CS 410 Technology Review Nov 2021

Document Context: Compare Text Analytics Tool / Framework available in the Market.

Why to Write on this:

As part of the course, we have cover most of the ground up in the Text Analytics. However, most of the Organization now-a-days are leveraging these features from the already available tool / framework. This document is an attempt to show why it is needed and what are the offerings from these tools / frameworks. This will give an overall idea of where we stand on the Text Analytics as part of these Tools, and I hope will be in good conjecture with the Course we studied in CS 410.

Document Writer:

Sushanta Panda : NetId - Panda5

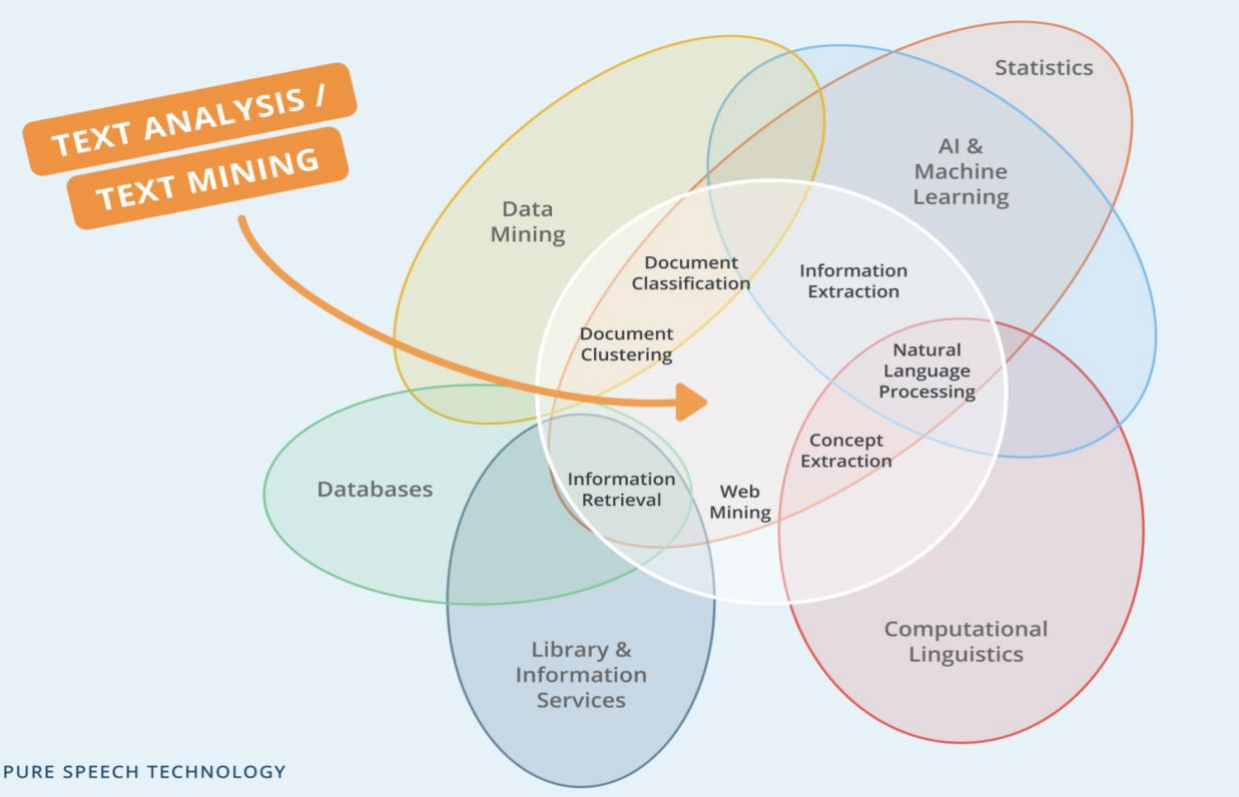
Introduction:

In the world of Text Analytics, top tech companies have created powerful text analytics tools, which helps the users to solve complex and critical problem. In this document, I described various Text Analytics tools available in the market for Text Mining and Text Analytics processing. Few of them are free and other are Paid. Now-a-days most of the companies leverage these software / tools available in the market rather build up from the scratch. These tools, which will plug into the data storage (structured or unstructured data) and will spit out the exact need of the analysis in the form of a report or dashboard as an output.

In the market lot of tools are available and their offering are very competitive in nature. This document depicts strength of these tools and how they are uniquely place in the market with respect to its competitiveness.

These ready-made tools well equipped to process structured / un-structured data, does the sentiment analysis, tokenization, text parsing chunking, named entity extraction and recognition etc.

On the overall the Text Analytics includes all or most of the components in all the text analytics tool as described in the below image. Though the below image shows the high-level depiction, we can further drill down into each component to explore further. However, the idea is to depicts what the tools have encapsulated as part of the process.

