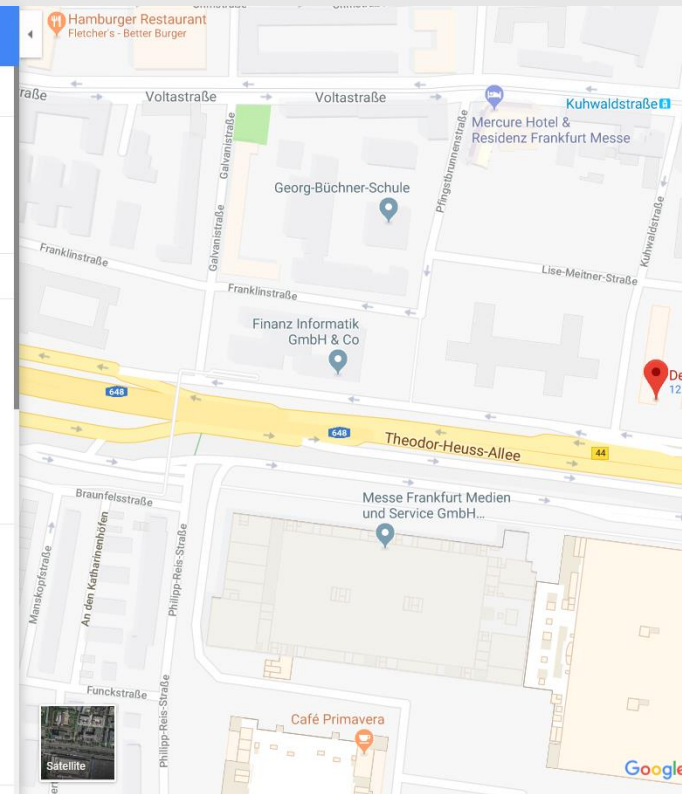
Like Share

A Angelika Wagner
Local Guide · 376 reviews · 279 photos

★★★★★ 9 months ago
(Translated by Google) Am a customer since the 80's and am very satisfied. Very good advice, good conditions. Now, however, you try to switch to online banking and customer must pay € 1 for a paper transfer. I do not like.

(Original) ... [More](#)

Like Share



SENTIMENT ANALYSIS IN R



- **Google Places API**
 - <https://developers.google.com/places/web-service/get-api-key>
 - <https://developers.google.com/places/web-service/details>
- **SentiWS Dictionaries (University of Leipzig)**
 - <http://wortschatz.uni-leipzig.de/de/download>
- **Dictionary of Emoticons**
 - <http://people.few.eur.nl/hogenboom/files/EmoticonSentimentLexicon.zip>
- **Regex Cheatsheet for R**
 - <https://www.rstudio.com/wp-content/uploads/2016/09/RegExCheatsheet.pdf>



Thank you!



Gabriela Alves Werb

gabriela.werb@wiwi.uni-frankfurt.de

Phone: +49 (0) 69 798 33 849



BACKUP



- Control for negation
 - “The service is awesome” vs “The service is far from awesome”
 - “The prices are the best” vs “The prices are not the best”
- Context
 - “I love this bank” vs “I love to wait forever in line in this bank”
- Natural Language Processing (NLP) is a rich and expanding field

