

Welcome to:

IBM's Watson as a Cognitive System



Unit objectives

After completing this unit, you should be able to:

- Understand the concepts of IBM Watson with other search engines
- Learn about preparing Watson for commercial applications
- Gain knowledge on Watson's software architecture
- Gain an insight into components of DeepQA architecture
- Learn about Building the Watson Corpus
- Understand the concepts of question analysis, semantic analysis
- Gain knowledge on question classification, hypothesis generation
- Understand scoring and confidence estimation

Define Watson

What is Watson?

A cognitive system that enables a new partnership between people and computers that enhances and scales human expertise.

Core Capabilities:

1. NLP – Natural Language Processing
2. Big Data Analysis and Machine Learning
3. Evidence Based Reasoning



https://youtu.be/Y_cqBP08yuA

<http://www.ibm.com/smarterplanet/us/en/ibmwatson/>

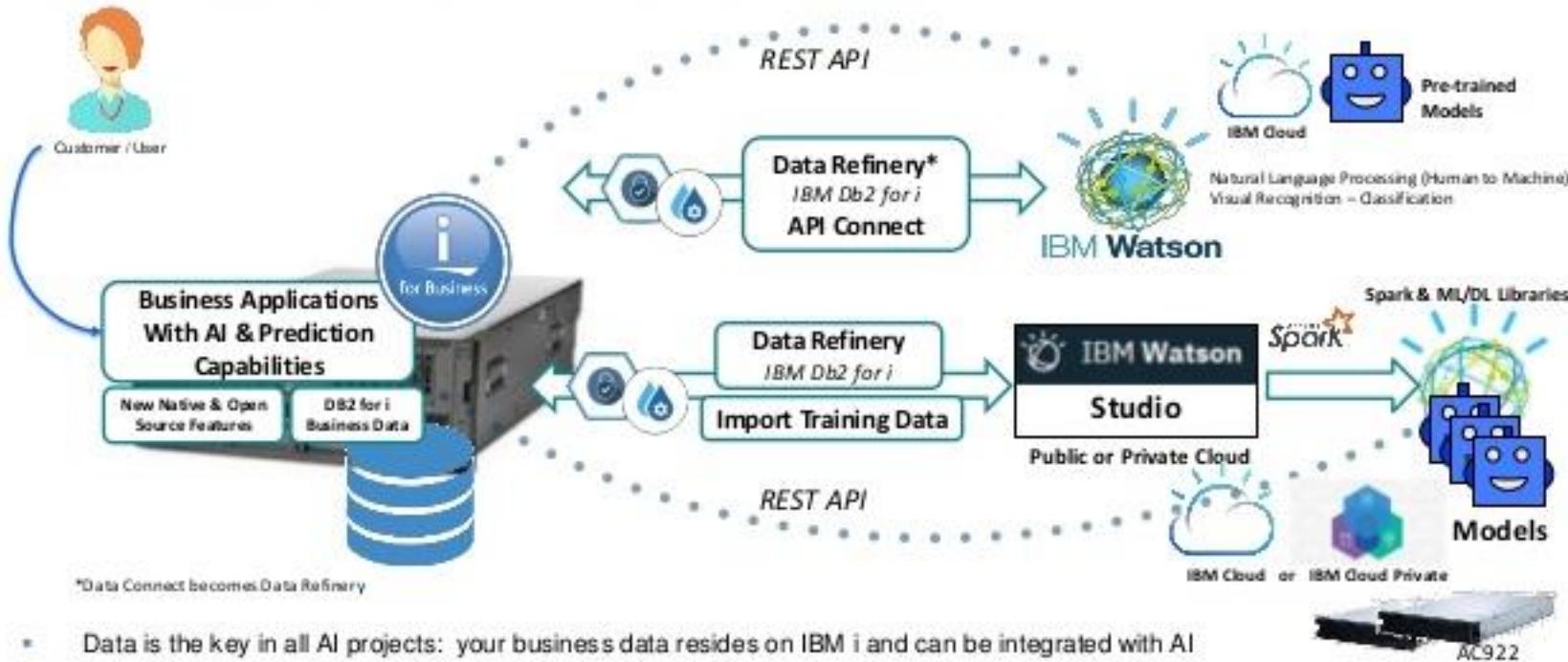
Figure: IBM Watson definition

Source: <https://images.app.goo.gl/idgA1cqUySvhYMNDA>

Current and future applications

IBM i & Artificial Intelligence

Approximate (AI) & precise (Transactional) computing together



- Data is the key in all AI projects: your business data resides on IBM i and can be integrated with AI
- Use pre-trained & customizable models with IBM Watson (Developer Cloud) services in IBM Cloud
- Build your own use case & business specifics models with IBM Watson Studio - IBM Cloud / on premises (DSX Local w/ Cloud Private)