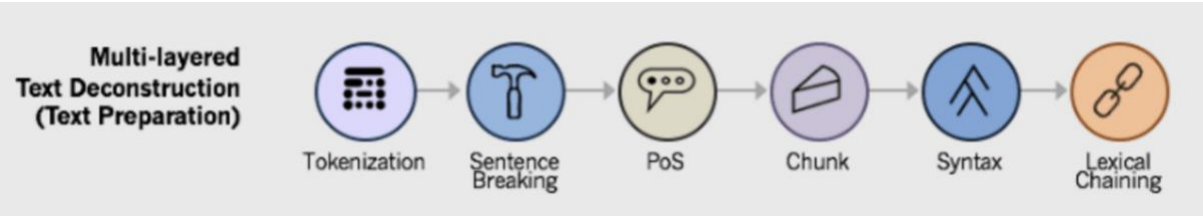


Courtesy: [Towards Data Science]

Build Tools from Ground Up Vs Use/Leverage Toolkit:

This question will be always come and keep coming for most of the organization, whether we should build the software from the ground up or should we leverage the tools already available in the market, and for Text Analytics it's no different. In the world of Technology, it's changing so fast, and every day new Research are happening around the Globe, and it's been implemented in the form of a product. I believe, Organization should leverage the already software / toolkit already available in the Market and let the organization focus on the problem they try to solve and let the technology companies try to enable them the right toolset for the business problem.

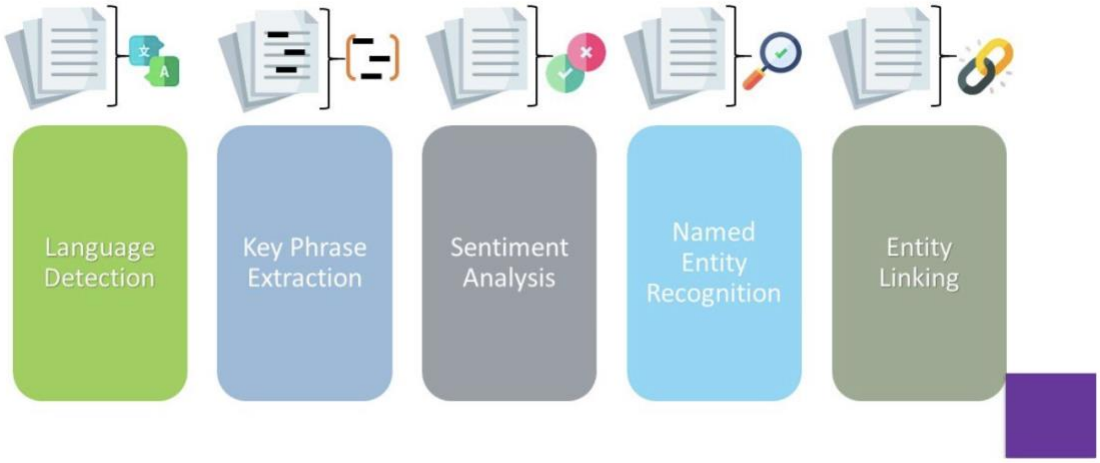
If we start in the Text Analytics (for unstructured data), where the steps like Tokenization, PoS, Chunk, Syntax etc are extensive and messy. However almost all the software tools now-a-days have these features readily available, and it's keep evolving.



Courtesy: [lexalytics]

All these steps will certainly be part of the tools and will extensively carry out these steps with millions and trillions of unstructured data (from tweets, surveys or anything from

www). So, certainly leveraging these tools will be like focusing on the main business problems, rather than building something from ground up which is already available. The cutting-edge technology starts after it, as now-a-days the above text processing seems basic everywhere, where the game changer is what information / knowledge we can extract out of it. These includes the Language detection / Key Phase Extraction / Sentiment Analysis, Named Entity Recognition etc. This is where the core of the Text Analytics resides



Though the process of the Text is extremely complex and critical before the next processing kicks off, the processing of the Text is heavy on Engineering side, whereas the next steps of the Knowledge is more of the Science to the Text Analytics.

Till now, we understood that these Tools are like the main power when it comes to the Text Analytics and if we want to solve some of the biggest problem in our business related to the Text Analytics we should go ahead and select one of these Tools. However, there are lot of tools and framework available in the Market, some of them from the Big Tech and others are from specific market segment. As part of the segment, let's compare the Text Analytics tools available in the Market with the offerings and pricing as the reference.

As part of the comparison of these tools from the Tech Companies, we will compare the following

- Organization developed the Text Analytics tool / Framework
- What are the offerings. These offerings are around the Text Analytics (Sentiment Analysis, Topic Modeling etc)
- What is the Pricing. However, the pricing gives as a refence and not a direct comparison
- Other key offerings. These key offerings are those other than the Text Analytics. For e.g., storage ML models, how many languages it supported, any domain specific offerings etc.

I will compare the major tech company's text analytics tool / framework and couple of others. The idea is to depicts what the tool are capable of and what it extends to, rather showing the extensive list of company's tool.

