

Sentiment analysis

SIMPLE TEXT ANALYTICS WITH R AND TIDYTEXT

PEETER TINITS 07.02.2018

A solid orange horizontal bar at the bottom of the slide.

Sentiment analysis of storylines

Words and phrases with known sentiment can be picked out in texts

The screenshot displays the Bitext API Test Tool interface. At the top, there's a red header with the text "API TEST TOOL". Below this, there are three dropdown menus: "English", "Sentiment", and "Graphical". The main area shows a text input field containing a paragraph about cameras. The text is analyzed and displayed with sentiment tags. A legend on the right side explains the color key for sentiment: "Sentiment topic" (white), "Positive sentiment text" (green), and "Negative sentiment text" (red). The text is also numbered 1 through 7, indicating text and topic links. At the bottom, there are two buttons: "ANALYZE TEXT" and "RESET".

English Sentiment Graphical

1 really enjoyed using the 1 Canon Ixus in Madrid on March 4. The
2 Panasonic Lumix 2 is a bit disappointing, but the 3 Canon 3 camera is
3 not bad at all. All I want when taking photos is point it and then just press the
4 really fair 4 price, this 5 camera is 5 perfect
for me. Besides, I have had a 6 good 6 customer 6 service 6 experience
7 John Faraday was 7 very nice!

LEGEND color key SENTIMENT

Sentiment topic
Positive sentiment text
Negative sentiment text
1 Text and topic link

