

# MineMyText

## User Manual

# What is MineMyText?

MineMyText is a cloud app that allows to discover, quantify and visualize **topics** and **sentiments** in large collections of **unstructured text documents**.



## Sentiments

- o I bought this for my 14 year old daughter as a gift. She received it in July. It works great - she lost 6 pounds in 2 weeks. The Fitbit makes staying in shape easy. The iPhone app works fine.
- +
- +

## Exemplary Customer Review about a Fitbit Flex

## Topics

**Birthday present**

**Loosing weight**

**Mobile app**

# Algorithms

## Topic Modeling

The idea behind topic modeling is that words that co-occur together in similar contexts tend to have similar meanings. Hence, sets of highly co-occurring words (e.g., easy, intuitive, use) can be interpreted as topics (e.g., usability) and used to cluster documents into thematic categories.

Latent Dirichlet Allocation (LDA) is a popular topic modeling algorithm that is able to discover topics running through a large collection of documents and to annotate individual documents with topic labels. As an unsupervised machine learning algorithm LDA is purely data-driven and inductively infers topics from given texts — neither necessitating any manual labeling of documents, nor the existence of predefined categories.

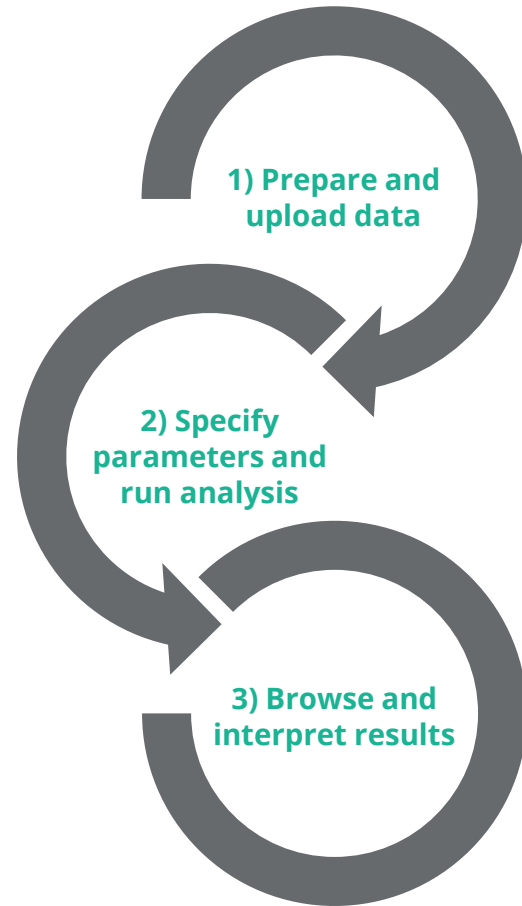
[http://en.wikipedia.org/wiki/Latent\\_Dirichlet\\_allocation](http://en.wikipedia.org/wiki/Latent_Dirichlet_allocation), <http://sentistrength.wlv.ac.uk>

## Sentiment Analysis

Sentiment analysis deals with the quantitative measurement of opinion, attitude and subjectivity in texts.

The SentiStrength algorithm estimates the strength of positive and negative emotions expressed in short texts. SentiStrength follows a dictionary-based approach to sentiment analysis and, therefore, can operate in many different domains. Besides relying on a dictionary of words with human sentiment polarity and strength judgments, it also exploits other (non-lexical) information, such as, negation, booster words, idioms, emoticons or punctuation.

**ONLY 3 STEPS**



# Step 1 – Prepare and Upload

MineMyText currently accepts two input formats.

1. **XLSX:** Microsoft Excel file with one document per row and the content and metadata in the columns
  - Mandatory text column: There must be a column labeled “text” containing the text to be analyzed
  - Optional date column: If your documents have timestamps, they have to be placed in a column labeled “date” and the column has to be formatted as a date datatype
  - Optional additional metadata: All other columns will be treated as metadata that can be used to filter and sort results
2. **JSON:** An array of JSON objects (one document per object), each with a “text” and “date” (YYYY-MM-DD) name/value pair. All other name/value pairs will be treated as additional metadata.

