



## Socratez™ White Paper

*Revolutionizing Unstructured & Structured Data Analytics*

### **Authors**

Naveen Ashish, PhD, Co-founder & Data Scientist

Kevin Payne, PhD, Co-founder & Lead Data Scientist

*TeraCrunch, June 2014*

## Table of Contents

<b>1. Abstract</b> .....	3
<b>2. Introduction</b> .....	3
<b>3. Socratez Text Analytics Engine</b> .....	5
<b>3.1 Decoding Opinions → Expressions</b> .....	6
<b>3.2 Socratez V1.3 Features</b> .....	6
3.2.1 Expressions.....	7
3.2.2 Sentiment .....	7
3.2.3 Contextual Sentiment .....	8
3.2.4 Comparative Sentiment .....	8
3.2.5 Entities.....	8
3.2.6 Aspects.....	8
3.2.7 Topics.....	9
3.2.8 Emotion.....	9
<b>3.3 Association Establishment</b> .....	10
<b>3.4 Technology Approach</b> .....	10
3.4.1 Natural Language Processing (NLP).....	11
3.4.2 Machine Learning based Classification .....	11
3.4.3 Knowledge Engineering .....	12
<b>4. Socratez Data Insightz™ Module</b> .....	13
<b>4.1 Applications of Socratez Insightz</b> .....	14
<b>4.2 Insightz Platform</b> .....	15
4.2.1 Universal Lexicon & Ontology System.....	15
<b>4.3 Technology Approach</b> .....	17
4.3.1 Statistical Modeling .....	17
4.3.2 Machine Learning .....	17
4.3.3 Computational Modeling & Simulation .....	17
4.3.4 Insightz Machine Learning Layer .....	18
<b>5. Socratez Data Extraction Module</b> .....	19
<b>6. TeraCrunch Professional Services</b> .....	20
<b>Author's Bio</b> .....	21

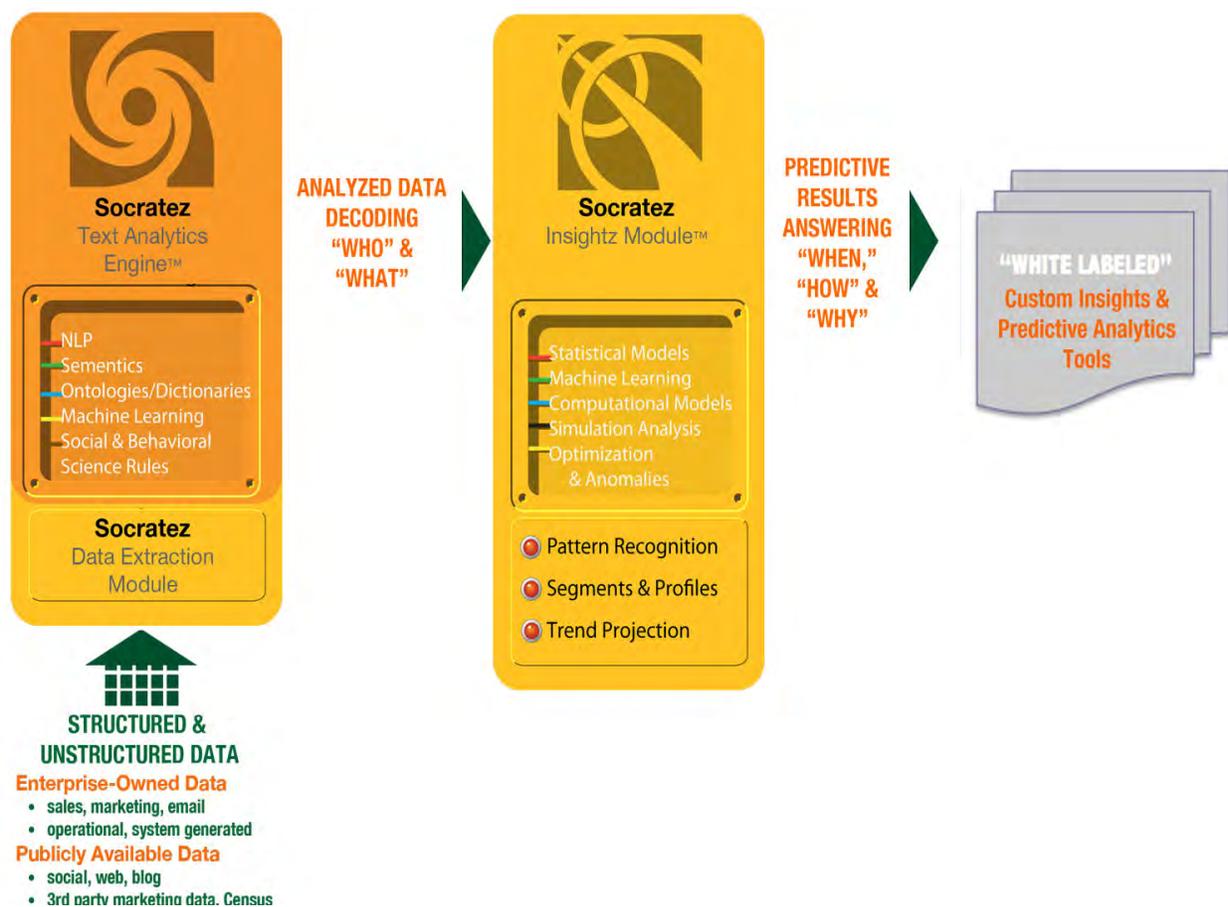
## 1. Abstract

Businesses today need tools to provide them insights over various types of data. This data is unstructured, semi-structured or structured and is either enterprise owned data (financial, sales, marketing, CRM, point of sale) or publicly available data (social, websites, blogs, census). Businesses are challenged with lack of tools, technology expertise, data scientists and the large investment required to bring such tools to bear. This white paper introduces the “TeraCrunch Soratez™ Platform”; a powerful, innovative way to extract, structure and discover Insights from big data.