ANALYZE TEXT RESET

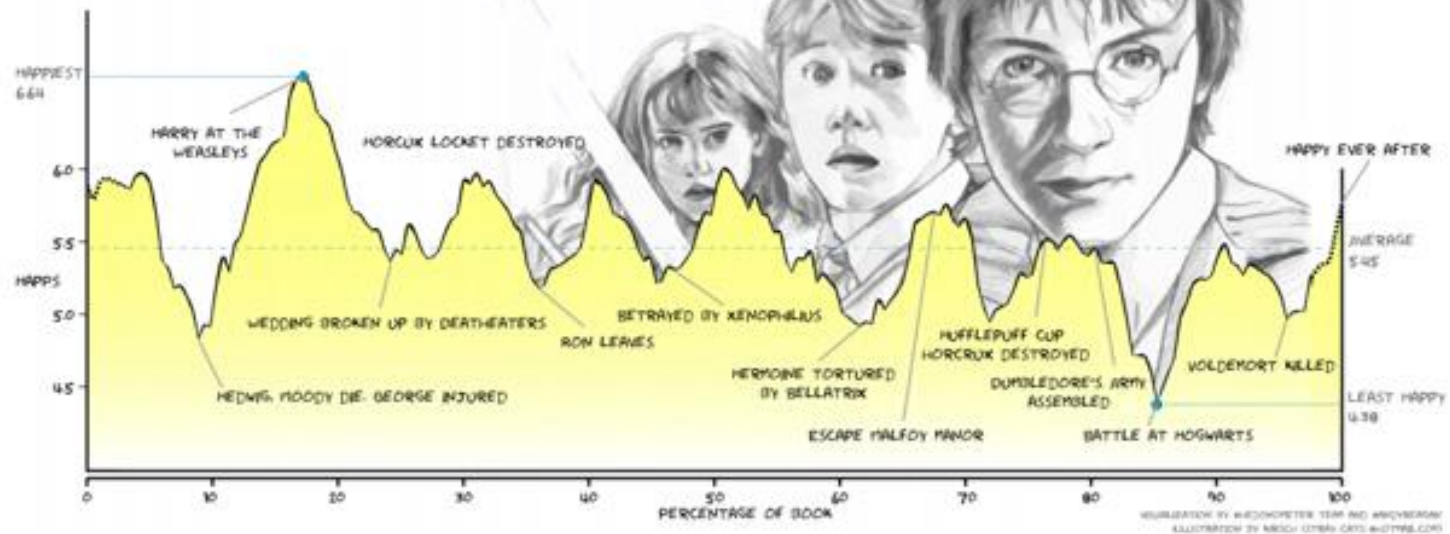
Bitext tool

Sentiment analysis of storylines

Words and phrases with known sentiment can be picked out in texts

API TEST TOOL

Harry Potter and the Deathly Hallows by J.K. Rowling



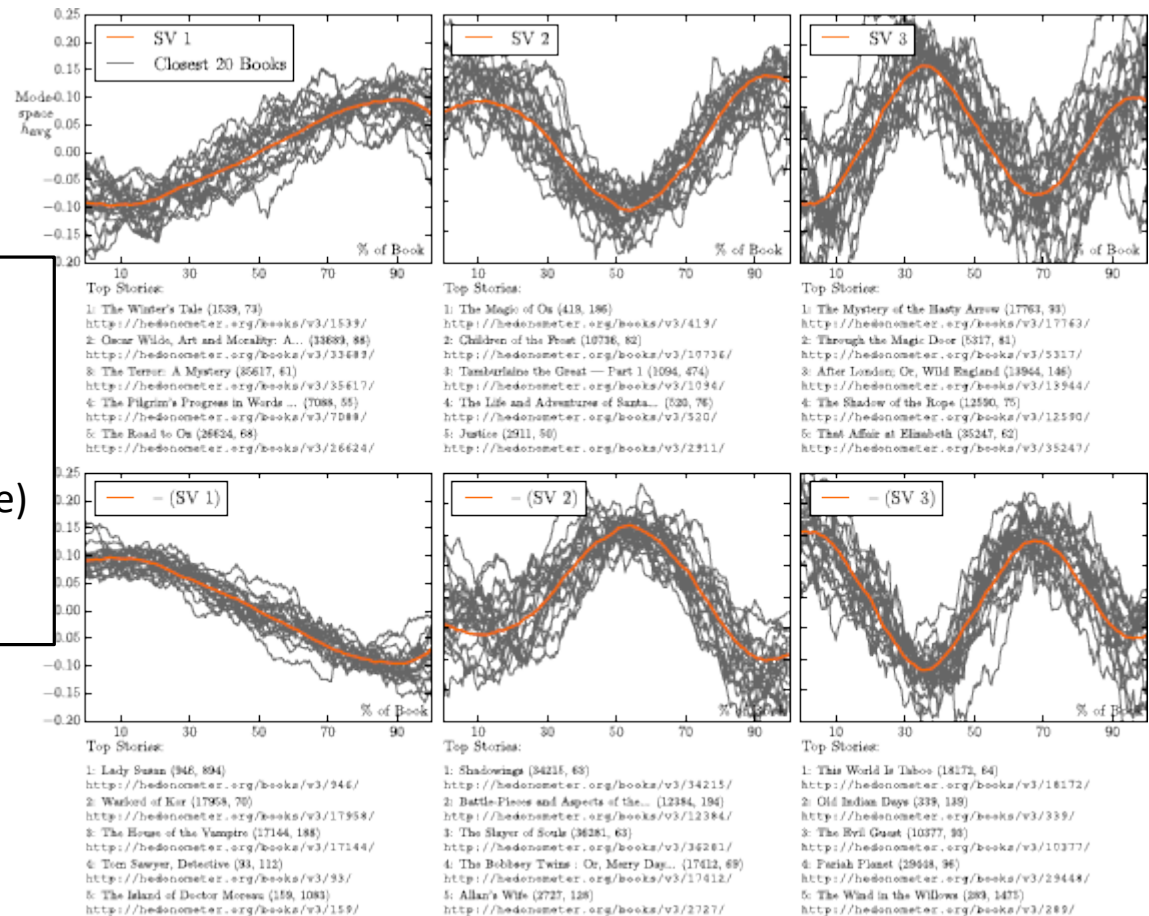
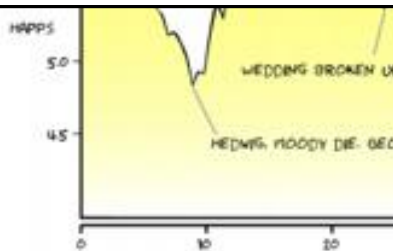
[The story arc of Harry Potter according to AI \(Hedonometer team and Andrew Reagan\)](#)