Figure: Current and future applications

Source: <https://images.app.goo.gl/iT8YeNrNNAEFJHhm8>

Watson healthcare services

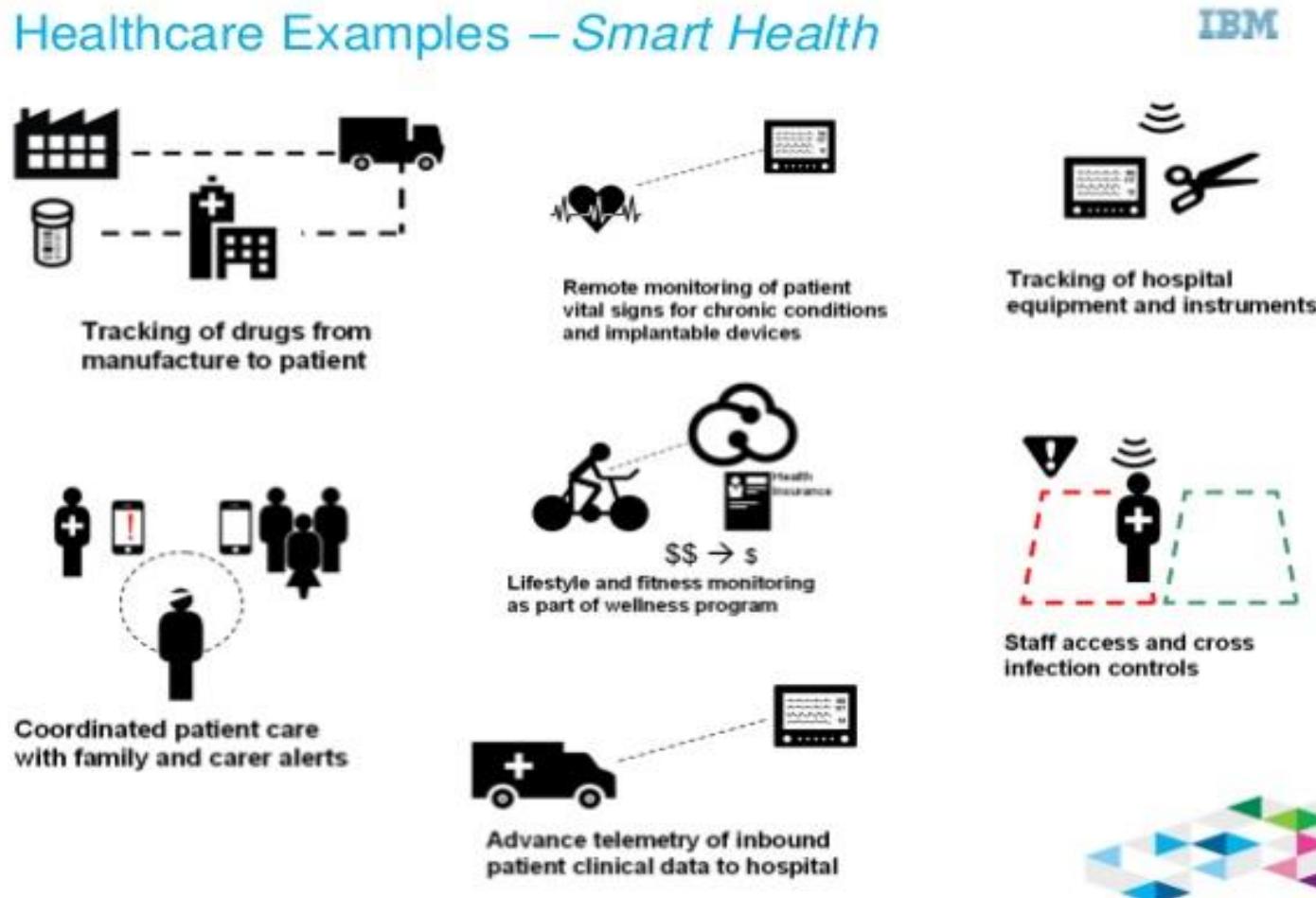
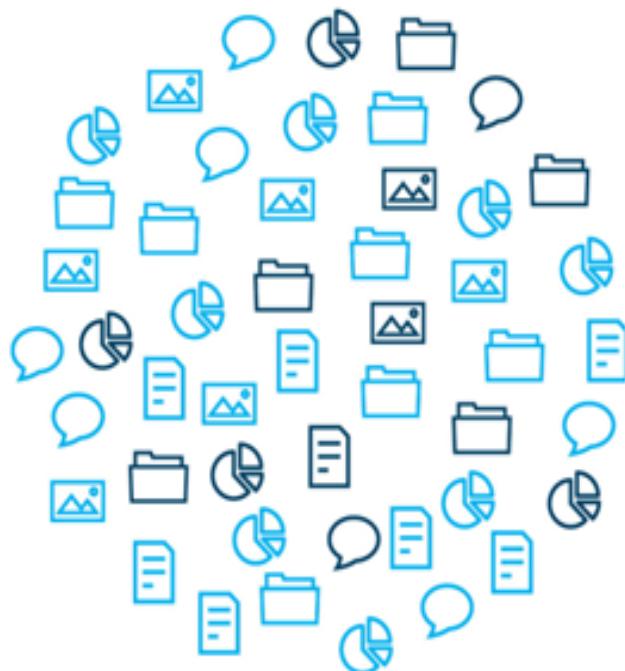


Figure: Watson Healthcare services
Source: <https://images.app.goo.gl/mgKViJxfrkk7VbDe6>

Cognitive intelligence by IBM Watson



Data, information, and expertise create the foundation

Examples include:

Analyst reports	Newspapers
Tweets	Blogs
Wire tap transcripts	Wiki
Battlefield docs	Court rulings
Emails	International crime database
Texts	Stolen vehicle data
Forensic reports	Missing persons data

Figure: Watson relies on collections of data and information

Source: <https://images.app.goo.gl/WwD75qbsLWXcmrMR6>

APIs for Watson: Watson build

The screenshot shows the IBM Cloud interface with the 'Watson' service selected in the sidebar. The main area displays various Watson services categorized under 'Infrastructure' and 'Platform'. Each service card includes a brief description, a small icon, and two buttons: 'Lite' and 'IBM'. A red circle with the number '5' is visible in the bottom right corner.

Infrastructure		Platform	
Compute	Discovery	Language Translator	Machine Learning
Storage	Add a cognitive search and content analytics engine to applications.	Translate text from one language to another for specific domains.	IBM Watson Machine Learning - make smarter decisions, solve
Network	Lite IBM	Lite IBM	Lite IBM
Security			
Containers			
VMware			
Platform			
Boilerplates	Natural Language Understanding	Personality Insights	Speech to Text
APIs	Analyze text to extract meta-data from content such as concepts,	The Watson Personality Insights derives insights from transactional data.	Low-latency, streaming transcription
Application Services	Lite IBM	Lite IBM	Lite IBM
Blockchain			
Cloud Foundry Apps			
Data & Analytics			
DevOps	Text to Speech	Tone Analyzer	Visual Recognition
Finance	Synthesizes natural-sounding speech from text.	Tone Analyzer uses linguistic analysis to detect three types of tones.	Find meaning in visual content! Analyze images for scenes,
Functions	Lite IBM	Lite IBM	Lite IBM
Integrate			
Internet of Things	Watson Assistant (formerly Conversation)	Watson Studio	
Mobile	Add a natural language interface to your application to automate	Embed AI and machine learning into your business. Create custom	
Network			
Security			
Watson			

