



Human Curated Linguistics

Technology behind Cognitive Analytics

Nikita Ivanov

nivanov@datalingvo.com

DataLingvo, Inc.

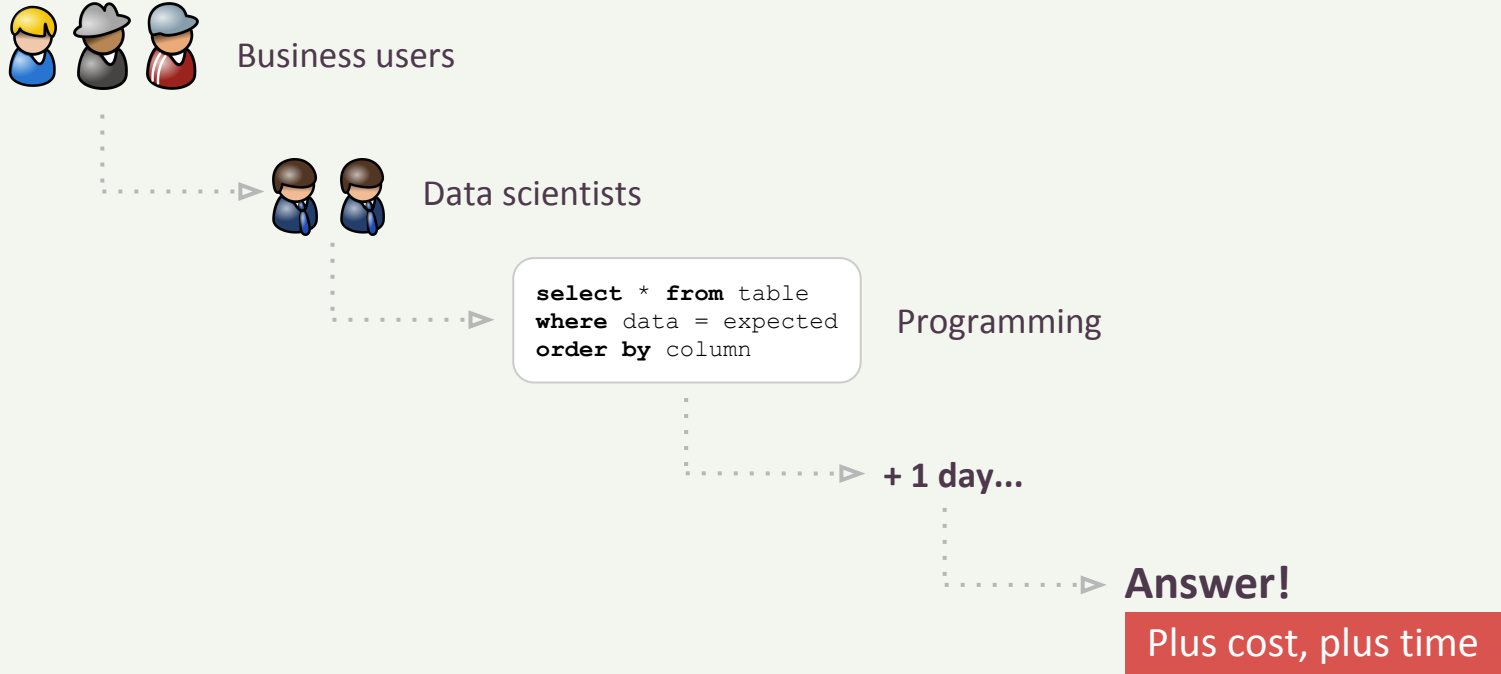
www.datalingvo.ai

The Problem

“**Corporate data is trapped in silos.** Rank and file don't have access to data or the skill set to turn that data to knowledge. For those that do - the process is time consuming and costly.”