Sentiment analysis of storyline

Words and phrases

API TEST TOOL

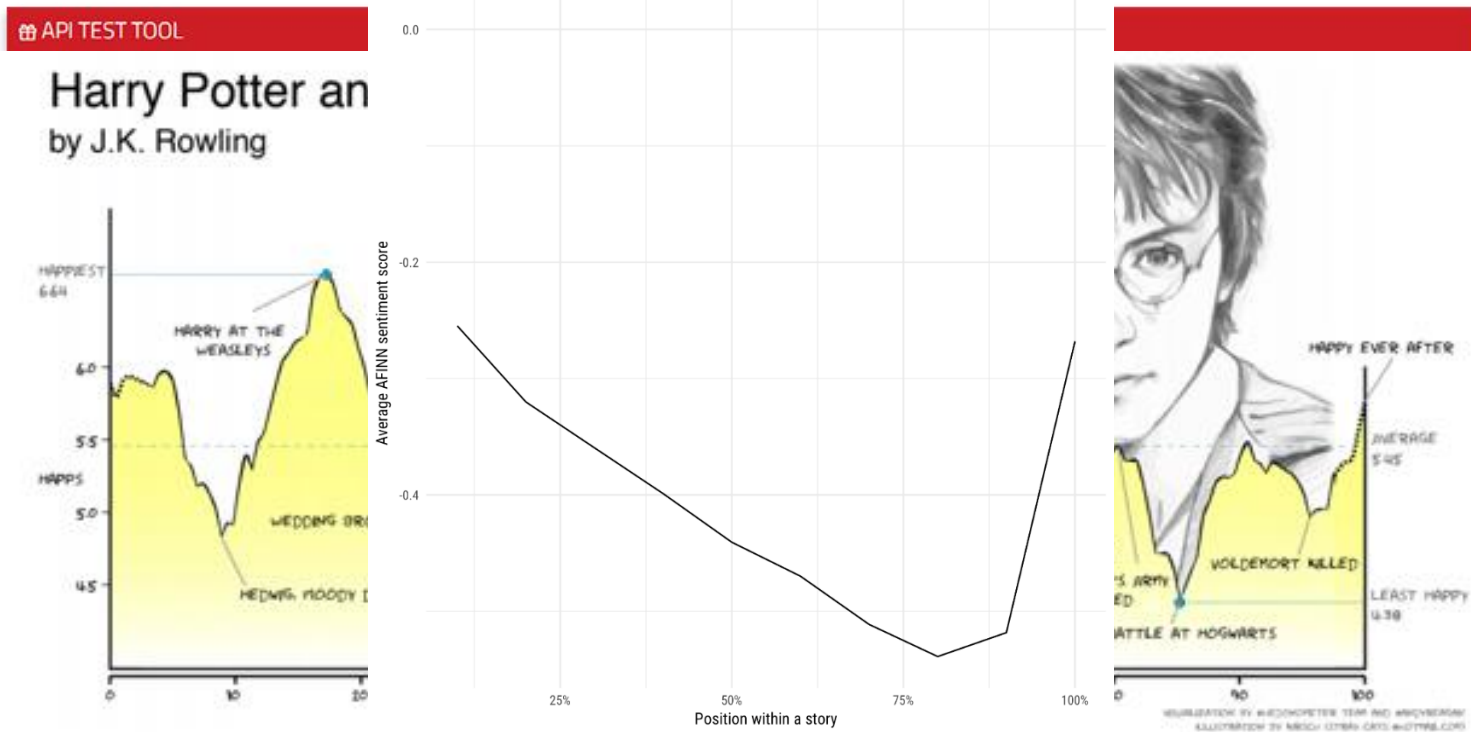
1. Rags to Riches (rise)
2. Riches to Rags (fall)
3. Man in a Hole (fall then rise)
4. Icarus (rise then fall)
5. Cinderella (rise then fall then rise)
6. Oedipus (fall then rise then fall)



Reagan et al. 2016 The emotional arcs of stories are dominated by six basic shapes