## 2. Introduction

TeraCrunch introduces the first platform in the world that structures, integrates, correlates and find insights from any data coming from multiple sources & formats, all from one Unified Socratez Platform. Socratez uses the latest technological innovations to find “deep” insights in any qualitative or quantitative data. If the data is qualitative, it quantifies it by decoding the “who” & the “what” in real-time. Socratez is the first platform in the world that has integrated a proprietary Text Analytics Engine with Predictive Insightz Module and this integration enables us to find insights with unparalleled accuracy & depth from quantitative or qualitative data.

**Figure 1 Socratez™ Platform**



This integration is realized by standardizing data structuring and automatically analyzing and extracting concepts such as Topics, Entities, Sentiment, Emotions & Expressions and also *Associations* between such concepts from any human or system generated data.

The platform comprises of three fundamental components or modules as shown in Figure 1 above, namely:

- 1) **Data Extraction Module:** To enable comprehensive and rapid collection of data from Internet, Social-Media and other data sources and prepare the data for further synthesis.
- 2) **Text Analytics Engine:** Structure the unstructured data with our patent-pending technology that decodes the “who and the “what” from the data.
- 3) **Insightz Module:** Data mining, analysis and predictive analytics over the (now) structured and other data.

Our IP is a result of decades of R&D by our PhD Data Scientists who come from Academia, NASA & Enterprise. Our innovative technology applies Natural Language Processing (NLP), Machine Learning Algorithms and Knowledge Engineering (for customizations).

Our aim is to automate unstructured, semi-structured and structured data integration & processing as the foundation of next generation analytics tools for search, classification, predictive analytics and other purposes. We also offer interactive dashboards to visualize these insights. The final results can be provided as a real-time tool or in the following ways:

- RESTful API: CSV, Excel or similar format
- Dashboard: real-time web based with full interactivity
- Integration of these results into an existing application
- Standalone Application or Tool

Businesses have targeted analytics needs. Often, without customization, off-the-shelf analytics are meaningless. Our Professional Services team of PhD Data Scientists and Engineers can customize our Platform to fit your business's unique needs.

### 3. Socratez Text Analytics Engine

Socratez has been designed with the aim of realizing a truly “next-generation” text-analytics engine: addressing the artificial intelligence (AI) frontier of “understanding” the content in natural language text. We go deep into understanding what has been expressed in the text content. We also realize that specific applications and clients requiring the applications have specific information and “signals” that they want to glean from the unstructured text.

On one hand, Socratez leverages the state-of-the-art in research and development in key areas such as machine learning, natural language processing (NLP), and knowledge engineering. On the other hand, we have designed the system keeping in mind the real-world requirements of our customers that have come to us with specific and challenging requirements on the particular signals that they want extracted from unstructured text data.

In a nutshell Socratez brings the following key innovations:

**DECODING OPINIONS** The vast majority of text analytics systems in the market today can at best determine Positive or Negative Opinions that have been expressed in text. However opinions themselves are of a variety of different kinds – for instance, negative opinions may be complaints, suggestions, requests, questions, etc.; positive opinions may be acknowledgements, endorsements, announcements etc. Socratez’s ability to identify the type and purpose of opinions from unstructured data is innovative.

**ASSOCIATION ESTABLISHMENT** Any text analytics system essentially extracts a set of concepts or elements of different kinds from the text. The analysis of the structured data (extracted from the unstructured data) becomes significantly more powerful if we can establish *associations* between the extracted elements. For instance which entity is associated with which sentiment, which topic or aspect is associated with which expression, etc. Socratez establishes *connection strength* between any pair of elements it extracts from the text.

**KNOWLEDGE ENGINEERING FOR CUSTOMIZATION** Socratez provides a *knowledge engineering* infrastructure for deep customization. Domain knowledge and its maximum leverage for information extraction is a key driver for being able to do in-depth information extraction and classification for specific applications. (We later elaborate on our knowledge engineering approach and capabilities.)

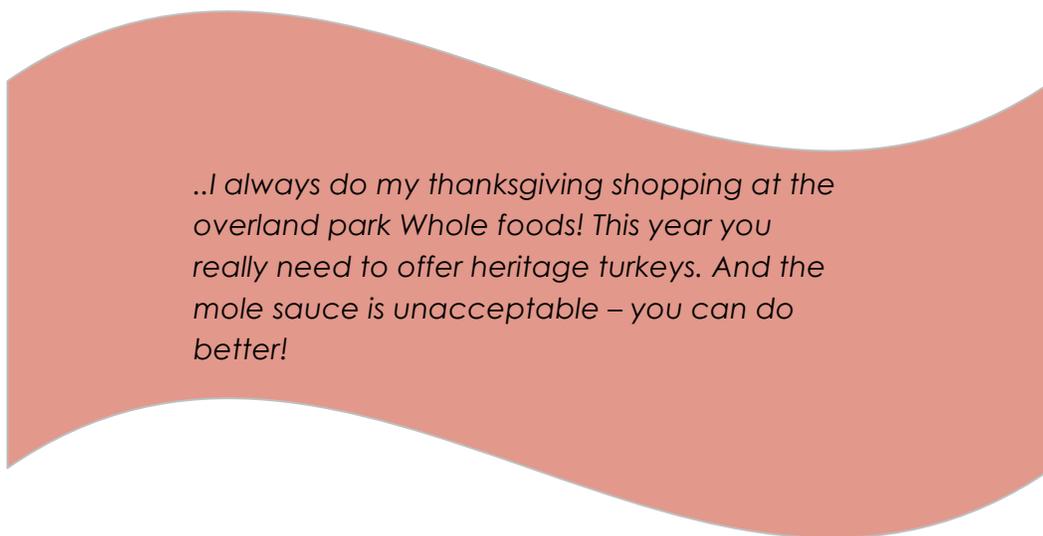
### 3.1 Decoding Opinions → Expressions

The notion of expressions in natural language text, from a text-analytics perspective, was introduced in [Ashish et al]. In conversations, people express themselves! They praise, they criticize, they complain, advocate, question, acknowledge or apologize, suggest, request, advise, relate experiences, inform, announce, etc.