Figure: Watson services in IBM Bluemix catalog

Source: <https://images.app.goo.gl/WoodwigvTSuJrKkv8>

IBM Watson applied to industries, businesses, and science



IBM ICE (Innovation Centre for Education)



Figure: Cognitive computer applies to industry, businesses, and science

- Ecommerce, also known as electronic commerce or internet commerce, refers to the buying and selling of goods or services using the internet, and the transfer of money and data to execute these transactions.
- Ecommerce is often used to refer to the sale of physical products online, but it can also describe any kind of commercial transaction that is facilitated through the internet.
- Whereas e-business refers to all aspects of operating an online business, ecommerce refers specifically to the transaction of goods and services.
- Types of Ecommerce models:
 - Business to Consumer (B2C)
 - Business to Business (B2B)
 - Consumer to Consumer (C2C)
 - Consumer to Business (C2B)

IBM Watson education

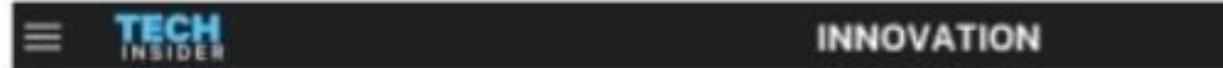
IBM ICE (Innovation Centre for Education)

- Watson element is designed to transform the classroom by providing critical insights about each student: Demographics, strengths, challenges, optimal learning styles, and more which the educator can use to create targeted instructional plans, in real-time. Gone are the days of paper-based performance tracking, which means educators have more face time with students, and immediate feedback to guide instructional decisions.
- Provide insightful information on how a class is doing overall and gives teachers a tool that encourages meaningful interactions with their students during class.
- Give teachers a centralized place to add and see helpful information about each student.
- Work with an easy to use tracker to help teachers mark student progress aligned with learning standards.

IBM Watson financial services

- Benefits of IBM Watson financial services:
 - Use modern systems to track and manage your finances, accounts and ledgers.
 - Deliver the standard financial reporting required as well as advanced analysis to identify issues and opportunities.
 - Track fixed assets effectively for financial accounting, preventative maintenance, and theft deterrence
 - Detect financial fraud using data analysis.
 - Manage developing specifications, buying, contracts inventory control, receiving and stores, and more.
 - Identify, assess and prioritize risk and ensure you meet the necessary governance requirements.

IBM Watson health (1 of 2)



IBM's Watson computer can now do in a matter of minutes what it takes cancer doctors weeks to perform

BI

Lauren F Friedman and Reuters, Business Insider
© May 5, 2015, 3:29 PM 160



FACEBOOK



LINKEDIN



TWITTER

Fourteen US and Canadian cancer institutes will use International Business Machines Corp.'s Watson computer system to choose therapies based on a tumor's genetic fingerprints, the company said on Tuesday, the latest step toward bringing personalized cancer treatments to more patients.

The advertisement features the text "IBM Watson for Oncology" prominently at the top. Below it, a descriptive paragraph reads: "Get oncologists the assistance they need to make more informed treatment decisions. Watson for Oncology analyzes a patient's medical information against a vast array of data and expertise to provide evidence-based treatment options." At the bottom of the ad, the words "Welcome to IBM" are visible on a banner.

IBM wants its Watson supercomputer to treat diabetes, fight cancer, and monitor pregnancy



AI doctors will become 'as ubiquitous as stethoscopes' (Wired UK)

One of the biggest problems facing doctors isn't patients' injuries or illnesses — it's the sheer quantity of data. But supercomputers are already helping to make sense of it.

<http://www.ibm.com/smarterplanet/us/en/ibmwatson/watson-oncology.html>

Figure: IBM Watson health

IBM Watson health (2 of 2)

- Special solutions are provided by Watson Health:
 - IBM Watson for genomics.
 - IBM Watson for drug discovery.
 - IBM Watson health patient engagement.
 - IBM Watson for oncology.
 - IBM Watson care manager.

Watson in the insurance industry

- Watson's ability to analyze structured and unstructured data, reference the right policy information and input documents, and then make insightful recommendations, can help employees determine whether a claim is eligible and what percentage of the claim should be paid. So with Watson, employees can make better decisions and get better results faster.
- How insurance is working with Watson:
 - A customer files a claim.
 - Employees collect relevant information about the incident.
 - Employees input this information into their system.
 - Watson works behind the scenes with employees to help determine claim eligibility and the percentage that should be paid.
 - Employees use a tab to access Watson's recommendation and decide how to best proceed with payment.
 - Employees assess claims more efficiently, saving hours of time each month.

IBM Watson Internet of Things (IoT)

The IBM Watson IoT Platform

Everything you need to Innovate with IoT

IBM Watson IoT Platform Connect

Attach, Collect, & Organize, Device Management, Secure Connectivity, Visualization

IBM Watson IoT Platform Information Management

Storage & Archive, Metadata Management, Reporting, Streaming data, Parsing and Transformation, Manage unstructured data
Weather APIs

IBM Watson IoT Platform Analytics

Predictive, Cognitive, Real-time, and Contextual

IBM Watson IoT Platform Risk Management

Security Analytics, Data Protection, Auditing/Logging, Firmware Updates, Key/Certificate Mgmt, Org Specific Security.
Blockchain (Beta)



Figure: IBM IOT

Source: <https://images.app.goo.gl/w3i54U1jT5oWTBAr>

IBM Watson cognitive video

- Watson video enrichment uses Watson's industry-leading cognitive abilities to analyze audio, textual, and visual data within multimedia content—deeply enriching the value of every video asset by creating automated data sets that are more detailed and searchable than is currently possible for large libraries of video.
- Watson analyzes multimedia content and builds deep, easily searchable metadata packages for every asset in your library.
- Content enriched by Watson is inherently more searchable, leading to improved discovery and increased usage.
- Watson goes well beyond today's metadata standards, automatically recognizing and tagging keywords, concepts, themes, objects and much more.
- Watson can even determine the dominant emotion of a video like joy, sadness, or anger.

