Who is **Basis Technology?**

Building innovative text analytic components **since 1998**
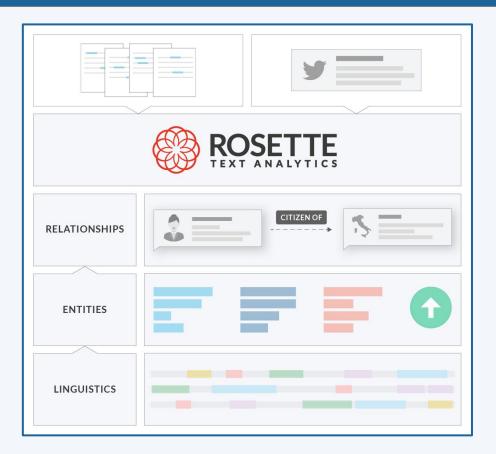
Robust capabilities

Prominent customers

Most use of our technology is done **unknowingly**
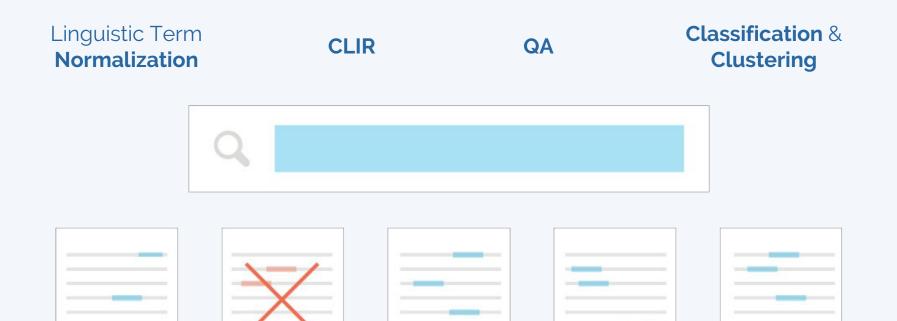
Select Customers

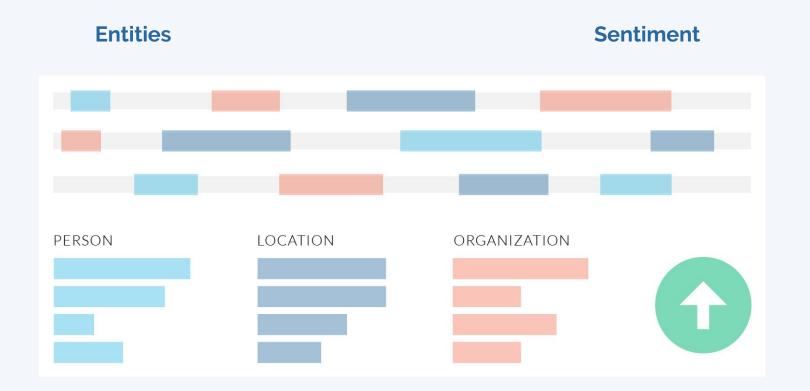# Text Analysis **History**

Three Eras Over the Past **Two Decades**

Finding
Documents

Extracting
Tidbits

PER
LOC
ORG

TOP EMPLOYEE OF

Connecting
the Dots

Era 2: **Extracting** Tidbits

Entities

Sentiment

PERSON

LOCATION

ORGANIZATION
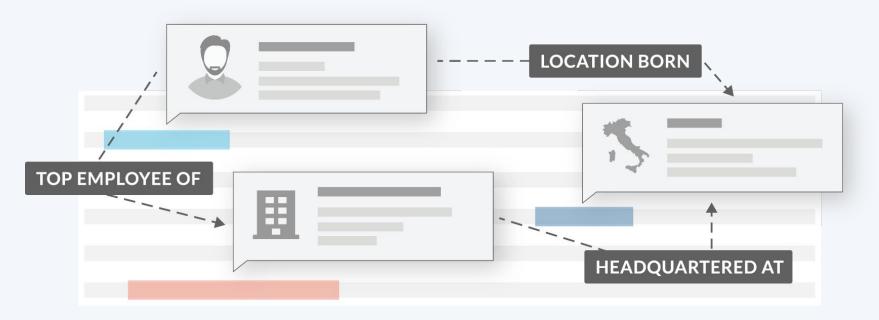
Era 3: **Connecting** the Dots

Relationships

Resolution/Linking

LOCATION BORN

TOP EMPLOYEE OF

HEADQUARTERED AT

Opportunities in **Regulatory** Data

8 months prior to Volkswagen's emissions scandal, Korean environmental regulators took action against Volkswagen for the same thing.