One of the key differentiating capabilities of Socratez is its ability to classify expressions in natural language text.

### 3.2 Socratez V1.3 Features

Table 1 provides a list of the concepts and signals that the current version of Socratez offers. We provide examples of the features in the context of the text snippet illustrated below – which is a (typical) snippet of unstructured text we analyze from a social-media forum, for one of our retail applications.



**Table 1 Socratez Offering**

FEATURE	BRIEF DESCRIPTION	EXAMPLE(S)
Expressions	Classifying and decoding Opinions	<ul style="list-style-type: none"><li>- There is a SUGGESTION regarding heritage turkeys</li><li>- There is a COMPLAINT regarding the mole sauce</li></ul>
Sentiment	Basic sentiment	<ul style="list-style-type: none"><li>- Positive sentiment regarding Whole Foods</li><li>- Negative sentiment regarding</li></ul>

		mole sauce
Entities	Particular entities of various kinds on the kinds	<ul style="list-style-type: none"> <li>- "Whole Foods" as an ORGANIZATION</li> <li>- "Overland Park" as a location</li> <li>- Other significant entities such as "turkey", or "mole sauce"</li> </ul>
Aspects	Very broad categorization of the text content	Broad categories such as "Food" (turkey)
Topics	(Possibly) previously unknown topics that are detected in text collection	For example a topic related to "heritage turkeys" or "organic turkeys" or "turkey sauce" may emerge in a text collection at a particular time.
Emotions	Decoding of sentiment in particular emotions	Further decoding negative sentiment as anger or grief or positive as ecstasy or love.
Association	Association strength score [0,1] between any pair of concepts extracted in a unit of text	AssociationStrength(Whole Foods, Positive) → 0.8 AssociationStrength(Complaint,mole sauce) → 0.85

We describe the elements in Table 1 in more detail below.

### 3.2.1 Expressions

As mentioned above, expressions are basically looking at opinions at a more granular level. The research and development work leading to Socratez was pioneering in introducing the notion of expression extraction from natural language text. We are one of the only text analytics engines to offer the feature of high accuracy expression extraction from text.

We offer extraction of a variety of "core" expressions that we have found common and prevalent in our applications – primarily in the retail and health verticals. We also offer a framework for custom development of new expressions particular to new applications.

### 3.2.2 Sentiment

The vast majority of text and sentiment analysis systems still determine text sentiment based on spotting *polar* words in the text. This simplistic approach fails to determine text sentiment in at least two major scenarios: 1) When the sentiment is *contextual*, and 2) When it is *comparative*.

SOCRATEZ provides semantically richer sentiment analysis capabilities as exemplified by its ability to correctly classify sentiment in the following examples:

### 3.2.3 Contextual Sentiment

*We waited over an hour to get our order* → Negative

*The turkey had too much salt* → Negative

### 3.2.4 Comparative Sentiment

*Most other places in town are better than this* → Negative

*The chicken here is better than anywhere else* → Positive

Socratez is one of the few systems to effectively address such complex sentiment extraction by identifying *multiple* elements in the text (sentence) that comprise the final sentiment. The multiple elements are then used to classify sentiment accurately. This is a significant advance over existing systems where sentiment classification is primarily done on the basis of individual polar words.

### 3.2.5 Entities

Socratez extracts a wide variety of entities from the text. We offer extraction of a core set of “standard” entity types including names, organizations and locations. We also offer extraction of custom entity types – for instance in health applications we have developed custom entity extractors for particular drugs or medical conditions, procedures, etc.

### 3.2.6 Aspects

Aspects are a (very) broad categorization of the text. For instance we may classify conversations in a health related forum (such as Medpedia amongst many such) into very broad categories of {Drug, Side Effect, Question, Personal Experience}.

The aspects can be particular to a domain or a dataset and are typically provided by the client.

### 3.2.7 Topics

In a domain and dataset the aspects are already known and fixed. We also dynamically uncover topics that “emerge” in a data collection that are not necessarily known in advance.

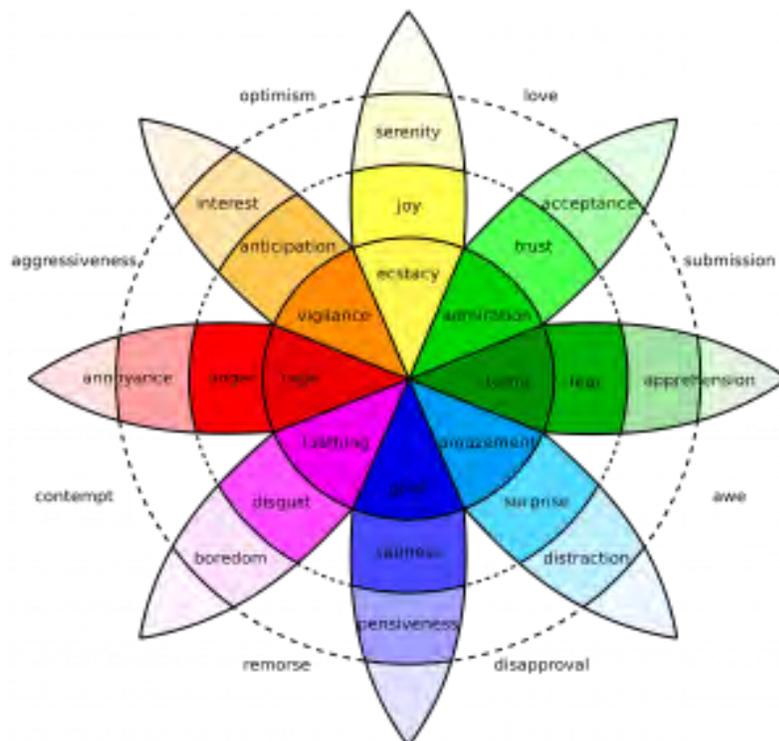
We have employed topic modeling algorithms such as “LDA” for this purpose. However we must mention that we do a significant amount of pre-processing of text to “normalize” it, as well as semantically categories parts of it – before sending it for topic modeling. This prior treatment of the text has been found to result in detecting significantly more meaningful topics in text collections.