IBM Watson for cyber security

- The volume of security incidents and available threat data far exceed the capacity of even the most skilled security professional. Enter Watson for cyber security. It augments a security analyst's ability to identify and understand sophisticated threats, by tapping into unstructured data (ex., blogs, websites, research papers) and correlating it with local security offenses.
- IBM QRadar advisor with Watson combines the cognitive capabilities of Watson and the industry leading QRadar security analytics platform to uncover hidden threats and automate insights, revolutionizing the way security analysts work.

Watson use cases

- OmniEarth: Cognitive computing shows Earth imagery water consumption patterns.

The page features the IBM logo and "Computer Services". The main content area includes a large aerial photograph of a residential neighborhood. A quote from Chelsea Minton, Senior Sales Engineer, OmniEarth, Inc., is displayed: "Analyzing things at the parcel-by-parcel level represents a paradigm shift for this industry." The "Business challenge" section discusses water conservation in drought-stricken California. The "Cognitive transformation" section explains how OmniEarth uses cognitive technology to analyze aerial images for water consumption patterns. The "Business benefits" section highlights 40x faster image processing and higher capacity for analyzing terrain. The "OmniEarth, Inc." section emphasizes cognitive computing for water consumption patterns. The "Solution components" section lists IBM Bluemix, IBM Watson Developer Cloud, and IBM Watson Visual Recognition. Social sharing icons for Twitter, LinkedIn, Google+, and Facebook are at the bottom right.

Business challenge
Water conservation is a top concern in drought-stricken California. The state imposed water restrictions to regulate consumption but relied on broad yearly or multi-year averages to understand usage levels, set targets and educate the public. OmniEarth, Inc. knew it could help the state monitor usage through aerial images, but its speed was limited by the fact that humans still had to interpret and tag the images manually before they could be analyzed.

Cognitive transformation
OmniEarth uses cognitive technology to recognize topographical features in unstructured aerial images—including lawns and agricultural zones—giving water districts insight into dynamic patterns of water consumption and the effects of weather and local initiatives. The system was trained to differentiate a pool from a pond, for example, with analysis that shows how water is used, how much can be saved and where to focus on customer education.

Business benefits

40x faster
Image processing than was possible using manual methods

Higher capacity
for analyzing terrain
on a massive scale, creating new business opportunities worldwide

Actionable insight
into satellite imagery
on a highly granular scale

OmniEarth, Inc.

Cognitive computing can show water consumption patterns from earth imagery

Founded in 2014 and headquartered in Arlington, Virginia, OmniEarth, Inc. builds scalable solutions for processing, clarifying and fusing large amounts of satellite and aerial imagery with other data sets. The results have a broad range of applications, from pipeline monitoring to precision agriculture and resource management.

Solution components

- IBM® Bluemix®
- IBM Watson™ Developer Cloud
- IBM Watson Visual Recognition

Share this

[Twitter](#) [LinkedIn](#) [Google+](#) [Facebook](#)

Figure: OmniEarth

Source: <https://images.app.goo.gl/BLVm2NHgmWriXxRRA>

Watson demonstrations link

<https://speech-to-text-demo.mybluemix.net/>

<https://natural-language-understanding-demo.mybluemix.net/>

<https://language-translator-demo.mybluemix.net/>

<https://conversation-demo.mybluemix.net/>

<https://tone-analyzer-demo.mybluemix.net/>

<http://visual-recognition-demo.mybluemix.net/>

<https://your-celebrity-match.mybluemix.net>

<https://facebook-photo-analyzer.mybluemix.net/>

<https://tradeoff-analytics-demo.mybluemix.net/>

<https://personality-insights-livedemo.mybluemix.net/>

<http://relationship-extraction-demo.mybluemix.net/>

<http://natural-language-classifier-demo.mybluemix.net/>

Figure: Watson demonstrations link

Self evaluation: Exercise 19

- To continue with the training, after learning the various steps involved in cognitive analytics and machine learning, it is instructed to utilize the concepts of machine learning algorithms to perform the following activity.
- You are instructed to write the following activities using Python code.
- Exercise 19: Speech Analytics

DeepQA system architecture (1 of 2)

