/media & digital technologies

/fh/// st.pölten

Reading Audio.

Sentiment Analysis of Comments about Music Videos on YouTube

Media Business Research Group | UAS St. Pölten
Andreas Gebesmair | Johanna Grüblbauer

SAMBA
Smart Data for Music Business Administration

Overview

- Brief introduction of SAMBA – The research project
- Demonstration of SambaVis – The interface
- Strengths and Weaknesses of Sentiment Analyses

On SAMBA

- Funded by FFG (6.COIN-Aufbau Call 2016), Project no. 856363
- From 01/2017 to 12/2019
- Collaboration of the Institute of Media Economy and the Institute of Creative Media Technology (St. Poelten UAS)
- Team: Andreas Gebesmair, Matthias Zeppelzauer, Johanna Grüblbauer, Christoph Musik, Christina Niederer, Alexis, Ringot, Victor Adriel de Jesus Oliveira
- Advisory Board: Bernhard Feichter (88.6.), Ulrich Raab (Universal Music), Andreas Rauber (TU Wien), Mario Rossori (Pate Records, VTMÖ), Markus Schedl (JKU Linz), Gerhard Wohlgenannt (WU)

Aim of the project

- The project aims at the utilization of un- and semistructured music related data from websites and Social Media for decision making in the music industries. We assess, adapt and optionally newly develop context-based methods of Music Information Retrieval (MIR) with regard to demands in the music business.
  - MIR algorithms are mostly used for B2C solutions (music recommendation), not as B2B instruments.
  - Standard instruments of Social Media Monitoring (SMM) are not very well adapted to special requirements in the music industry.

Research

- Research of existing business solutions, interviews with music managers, workshops with experts
  - All experts from the industry have a strong interest in social media marketing to promote their contracted artists and to monitor and evaluate the artists on social media platforms
  - They report using Google to search for what kind of online presence artists have. And in addition to Facebook, Instagram, YouTube, Soundcloud, they also use commercial SMM tools (e.g. ForTunes, Social Blade, Hootsuite, NextBigSound)
  - SMM is very time-consuming and emphasize that tools for social media monitoring need to be easy to use

Our solution

- Visualization of social media life cycles of songs based on comments to music videos on YouTube (database with over 600,000 comments about Austrian artists). Shows
  - social media metrics
  - language detection
  - sentiment analysis (AFINN, NLTK Vader, TextBlob)
  - Topic extraction (Gensim – Latent Dirichlet Allocation), Biterm Topic Model)