### 3.2.8 Emotion

Similar to how we decode opinions into expressions, we further decode sentiment (positive or negative) into emotions. Figure 1 shows the Plutchik’s wheel of emotions that is essentially a deeper categorization of human emotions.

We have formally incorporated the ontology of emotions and can classify sentiment as specific emotions from this categorization.

**Figure 3 Plutchik’s Wheel of Emotions**



### 3.3 Association Establishment

Another major differentiating aspect of Socratez is that it establishes associations between the various concepts and other elements that it extracts.

We provide an association strength score in [0,1] between any pair of elements extracted in a piece of text. For instance across any entity and sentiment extracted, across expressions and sentiment, across aspects and expressions etc.,

ELEMENT1	ELEMENT2	STRENGTH
Turkey Sauce	Negative	0.75
Turkey	Complaint	0.53
Heritage Turkey	Suggestion	0.82
Whole Foods	Positive	0.9

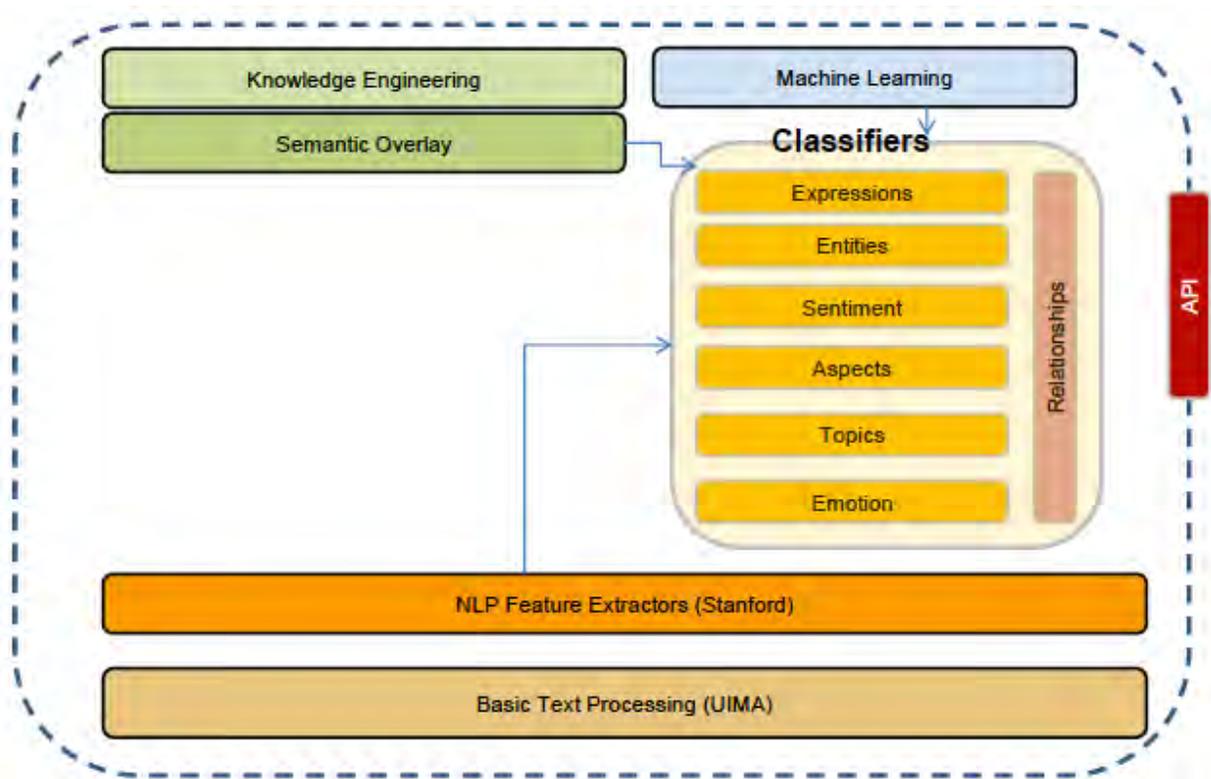
### 3.4 Technology Approach

The SOCRATEZ approach to text distillation is built on three fundamental pillars, namely:

- Natural Language Processing (NLP)
- Machine Learning
- Knowledge Engineering (Semantics)



**Figure 2 Socratez Text Analytics Engine**



### 3.4.1 Natural Language Processing (NLP)

Text processing at all levels ranging from basic string operations on text to part-of-speech (POS) tagging to named-entity recognition to natural language parsing of sentences are all employed in the text distillation process. We have extensively leveraged available natural language processing (NLP) toolkits, particularly the Stanford CoreNLP software for this purpose. We have also developed and applied libraries for specific tasks such as text and sentence segmentation.

Our work has resulted in a higher level library of capabilities for text and NLP that is readily available to other modules in Socratez for feature extraction and classification.

### 3.4.2 Machine Learning based Classification

The SOCRATEZ approach is to make *inferences* from text as opposed to just spotting terms in the text. Inference is done using machine learning based classification

A rich gamut of features is employed to achieve high accuracy and scalable classification. Some of the features include RNN Deep Learning, TFIDF, Naïve-Bayes, MaxEnt, CRF, Word Distribution, SVM, Regression. For determining expressions we

employ features based on indicative language phrases in the conversation, part-of-speech (POS) tags, punctuation marks or other symbols or expressions, the (classes of) entities in the text, etc.