DeepQA Architecture

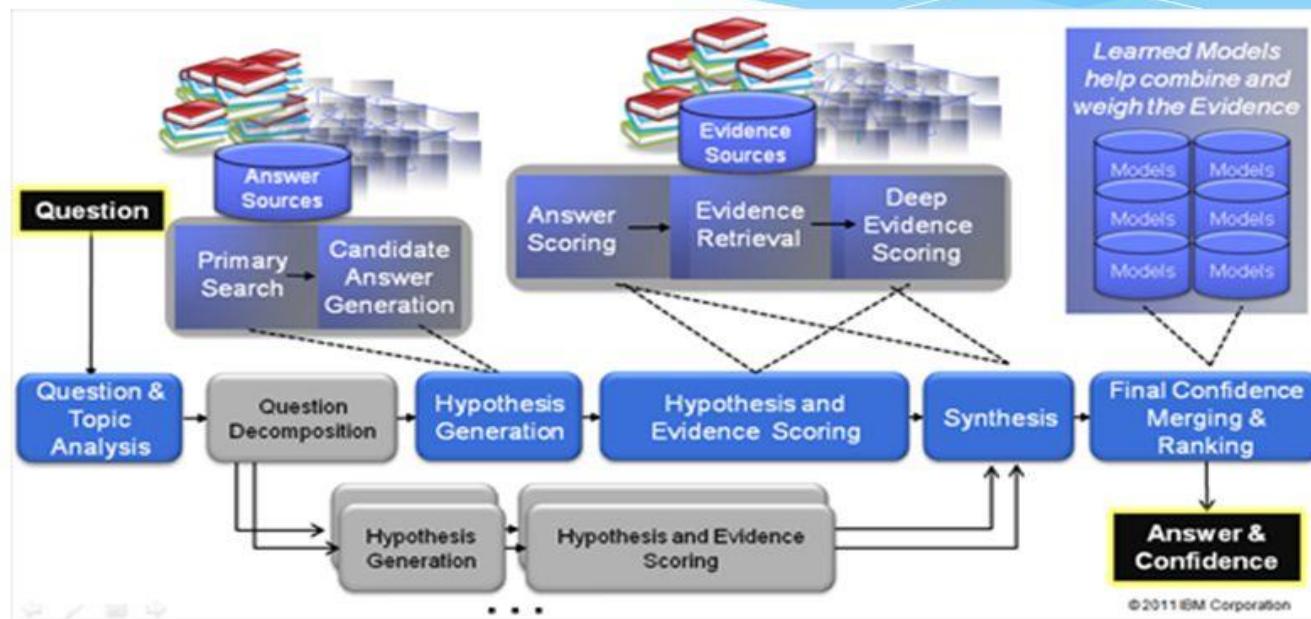


Figure: DeepQA Architecture

Source: <https://researcher.watson.ibm.com/>

DeepQA system architecture (2 of 5)

- Answers and evidence sources: Structured and unstructured data.

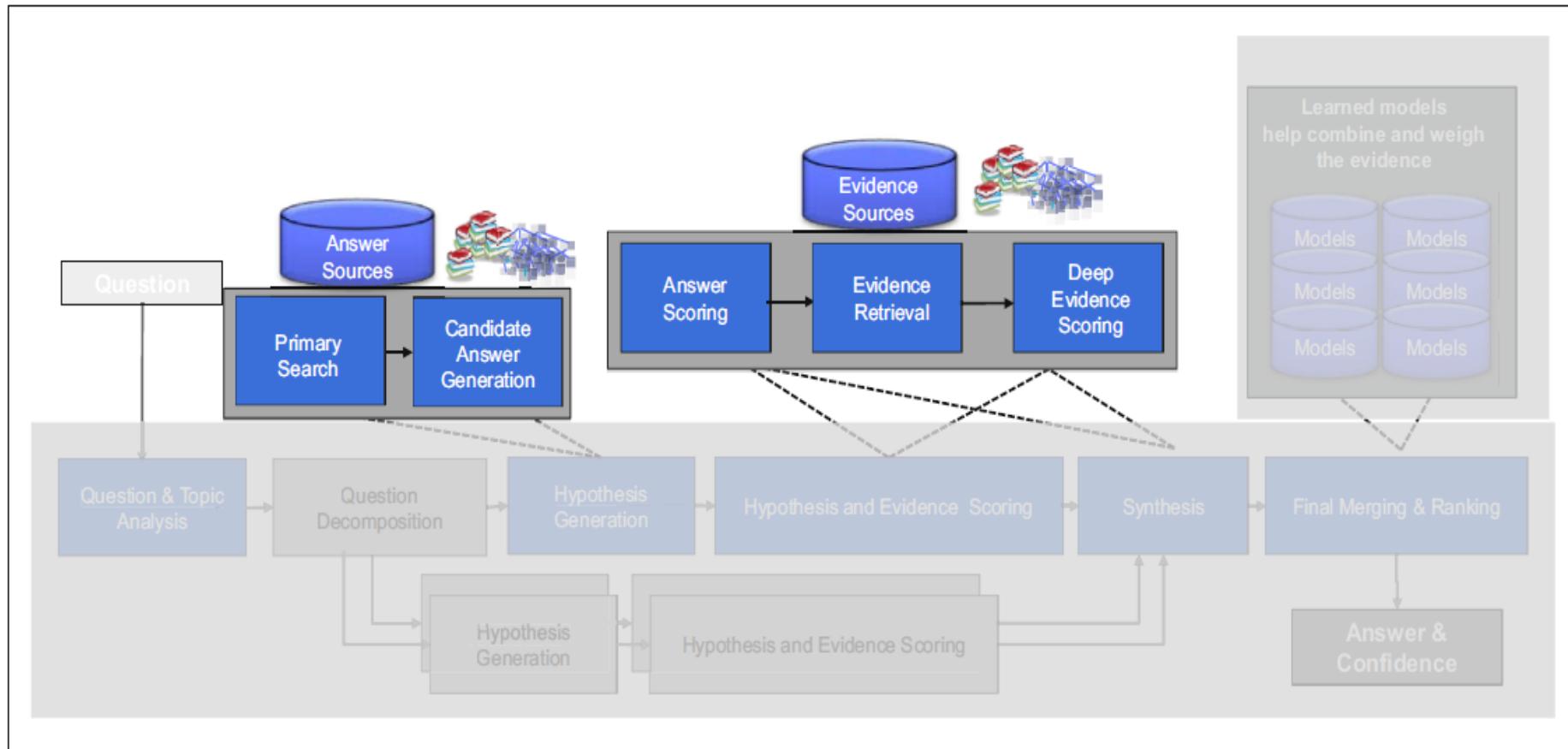


Figure: Answers and evidence sources-structured and unstructured data

DeepQA system architecture (3 of 5)

- Learned models help combine and weigh the evidence.

