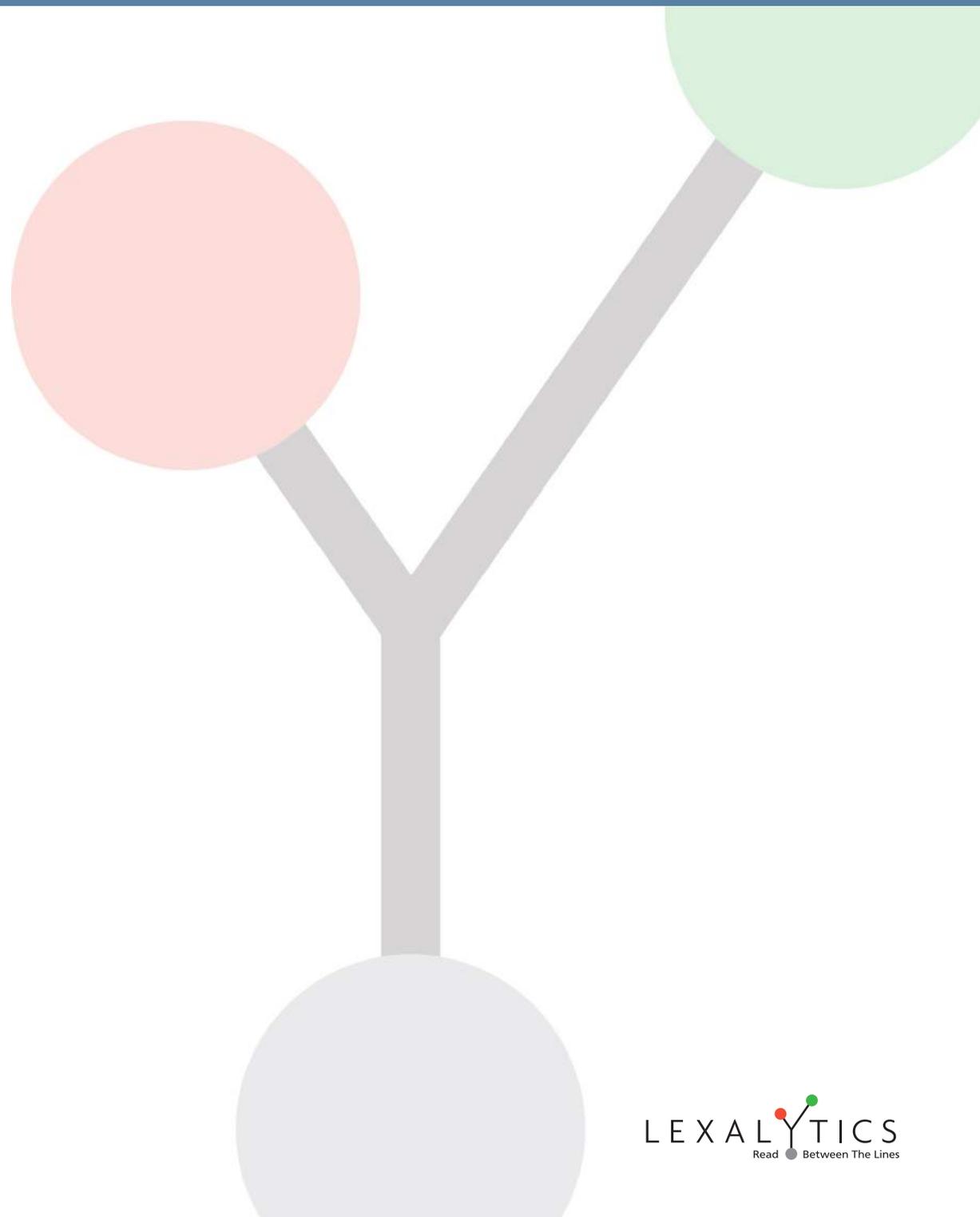


# Natural Language Processing

Whitepaper



Natural Language Processing research is a complex matter. For you, the professional working in the NLP field, time and effort are your two most precious resources — and you shouldn't waste them on automating the low-level technicalities of NLP. One of Lexalytics' many branches is working with NLP experts to cover their low-level automation; by utilizing our already-programmed functions, you free up your own time and energy to focus on developing your higher-level processes.