The integration of machine learning based classification with the use of domain specific ontologies facilitates a true combination of supervised and unsupervised techniques for classification.

### 3.4.3 Knowledge Engineering

Semantics i.e., knowledge about the domain is the third fundamental cornerstone in our approach. Leveraging domain specific ontologies and vocabularies has proven to be extremely useful in any SOCRATEZ classification – be it for expressions, sentiment or aspects.

We have assembled several ontologies in key domains such as retail and health and have successfully integrated them into Socratez. In addition, we have also created a comprehensive ontology of *emotion* terms, along the lines of the emotion categorization recommended in the Plutchik wheel of emotions. Our overall approach is to first leverage existing ontologies where available – and indeed in domains such as health this has been a possibility with the availability of existing and very comprehensive medical ontologies such as the UMLS and SNOMED. The next option is to rapidly assemble required ontologies from open knowledge sources such as DMOZ, Freebase or Wikipedia. The knowledge engineering component of Socratez is equipped with modules for rapid assembly of ontologies from sources such as the above on an on-demand basis.

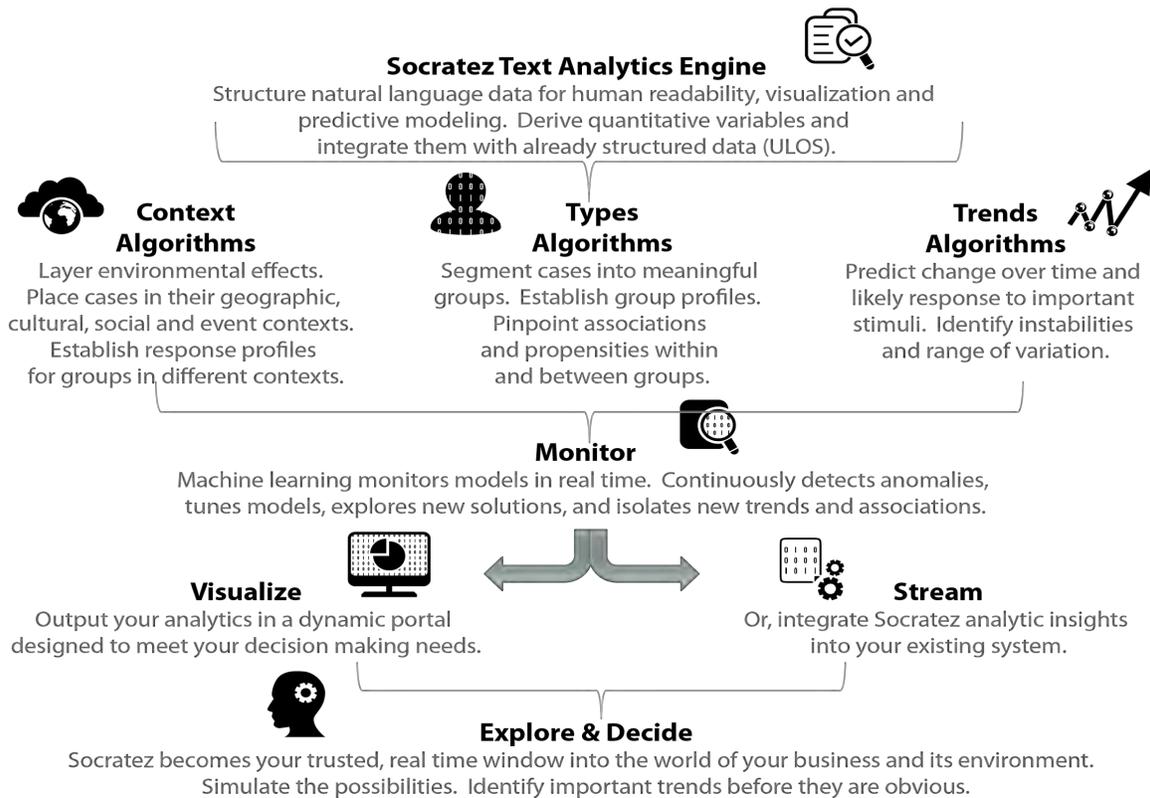
## 4. Socratez Data Insightz™ Module

Socratez™ Data Insightz Module consists of a gamut of algorithms that can extract deep meaning and insights from structured or quantitative data,

**Socratez Insightz can work on qualitative or quantitative data**

The power of having it pre-integrated with our Text Analytics Engine is its ability to find predictive insights in real-time also from any qualitative unstructured data because the Text Analytics Engine quantifies the data with the “who” & the “what”. These insights, drawn from deep, real time analysis of the data, offers new opportunities for improving business. Our professional services team can customize our Data Insightz Module to build custom solutions.

**Figure 4 Socratez Insightz Overview**



## 4.1 Applications of Socratez Insightz

Socratez Text Analytics not only extracts the expressions, sentiment, associations, entities and topics from the Turkey example in Table 1. It also recognizes that the sender of this message has a high commitment to healthy, organic foods. Socratez Insightz is able to automatically factor this into relevant predictions about what types of people will buy the turkey product, what marketing messages would be most successful, etc. In this example, Socratez finds patterns, trends and correlations to predict that:

- Millennials are growing more skeptical on Producers claims of “natural” and they wont buy a product just because it is marketed as natural.
- Using elements of location & proximity, we predicted that Moms in low income suburban areas are becoming more health conscious and show strong motivation to buy organic products. If not for this Insight, this group is not the target group of a marketing organization
- Professional men are increasingly selecting “natural” foods, even though this has not historically been a targeted group.