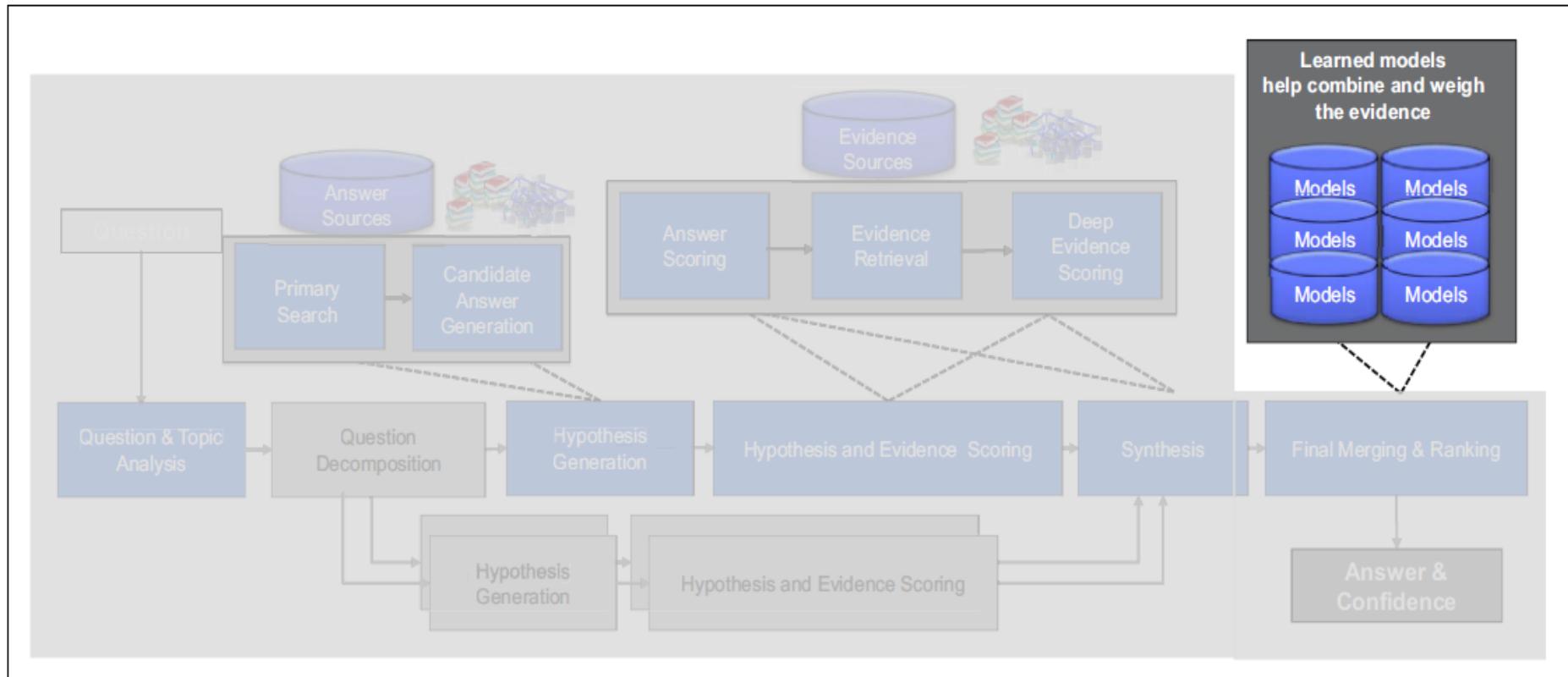


Figure: Learned models help combine and weigh the evidence

DeepQA system architecture (2 of 2)

- Minimum DeepQA pipeline.

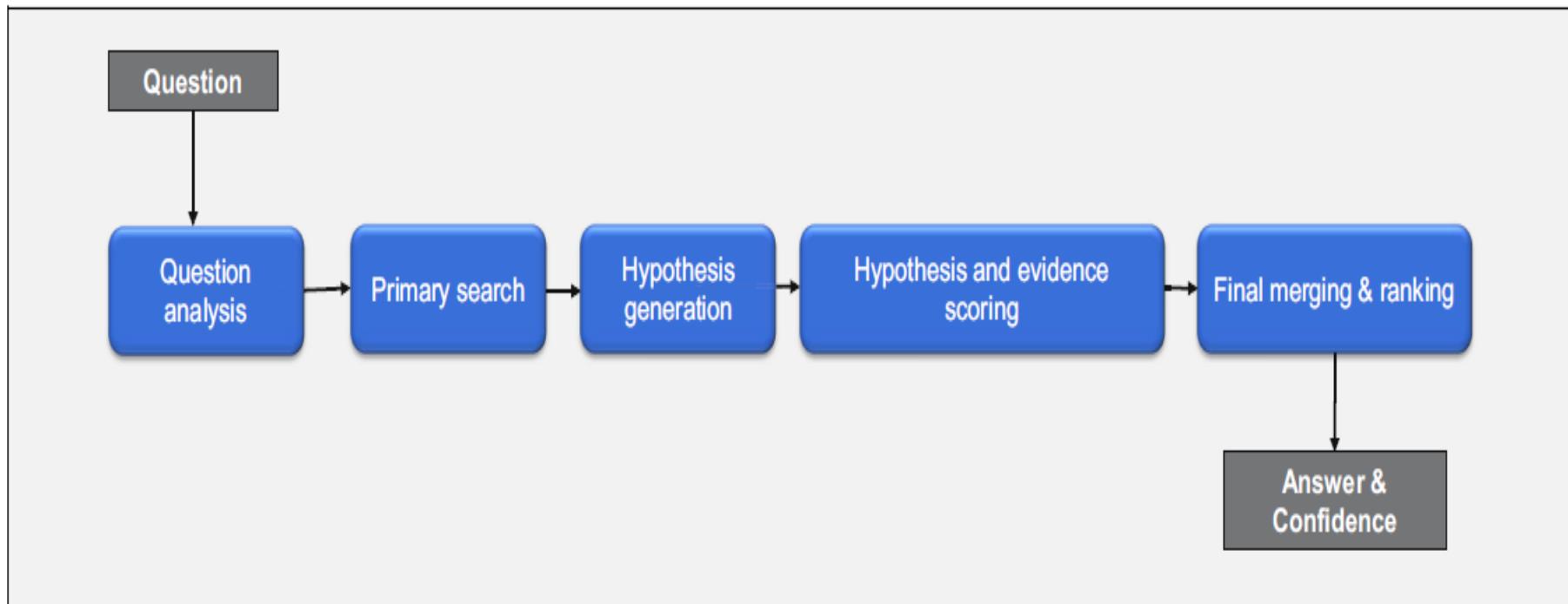


Figure: Minimum DeepQA pipeline

DeepQA system architecture (5 of 5)

- Example of how Watson answers a question following the DeepQA architecture.