Our technology is tried and refined; in this whitepaper, you'll see why Lexalytics is the best choice to make your job easier.

With over ten years of development behind us and a bright future ahead, our text analytics offerings contain cutting-edge features for NLP routines, including:

- Named entity extraction
- Part of speech tagging
- Categorization and classification
- Topic, theme, and concept identification
- Sentiment analysis of both entities and themes
- Relationship definition between parts of speech
- Multilingual support (more than 20 languages)

These features increase your productivity by relieving you of standard programming tasks. Simple to use straightaway and boasting unprecedented customizability, our on-premise and in-the-cloud text analytics offerings will always fit your needs.

Want to know more? Visit our webpage to learn more about our software and contact us, or read on for highlights of why our accuracy, reliability, and customization options make Lexalytics your best choice for automating low-level functions and quickly moving onto developing your higher-level NLP processes.





Lexalytics' text mining tools are completely transparent in every language, so that you can monitor and adapt them to your needs. Two of our options in particular are important for adapting to your NLP.

First, our models for **part of speech tagging** include systems for everything from Tweets to academic papers. We've spent countless hours developing these models to be accurate and reliable so you don't have to worry. If you need customization, though, the PoS tagger is always within your control: one of our numerous advantages is the unparalleled customizability we deliver.

Meanwhile, **entity extraction** tools have never been better. We offer a host of options for extracting named entities, so that you can choose what's best for you. You can:

- Process normalized lists of entities
- Use part-of-speech patterns to define entity types
- Use regular expressions to define atypical entity types (such as phone numbers and addresses)
- Utilize our exhaustively-trained conditional random field model to recognize companies, people, places, and products
- Or train your own maximum entropy model for recognizing new entity types that are not already included
- Or combine systems into a hybridized model optimized to your content
- Our software even determines and assigns sentiment to every entity it finds for better understanding of their context

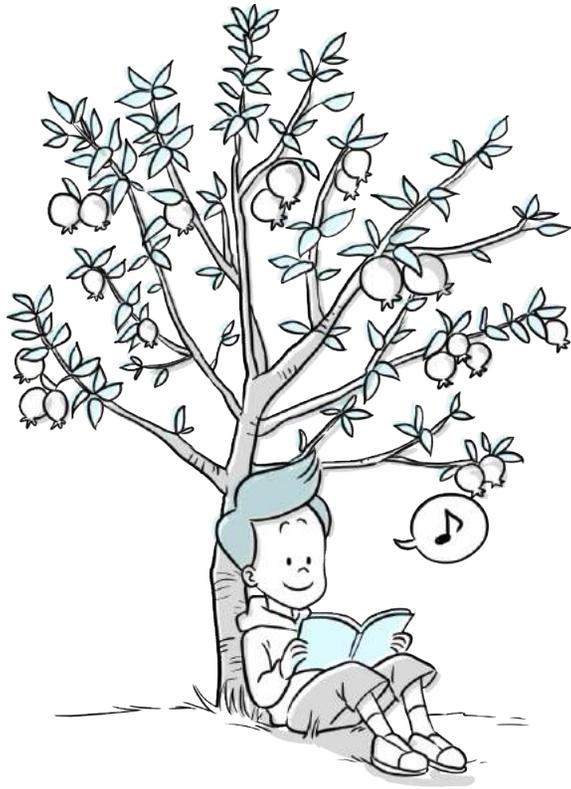
Need to define topics in a set of text documents? Lexalytics supplies several options for classification so you can choose the best fit:

Boolean search queries are simple and easy to use, great for picking out very specific words or terms. Lexalytics provides support for the AND, OR, NOT, and NEAR operators. If you're looking for any depth, however, or if you have a large list of terms you wish to find, query topics will take time to set up.

