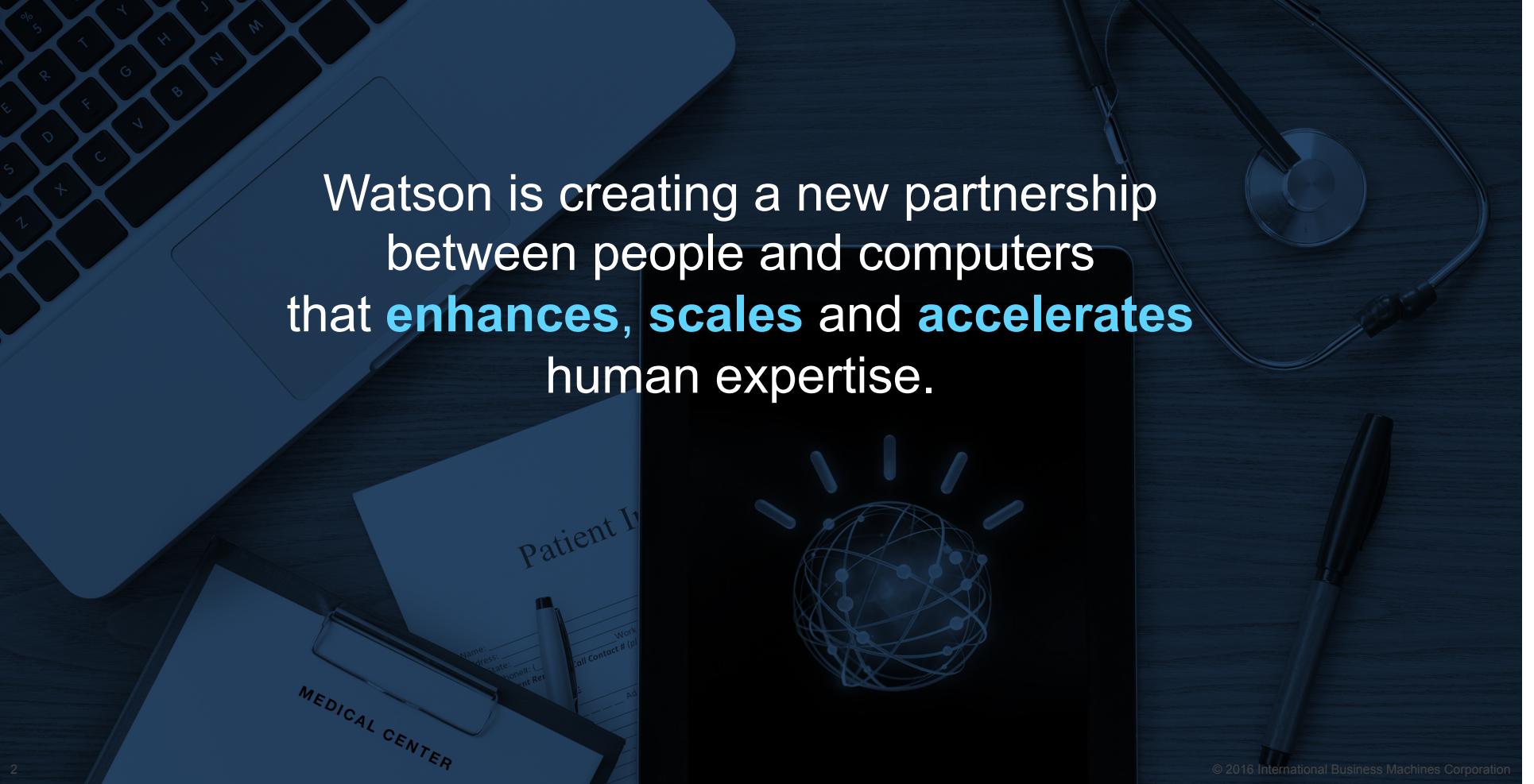


# Watson Developer Cloud

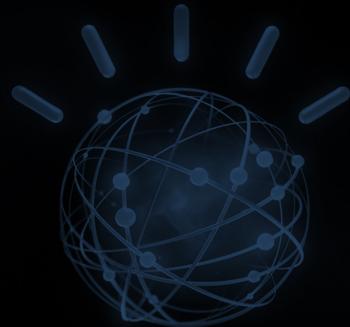
Overview



IBM Watson



Watson is creating a new partnership  
between people and computers  
that **enhances**, **scales** and **accelerates**  
human expertise.



# What is Cognitive Computing

*Cognition is the set of all mental abilities and processes related to knowledge: attention, memory and working memory, judgment and evaluation, reasoning and "computation", problem solving and decision making, comprehension and production of language, etc. Human cognition is conscious and unconscious, concrete or abstract, as well as intuitive (like knowledge of a language) and conceptual (like a model of a language). Cognitive processes use existing knowledge and generate new knowledge.*

*Wikipedia: <https://en.wikipedia.org/wiki/Cognition>*

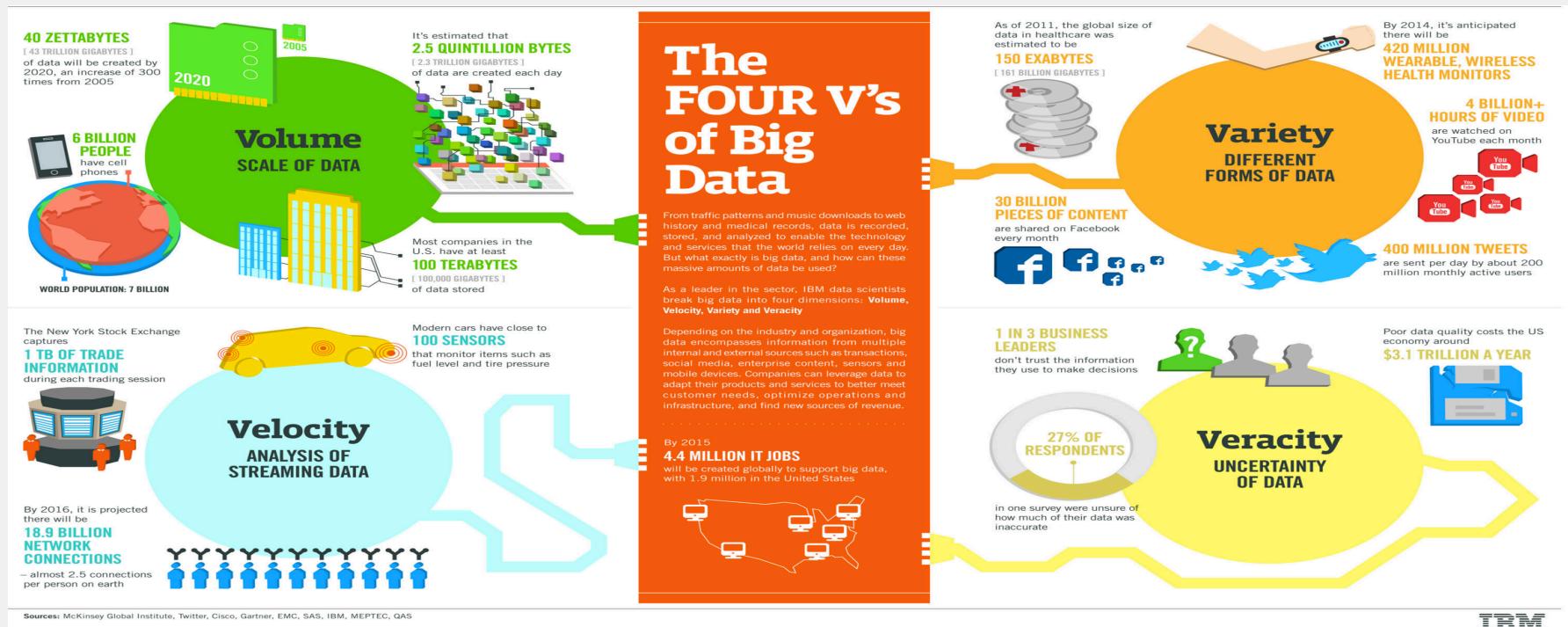
*Cognitive computing systems learn and interact naturally with people to extend what either humans or machine could do on their own. They help human experts make better decisions by penetrating the complexity of Big Data.*

