

# Sentiment Analysis with AlchemyAPI: A Hybrid Approach

---

9/2013

## Methodology

### Advantages of a Hybrid Approach

AlchemyAPI's approach to natural language processing incorporates both linguistic and statistical analysis techniques into a single unified system. This hybrid approach provides a unique advantage since both techniques have benefits and drawbacks depending on the content and specific use-cases. Linguistic analysis takes a basic grammatical approach to understand how words combine into phrases, and how those phrases combine into sentences. While this approach works well with editorialized text (e.g., news articles and press releases), it does not perform as well when it comes to user-generated content, often filled with slang, misspellings and idioms. Statistical analysis, however, understands language from a mathematical standpoint and works well on "noisy" content (e.g., tweets, blog posts, and Facebook status updates). The combination of these two approaches allows for increased accuracy on a variety of content.

### Massive Automated Learning

Large-scale learning provides AlchemyAPI with a true data advantage. AlchemyAPI continues to crawl the entire web, resulting in a training set of 200+ billion words (285 times larger than Wikipedia). Since AlchemyAPI's system is based on learning algorithms, the broader the domain coverage of data fed into the system, the more it's able to learn about new figures of speech, idioms, and nuanced expressions.

Because language constantly evolves, the system needs to be continually retrained on a wide variety of data to remain up to date. In addition to recrawling the web every month, AlchemyAPI monitors data feeds from the top social media platforms. For instance, over 1 billion tweets are added to AlchemyAPI's data sets every month. AlchemyAPI also leverages data from its community of more than 30,000 API users. This massive data set, generated from 6 continents and dozens of industries, allows AlchemyAPI to more accurately analyze entities and topics the system has never seen before.

AlchemyAPI continues to crawl the entire web, resulting in a training set of 200+ billion words (285 times larger than Wikipedia).

## Sentiment Analysis

In text analysis, sentiment is the attitude or opinion expressed towards something, such as a person, product, organization or location. Sentiment can be positive, "this car is great," negative, "this car is overpriced," or neutral, "this car is red." For computers, the process of determining the sentiment of text requires highly-trained algorithms.

### Increased Accuracy

Creating software that can analyze text accurately and at scale is a complex undertaking. AlchemyAPI's sentiment analysis is uniquely positioned because its system is based on a hybrid approach and large scale learning technology trained on massive data from a variety of domains. This leads not only to a deeper understanding of sentiment and a broader set of features, but ultimately, to a more accurate system.

Third-party research demonstrates AlchemyAPI's superior sentiment accuracy. For example, a paper by [Meehan et al \(2013\)](#), "Context-Aware Intelligent Recommendation System for Tourism," showed an 86.01% accuracy level after manual testing performed on a corpus of 5370 tweets. Since accuracy depends on the types of documents and specific use-cases, customers should evaluate sentiment using their documents. AlchemyAPI provides free software evaluations to encourage customers to assess accuracy figures for specific implementations.

### A Deeper Understanding of Sentiment

AlchemyAPI provides multiple targeting options for sentiment analysis, including: document-level, entity-level, keyword-level, directional, relational and user-specified. This finer grained analysis of sentiment enables more actionable business intelligence.

## Document-Level Sentiment Analysis

Document-level sentiment is the most general form of sentiment analysis, where a document or block of text is analyzed as positive, negative or mixed. This is a rather crude metric since documents often mention more than one thing and include mixed sentiment.

## Entity-Level Sentiment Analysis

The addition of entity-extraction takes sentiment analysis a step further so that sentiment data is included for every detected entity. For example, a statement can be negative about a person or product and positive about another. Third-party research, such as [Rizzo and Troncy \(2011\)](#) and [Saif et al \(2012\)](#), demonstrate AlchemyAPI's superior accuracy when it comes to entity extraction.

## Keyword-Level Sentiment Analysis

Sentiment can be calculated for each keyword extracted, for example, allowing a user to understand the sentiment associated with a product or product feature.

## Directional-Level Sentiment Analysis

Directional sentiment reveals who is emitting the sentiment. For example, if a person spoke negatively about a product, determine not only that the product was mentioned negatively, but who mentioned the product negatively.

## Relational Sentiment Analysis

The relations call identifies the subject-action-object relations within text and sentiment can be calculated for each subject or object. Combined with entity and keyword extraction, relational sentiment can be the basis for powerful search and trend analysis applications.

## User-Specified Targets

If the above options for targets were not specific enough for your application, AlchemyAPI supports user-specified targets. This allows sentiment to be calculated towards any arbitrary target.

## Sentiment Intensity

AlchemyAPI's dynamic approach enables the system to have a much deeper understanding of text. For example, AlchemyAPI understands the following:

- **Negations:** Words that reverse the polarity of the sentiment, for example “this phone is good,” versus “this phone is not good.”
- **Amplifiers:** Words that make the sentiment more intense, such as, “very.”
- **Diminishers:** Words that make the statement less intense, for example, “somewhat.”
- **Intensity:** Negations, amplifiers and diminishers lead to an intensity score for each detected sentiment. Since there is a big difference between, “This coffee is a little sweet,” and, “This coffee is totally disgusting,” an intensity score shows the intensity of each negative statement.

## Looking Forward

Performing sentiment analysis well is a complex task, but with AlchemyAPI's unique natural language processing approach and large-scale learning technology, it is able to maintain high accuracy across a wide range of industry verticals and applications, not just today, but well into the future. With an upcoming update of AlchemyAPI's sentiment engine powered by deep learning, the massive data sets will be leveraged even further leading to a smarter, more accurate system. Additional languages are also in development so the system will find uses in even more industry verticals and locations.

AlchemyAPI's dynamic approach enables the system to have a much deeper understanding of text.

## About AlchemyAPI

The product of over 75 person years of engineering effort, [AlchemyAPI](#) is a text mining platform providing the most comprehensive set of semantic analysis capabilities in the natural language processing field. Used over 3 billion times every month, AlchemyAPI enables customers to perform large-scale social media monitoring, target advertisements more effectively, track influencers and sentiment within the media, automate content aggregation and recommendation, make more accurate stock trading decisions, enhance business and government intelligence systems, and create smarter applications and services. If you would like to learn more about our company and services, please call us at 1-877-253-0308 or email [info@alchemyapi.com](mailto:info@alchemyapi.com).

Copyright Notice

© 2013 - AlchemyAPI, Inc. - Any information from this paper that is to be used in any way requires prior written approval from AlchemyAPI.