Gartner
BI & Analytics Summit, 2013

Old Way



Cognitive Analytics

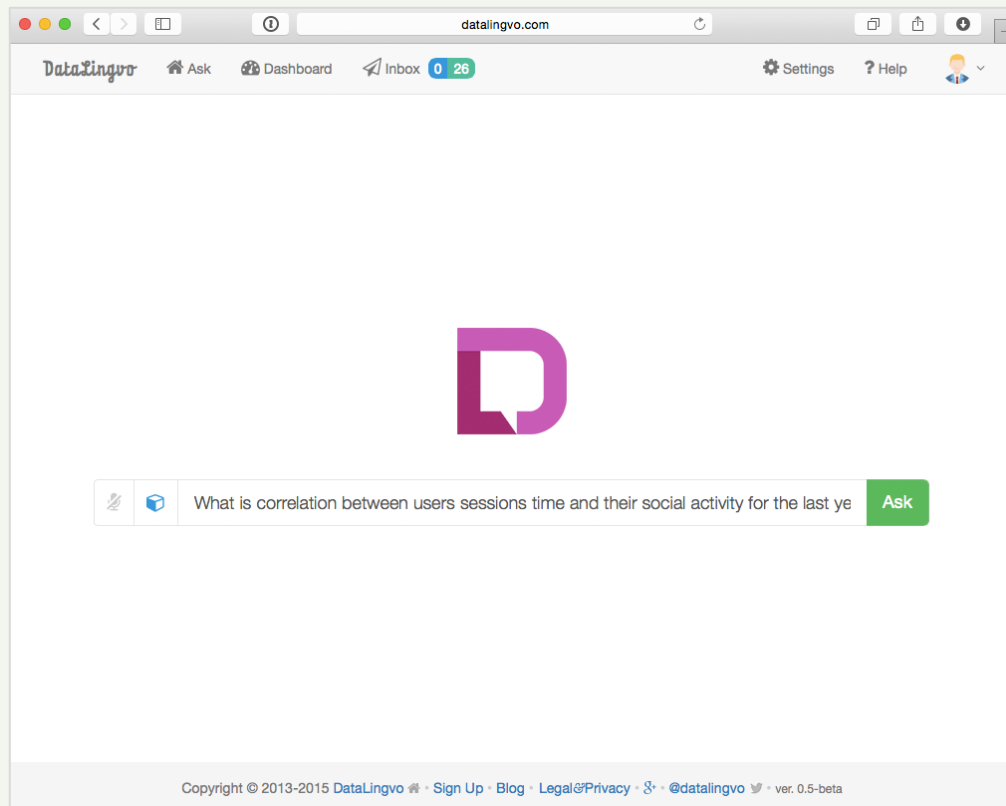
Cognitive Analytics lets anyone in an organization ask any question about company's data and get a real-time answer by using a natural language.

Think Siri for BI and data analytics.

See it live at www.datalingvo.ai
ver. 0.5-beta

New Way

Just **say or type** a question using natural language

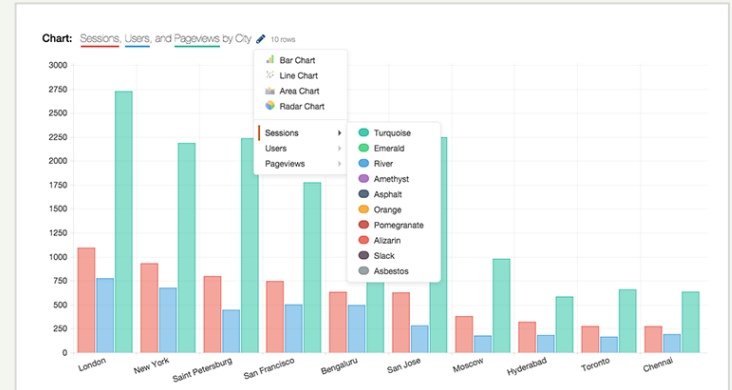
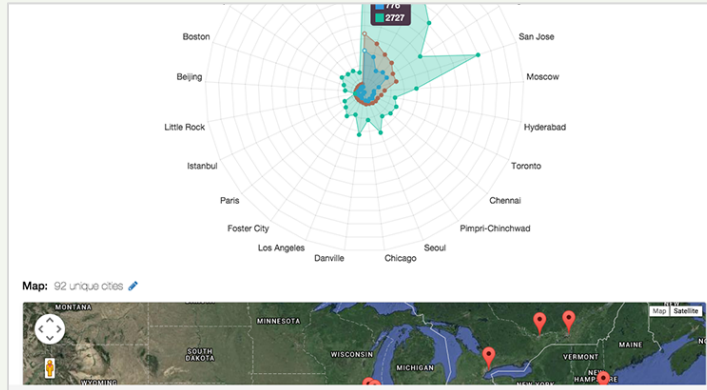
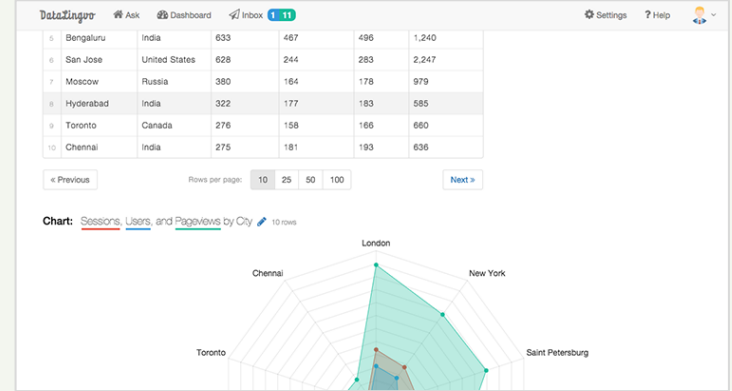
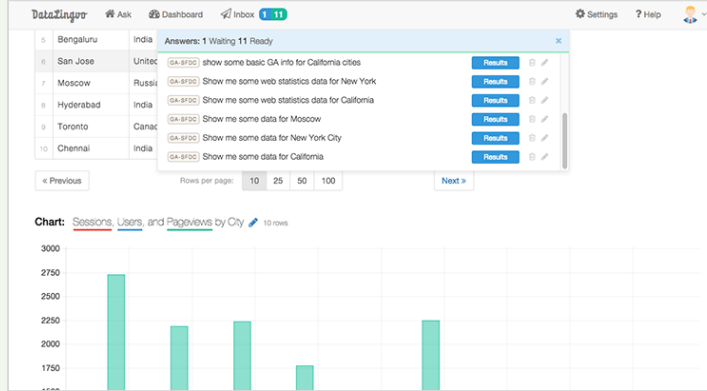


