## What is sentiment analysis - A practitioner's perspective:

Essentially, sentiment analysis or sentiment classification fall into the broad category of text classification tasks where you are supplied with a phrase, or a list of phrases and your classifier is supposed to tell if the sentiment behind that is positive, negative or neutral. Sometimes, the third attribute is not taken to keep it a binary classification problem. In recent tasks, sentiments like "somewhat positive" and "somewhat negative" are also being considered. Let's understand with an example now.

Consider the following phrases:
1. "Titanic is a great movie."
2. "Titanic is not a great movie."
3. "Titanic is a movie."

The phrases correspond to short movie reviews, and each one of them conveys different sentiments. For example, the first phrase denotes positive sentiment about the film Titanic while the second one treats the movie as not so great (negative sentiment).

Take a look at the third one more closely. There is no such word in that phrase which can tell you about anything regarding the sentiment conveyed by it. Hence, that is an example of neutral sentiment.

Data:
The dataset is comprised of tab-separated files with phrases from the Rotten Tomatoes dataset. The train/test split has been preserved for the purposes of benchmarking, but the sentences have been shuffled from their original order. Each Sentence has been parsed into many phrases by the Stanford parser. Each phrase has a PhraseId. Each sentence has a SentenceId. Phrases that are repeated (such as short/common words) are only included once in the data.

train.tsv contains the phrases and their associated sentiment labels. We have additionally provided a SentenceId so that you can track which phrases belong to a single sentence.
test.tsv contains just phrases. You must assign a sentiment label to each phrase. The sentiment labels are:

## Problem Statement:

Input:
A document d

A fixed set of classes C = $\{c_1, c_2, .., c_n\}$

Output: A predicted class c $\in\in$ C

## Evaluation:

Submissions are evaluated on classification accuracy (the percent of labels that are predicted correctly) for every parsed phrase. The sentiment labels are:

0 - negative
1 - somewhat negative
2 - neutral
3 - somewhat positive
4 - positive

## Submission Format

For each phrase in the test set, predict a label for the sentiment. Your submission
should have a header and look like the following:

```
PhraseId, Sentiment
156061,    2
156062,    2
156063,    2
…
```