*IBM Research: <http://www.research.ibm.com/cognitive-computing>*

*Cognitive systems are fundamentally different... Cognitive systems are capable of learning from their interactions with data and humans—essentially continuously reprogramming themselves... the machines of the future will do much more than compute. They will be able to sense, learn and better predict the consequences of actions. In the years ahead, machines will cull insights from the vast amounts of information being gathered to help us learn how the world really works, and make sense of all of that complexity, and provide trusted advice to humans ...*

*Dr. John E. Kelly III (IBM Senior Vice President and Director of IBM Research)  
<http://asmarterplanet.com/blog/2012/05/welcome-to-the-era-of-cognitive-systems.html>*

# Why Cognitive Computing



*The volume, velocity, variety, and veracity of data is creating an unprecedented opportunity and a need for real-time insights and evidence based decision making.*

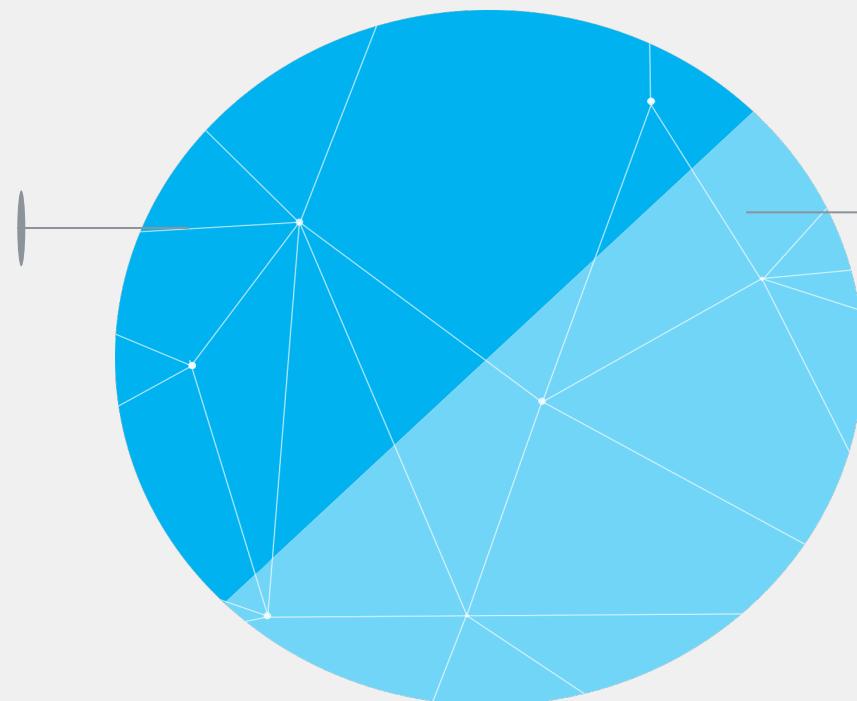
# A New Partnership Between Humans and Technology

**Humans  
excel at:**

- COMMON SENSE
- MORALS
- IMAGINATION
- COMPASSION
- ABSTRACTION
- DILEMMAS
- DREAMING
- GENERALIZATION

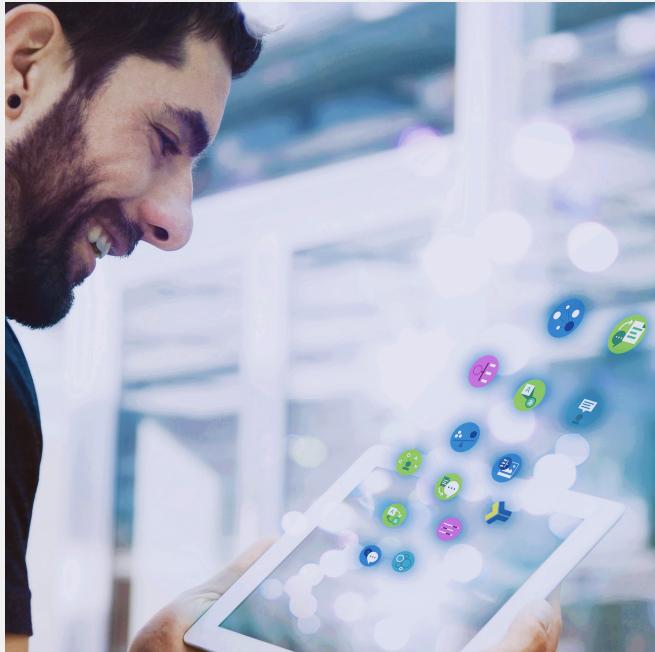
**Cognitive Systems  
excel at:**

- LOCATING KNOWLEDGE
- PATTERN IDENTIFICATION
- NATURAL LANGUAGE
- MACHINE LEARNING
- ELIMINATE BIAS
- ENDLESS CAPACITY



# Bluemix and Developer Cloud

# Watson Developer Cloud



Rapid Innovation in Cognitive Solutions



Make your apps Read, Hear, Talk, See & Learn



Self-Serve environment

- » Cloud access to API's
- » Developer environment
  - API's, SDK's, Demos, Code, App Gallery, Forum

As the Watson technology evolves and deepens,  
so are the ways it's being put to work in the world.