Sentiment analysis of storylines

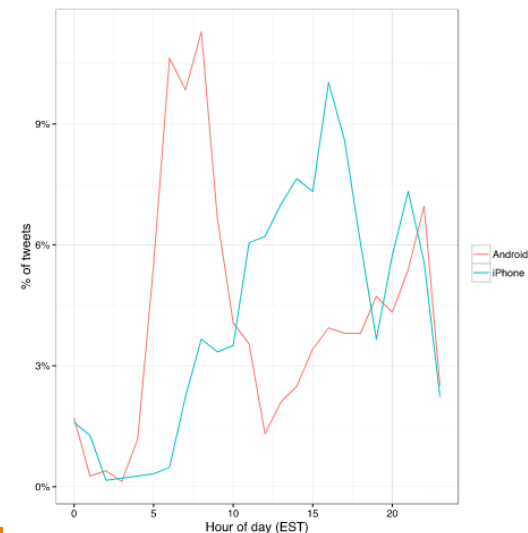
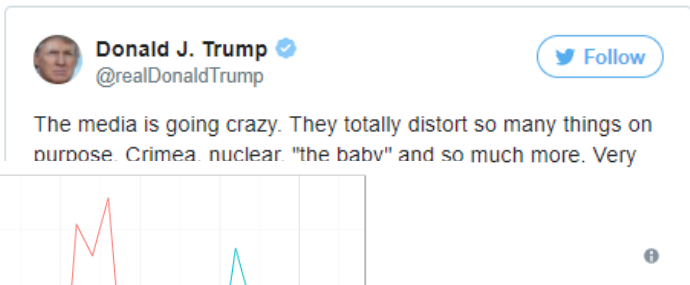
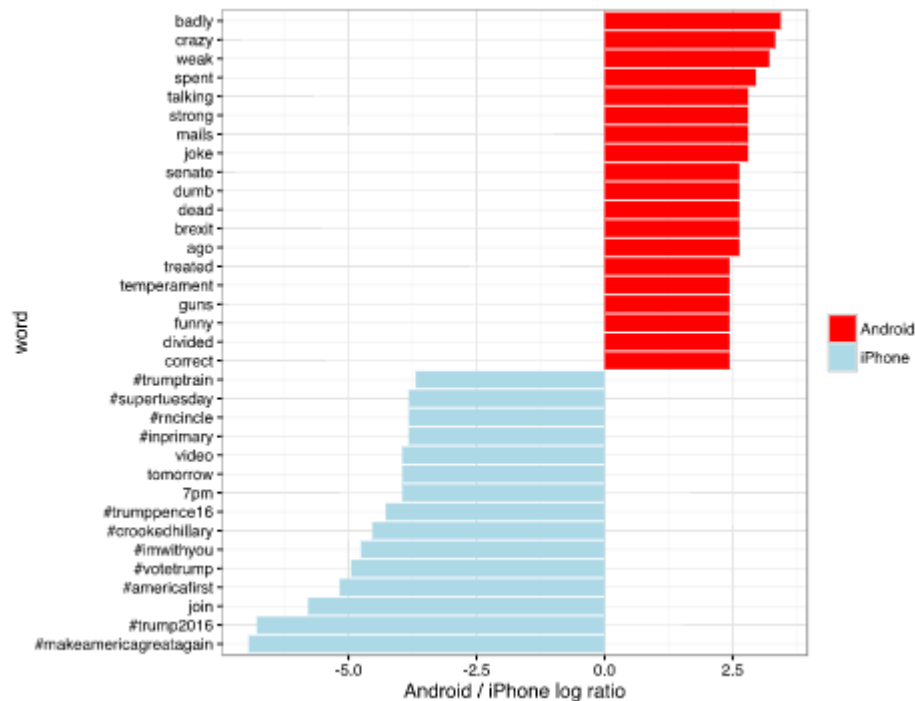
Words and phrases with known sentiment can be picked out in texts



David Robinson 2017, Examining the arc of 100,000 stories: a tidy analysis

Text analysis of Trump's tweets confirms he writes only the (angrier) Android half

I don't normally post about politics (I'm not particularly savvy about polling, which is where data science has had the largest impact on politics). But this weekend I saw a hypothesis about Donald Trump's twitter account that simply begged to be investigated with data:



