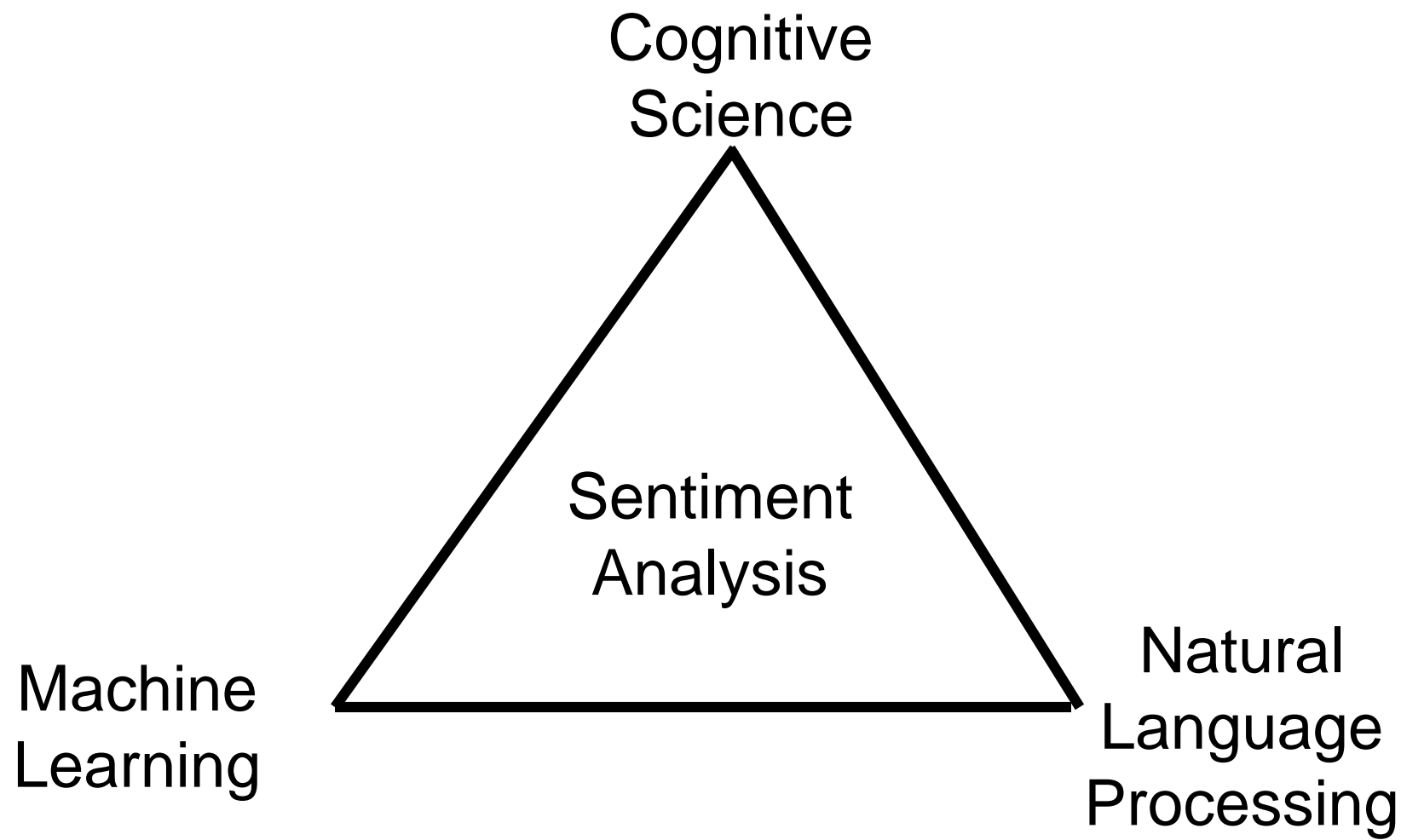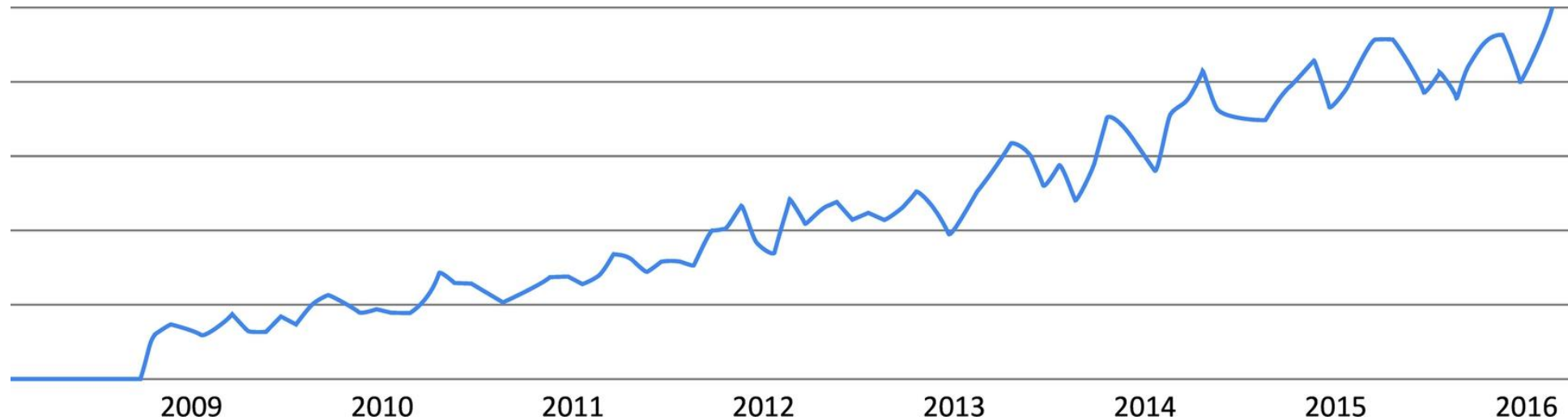# Sentiment Analysis