Get an instant **real-time answer!**




Comprehensive Answers


Easy entry -
comprehensive
answers





Example User Questions


Show me top 10 most active cities for the last 90 days. 



Which city produces more activity: San Francisco or New York? 



Show me the countries where users' number has increased for the last 3 months. 



Give me data by states where users' number increased at least 20% for the last month comparing to the previous one. 



Show me USA cities and their the most popular browsers. 



Give me data for German cities, where users appeared only this year. 

 Show me the amount of sales by states with the most active visitors for this year. 

 Show me the number of visitors for the most active sales months in US. 

 Show me site statistic by cities where the sales were closed this year. 

 What is a correlation between new users and opportunities created for each state? 

 What are the most visited pages in the best sales month? 

Natural Language



Uniform interface for any data sources

- Zero learning curve & adoption barrier
- Enables millions of business users
- Ideal for mobile, embedded and wearable

HCL - Human Curated Linguistics

HCL - Core DataLingvo Technology *

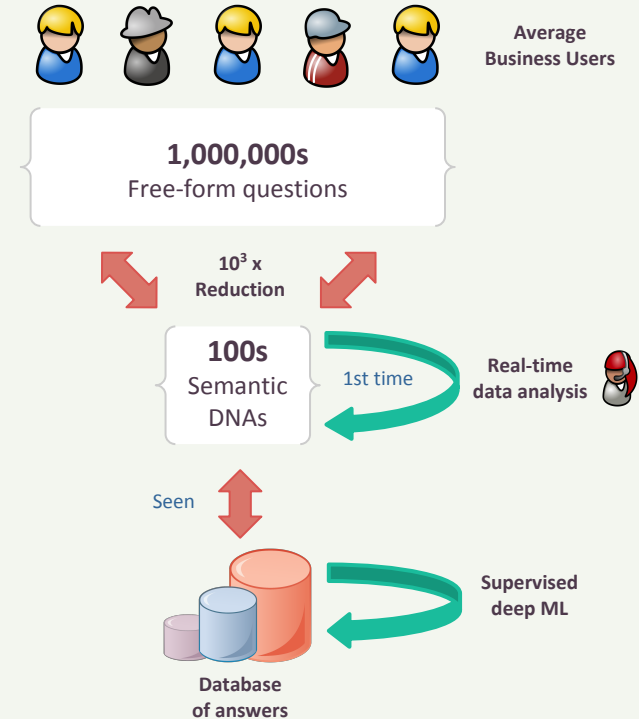
HCL Principles:

- 10,000x reduction into Semantic DNAs
- New DNAs answered by a data analyst **only once**
 - New answer triggers AI-based self-learning
- Previously seen DNAs answered automatically

HCL Differentiators:

- Free-form natural language comprehension
- Guaranteed result correctness (determinism)

* Patent Pending



Back-End Interface

Process 13a74a3cc41-ef548df7-44ab-4064-8c07-2de91ef9feb7

Text: what is the correlation between sessions and week days for the last 3 months ?