The Insightz Types module generates consumer profiles and segments of the potential purchasers. The Insightz Context module can place each consumer in their geolocation (to account for local market choices), their social network (to account for the influence of important others on their buying decisions), or their cultural groups (to account for the impact of beliefs and values on their decisions). And the Insightz Change module can formulate predictions about timing and quantity of purchase, given their recent history.

Depending on the problem, Socratez Insightz not only incorporates the additional, rich information from the unstructured data, it also brings in the already structured data that are available. In this case, the company’s records of this consumer’s purchase patterns or bureau data capturing demographic, financial and behavioral characteristics. Together, the natural language data (structured by the Socratez Text Analytics engine) and the more traditional, already structured data, for a much richer and more nuanced consumer profile.

Socratez Insightz also applies to operational problems, like predicting server load & performance issues, optimizing demand response, etc.. It is applicable to system-generated data as well as human generated data.

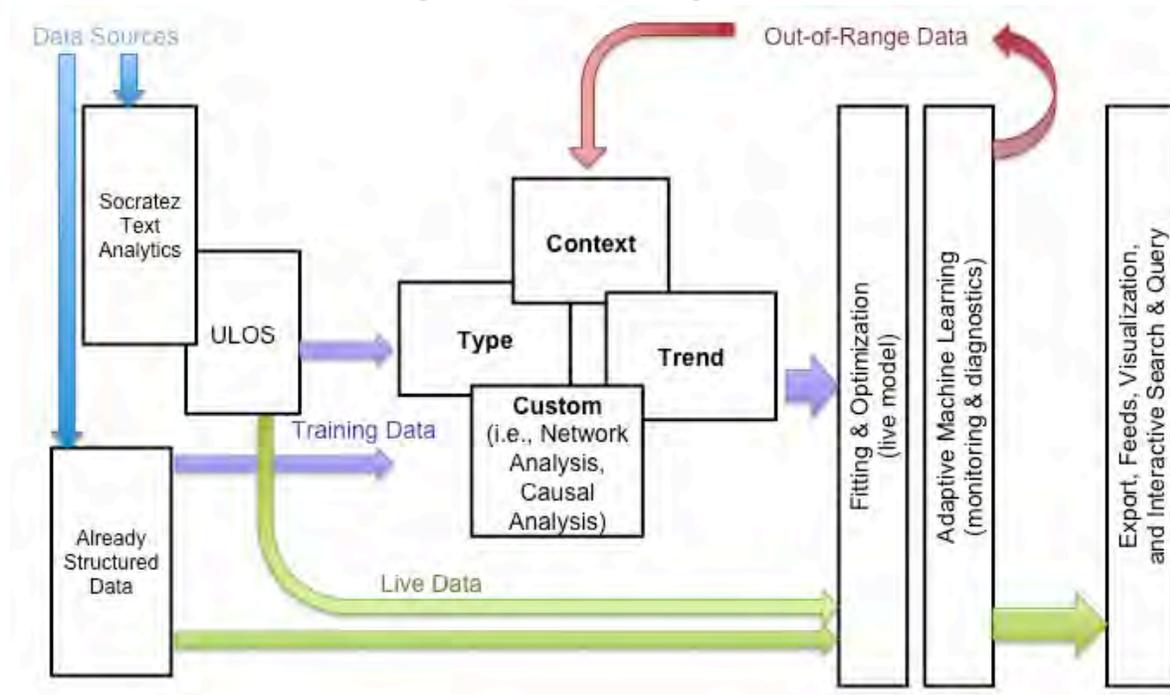
## 4.2 Insightz Platform

To achieve the most accurate Predictions we have designed a robust framework that insures reliability and accuracy in our predictive results.

### 4.2.1 Universal Lexicon & Ontology System

We have built a proprietary “Universal Lexicon & Ontology System (ULOS)” as a base layer ontology. This ontology is designed to encode relations for more than half a million real world entities and actions. It detects, extracts and quantifies underlying characteristics of the message, its author, its intended audience, and its context. We’ve mapped linguistic properties among hundreds of thousands of terms for the cognitive, emotional, evaluative, social, and cultural clues they reveal about the sender, receiver, message and context. Those factors are designed to build quantitative variables that are ready for inclusion in predictive models.

**Figure 5 Socratez Insights Platform**



We have identified and built models (shown in Table 2) in our predictive core that address many of the most common, and pressing, business questions. Each of these modules works independently or together, as the business problem dictates. Socratez not only generates insight, but it does this dynamically for particular target groups in identified environments.

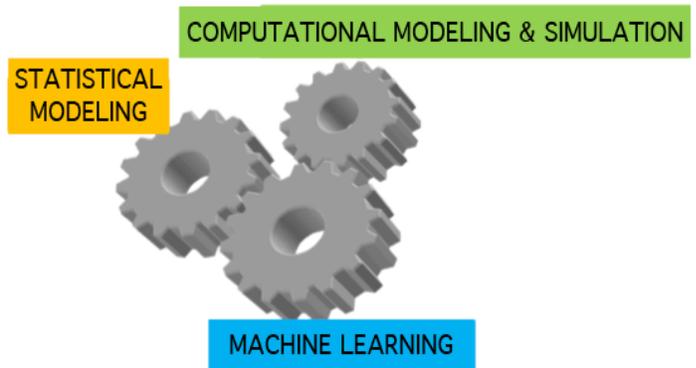
**Table 2 Socratez Insightz Predictive Models**