In Social Networks

- Sentiment analysis, which is also called opinion mining, has been one of the most active research areas in natural language processing since early 2000.

- The aim of sentiment analysis is to define automatic tools able to extract subjective information from texts in natural language, such as opinions and sentiments, so as to create structured and actionable knowledge to be used by either a decision support system or a decision maker.
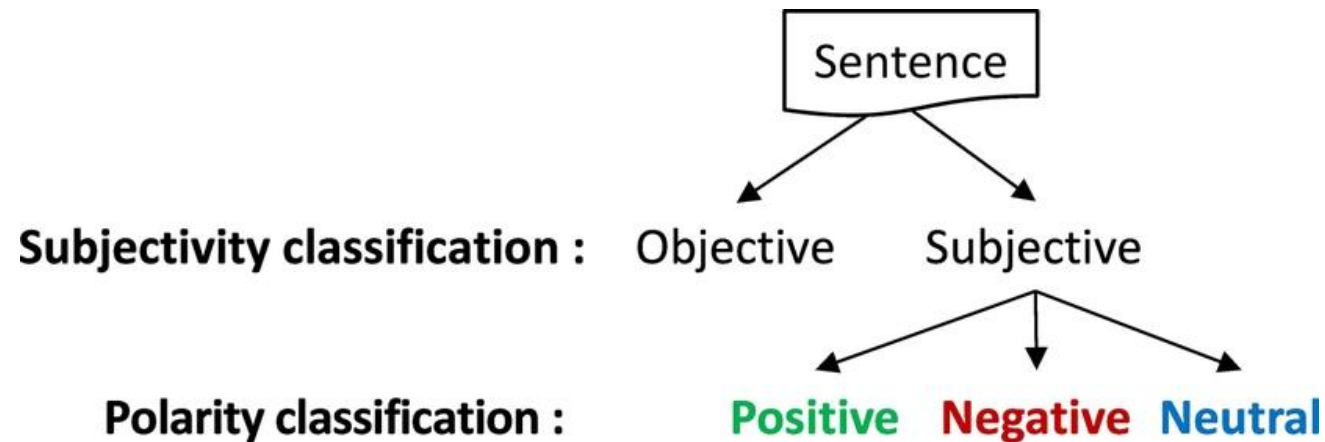
Tripod of Sentiment Analysis

Google Trends data related to the keywords sentiment analysis.