NLP Keys:

Type	Matched	Value
Stem	0	correl session and week dai for the last 3 month
Lemma	0	correlation session and week day for the last 3 month
Token	0	correl GA > SESSIONS and GA > DAYOFWEEK DATE
Token GA	0	correl GA > METRIC and GA > DIMENSION DATE

Metrics: ☒ Auto Fill ☒ 10 metrics

Date Range: From: 2015-01-31 To: 2015-04-23

Sort: ga:dayOfWeek, -ga:sessions

Dimensions: ga:sessions ga:dayOfWeek

Metrics List:

- ga:adSenseAdUnitsViewed AdSense, INTEGER
The number of AdSense ad units viewed.
- ga:adSenseExits AdSense, INTEGER
The number of sessions that ended due to a user clicking on an AdSense ad.
- ga:adSenseRevenue AdSense, CURRENCY
The total revenue from AdSense ads.
- ga:adSenseCoverage AdSense, PERCENT
The percentage of ad requests that returned at least one ad.
- ga:adSenseCTR AdSense, PERCENT
The percentage of page impressions that resulted in a click on an AdSense ad.
- ga:adSenseAdsViewed AdSense, INTEGER
The number of AdSense ads viewed
- ga:adSensePageImpressions AdSense, INTEGER

Process 13a74a3cc41-ef548df7-44ab-4064-8c07-2de91ef9feb7

Text: what is the correlation between sessions and week days for the last 3 months ?

☐ Chart ☒ Submit 2 keys ☐ GA Values ☐ Dry Run ☒ Validate ☐ Require review

NLP Sentence:

Origin: 50.136.202.204 at 4:14:36 PM
User: Beta Tester, beta@datalingvo.com
Raw text: what is the correlation between sessions and week days for the last 3 months?
Server request ID: 13a74a3cc41-ef548df7-44ab-4064-8c07-2de91ef9feb7

#	Text	Stop	Stem	Lemma	POS Description	Dict	GA	SESSIONS	DATE	NUM	DATE	JSON
0	what is the	✓	what is the	what is the	Synthetic tag							<input type="checkbox"/>
1	correlation		correl	correlation	NN Noun, singular or mass	✓						<input type="checkbox"/>
2	between	✓	between	between	IN Preposition or sub- conjunction							<input checked="" type="checkbox"/>
3	sessions		session	session	NNS Noun, plural		✓					<input type="checkbox"/>
4	and		and	and	CC Coordinating conjunction	✓						<input type="checkbox"/>
5	week days		week dai	week day	Synthetic tag		✓					<input type="checkbox"/>
6	for the last 3 months		for the last 3 month	for the last 3 month	Synthetic tag				✓			<input type="checkbox"/>
7	?	✓	?	?	Synthetic tag							<input type="checkbox"/>

```
{
  "index": 2,
  "notes": [
    {
      "noteType": "NLP",
      "from": 2,

```

Real-Time Human Curated Linguistics

System Architecture

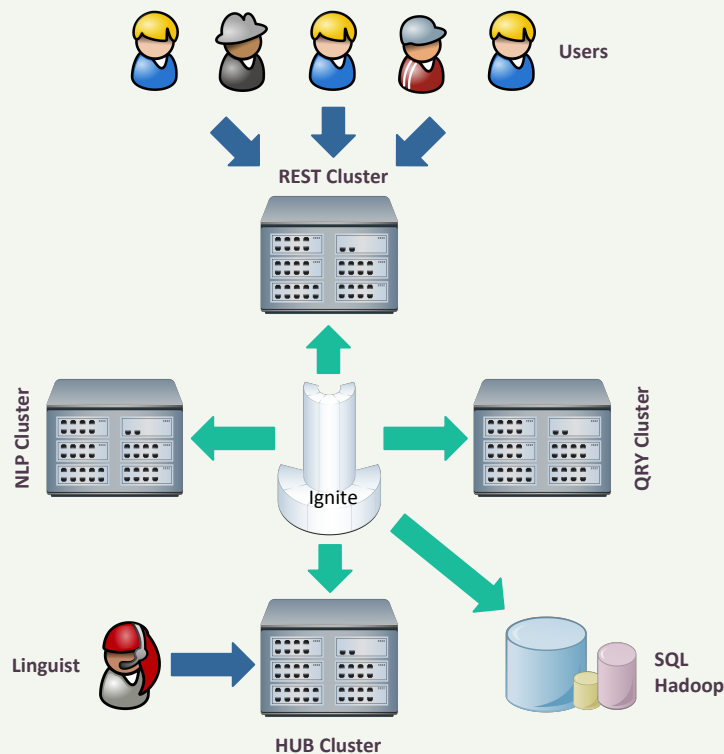
Distributed In-Memory Architecture

Systems stack:

- Apache Ignite In-Memory Data Fabric
- Stanford NLP + WordNet NLP toolset
- Scala backend
- Node.js REST/HTTP servers
- AngularJS + HTML5 + CSS3 web clients
- Cordova + Ionic iOS + Android clients
- Postgres + Hadoop storage

Marketing automation stack:

- AutopilotHQ inbound lead nurturing
- JIRA + Slack + 15Five team management
- Stream.io + Mixpanels user & app analytics
- Salesforce CRM
- ZenDesk customer support
- Google Analytics web analytics
- HootSuite social management
- AdWords + TwitterAds outbound



Demo