<b>PREDICTIVE MODELS</b>	<b>BRIEF DESCRIPTION</b>	<b>EXAMPLE(S)</b>
Types (segmentations, profiles, personas)	Ability to classify people, statements, businesses and other identifiable cases into meaningful groups. Socratez Types Insightz module evaluates the data for these meaningful points of distinction using a variety of classification and clustering algorithms.	<ul style="list-style-type: none"> <li>• How are my customers different from one another?</li> <li>• How do my salespeople act differently from one another?</li> <li>• Which products tend to go together, and for whom?</li> </ul>
Change (trend, cohort, longitudinal, and time series analyses)	In a fast-paced, dynamic world, business needs a real time understanding of what's changing and why. Socratez Trends Insightz detects what's changing, how it's changing, and offers understanding of why this is the case.	<ul style="list-style-type: none"> <li>• Where are my sales likely to be next week? Next quarter?</li> <li>• How are operational costs changing as we move into a new market?</li> </ul>
Context (physical, cultural, social, network, organizational, and market environments)	Socratez Context Insightz module uses the latest in hierarchical and variance components models to identify these environmental effects and partition the variance accordingly so that you can choose the best strategy for the environment your business currently operates in.	<ul style="list-style-type: none"> <li>• Do people who follow sports also like my products?</li> <li>• How do people use my services differently as they age?</li> </ul>
Anomaly Detection (ML net)	Monitoring algorithms, improving model specification, and identifying unanticipated factors	Detecting new topics of conversation associated with "heritage turkey"

## 4.3 Technology Approach

The SOCRATEZ approach to Predictive Insights is built on three fundamental pillars and Insightz Machine Learning Layer:

- Statistical Modeling
- Machine Learning
- Computational Modeling
- Insightz Machine Learning Layer



### 4.3.1 Statistical Modeling

Socratez Insightz draws on a full range of statistical techniques: the General Linear Model (GLM), Generalized Linear Model (GLZ), Multivariate Adaptive Regression Spline models (MARS), Variance Components and Hierarchical Linear Models (VCM & HLM), Structural Equation Models (SEM), and Fourier Analysis and Wavelets, as well as clustering, factorization, and other latent variable models. We also apply a number of nonlinear data transform, spline and subsampling techniques to preprocess the data and optimize it for model fit.

### 4.3.2 Machine Learning

Socratez Insightz employs supervised and unsupervised learning techniques for both data mining for unknown features in the data and for supplementing, refining and monitoring our statistical predictive algorithms. We have integrated Ensemble methods, Decision Trees, Neural Networks, and Support Vector Machines (SVN), as well as regression, clustering and classification, dimensionality reduction, and anomaly detection routines.

### 4.3.3 Computational Modeling & Simulation

Finally, Socratez Insightz employs a range of techniques that allow for sophisticated, “what if” predictions. These include static and adaptive game theoretic models; decision theoretic, social choice, and agent-based simulations; and equilibrium, complex, chaotic, and catastrophic models. Our computational and simulation

models can work within contextual and dynamic parameters established through our statistical and machine learning estimates.

The Socratez Insightz modules suite has been built on an integrated framework of R, Weka, and Octave with Python and Java connective scripts. We have leveraged the best in open source platform technologies and added significant proprietary code that further integrates and optimizes performance. We have also encoded a significant amount of social and behavioral science providing logical context for operation.

#### **4.3.4 Insightz Machine Learning Layer**

Predictive models are built by training them against existing data. That means every predictive model is limited to predicting the future based on the past. This works well in an unchanging world — or in a world without surprises where change consistently occurs in the way it always has. But the data constantly changes with time and therefore this approach does not always continue working well.

Socratez Insightz throws a machine learning “net” around its predictions. Whether the model is based on qualitative (unstructured, text) or quantitative (structured, numeric) data, Socratez scans the space around the model to detect the leading signals of unanticipated change. It identifies new trends, patterns and correlations so that you have the maximum lead-time for strategic planning and response.

As new data are processed through Socratez Insightz, they are compared against the best model already fitted by the system. When new data are out of the predicted range, they are automatically analyzed to establish whether there are likely to be new factors influencing these previously established relations. If enough data are out of the predicted range, the model is automatically re-parameterized and our data scientists are notified in case further expert diagnostics are needed. Socratez also performs this continuing comparative analysis with qualitative (natural language) data.

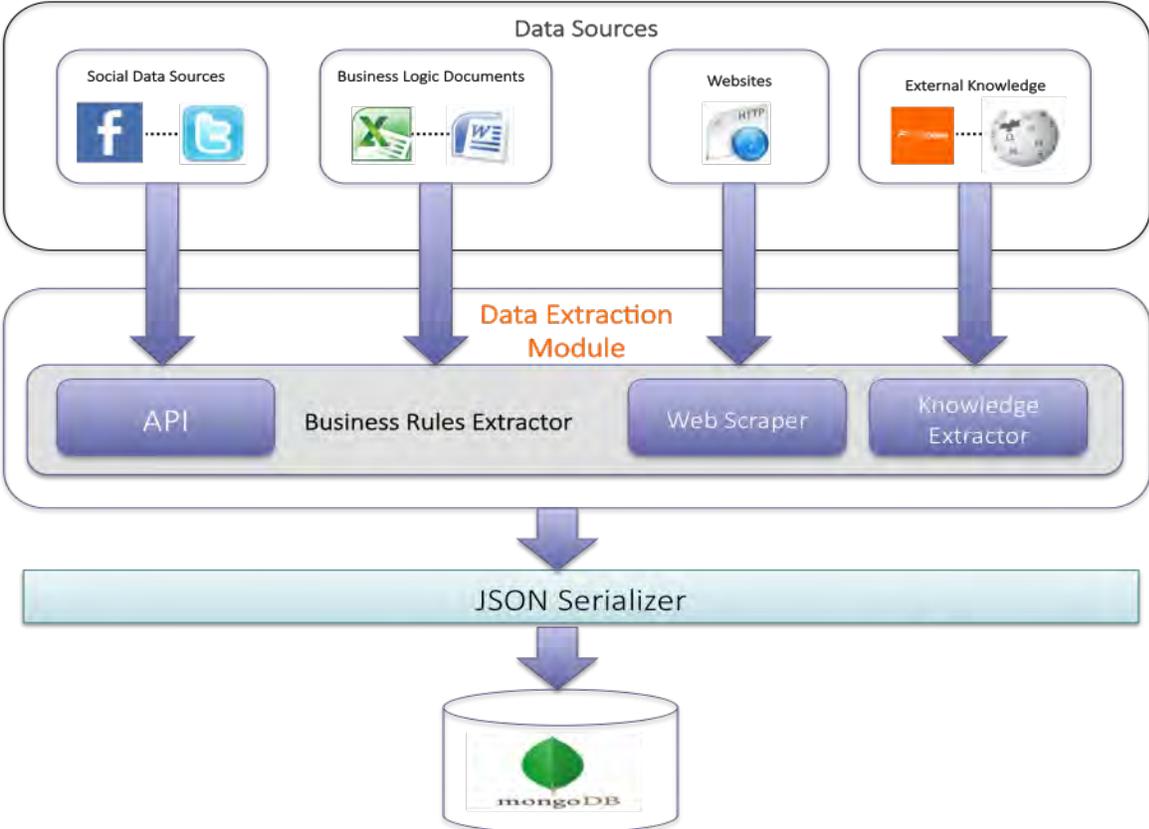
So, for example, while Socratez is tracking existing conversations about “heritage turkey” and “organic meat,” it also detects that there is the beginning of a rise in conversation about “stunning” and the use of “ammonia” in meat production. The system notifies the client that those topics are gaining traction among the thousands of conversations before these changes are noticeable to human attention.