Thanks to its strong applicability and interest in both the academic field and the industrial field, sentiment analysis is nowadays a trending topic.
The above figure represents the Google Trends data related to the keywords *sentiment analysis*, clearly demonstrating the continuous and increasing interest in this field.

Sentiment analysis workflow

The first aim when one is dealing with sentiment analysis usually consists in distinguishing between subjective and objective sentences.

If a given sentence is classified as objective, no other fundamental tasks are required, while if the sentence is classified as subjective, its polarity (positive, negative, or neutral) needs to be estimated

Levels of Analysis

- The aim of sentiment analysis is to "*define automatic tools able to extract subjective information **from texts** in natural language*."

- The first choice when one is applying sentiment analysis is to define what *text* (ie, the analyzed object) means in the case of study considered.

- In general, sentiment analysis in social networks can be investigated mainly at three levels, message level, sentence level and entity level.

Different levels of analysis.
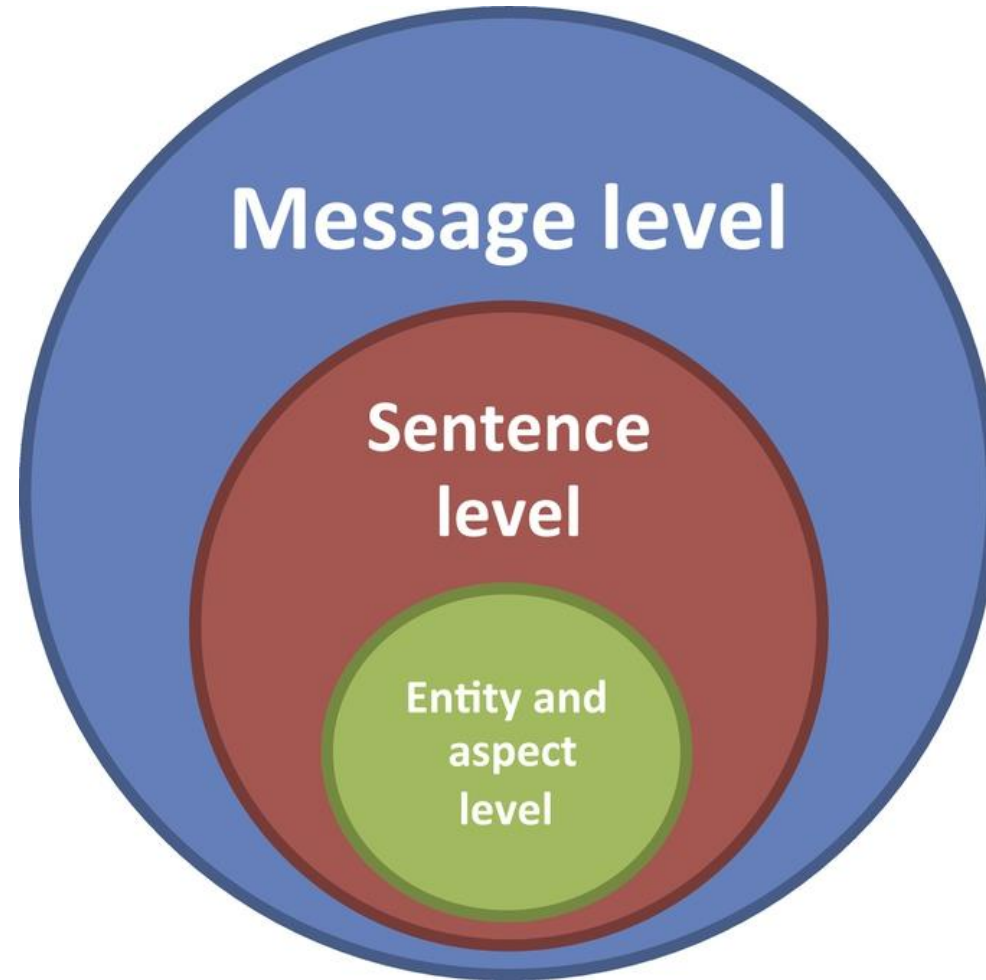
- **Message level**: The aim is to classify the polarity of a whole opinionated message.