**20**  
Industries



**45**  
Countries



**50,000**  
Students  
in Melbourne



**80K**  
Developers  
building with Watson



**500+**  
Partners  
Powered by Watson



**1.1M**  
Patients  
at Bumrungrad



**8**  
Languages  
Learned by Watson



**200**  
Universities  
offering Watson courses



**5.5M**  
Citizens  
in Singapore

IBM Watson

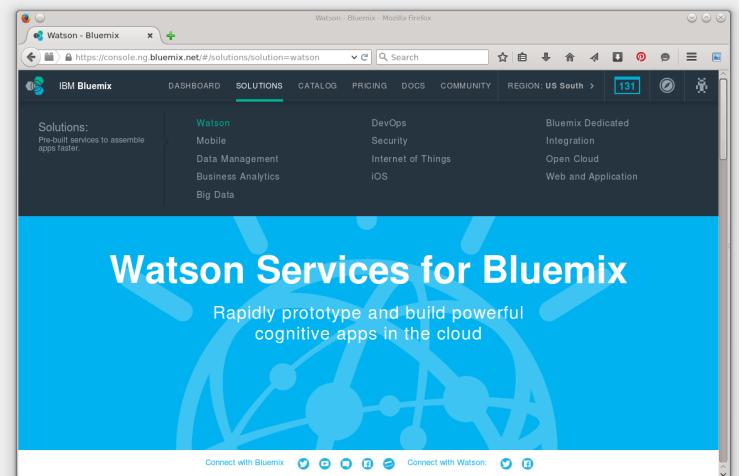
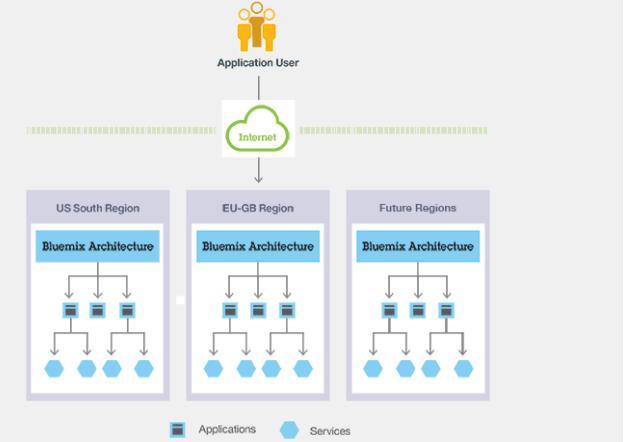
# What is Bluemix

IBM® Bluemix™ is the IBM open cloud **platform** that provides mobile and web **developers** access to IBM software for integration, security, transaction, and other key functions, as well as software from business partners.

- Built on the **Cloud Foundry** open source technology

## Features

- A range of services that enable you to build and extend web and mobile apps fast.
- Processing power for you to deliver app changes continuously.
- Fit-for-purpose programming models and services.
- Manageability of services and applications.
- Optimized and elastic workloads.
- Continuous availability.



# Bluemix Concepts

**Applications** are artifacts built by a developer

- Mobile and Web

**Services** are ready-for-use cloud extensions hosted by Bluemix

- Watson cognitive services, Database, Messaging, etc.
- Developers may easily create custom services

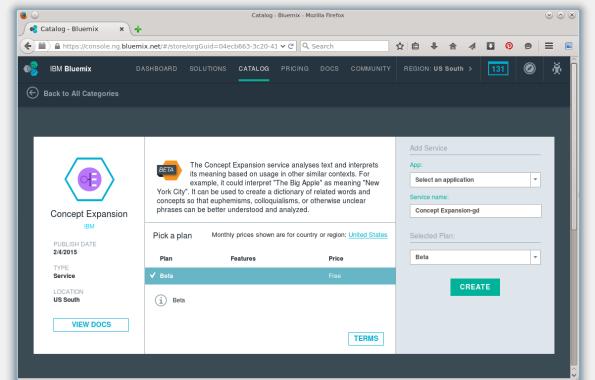
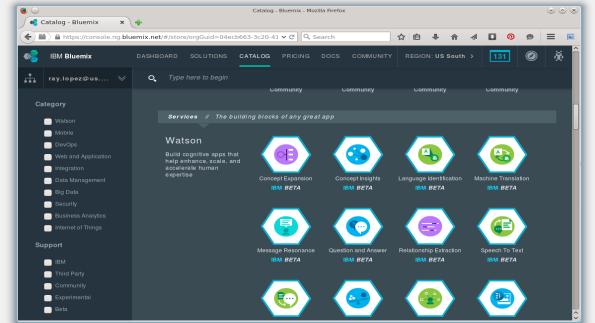
**Starters** are templates that includes predefined services and application code.

- **Boilerplate**: container for application (associated runtime environment and predefined services for a particular domain).
- **Runtime**: Set of resources that is used to run an application (node.js, Liberty, etc).

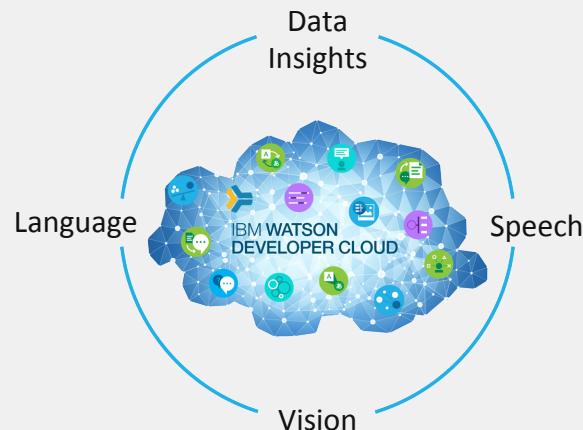
**Regions** are defined geographical territories where applications and service instances can be deployed.

- Data may be deployed in regions separate from apps
- Accessible via Bluemix web interface or CF CLI

Region name	Region prefix	CF api endpoint	UI console
US South region	us-south	api.ng.bluemix.net	console.ng.bluemix.net
Europe United Kingdom region	eu-gb	api.eu-gb.bluemix.net	console.eu-gb.bluemix.net



# Gives Developers Easy Access to Cognitive Building Blocks



[Watson Developer Cloud](#)

Watson cognitive services are grouped into four categories

#### Language

- AlchemyLanguage
- Personality Insights
- Natural Language Classifier
- Retrieve and Rank
- Language Translation
- Tone Analyzer
- Relationship Extraction
- Document Conversion

#### Data Insights

- AlchemyData News
- Tradeoff Analytics

#### Speech

- Speech to Text
- Text to Speech

#### Vision

- AlchemyVision
- Visual Recognition

# Embed Watson APIs into Current Solutions

Startups and large enterprises embed Watson cognitive capabilities into existing solutions.

IBM Watson Explorer users leverage integrated Watson APIs to expand cognitive capabilities

The screenshot shows a customer profile for Brandon Hatcher. On the right side of the screen, there is a red box highlighting the "Ask Watson" feature. This feature includes a Watson logo, a greeting ("Watson Hi Eric How can I help you today?"), a text input field ("Type your question here. Use the same wording as you would use to ask a person."), and a blue "Ask" button. The rest of the page displays various sections like Customer Information, Demographic Information, Family Members, Vehicles, Interactions, and Latest Activity.