By using simple entity ID, investors and regulators would have **already known** about VW emissions issues.

EPA — United States Environmental Protection Agency

Environmental Topics   Laws & Regulations   About EPA   Search EPA.gov
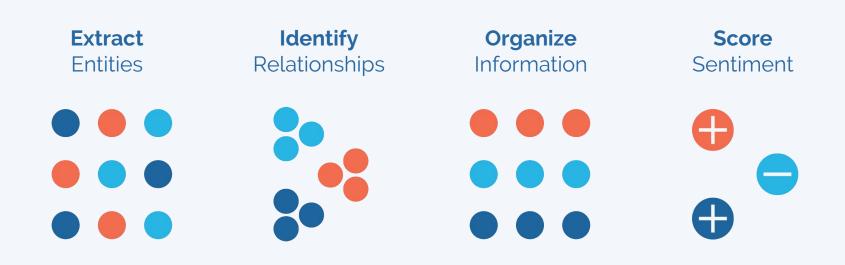
News Releases from Headquarters

Volkswagen to Spend Up to $14.7 Billion to Settle Allegations of Cheating Emissions Tests and Deceiving Customers on 2.0 Liter Diesel Vehicles

Settlements Require VW to Spend up to $10 Billion to Buyback, Terminate Leases, or Modify Affected 2.0 Liter Vehicles and Compensate Consumers, and Spend $4.7 Billion to Mitigate Pollution and Make Investments that Support Zero-Emission Vehicle Technology
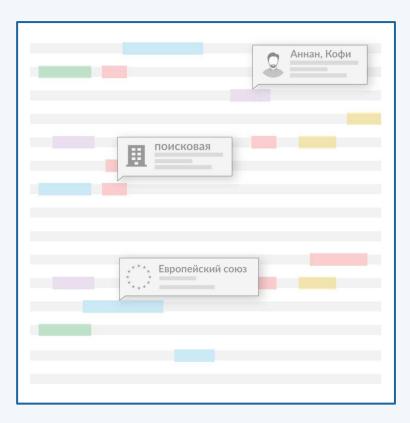
06/28/2016

Entity **Identification** and **Extraction**

Using Basis Technology, Prattle **identifies** and **extracts** all mentions of publicly traded corporations by all developed world regulators.

Named entity recognition will be necessary for **foreign language** regulators.

Document **Scraping** and **Storage**

All U.S. regulatory agency **documents** scraped via API

All U.S. regulatory **notices and comments** scraped via Regulations.gov

Rosette API (entity extraction function) **processes** raw text

Scraping and Storage cont.