- For example, given a product review, the system determines whether the text message expresses an overall positive, negative, or neutral opinion about the product.

- The assumption is that the entire message expresses only one opinion on a single entity (eg, a single product).

- **Sentence level**: The aim is to determine the polarity of each sentence contained in a text message. The assumption is that each sentence, in a given message, denotes a single opinion on a single entity.

- **Entity and aspect level**: Performs a finer-grained analysis than message and sentence level. It is based on the idea that an opinion consists of a *sentiment* and a *target* (of opinion).

- For example, the sentence "*The iPhone is very good, but they still need to work on battery life and security issues*" evaluates three aspects: iPhone (positive), battery life (negative), and security (negative).

Regular Versus Comparative Opinion

- **Regular opinion**: A regular opinion is often referred to in the literature as a *standard opinion* and it has two main subtypes:

    - **Direct opinion**: A direct opinion refers to an opinion expressed directly on an entity (eg, "*The **screen brightness** of the iPhone is awesome*").

    - **Indirect opinion**: An indirect opinion is an opinion that is expressed indirectly on an entity on the basis of its effects on some other entities.
    - For example, the sentence "*After I switched to the iPhone, I lost all my data!*" describes an undesirable effect of the switch on "my data," which indirectly gives a negative sentiment to the iPhone.

Comparative opinion

- A comparative opinion expresses a relation of similarities or differences between two or more entities and/or a preference of the opinion holder based on some shared aspects of the entities.

- For example, the sentences, *"iOS is better performing than Android"* and *"iOS is the best performing operating system"* express two comparative opinions.

- A comparative opinion is usually expressed with use of the comparative or superlative form of an adjective or adverb.

The Role of Semantics

- The semantics of the language used in social networks is fundamental to accurately analyze user expressions.

- The context of a textual expression is therefore a crucial element that should be taken into account to properly deal with the underlying sentiment.

- A sentence "taken as it is" can appear as negative or positive, but if it is properly analyzed from a semantic point of view it can be completely different.

- For instance, the sentences "I watched the most terrific horror movie. It was like a real nightmare! PAAAANIIIICCC" can be initially interpreted as negative, but taking into account the context where these kinds of opinions are expressed (ie, a community of horror-movie lovers) and some lexical cues that are typical of the social network language, we should derive a (real) positive judgment.

- Lexica, corpora, and ontologies need to be properly constructed and used for us to have a deep understanding of the semantics of the natural language in online social networks.

Dealing With Figures of Speech

- A figure of speech is any artful deviation from the ordinary mode of speaking or writing.

- In the tradition of Aristotle, figures of speech can be divided into two groups: *schemes* and *tropes*. The function of schemes and tropes is to carry out a transference of some kind; schemes are characterized by a transference in order, while tropes are characterized by a transference in meaning.

- For example, the most problematic figures of speech in natural language processing are **irony** and **sarcasm.**

Indexes and Metrics to Analyze Data Collected Through Online Social Networks

- **Engagement metrics**: numerically quantify a phenomenon and the features that led to its spread. They include:

  - **Amplification metrics**: computed counting of the number of shares for Facebook and retweets for Twitter. An analysis over time of metrics of this type allows feedback to assess the content shared by a user within that user's social network.

  - **Applause metrics**: represent an approval rating from the audience of a particular content; it is expressed on Twitter, Facebook, and YouTube as "like."

- **Conversation rate**: number of the conversations per post. On Facebook, YouTube, and LinkedIn they consist of comments, and on Twitter they consist of replies.

- **Influence metrics**: analyze quantitatively users who participated in the conversations. This is defined as the ability to generate a multitude of content in a limited period of time.

- **Reach**: number of unique individuals (account) that have been exposed to the content analyzed and who have had the opportunity to engage with it.

- **Impression**: the number of times that some content has had the opportunity to be seen within the social network platform.