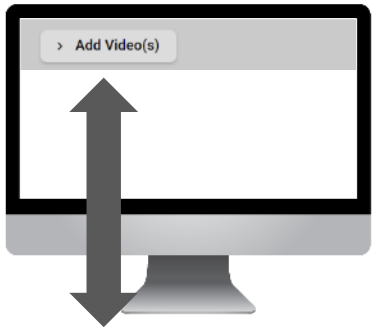
- https://jukebox.fhstp.ac.at/

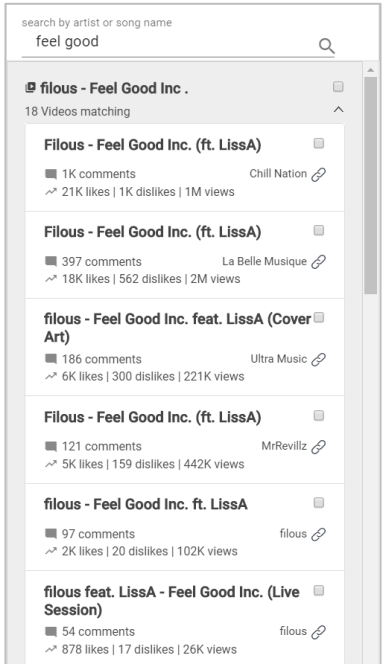
Our solution

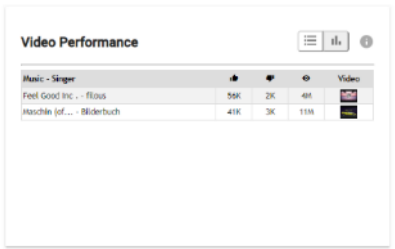
media & digital technologies

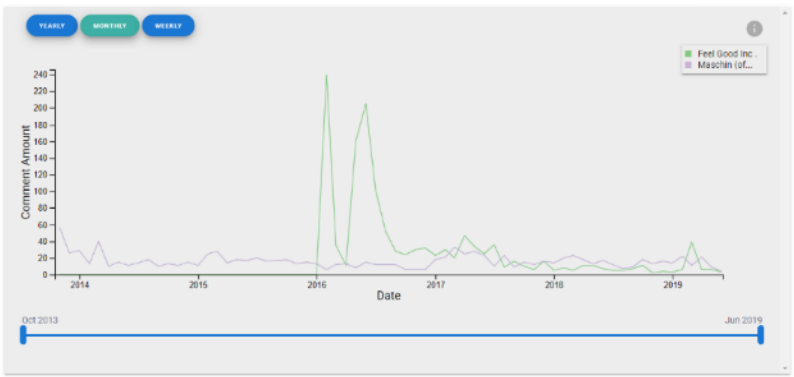
fh st.pölten

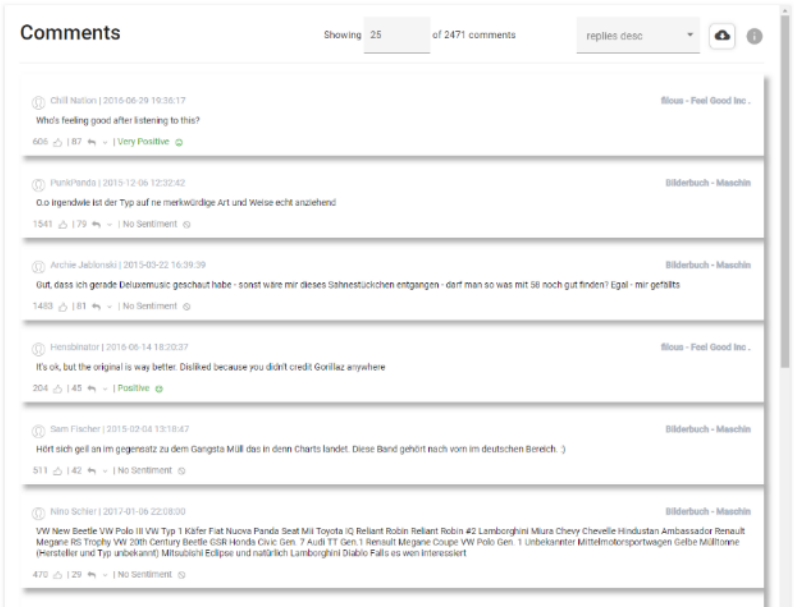
SEARCH → VIDEOS → COMMENTS → ANALYTICS

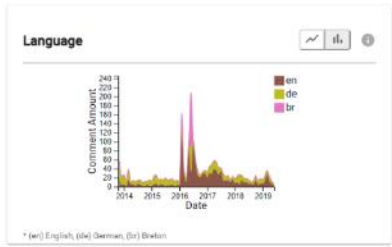


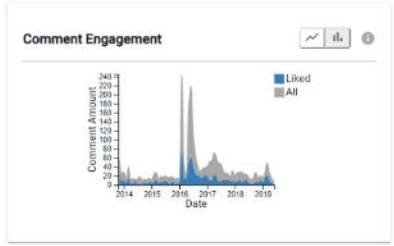


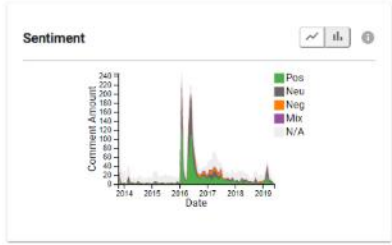


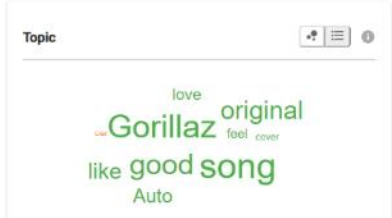






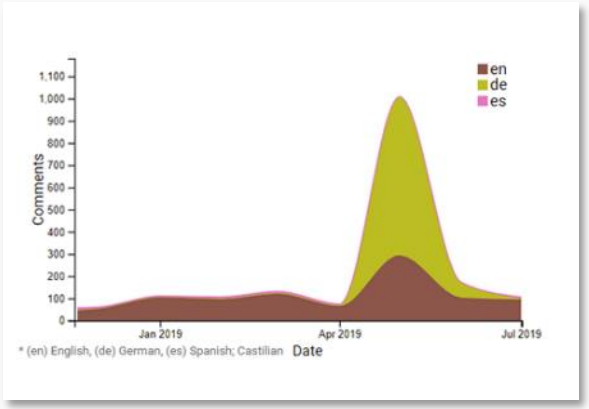






Example

“We're Going to Ibiza”
by Vengaboys
vs.
The Ibiza Affair



peak of German comments related to the topic

Sun Jan 17 2010 — Mon Feb 24 2020

What is sentiment analysis?