Q&A through Dialog and Natural Language Classifier

Understand your customer with Personality Insights

# Combine Watson APIs for Higher Value Apps

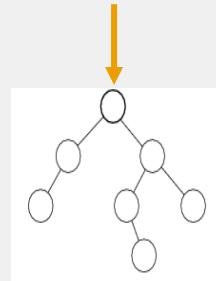
*Example: Interactive Q&A with greater understanding and personality*

Speech to Text,  
Text to Speech



Natural  
Language  
Classifier

Dialog



How do I open an account?  
Convert Speech to Text

Identify Intent  
-Intent = "Open\_Account"  
-Confidence: 0.876655900

Dialog tree walks user to desired outcome  
- Intent="Open\_Account" "  
- Context = "Online Banking"  
- Dialog could call a DB or CMS

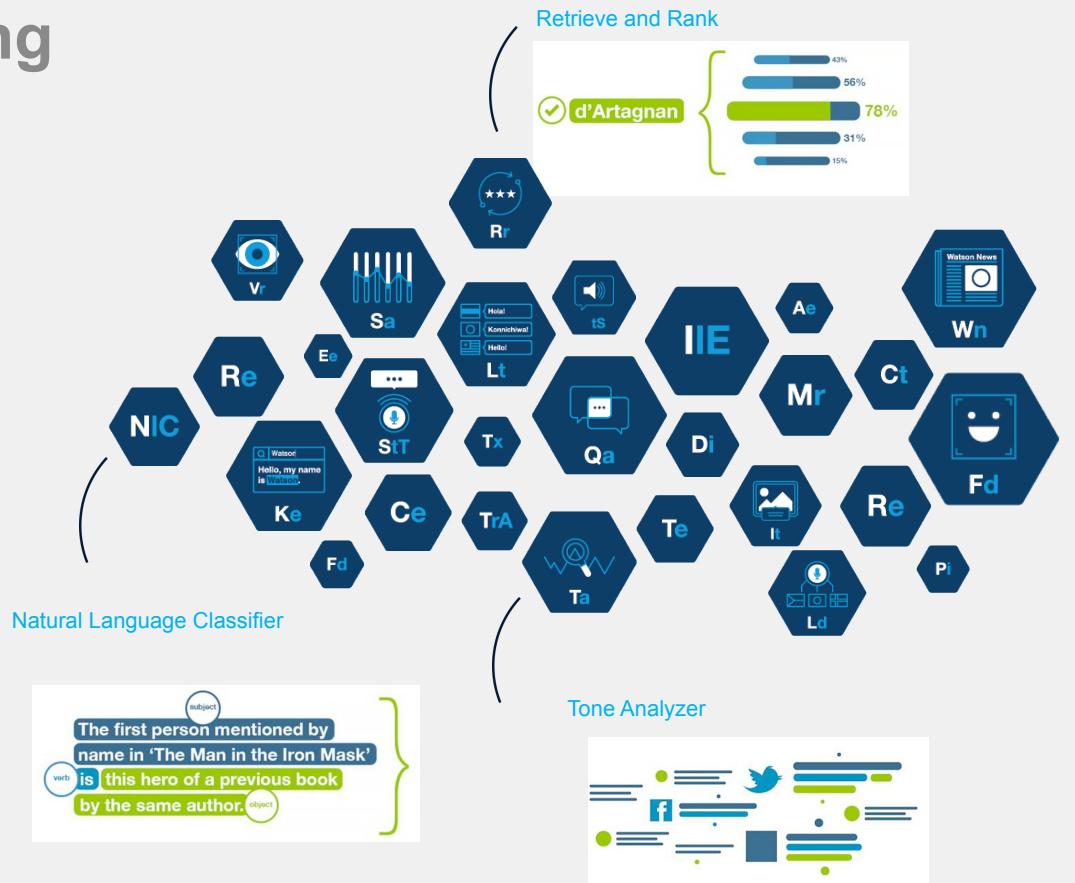
# Watson Services

IBM Watson

# Watson Platform Is Growing

Watson's APIs are the cognitive building blocks that harness our data.

Message Resonance	Entity Extraction
Face Detection	Tone Analyzer
Natural Language Classifier	Personality Insights
Speech to Text	Taxonomy
Text to Speech	Author Extraction
Language Translation	Concept Tagging
Language Detection	Relationship Extraction
Sentiment Analysis	Relationship Extraction
Retrieve and Rank	Feed Detection
Image Link Extraction	Keyword Extraction
Tradeoff Analytics	Visual Recognition
	Image Tagging
	Text Extraction



# Functional Clusters and Use Case Samples

## Data Enrichment, Augmentation & Tagging

- Enriching & organizing unstructured data

## Sense making, Interpretation, and Understanding

- Understanding unstructured data at scale

## Interact, Respond and Converse

- Engaging and interacting with humans; AI / bots / augmentation

## Translation

- Translating languages, intents, formats, ideas

## Visual Recognition and Image Analysis

- Seeing & interpreting visual data

## Assistant / Advisors

- Personal Assistant – virtual assistant to end user
- Agent Assistant – Help service/domain agents.
- Expert Advisor – Help users through an expert trained system
- Professional Advisor – Help professional with data, decisions, etc.
- *Q&A systems leveraging Retrieve & Rank, NLC, Alchemy Language, Conversations*

## Automation - Enable processing of routing tasks

- *Alchemy Language, Personality Insights, Tone, NLC*

## User/Customer Understanding and Recommendations

- Marketing Campaign Analytics Solution, Targeted Marketing, Customer Acquisition, Personal Connections, R
- *Sentiment Analysis, Visual Recognition, Personality Insights, Tone Analyzer, Alchemy Language*

## Data Insights / Analytics – extract features, meta data, tags, etc from information

- *Alchemy Language, AlchemyData News, Visual Recognition*

## Engagement / Robotics

- *Speech-to-Text, Text-to-Speech, Language Translation, Conversation*

# Natural Language Classifier

## Description

- Provides a developer the ability to integrate a natural language interface into their application that is able to interpret a user's question and result in a corresponding action.
- State of the art classifier tailored & tuned for short text
- Accepts short string of text and is able to classify text into predefined classes



## Use Cases / Benefits

- FAQ style Question and Answer
- Classify tweets/SMS
- Sentiment Analysis

# Retrieve and Rank



## Description

- Helps users find relevant information for a query by using a combination of search & machine learning algorithms to detect "signals" in the data.
- Ability for organizations to sync content between their Enterprise and the Cloud service
- Re-ranks search results based on a trained machine learning model