# Step 1 – Prepare and Upload (contd.)

## XLSX Example

	A	B	C
1	text	date	rating
2	Arrived not working (fully charged, but won't turn on). Replace	5/12/2015	2
3	My Fitbit flex failed after 6 days with no water contact or anyth	5/12/2015	1
4	I absolutely love my Fitbit! It encourages me to get up and mov	5/12/2015	5
5			

# Step 1 – Prepare and Upload (contd.)

## JSON Example

```
[{
  "date": "2015-05-12",
  "rating": 2.0,
  "text": "Arrived not working (fully charged, but won't turn on). Replacement on the way - will update once received."
},
{
  "date": "2015-05-12",
  "rating": 1.0,
  "text": "My Fitbit flex failed after 6 days with no water contact or anything! Now it only works when plugged in! No amount of flicks or resets will save it.."
},
{
  "date": "2015-05-12",
  "rating": 5.0,
  "text": "I absolutely love my Fitbit! It encourages me to get up and move around every day. My friend and I compete to see who can get the most steps."
},
...
]
```

# Step 2 – Run Analysis

- Two sets of parameters have to be defined
  - Topic Modeling
    - **Number of topics:** Choosing an appropriate number of topics is crucial for getting good analysis results. Choosing too many topics can lead to many very similar topics; choosing too few topics unnecessarily constrains the exploratory potential of the method. The optimal number of topics also depends on the size and diversity of the corpus. Therefore, we advise to experiment with different topic numbers.
  - Natural Language Pre-Processing
    - The quality of the results also depends on the preprocessing options. Natural language typically contains a lot of noise that should be removed, e.g., through
      - **Stopword filtering** ([http://en.wikipedia.org/wiki/Stop\\_words](http://en.wikipedia.org/wiki/Stop_words)),
      - **Lemmatizing** (<http://en.wikipedia.org/wiki/Lemmatisation>),
      - **Stemming** (<http://en.wikipedia.org/wiki/Stemming>), or
      - **Removing HTML tags and digits.**
- Performing topic modeling on a large collection of documents (10.000+) can take several minutes or even hours. Be patient, you will receive an e-mail when your analysis is done.



# Step 3 - Interpret Results

## Project Overview

What topics are people writing about?

### Topic 1

day step week work walk time steps\_day  
walking couple end

### Topic 2

wear shower water time band love charge  
comfortable wear\_shower swimming

### Topic 3

weight lost pound lose loss week lb  
lost\_pounds month weight\_loss

### Topic 5

wrist wearing wear time zip pedometer  
put thing lost clip

### Topic 6

heart rate heart\_rate monitor rate\_monitor  
heart\_rate\_monitor blood pressure blood\_pressure  
pedometer

### Topic 7

sleep tracking time night step sleep\_tracking  
day pattern feature hour

### Topic 9

gift love christmas bought husband daughter  
received birthday gave present

### Topic 10

motivated move love day walk make step  
moving motivate keeps\_motivated

### Topic 11

product great recommend great\_product love  
good recommend\_product excellent not\_recommend  
good\_product

### Topic 13

sleep mode sleep\_mode put time tap  
tapping put\_sleep forget turn

### Topic 14

track sleep step keep\_track track\_steps love  
activity track\_sleep great keeps\_track