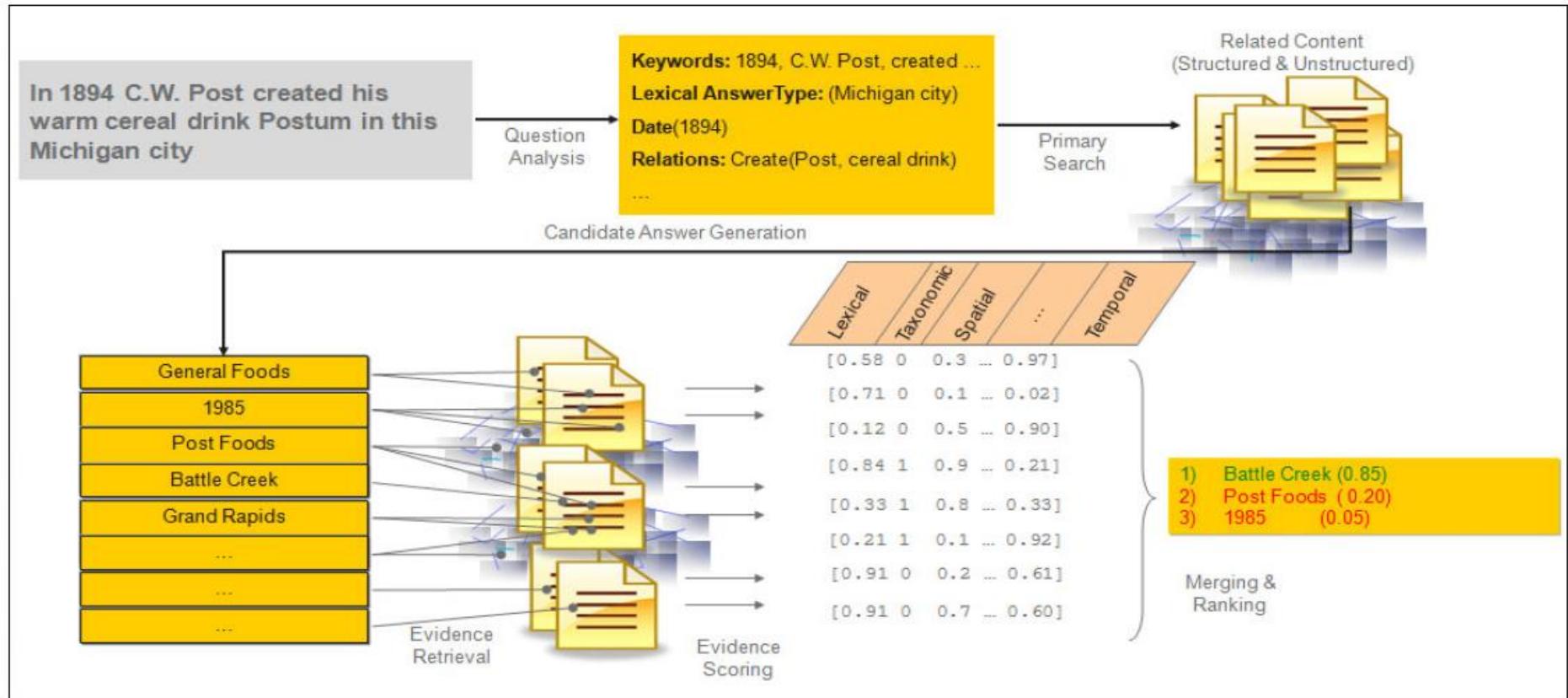


Figure: Example of how Watson answers a question following the DeepQA architecture

Question analysis

- Primary search.

- The keywords (1894, C.W. Post, created, warm, cereal, drink, Postum, Michigan, city) are used to search over millions of documents to find relevant hits.
- 55 documents are found, and 30 passages are found.

General Foods

General Foods Corporation was a company whose direct predecessor was established in the USA by Charles William Post as the Postum Cereal Company in 1895. The name General Foods was adopted after an unsuccessful acquisition. In 1922, General Foods was acquired by Philip Morris Company for \$14.5 billion, the largest acquisition up to that time. In December, 1969 Philip Morris acquired Kraft, Inc., and in 1970 combined the two food companies as Kraft General Foods (KGF). "General Foods" was dropped from the corporate name in 1975 and now exists only as part of a brand name for a flavored coffee-based beverage. [General Foods](#)

Post Foods

Post Foods, LLC, also known as Post Cereals formerly Postum Cereal, was founded by C.W. Post in 1895 with the first Postum, a "cereal beverage", developed by Post in Battle Creek, Michigan. The first cereal, Grape-Nuts, was developed in 1897. Post has its headquarters in the Bank of America Plaza in downtown St. Louis, Missouri. The Postum Cereal company, after acquiring Jell-O gelatin in 1929, Bakers' chocolate in 1931, Maxwell House coffee in 1932, and other food brands, changed its name to General Foods Corporation in 1939. General Foods was acquired by Philip Morris Companies in 1970.

In 1995, Philip Morris merged General Foods with Kraft Foods, which it had acquired in 1987 to form the Kraft General Foods division. The core brands of Postum were acquired in 1998. In 1995, Kraft General Foods was reorganized and renamed Kraft Foods. On November 15, 2007, Kraft announced that it would spin off Post Cereals and merge its business with Kellogg Holdings.²² That merger was completed August 4, 2008.²³ The official website of this acquisition is [Kraft Foods](#).²⁴

Will Keith Kellogg

Britannica Concise Encyclopedia: Will Keith Kellogg

(born April 7, 1860, Battle Creek, Mich., U.S.—died Oct. 6, 1951, Battle Creek) U.S. breakfast-cereal manufacturer and philanthropist. After working with his brother John Harvey Kellogg, a physician and health-food pioneer, he founded (1906) the W.K. Kellogg Co. to manufacture dry breakfast cereals, cornflakes being its sole product in the early years. It soon became a leading U.S. producer of these and other convenience foods. By the early 21st century, its annual sales exceeded \$9 billion. The W.K. Kellogg Foundation is one of the country's largest philanthropic institutions.

For more information on [Will Keith Kellogg](#), visit [Britannica.com](#).

Indri Passage Search

Lucene Passage Search

Document Search Results

Rank	Title
0	General Foods
1	Battle Creek
2	Post Foods
3	Will Keith Kellogg
4	Breakfast Cereal
5	John Harvey Kellogg
6	C. W. Post
7	Kellogg Company
8	Postum
	...