Processed documents **uploaded** to Elasticsearch

Updates to Federal Register automatically **ingested** and **formatted**

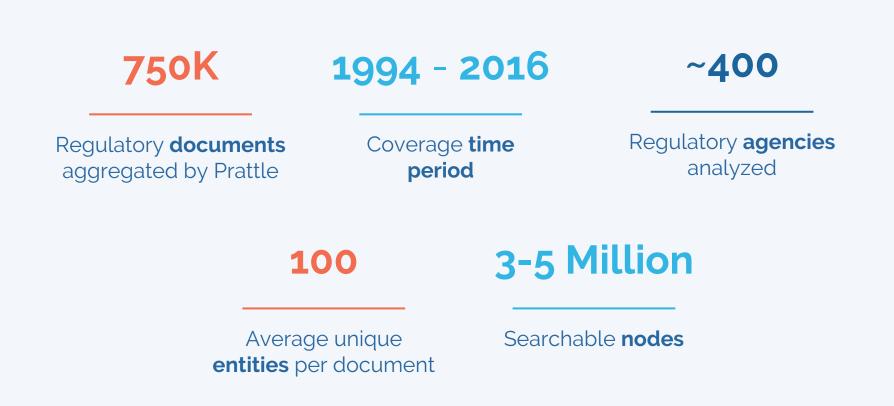Data **stored** as nodes on a graph including Document, Agency, Docket, Comment, and Entity
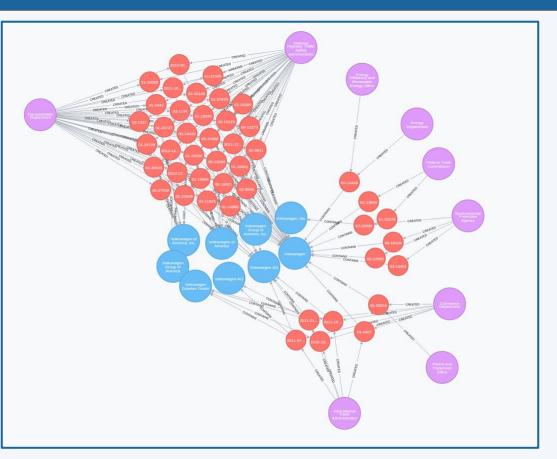
Volume of Communications

BASIS TECHNOLOGY | prattle

750K
_____
Regulatory **documents** aggregated by Prattle

1994 – 2016
_____
Coverage **time period**

~400
_____
Regulatory **agencies** analyzed

100
_____
Average unique **entities** per document

3-5 Million
_____
Searchable **nodes**

Identifying **Relationships**

BASIS TECHNOLOGY   prattle

Agencies and **Regulations**
mentioning Volkswagen

● **Agencies**

● **Documents**

● **Mentions**
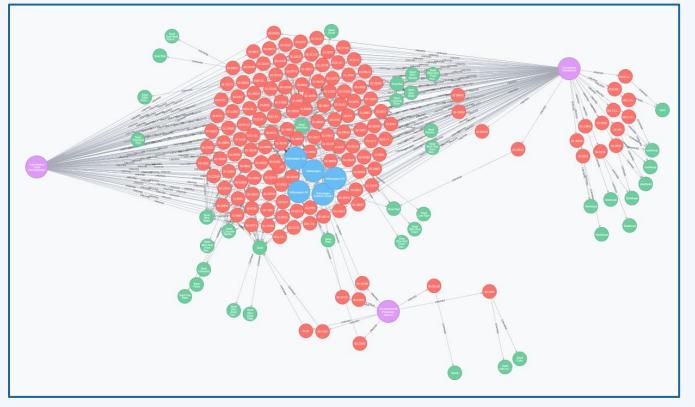
Going a Step **Further**

Agencies and **Regulations** mentioning Volkswagen and steel

- 🟣 **Agencies**
- 🔴 **Documents**
- 🔵 **Volkswagen**
- 🟢 **Steel**

Constructing **Lexicons**

Historical regulatory actions are used as a **baseline** for gleaning sentiment from regulators.

This allows for the construction of a completely **unique lexicon** for each regulator.
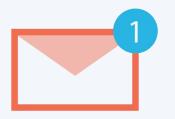
Moving Forward

Alert-based system
that identifies new regulatory mentions of specific entities

Relationship-based **suggestion algorithm** to identify mentions of related entities

Prediction algorithm to identify regulatory risk due to relationships

The **Future** of NLP

Sentiment analysis is now able to **effectively** analyze complex, nuanced language, making robo analysts an effective option.

However, the future isn't artificial intelligence eliminating human analysis; it's artificial intelligence **augmenting** human analysis.

THE WALL STREET JOURNAL.

LIFE | IDEAS | ESSAY

Learning to Love Intelligent Machines

Twenty years after famously losing to Deep Blue, chess champion Garry Kasparov says that it's time to embrace AI and its liberating potential

ILLUSTRATION: PEP MONTSERRAT