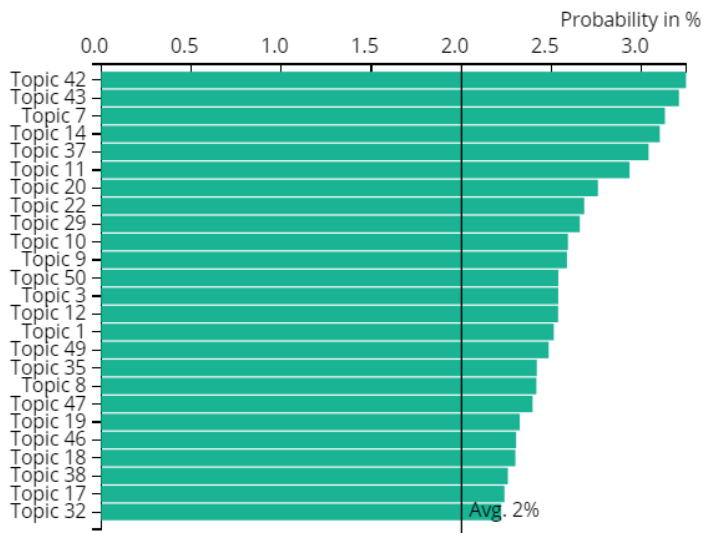
### Topic 15

stair ultra count track climbed step big  
flight deal floor

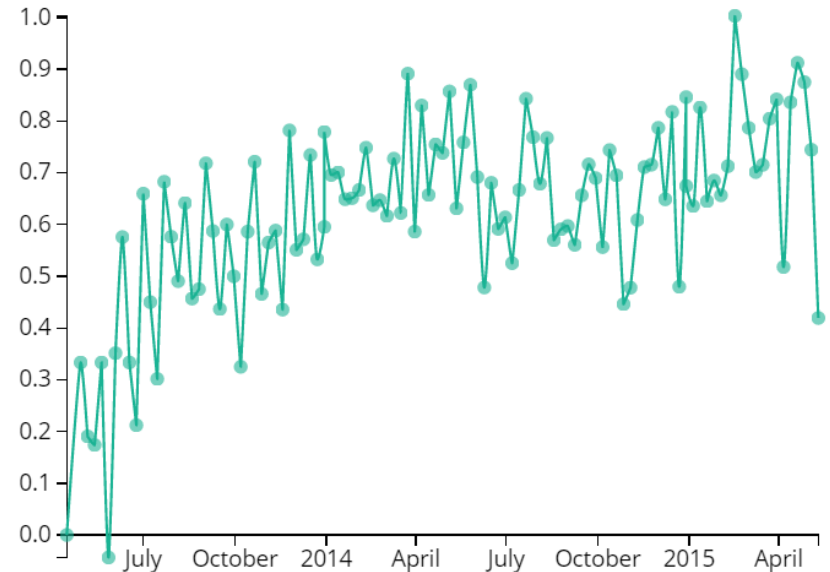
# Step 3 - Interpret Results

## Project Overview

How frequent are topics?



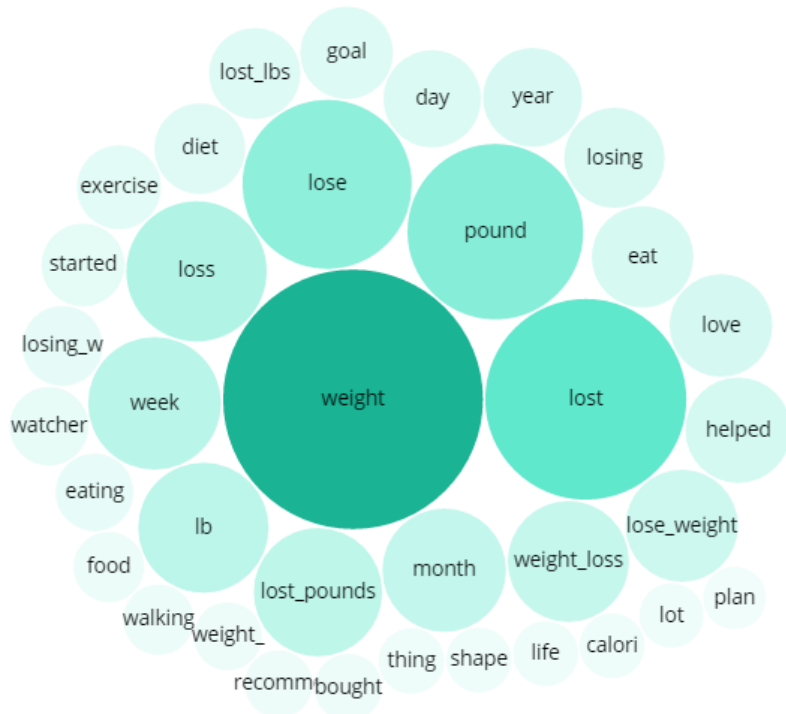
How is the sentiment developing?