If you're looking to identify everything that falls within a specific category of term that isn't already included (such as "types of disease"), contact us about developing a **model-based classifier**. We'll input a set of training content, and create a classifier that works to your specifications. But for immediate and actionable categories, look no further than our Wikipedia™-based **Concept Topics**. We processed over 640 thousand of the top articles on Wikipedia™, identifying 1.1 million keywords and bi-grams with 56 million links between them. Based on the strength of these links, we developed our Concept Matrix: a web of associations and relationships that forms the basis for our Concept Topics system.

Concept Topics are just that: topics that represent broader concepts. Each Concept Topic contains a number of keywords that relate to the larger idea. Each of those keywords is associated with more keywords, and so on: these relationships come from the Concept Matrix, and the strength of an association is based on how close a keyword is to the original topic. Instead of building a statistical classifier or a huge taxonomy of words, a few keywords define an entire concept — and we're as accurate or better than human analysts, without the massive manpower and time investments you'd need to analyze it yourself.





Lexalytics offers the most accurate, most reliable sentiment analysis out there, but what makes our systems special is their flexibility.

For one, we'll report sentiment at whatever level you choose: document, entity, theme, classifier, or topic. Phrases are completely configurable for sentiment, and you can utilize multiple sentiment dictionaries to better adjust for different types of documents.

We also support the tuning of intensifiers such as "very" and "really", and negators such as "not" and "never". Intensifiers and negators act in concert to develop a modifying score for all of the sentiment-bearing phrases.

And you have total control over which sentiment patterns are used for applying sentiment to which entities.

## PARTS-OF-SPEECH SHOW PATTERNS

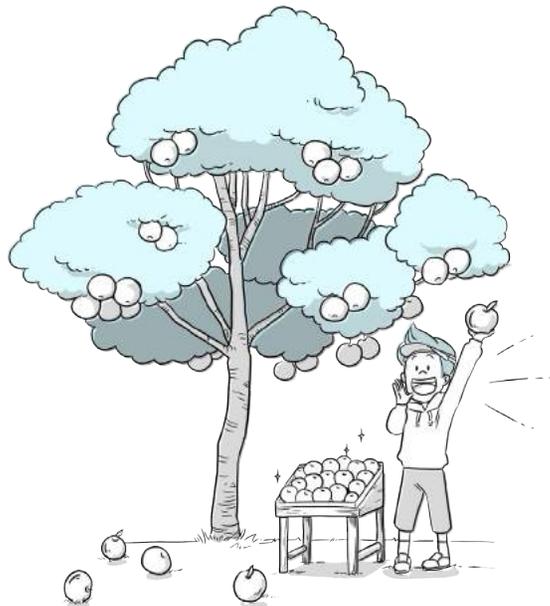
Our algorithms can define any part-of-speech relationship, and any action can be taken when that relationship is discovered. Relationships allow you to define patterns between parts of speech, including entities, so that you can see what "a" is doing to "b". These relationships provide enormous insight. For example, you can parse for indications of a

merger or acquisition (X organization is acquiring Y organization), identify patterns that indicate job changes (X person is leaving Y job and moving to Z company in Q role), or see when someone is asserting their opinion (X person is stating Y opinion on Z entity).

These are just a few of the many functionalities that come with our text mining tools, functionalities that will make your work much, much easier. Give us a call (or send an email) to get started today.

Lexalytics® is the industry leader in translating text into profitable decisions. Lexalytics deploys state-of-the-art on-premise and in-the-cloud text and sentiment analysis technologies that process billions of unstructured documents every day globally, transforming customers' thoughts and conversations into actionable insights. The on-premise Salience® and SaaS Semantria® platforms are implemented in a variety of industries for social media monitoring, reputation management and voice of the customer programs.

Lexalytics is based in Boston, MA, and has offices in the U.S. and Canada. For more information, please visit [www.lexalytics.com](http://www.lexalytics.com), email [sales@lexalytics.com](mailto:sales@lexalytics.com) or call 1-617-249-1049. Follow Lexalytics on Twitter, Facebook, and LinkedIn for updates and insights into the world of text mining.



320 Congress St  
Boston, MA 02210

General Inquiries  
1-800-377-8036

Sales  
[sales@lexalytics.com](mailto:sales@lexalytics.com)  
1-800-377-8036 x1

International  
1-617-249-1049