## Use Cases / Benefits

- Level 2 support agent needs to retrieve answers from large quantity of content
- Need to provide search results for "long-tail" questions that don't map directly to a known answer
- Integrate a traditional search function into an app
- Legal assistant looking for precedents for a case
- Patent search

# Conversation



## Description

- Enables Developers with Business users to create natural, human-like conversational experiences across all channels (e.g. mobile, messaging, robots, etc.)
- Combines Intents, Entities and Dialog into a seamless experience
- Minimal coding / Machine Learning expertise required

## Use Cases / Benefits

- Enables customers to self-serve on their terms with a simple experience.
- Delivers information and services with a consistent, on-brand and engaging experience
- Reduces costs through deflection of calls to Contact Centers

“ I’m frustrated, I haven’t been able to login into your online billing system ”

**Intent** Password Reset

**Entities** Online Billing System

**Emotional Tone** Anger

**Context** Bill Smith, 47,  
Gold Member, High Value

**Context** Mobile

# Alchemy Language



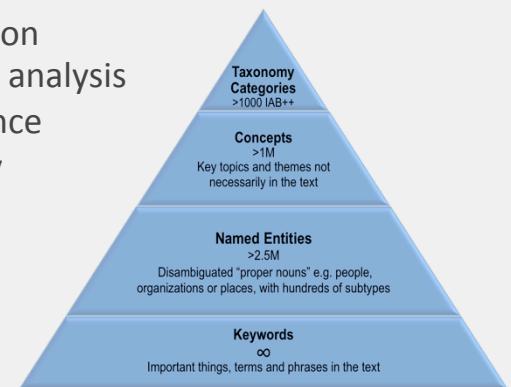
## Description

- Group of 12 services that offer text analysis through natural language processing.
- Deep learning techniques power services such as sentiment analysis, entity extraction, keyword extraction, content tagging and more for holistic insights.
- Custom models allow for domain specific knowledge to be trained.

Entity Extraction | Sentiment Analysis | Keyword Extraction | Concept Tagging | Relationship Extraction | Taxonomy Classification | Author Extraction | Language Detection | Text Extraction | Microformats Parsing | Feed Detection | Linked Data Support | Emotion Analysis

## Use Cases / Benefits

- Content Discovery
- Social Media Monitoring
- Business Intelligence
- Public Relations
- Ad Targeting
- Content aggregation
- Twitter sentiment analysis
- Business intelligence
- Content discovery



© 2016 International Business Machines Corporation

# Personality Insights

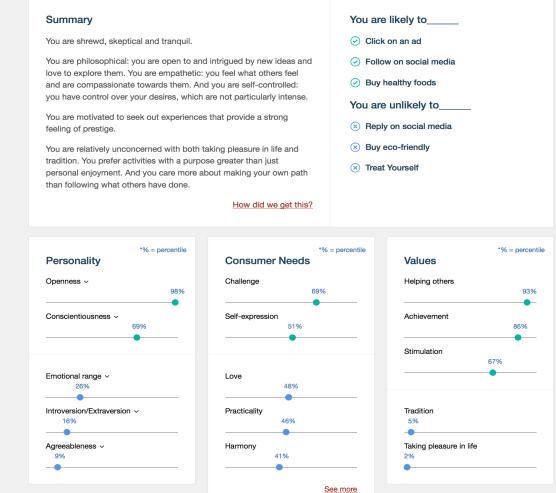


## Description

- Analyzes text to infer personality characteristics, needs and values to engage users in their own terms.
- Service takes words from a single author and outputs percentile scores for the author's needs, values and the Big Five personality traits.

## Use Cases / Benefits

- Market Segmentation, Campaign Management, or Promotion
- Customer Care
- Matching / Recruitment



# Tone Analyzer



## Description

- Uses linguistic analysis to detect and interpret emotional, social, and writing cues found in text.
- Derives several characteristics about each tone, such as conscientiousness, agreeableness, and openness for the social tone.
- Offers rhetorical suggestions for an author to improve the intended tone of their message.

## Use Cases / Benefits

- Analyzing email messages, presentations, or other customer communications before sending them
- Examining how readers might perceive your blog posts
- Help understand the potential impact of the word choice in any other text-based communications



# Alchemy Data News



## Description

- Alchemy Data News provides developers access to a hyper-relevant, curated dataset of news and blogs that has been enriched with Alchemy Language.
- Alchemy Data News indexes 250,000 news and blog articles every day enabling highly targeted search and trend analysis.
- Aggregates data from a variety of sources, enriches it with natural language processing and stores it for single-query consumption.

## Use Cases / Benefits

- Removes storage management burdens and system performance headaches Faster response time by caching results of already-processed content
- Gain new insights from past data by applying NLP functions on the entire dataset as a batch
- Signal detection
- News alerting
- Event detection ("acquisitions")
- Trending topics, concepts, people in the news

# Language Translation



## Description

- Statistical Machine Translation
- High-quality, **domain-specific** text translation from one language to another.
  - Language Identification (included) identifies the text input language.
- Model customization enables support for specific translation needs.

## Use Cases / Benefits

- Enables a help desk representative to assist international customers through a chat

### Today's Models:

- News (English to Spanish, French, Portuguese, Arabic)
- Conversational (English to Spanish, French, Portuguese, Arabic)
- Patent (Spanish, Portuguese, Chinese, Korean to English)

# Speech to Text



## Description

- Converts human voice into the written word, using machine intelligence to combine information about grammar and language structure with knowledge of the composition of the audio signal to generate a more accurate transcription.
- The transcription is continuously sent back to the client and retroactively updated as more speech is heard.
- Recognition models can be trained for different languages, as well as for specific domains.

## Use Cases / Benefits

- Mobile app speech interaction
- Enterprise search in media files
- Call center transcription
- Voice-control of devices or applications
- Dictation of emails

### Today's Models:

- English, Spanish, Japanese, Mandarin Chinese, Brazilian Portuguese, Modern Standard Arabic
- Broadband (live speech), Narrowband (telephony)

# Text to Speech



## Description

- Generates an audio file that has a verbal representation of the input text – complete with appropriate cadence and intonation.
- Supports a subset of SSML (Speech Synthesis Markup Language)

## Use Cases / Benefits

- Mobile apps.
- Assistance tools for the vision-impaired.
- Read texts / emails aloud.