- [Sentiment Analysis] deal with the computational treatment of [..] opinion, sentiment, and subjectivity in text. (Pang and Lee 2008, S. 5-6)
- Sentiment Analysis is the computational study of opinions, sentiments and emotions expressed in text. (Liu 2010)
- Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity. The entity can represent individuals, events or topics. (Walaa et al. 2014)
- [Sentiment analysis] refers to determining one's attitude towards a particular target or topic. Here, attitude can mean an evaluative judgment, such as positive or negative, or an emotional or affectual attitude such as frustration, joy, anger, sadness, excitement, and so on. (Mohammad 2016)

Our definition

- Sentiment analysis evaluates,
  - with the help of human coders and/or algorithms,
  - in real time or afterwards,
  - the opinions, feelings, evaluations and emotions,
  - published in analog or digital form,
  - at the level of symbols, words, sentences, texts, spoken language, images, etc.
- The values can refer to different entities: e.g. products or services, organizations or brands or individuals, topics or events and associated attributes.

Human annotations

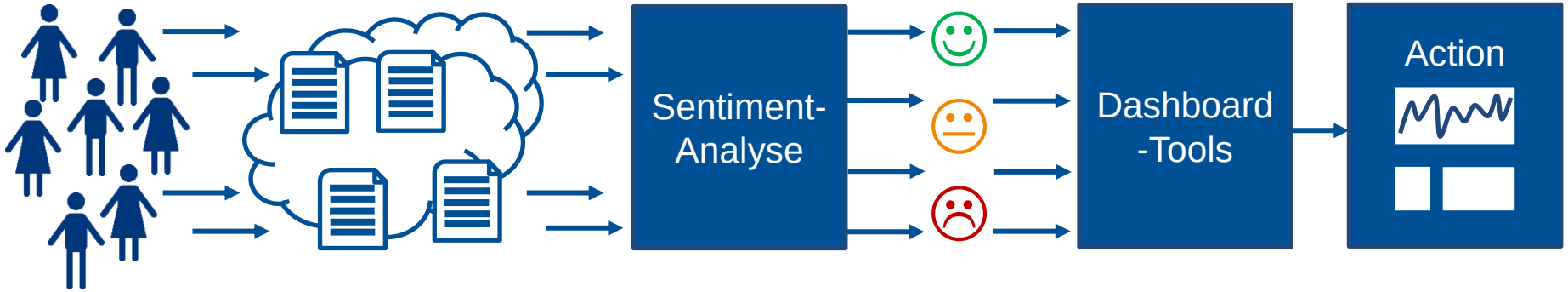
/media & digital technologies

/fh/// st.pölten

Comment	Human annotations					
I like both versions of the song this and the orginal but they kinda depress me lol	5	2	3	4	4	4
good but not my style.	4	4	3	4	2	4
I like da original better. Its still good tho.	4	4	4	4	2	4

5 = very positive … 1 = very negative

Sentiment analysis process



Numerous news & social media publishers constantly create new content

Sentiment analysis examines documents in relation to specific topics/events

Delivers real-time, actionable information on dashboards and in reports

Use cases

- Business Intelligence, Stock Exchange
- Marketing Intelligence, Brand Management & CRM
- Politics / Political Science
- Sociology
- Psychology
- Healthcare
- HR, personality trait recognition
- ...

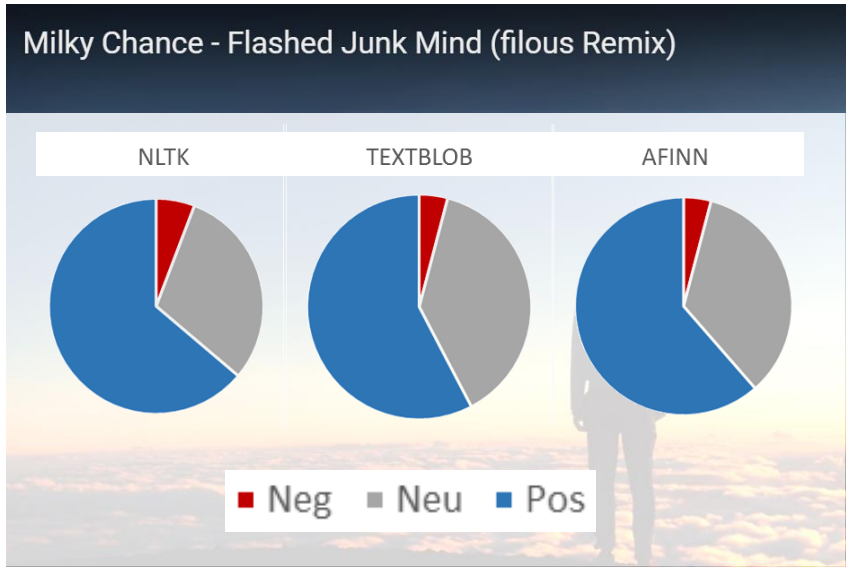
How do algorithms work? And how well?