Birth of the cool

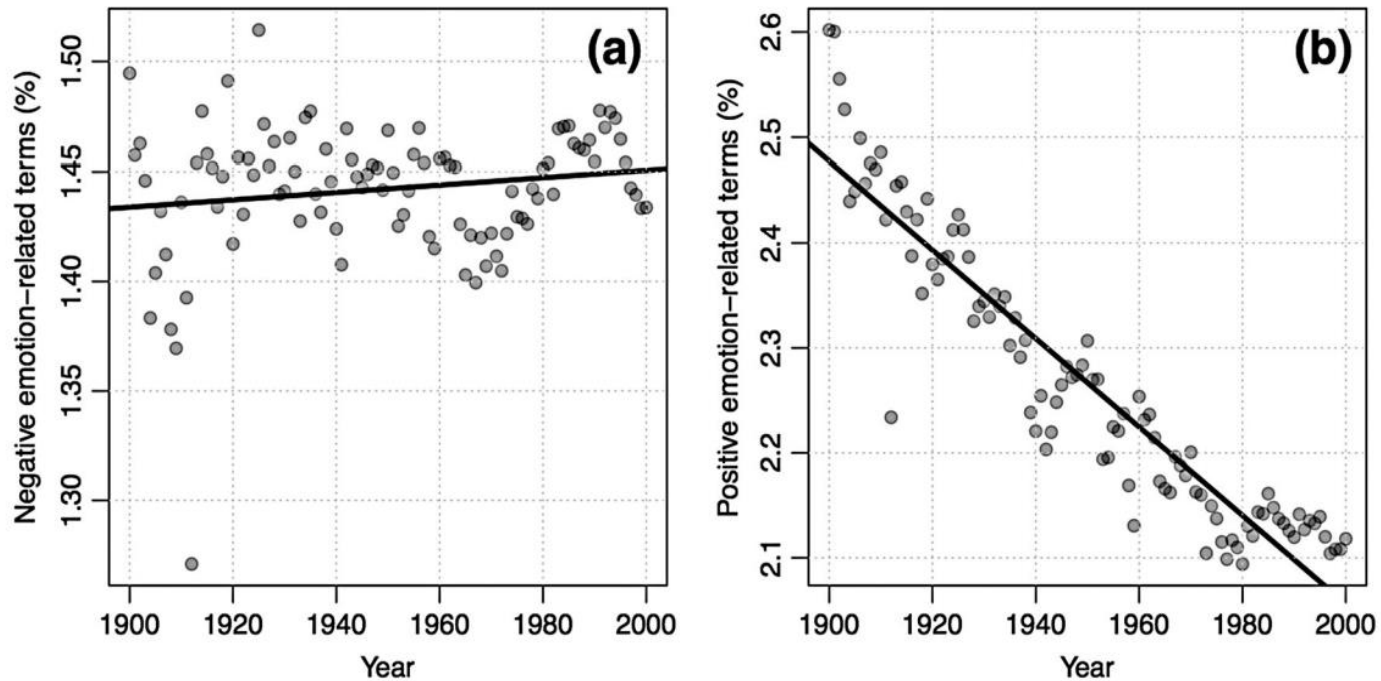


Figure 3. Emotionality changes in Anglophone literature, for the Google Books corpus. (a): negative emotions-related terms. (b): positive emotions-related terms. Solid lines represent linear regressions of the data.

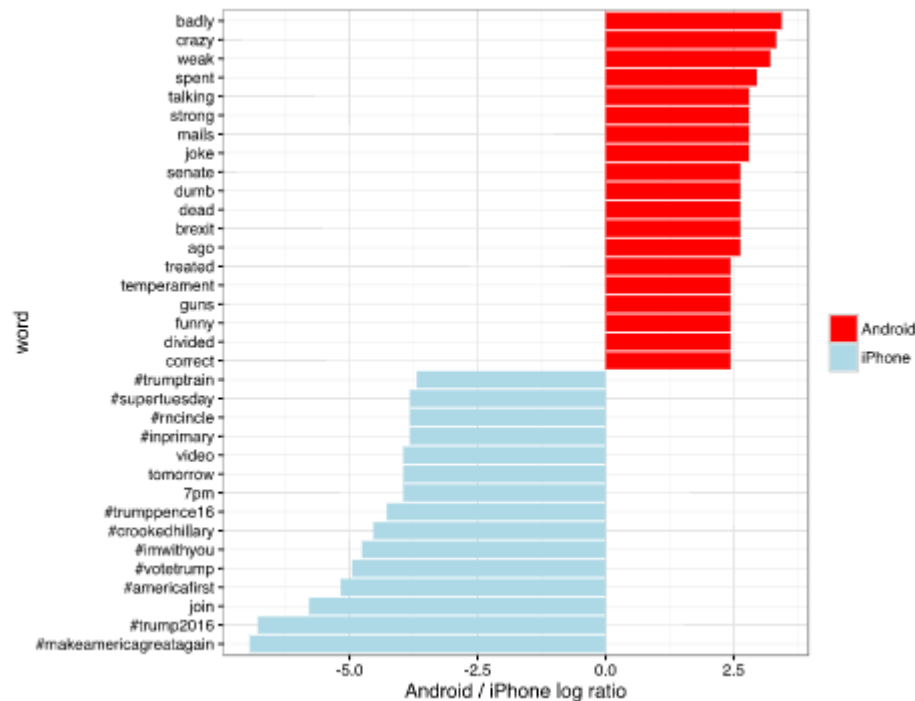
Sentiment analysis in Lord of the Rings.



[Jakub Glinka 2017. Sentiment Analysis of the Lord of the Rings in tidytext](#)

Keywords in 2 groups of texts

E.g. Trump twitter Android & iPhone



Basic operations on texts

- `mutate(new_var = op(old_var))` – make/change column
- `filter(var==what_you_want)` – get only the rows with right value
- `count(var)` – count unique values
- `arrange(column)` – sort by column
- `arrange(desc(column))` – reverse
- `group_by(what)`
- `ungroup()`
- `rename()` – just when needed
- `summarize()` – based on `group_by`
- `unnest_tokens(words,line)` – make lines into words
- `str_detect(where, „what“)` – check if it contains „what“
- `anti_join(with_what, „var“)` – remove matching values
- `inner_join(with_what, „var“)` – keep only matching values
- `left_join(with_what, „var“)` – just add where possible
- `bind_df_idf(of, by, value)` – get keywords