### Today's Models:

- Brazilian Portuguese, English, French, German, Spanish, Italian and Japanese

# Visual Recognition



## Description

- Uses advanced visual recognition to see complex scenes in their entirety
  - without needing any textual clues
  - to understand objects and surroundings
- Includes an unmatched number of preset classifier and trained labels (2,000+).
- Ability to train custom visual classifiers

## Use Cases / Benefits

- Content recommendation
- Social media monitoring
- Ad targeting
- Facial recognition
- Organize image libraries

# Trade-off Analytics

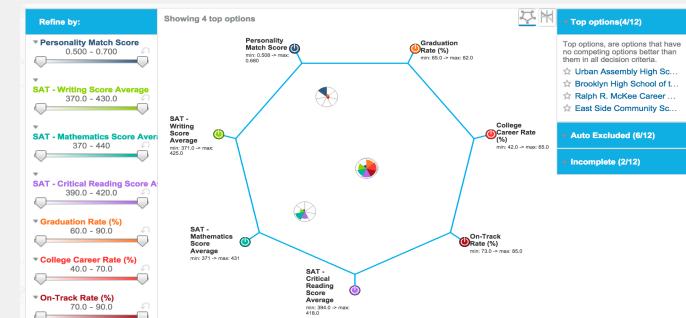


## Description

- Helps people make better choices while taking into account multiple, often conflicting, goals that matter when making that choice.
- Uses Pareto filtering techniques in order to identify the optimal alternatives across multiple criteria.
- Uses various analytical and visual approaches to help the decision maker analyze the tradeoffs.

## Use Cases / Benefits

- Investment plan selection
- Online shopping
- Medical treatment selection
- Call center – Call Scout
- Allow consumers to compare and contrast competitive products or services



© 2016 International Business Machines Corporation

# Document Conversion



## Description

- API based solution to convert documents with input format (PDF, Word, HTML) to output formats (HTML, Text, JSON needed for other Watson Services) synchronously or in a batch.
- Automatically converts customer documents in varied formats to the format necessary for consumption by Watson services

## Use Cases / Benefits

- Company bulk imports corporate training material documents to train Watson services such as Retrieve and Rank and NLC

# Getting Started

IBM Watson

# Get Started: Explore, Experiment, Build

1

Explore the Watson services

*Demos. Concept Videos. Docs. IBM.com/WatsonDeveloperCloud*

2

Experiment with the Labs

*ibm.biz/node-red-basics   ibm.biz/node-red-alchemy-vision-lab*

3

Build Apps

*Leverage Sample Code, App Gallery, github...*

*Join a Hackathon. Ask questions in the Forum.*

*Try out the APIs free of charge.*

# Bluemix Basics

## **Signing up for Bluemix**

- Sign up for Bluemix with a valid email address. You will need to read an email sent to the address used to register to activate your Bluemix account.
- Corporate subscriptions for Bluemix and Watson services are available

## **Working with Code**

- Most Bluemix projects are done out on GitHub. Many are open, some are not.

## **Not going broke**

- Bluemix allows limited use of the services for free for the first month. After 30 days, you may begin to incur charges – you will need a credit card to keep your Bluemix account active after the trial period expires.
- Only services in your Bluemix account incur charges – your account can work on projects in “corporate” account areas without YOU being charged.

# Where do I go?

## Exploring the Watson Services

- Each service has it's own tile in the Bluemix catalog – and if you hit the “View Docs” button, you will see API Explorers and demo applications

## Doing the Labs

- Do some of the demo applications provided for each service that interests you. These are short, and have code available on GitHub.
- There are some additional labs that you can do to familiarize yourself with Bluemix and Watson:
  - Bluemix developer tutorials (<https://hub.jazz.net/tutorials/>)
  - Getting started with Watson on Bluemix ([http://www.ibm.com/watson/developercloud/doc/getting\\_started/](http://www.ibm.com/watson/developercloud/doc/getting_started/))

## Building Applications

- Get a Bluemix account. Your usage is free for the first month.
- Be familiar with the Application Starter Kits (<http://www.ibm.com/watson/developercloud/starter-kits.html>). These will often jumpstart your application development, since they cover some of the basic cognitive use cases.

SDKs Available for Node, Java, Python, iOS, Unity - <https://github.com/watson-developer-cloud>

# Resources

# Resources

## Watson Services – Demos / Documentation / API Explorer

- <http://www.ibm.com/watson/developercloud/services-catalog.html>

## Watson Developer Cloud Github – SDKs and Sample Code

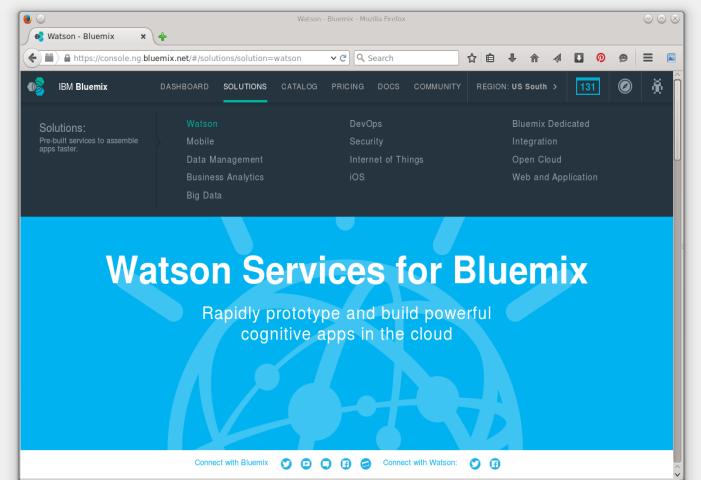
- <https://github.com/watson-developer-cloud>

## Sample Application

- Application Starter Kits: <http://www.ibm.com/watson/developercloud/starter-kits.html>

## Developer Handbooks

- Watson Developer Cloud Handbook: <https://ibm.box.com/s/nav52vt6q2xwib5zqwupwjf78mxtgems>
- Personality (PI) Handbook <https://ibm.box.com/s/6h8dxsc3pq5idtgehjb6fwh7vbejsvtc>
- Natural Language Classifier (NLC) Handbook: <https://ibm.box.com/s/rdlog2sue79178816s0rabkbi7ifu5vg>
- Retrieve and Rank (R&R) Handbook: <https://ibm.box.com/s/n0lqowt0v97nxb5mtei6qrbkkbt0t2dm>
- Cognitive Patterns: <https://ibm.box.com/s/x36om0iumv049f8ns4udh4fjkkb94rvt>



# Resources

## Support / Questions

- Stack Overflow - <http://stackoverflow.com/questions/tagged/ibm-watson-cognitive>
- DeveloperWorks Answers -  
<https://developer.ibm.com/answers/smart-spaces/25/watson.html>

## Watson Developer Community Slack

- [wdc-community.slack.com](https://wdc-community.slack.com/) |

## Blog and Other Links

- Watson Blog - <https://developer.ibm.com/watson/blog/>
- Resource Summary (Ryan Anderson) -  
[https://dreamtolearn.com/ryan/r\\_journey\\_to\\_watson/13](https://dreamtolearn.com/ryan/r_journey_to_watson/13)
- Watson Academy - <https://www.watson-academy.info/>

# Backup

IBM Watson

# Dialog



## Description

- Integrate a natural language step by step process within a given application. The process could range from a password reset to a more complex process like an insurance quote.
- Keeps track of user conversations, can gather user information and make decisions based on user profile information

## Use Cases / Benefits

- Execute a task on behalf of an end user (Resetting a password, activate a credit card, add a new service to an existing plan)
- Guiding a user through the registration of an event or application (register for a conference, apply for an insurance Quote etc)
- Personalizing information and helping guide users through a decision making process
- Disambiguating questions that lack necessary context

Order a phone with Watson

To begin, tell Watson that you'd like to order a phone!