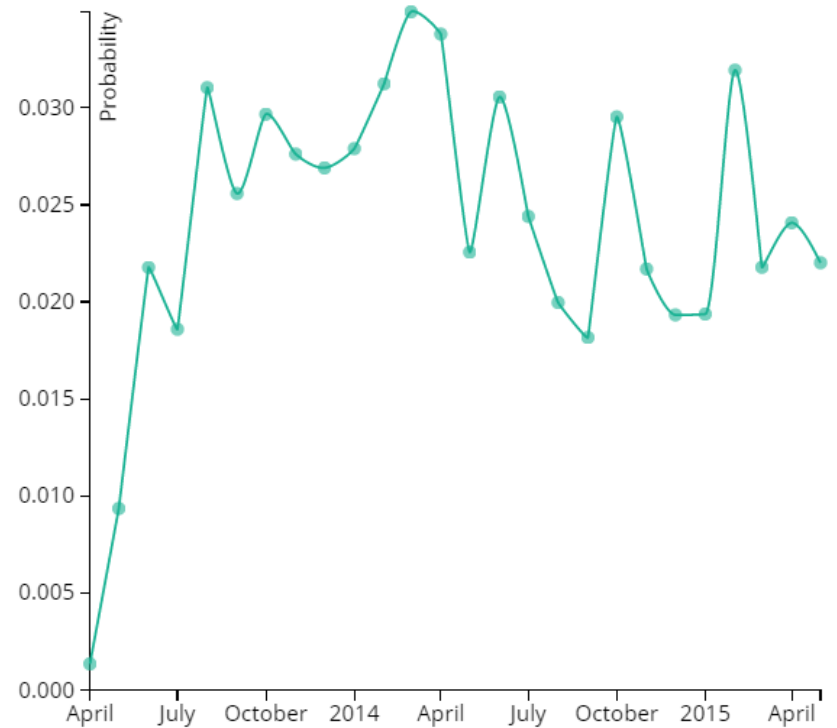
# Step 3 - Interpret Results

## Topic Details

What is a given topic about?



How is a topic developing over time?



# Step 3 - Interpret Results

## Topic Details

Which documents are about a given topic?

Probability ▾	Rating	Text
81.1%	5.0	Helped me reach my weight loss goals and maintain my weight for 6 weeks now. Has become part of me 24/7.
78.17%	5.0	The FitBit has helped my make a total lifestyle change. I've lost 27 pounds so far, and counting.
78.01%	5.0	I AM TRYING TO LOSE WEIGHT FOR 3 EVENTS THIS SUMMER. THE FITBIT HAS HELPED WITH MY MOTIVATION A LOT. I HAVE NOT LOST A LOT OF WEIGHT. BUT WHEN SOMEONE SEES ME AND THEY HAVE NOT SEEN ME FOR A WHILE THEY SAY "YOUR GETTING SKINNY". I WOULD RECOMEND THIS ITEM TO ANYONE WANTING TO GET MOTIVATION TO MOVE AND GET FIT.
76.33%	5.0	I love mine. It helped me lose 20 lbs.
72.31%	5.0	I bought this at the end of July 2014. I wanted to track what I was eating, loose about 15 pounds, and increase my general fitness. I will be 60 years old in May. I have changed my lifestyle, I eat less food and things that are more healthy. I have lost about 33 pounds as I just kept going. I feel better than I have in 20 years, my general fitness is outstanding now since I have paid attention. The Fitbit Flex has been an integral part of my program. It's an outstanding value to accomplish all of this for me. The only problem, I need a lot of new clothes, my waist size decreased 4 inches, in a little over 3 months. I walk very briskly twice a day for my exercise. If you will learn how to use the Fitbit, it is really quite simple and very worthwhile.

**Click [here](#) to see  
MineMyText live**