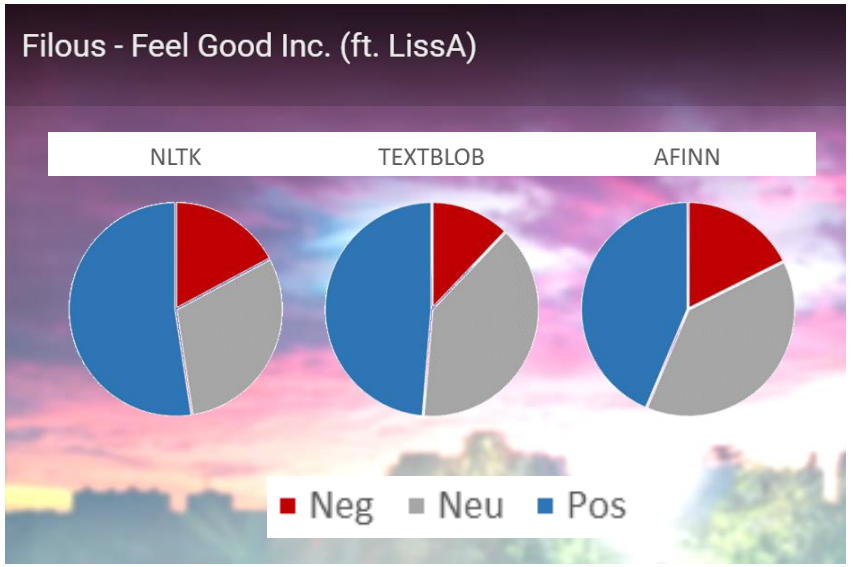
- Lexicon-based (incl. emoticons and emojis)
- Rule-based: Negations, emphasis by punctuation, intensifier (e.g. "really" as booster, "kind of" as dampener) are considered
- Unit of analysis: word or phrase; sentiment is shown for commentary
- Ambivalence can be assessed by parallel evaluation of positive and negative sentiment
- Distinction between polarity and intensity (subjectivity)
- Algorithms tend to judge more consistently than humans (agreement of ratings between algorithms is greater than between humans)
- 90% agreement in terms of direction between humans and algorithms

Eg. Bilderbuch „Maschin“

Kommentar	NLTK pos	NLTK neg	NLTK neut	NLTK compound	Textblob polarity	Textblob subjectivity	Afinn Result
Best German band from Austria!	0,529	0,000	0,471	0,670	0,500	0,150	3,000
Super!	1,000	0,000	0,000	0,636	1,000	1,000	3,000
Love this song so much from indo!	0,428	0,000	0,572	0,670	0,375	0,400	3,000
So ne scheisse 😂	0,000	0,000	1,000	0,000	-0,200	0,800	-4,000
I don't speak that much German, so even though I hear this as "na na na na nanana mesdfghjklqmeeeeeeennnn MASCHIIIIIINNNNNN", it's still a fucking great song :D	0,283	0,000	0,717	0,865	0,600	0,583	4,000
I don,t understand one word they saying,But, i like the beat good job.	0,375	0,000	0,625	0,660	0,700	0,600	5,000
nice car	0,737	0,000	0,263	0,422	0,600	1,000	3,000

Eg. Filous





Credit Gorillaz → Cover

Limitations

- No German-language lexicons, machine translation necessary in advance
- SA and Machine Learning: problem of domain-specific training data (movie review vs. YouTube commentary)
- No distinction between "neutrality" and "non applicable" (sentiments are always calculated - even for meaningless comments)
- No consideration of the broader context such as the object to which the "sentiment" refers, other comments, general knowledge about the object domain
- In this respect also no understanding of irony, sarcasm, off-topic comments

Eg. Bilderbuch „Maschin“

Kommentar	NLTK pos	NLTK neg	NLTK neut	NLTK compound	Textblob polarity	Textblob subjectivity	Afinn Result
2:17 is killing me	0,000	0,595	0,405	-0,660	0,000	0,000	-3,000
Gibt's irgendwo eig auch die lyrics? würd mich interessieren was er wirklich singt...versteh nur die hälfte :D aber hammer lied :P	0,213	0,310	0,477	-0,500	0,339	0,644	2,000
my midas finance circle,whom i sent to the EZB,also,as nick, to the white house and the german finance ministry, instruct,that through the national bank, the "netto"earn of the whole society members rise variablable, for example ten times, or hundredfifty as the original sum (...)	0,109	0,008	0,883	0,945	0,153	0,419	1,000

0 = sehr objektiv
1 = sehr subjektiv

Thank you for your attention!

- Andreas Gebesmair & Johanna Grüblbauer
  Media Business Research Group
- St. Poelten University of Applied Sciences
  Matthias Corvinus-Straße 15, A-3100 St. Pölten,
  mb.fhstp.ac.at/