Similarly, while the system is monitoring and predicting sales by demographic group and geographic region, it notices that an anomalous trend is beginning as existing customers from one group in one region is purchasing less “natural turkey” in supermarkets than before. Further analysis allows Socratez to determine that there is a new market force at work: a local farmer’s coop introducing locally grown turkeys into that market. The client is able to address this competition far sooner than they would otherwise.

### 5. Socratez Data Extraction Module

Socratez Data ingestion module is designed to extract any type of data from multiple sources in different formats. This includes extracting data from Application Programming Interfaces (API), web pages using scrapers, business logic documents (word, excel), external knowledge sources and any enterprise owned data (CRM, ERP, POS data, Sales & Marketing, etc.). The data extracted is converted to JSON format and stored in a highly efficient and scalable MongoDB datastore. Each approach and the corresponding data sources have been described in detail below

Figure 6 Socratez™ Data Extraction Module



- **Application Programming Interface (API's)** Socratez Data ingestion engine can extract data from any social site or website offering an API. We have API's build into Facebook and Twitter, but can do the same with other sites like Instagram, Pinterest, etc. The most common approach for extracting data is using keywords. We can also extract data using a specific Facebook or Twitter username or page name. Socratez also supports Freebase API's which is a knowledge graph of connected entities on the web.
- **Web Scrapers** Socratez Data Ingestion module can also extract data from any web pages or blogs using our proprietary scrapers. The scrapers are flexible enough to adapt to any new web page content and extract relevant pieces of information.
- **Business Rules Extractor** Socratez Engine can customize data extraction based on custom Client Business Logic and Rules. This allows for filtering and cleaning of the data even before it get into a database. This allows Socratez to achieve high accuracy as less unwanted (junk) data is extracted to begin. It contains rules defining what pieces of extracted data should be preserved vs discarded, contains client specific rules for data analysis and the like.
- **Knowledge Extractor** Socratez engine leverages external data sources such as Freebase and Wikipedia to enrich the internal data models (knowledge graphs/ontologies). While Freebase is a knowledge graph of connected entities on the web, Wikipedia is a source of community-generated content.

## 6. TeraCrunch Professional Services

TeraCrunch's Professional Services offering enables customization of Socratez Platform to fit niche custom needs of our customers. Our Professional Services team consists of highly experienced PhD Data Scientists and Engineers. The customizations includes:

1. Backend Engineering
  - Custom data extraction work (any website, blog, social site, documents...)
  - Data integration & Data mining
2. Data Science Customizations
  - Developing vertical specific Ontologies/Dictionaries
  - Algorithm selection and modeling
  - Profiling & Segmentation
3. Front-end Development of the Tool
  - Dashboards or widgets for visualizing analyzed data
  - Real-time API that can be integrated with existing tools & systems

## Author's Bio

### Dr. Naveen Ashish



Dr. Naveen Ashish expertise and interests are in the areas of data integration and information extraction from unstructured data employing techniques from Machine Learning, NLP, and Semantics.

He has also worked as a Scientist at NASA Ames Research. He has published over 75 papers in top journals and conferences in the area on the above topics. He has authored a book on Geospatial Semantics: Foundation, Algorithms & Applications (<http://alturl.com/x6cgi>). His other books are in the space of

Information Mediators and NASA & the Semantic Web (<http://alturl.com/4m6ho>). He has been a Principal or Co-Investigator on several R&D awards from institutions such as the NIH and NSF. Naveen received his Bachelors from IIT, Kanpur in India and a MS & PhD in Computer Science from the University of Southern California in 2000

### Dr. Kevin Payne



Dr. Kevin Payne has been a research methodologist and data scientist for over twenty years. He has a broad background across statistical, computational and machine learning techniques. But as a sociologist and social psychologist, and is focused on the real human meaning that can be discovered in the data. In his previous career, as a professor, he led or collaborated on dozens of research projects, published or presented 40 research papers in international, national and regional venues, and grew the Park University Sociology Program over 2 1/2 times (to \$5.5M in annual enrollment revenue).

Kevin did his undergraduate work at William Jewell College and Oxford, UK. His Master's and PhD (in Sociology, with a collateral field in Psychology and specializations in Social Psychology, Theory and Research Methodology) are from the University of Missouri-Columbia.

Contact [services@teracrunch.com](mailto:services@teracrunch.com)