I want a new phone

Do you want the new X-Phone ?

sure

© 2016 International Business Machines Corporation

A screenshot of a web-based Watson Dialog interface. It shows a conversation between a user and Watson. The user says 'I want a new phone'. Watson asks 'Do you want the new X-Phone ?'. The user replies 'sure'. A red stamp with the number '9' is visible on the right side of the screen.

# Concept Insights



## Description

- Service that maps input text to a Wikipedia-based concept graph.
- The service not only identifies explicit links when a concept is directly mentioned, but it also provides **implicit** links to concepts that are not directly mentioned.

## Use Cases / Benefits

- Increase user engagement in digital media by suggesting relevant related people and topics.
- Improve engagement on any external website

**DEPRECATED**

# Natural Language Classifier

Weather example: ‘will it be foggy’ is classified with 91% confidence that it is condition

- Training data does not contain the words “fog or foggy”. Yet the classifier is 91% confident that it’s a condition

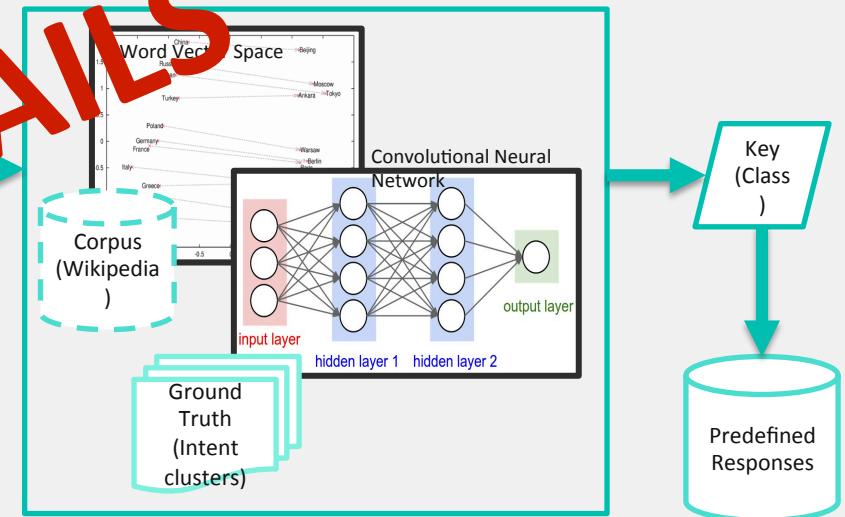
How does it do this?

- It’s a deep learning classifier
- It has ‘read’ wikipedia and ‘understands’ that fog and rain and sun are all similar
- It even understand “thunderstorms” are a condition
  - will there be any thunderstorms today? -> 98% confident in condition (again no mention of thunderstorm in the training data)

Research Dataset targeted at short text – TREC 10 (2000 Training Qs, 500 Validation Qs)

**DEEPER DIVE DETAILS**

Question	Intent Class
Does the Natural Language Classifier train using all of the data, or does it partition it in some way?	all_training_data_used
What kind of preprocessing does the classifier perform on its input?	all_training_data_used
Does the classifier train using all of the data, or does it partition it in some way?	all_training_data_used
does the classifier train on all the data or does it hold some out?	all_training_data_used
does nClassifier perform random sampling to the training datasets?	all_training_data_used
Where can I find documentation on the Natural Language Classifier API?	api_documentation
Where can I find documentation?	api_documentation
What is the API for the classifier?	api_documentation
Where can I find documentation for the NL classifier API?	api_documentation
Where can I find REST API documentation for the NL classifier?	api_documentation
Where can I find documentation for the classifier API?	api_documentation
How do I access the classifier API?	api_documentation
How many instances of the Natural Language Classifier can I create and what does it cost?	api_documentation
How will classifier usage be priced after GA?	billing_and_usage
How much does it cost to use the Classifier?	billing_and_usage
How much does it cost?	billing_and_usage
How much does it cost?	billing_and_usage
How many classifiers can I train simultaneously?	billing_and_usage
Is there a limit to the number of classifiers I can create?	billing_and_usage



## Retrieve and Rank

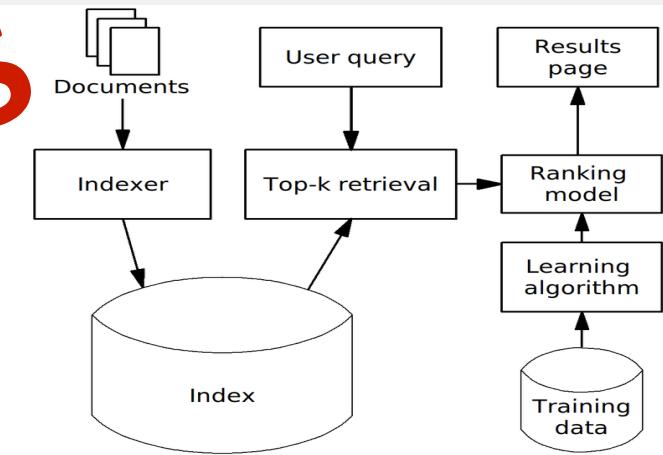
Based on well-known research output – Learn to Rank

- Machine learned ranking using supervised or reinforcement learning.
- Training data consist of items with partial ordering using numerical “relevance” score
- Ranker will apply an ordering to new/unseen lists in way that is similar to ranking in training data.

Different from general classification

- Optimization is list-wise
  - All list items considered when learning model parameters (rather than each item separately)
- Allow for graded relevance
  - Multiple levels (0, 1, ...) represent preferential ordering

**DEEPER DIVE DETAILS**



## Retrieve and Rank - Example

Simple Support Center Example

Solr contains 3 key fields: Short Description, Long Description, and Tech Notes

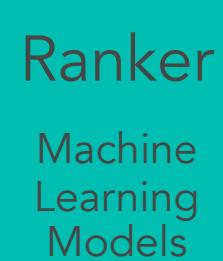
"What does the X221 error message mean?"