Comparison of the Text Analytics Tool / Framework:

Below are the few top tech company's text analytics tool / framework and other couple of other tools, includes it's offering and extended offerings. As stated above, it's not the exhaustive list of tools / framework available in the market, rather to show what are these tools capable of and what are the features its extended.

Organization	URL	Offerings	Paid / Free	Other Key extend Offerings
Azure	https://azure.microsoft.com/en-in/services/cognitive-services/text-analytics/	• Syntax Analysis • Sentiment Analysis • Entity Analysis • Text Classification • Content Classification • Topic Modeling • Document Summarization	• Pricing	Azure offers variety of the solution including to the Azure Text Analytics such as the following (and not limited to) • Storage of your data • Machine Learning Model on the Fly • Dashboard using Power BI (Directly Plug your datasets) • Merging of the Output of your Text Analytics dataset with other structured Datasets to even fine-grained detailed report • Highest Quality Data Privacy, Compliance and Security to your data • 120 Language Supported
Google Cloud	https://cloud.google.com/natural-language/	• Syntax Analysis • Sentiment Analysis • Entity Analysis • Text Classification • Content Classification • Topic Modeling	Pricing	Google Cloud also offers variety of solution stack pack includes to the Text Analytics • Storage • Machine Learning Models on the Fly • Other than the Natural Language API, Healthcare Natural Language API also available • Good Customer based; hence the usage seems to be good • Highest Quality Data Privacy, Compliance and Security to your data

				• 110+ Language Supported
	https://www.ibm.com/watson	• Text Analytics • Machine Learning • Sentiment Analysis • Named Entity Extraction • Categorization • Topic Modeling	Pricing	Other than the Market Leader in Innovation, some of the key features other than the IBM's Text Analytics platform (Watson) • +100 Million user bases • Highest Quality Data Privacy, Compliance and Security to your data • Able to understand Industry specific jargon from your data other than the processing / analyzing structured / unstructured data • End-To-End Automation • Can Plug-in the data anywhere including AWS S2 / S3, Azure or Google Cloud
	https://aws.amazon.com/comprehend/	• Text Analytics • Machine Learning • Sentiment Analysis • Named Entity Extraction • Categorization • Topic Modeling	Pricing	AWS is the market leader in the Cloud business, hence it has a wide offering including the Text Analytics • Able to store your data • Machine Learning Models on the Fly • Specific use case related to call center analytics, Index & Search Product Views, Categorize Financial Documents. These Use cases are very robust and detailed view, any organization with similar use case will directly plug-n-play. • Good Customer Base • 100+ Language Supported
	https://www.meaningcloud.com/	• Text Clustering • Text Classification	Pricing	Though the Google, Microsoft and Amazon is very widespread across multiple domain and multiple business domains, Cloud meaning is

		• Sentiment Analysis • Topics Extraction • Document Structure Analysis • Deep Categorization • People Analytics • Social Media Analytics • Voice of customers		focus on some of the specific domain, as highlighted below as part of the offerings • Good understanding on analyzing the text data with respect to call center data to help the KPI • Analyzing the Customer feedback unstructured data (email, call notes, feedback, social media, surveys etc) and gather the report with as part of the what the customer is saying • Content publishing and Document Analytics
LEXALYTICS AN INMOMENT COMPANY	https://www.lexalytics.com/	• Text Analytics • Machine Learning • Sentiment Analysis • Named Entity Extraction • Categorization	-	As stated above for meaning cloud, Lexalytics solve some of the specific problem as part of the offerings as stated below • Storage and supports visualization • 20+ language native processing • Specific Industry Use Case • Social Media Monitoring • People Analytics & Voice of Employee • Solve Regulatory compliance problems that involve complex text documents • Transfer mountain of unstructured customer feedback into userfull data.

Citation / Reference:

Although the citation / Reference is already stated in the document wherever applicable, below is the all links where the citation / reference is done

- Towards Data Science - https://towardsdatascience.com/
- Towards Data Science (Text Analytics / Text Analysis) - https://towardsdatascience.com/a-guide-text-analysis-text-analytics-text-mining-f62df7b78747
- Azure Synapse (Text Analytics) - https://azure.microsoft.com/en-in/services/cognitive-services/text-analytics/
- Google Cloud Natural Language Processing - https://cloud.google.com/natural-language/
- IBM Watson - https://www.ibm.com/watson
- Amazon Comprehend (Text Analytics) - https://aws.amazon.com/comprehend/
- Meaning Cloud - https://www.meaningcloud.com/
- Lexalytics - https://www.lexalytics.com/
- Medium - https://medium.com/
- Text Analytics of Google Cloud Vs Microsoft Vs Amazon Vs IBM - https://medium.com/kontikilabs/comparing-machine-learning-ml-services-from-various-cloud-ml-service-providers-63c8a2626cb6
- Text Analytics Image Courtesy - https://connectjaya.com/azure-cognitive-servicestext-mining-and-sentiment-analysis/