**DEEPER DIVE DETAILS**

	Field 1 - Short Description			Field 2 - Long Description			Field 3 - Tech Note		
	Feature Score 1	Feature Score 2	Feature Score 3	Feature Score 1	Feature Score 2	Feature Score 3	Feature Score 1	Feature Score 2	Feature Score 3
Answer ID #1	0.53	0.10	0.61	0.70	0.40	0.15	0.82	0.09	0.32
Answer ID #2	0.45	0.02	0.34	0.57	0.39	0.83	0.22	0.94	0.21
Answer ID #3	0.95	0.15	0.86	0.82	0.02	0.23	0.13	0.01	0.52
Answer ID #4	0.29	0.01	0.68	0.21	0.83	0.19	0.68	0.27	0.15
Answer ID #5	0.16	0.62	0.11	0.93	0.06	0.93	0.81	0.68	0.89
Answer ID #6	0.24	0.06	0.69	0.78	0.31	0.34	0.70	0.03	0.73
Answer ID #7	0.29	0.57	0.76	0.46	0.71	0.61	0.29	0.34	0.53
Answer ID #8	0.81	0.87	0.07	0.14	0.45	0.11	0.74	0.25	0.37
Answer ID #9	0.79	0.24	0.44	0.13	0.29	0.52	0.82	0.79	0.51
Answer ID #10	0.42	0.60	0.59	0.96	0.64	0.77	0.90	0.36	0.72
.....									

Answer ID #7
Answer ID #2
Answer ID #8
Answer ID #4
Answer ID #1
Answer ID #25
Answer ID #33
Answer ID #18
...



\*IBM Watson Solr Plugin contains a set of 'scorers' that scores each data attribute (Feature Score)

# Alchemy Language: Entities & Concepts

## Entity Extraction

- What is it:** AlchemyAPI's named entity extraction is capable of identifying people, companies, organizations, cities, geographic features and other typed entities from your HTML, text or web-based content.
- How it works:** Entity extraction can add a wealth of semantic knowledge to your content to help you quickly understand the subject of the text. It is one of the most common starting points for using natural language processing techniques to enrich your content.

## Concept Tagging

- What is it:** The concept tagging API is capable of making high-level abstractions by understanding how concepts relate, and can identify concepts that aren't necessarily directly referenced in the text.
- How it works:** AlchemyAPI employs sophisticated text analysis techniques to concept tag documents in a manner similar to how humans would identify concepts. If an article mentions CERN and the Higgs boson, it will tag Large Hadron Collider as a concept even if the term is not mentioned explicitly in the page. By using concept tagging you can perform higher level analysis of your content than just basic keyword identification.

DEEPER DIVE

DETAILS



# Alchemy Language: Keywords & Relationships

## Keyword Extraction

- What is it:** Keywords are the important topics in your content and can be used to index data, generate tag clouds or for searching. AlchemyAPI's keyword extraction API is capable of finding keywords in text and ranking them. The Keyword extraction API works on URLs, HTML documents and plain text. Just like every other feature, AlchemyAPI automatically detects the language of the content and then performs the appropriate analysis.
- How it works:** AlchemyAPI's keyword extraction algorithm employs sophisticated statistical algorithms and natural language processing technology to analyze your content and identify the relevant keywords. Keyword extraction is supported in over a half-dozen different languages, enabling even foreign-language content to be categorized and tagged.

DEEPER DIVE DETAILS



# Alchemy Language: Sentiment and Taxonomy

## Sentiment Analysis

- **What is it:** AlchemyAPI's sentiment analysis API provides easy-to-use mechanisms to identify the positive or negative sentiment within any document or webpage. The sentiment analysis API is capable of computing document-level sentiment, sentiment for a user-specified target, entity-level sentiment, quotation-level sentiment, directional-sentiment and keyword-level sentiment. These multiple modes of sentiment analysis provide for a variety of use cases ranging from social media monitoring to trend analysis.
- **How it works:** AlchemyAPI's sentiment analysis algorithm looks for words that carry a positive or negative connotation then figures out which person, place or thing they're referring to. It also understands negations (i.e. "this car is good" vs. "this car is not good") and modifiers (i.e. "this car is good" vs. "this car is really good"). The sentiment analysis API works on documents large and small, including news articles, blog posts, product reviews, comments and Tweets.

## Taxonomy

- **What is it:** AlchemyAPI automatically categorizes your text, HTML or web-based content into a hierarchical taxonomy. Using complex statistics and natural language processing technology, the taxonomy API can classify your content into its most likely topic category up to five levels deep.
- **How it works:** Deeper levels allow you to classify content into more accurate and lucrative subsegments. For instance, an application focused on identifying content discussing personal lending practices can narrow its classification into sub topics that target decisions with finer resolution.

**DEEPER DIVE DETAILS**

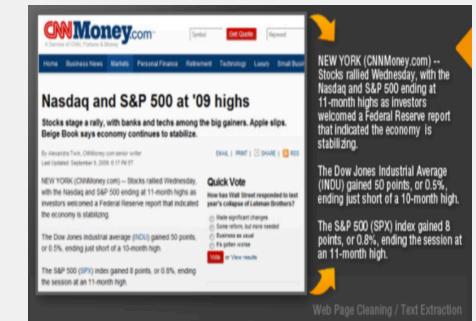


# Alchemy Language: Text Extraction and Language Detection

## Text Extraction

- What is it:** Can automatically extract the important information from a webpage, removing navigation links, advertisements and other undesired content. Use AlchemyAPI's text extraction API to just focus on the key text to improve website indexing, increase contextual advertising relevancy and simplify analysis.
- How it works:** Text extraction can remove the embedded links within the important content, which makes it possible to use for web-crawling applications.

DEEPER DIVE DETAILS



## Language Detection

- What is it:** provides a robust language detection facility capable of detecting the language of any text, HTML or web-based content. AlchemyAPI identifies more languages than any other text analysis service at extremely high rates of accuracy.
- How it works:** With language detection, you can easily categorize or filter any content based on the language it was written in.

# Alchemy Language: Feed Detection, Linked Data, and Microformats

## Feed Detection

Feeds are often embedded into websites to allow visitors and automated feed readers to access syndicated content. AlchemyAPI's feed detection API can find the feeds within webpage and return the links.

- Use feed detection to discover new content, including blog posts, news articles and comment streams

## Linked Data

Linked Data is a method of exposing, sharing and connecting data on the web via dereferencable URLs. Linked Data aims to extend the Web with a data commons by publishing various open datasets as RDF on the Web and by setting RDF links between data items from different data sources

- if Apple, Inc. is identified as an entity in your content using Linked Data makes it possible to get the following information: number of Apple employees, number of worldwide Apple locations, or last year's Net Operating Income. If AlchemyAPI finds a member of Congress in the content, linked data can access his or her picture, birth date, party affiliation, alma Mater or successor.

## Microformats Parsing

Microformats are included in the HTML of webpages to add semantic information. These microformats allow the webpage to be more easily scanned and processed automatically through software. They are typically used for contact information, geographic coordinates, license information and similar information. AlchemyAPI's microformat parsing API can automatically detect and parse the microformats embedded within a webpage.

- Use microformat parsing to enhance webpage categorization and indexing and to perform content discovery tasks.

DEEPER DIVE  
DETAILS