Passage Search Results

Rank	Passage
0	C.W. Post came to the Battle Creek sanitarium to cure his upset stomach. He later created Postum, a cereal-based coffee substitute
1	The caffeine-free beverage mix was created by The Postum Cereal Company founder C. W. Post in 1895 and produced and marketed by Postum Cereal Company as a healthful alternative to coffee
2	1895: In Battle Creek, Michigan, C.W. Post made the first POSTUM, a cereal beverage. Post created GRAPE-NUTS cereal in 1897, and POST TOASTIES corn flakes in 1908
3	1854 C. W. Post (Charles William) was born. He founded the Postum Cereal Co. in 1895 (renamed General Foods Corp. in 1922) to manufacture Postum cereal beverage
4	The company was incorporated in 1922, having developed from the earlier Po stum Cereal Co. Ltd., founded by C.W. Post (1854-1914) in 1895 in Battle Creek, Mich. After a number of experiments, Post marketed his first product-the cereal beverage called Postum-in 1895
5	...

Figure: Primary search

Generation of hypotheses (1 of 2)

- A hypothesis is an educated "guess" about the source of the outbreak I.
- Generating hypotheses enables the investigators to test these hypotheses in an analytic study m The success of the investigation depends upon the quality of the hypotheses.

Generation of hypotheses (2 of 4)

- Hypothesis and evidence scoring.
- Candidate answers that pass the soft filtering threshold undergo a rigorous evaluation process that involves in steps.
 - Evidence retrieval:
 - Gathers additional supporting evidence for each candidate answer, or hypothesis.
 - E.g., Passage search: gathering passages by adding CA to primary search query.
 - Scoring:
 - Deep content analysis - includes many different components, or scorers, that consider different dimensions of the evidence.

Generation of hypotheses (3 of 4)

- Clue and passage.
- eClue: In May 1898, Portugal celebrated the 490th anniversary of this explorer's arrival in India.
- Passage: On the 27th of May 1498, Vasco da Gama landed in Kappad Beach
- That clue and passage will provide great support for the answer Vasco Da Gama.

Generation of hypotheses (2 of 2)

- Final merging and ranking.

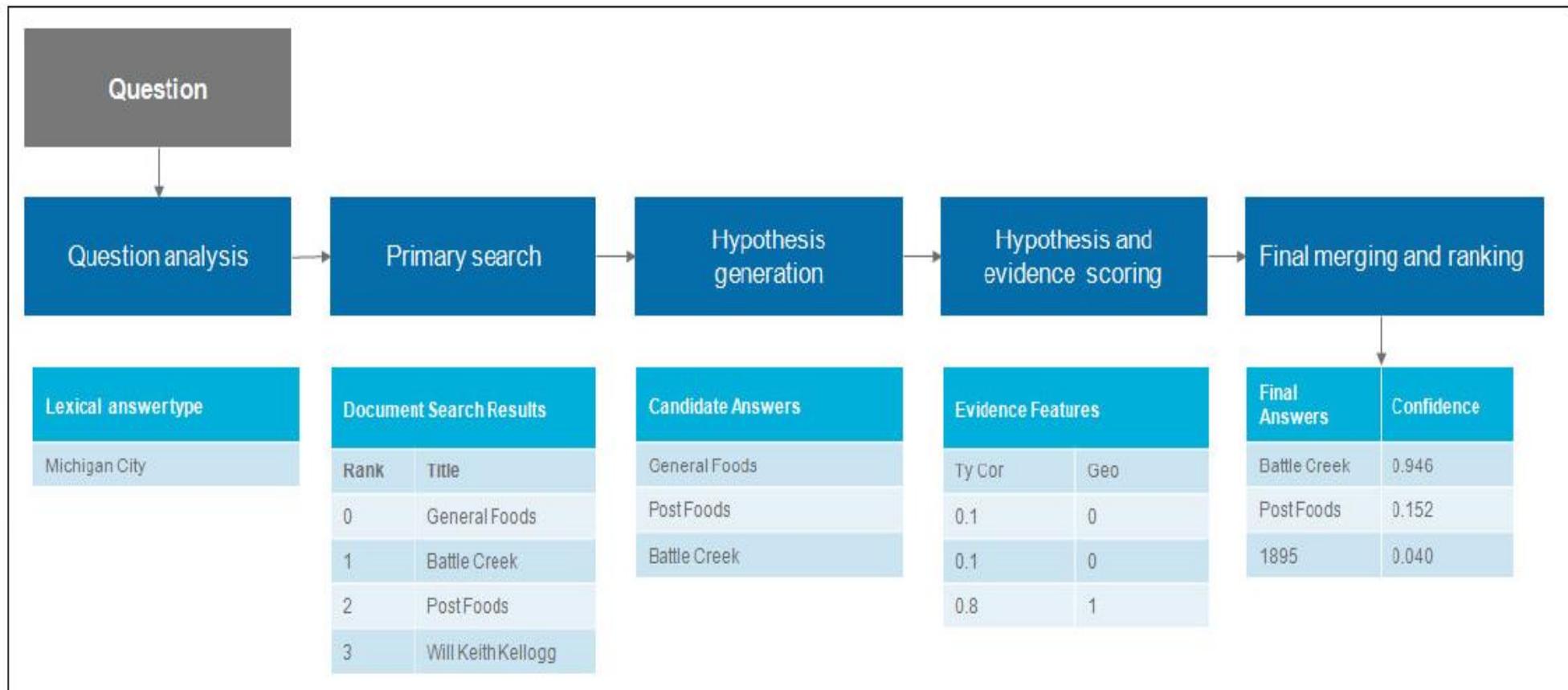


Figure: Final confidence scores

Evolution from DeepQA to Watson developer cloud



IBM ICE (Innovation Centre for Education)



The powerful capabilities and outcomes brought on by cognitive computing.



The ability to build business in code with the API economy.



The proliferation of different types of data.

Figure: Disruption is fueled by three forces

Source: <https://images.app.goo.gl/e6nGwcvPfVCmvGnA7>

Why commercialize Watson (1 of 2)



Figure: Data and industry sectors

Why commercialize Watson (2 of 2)

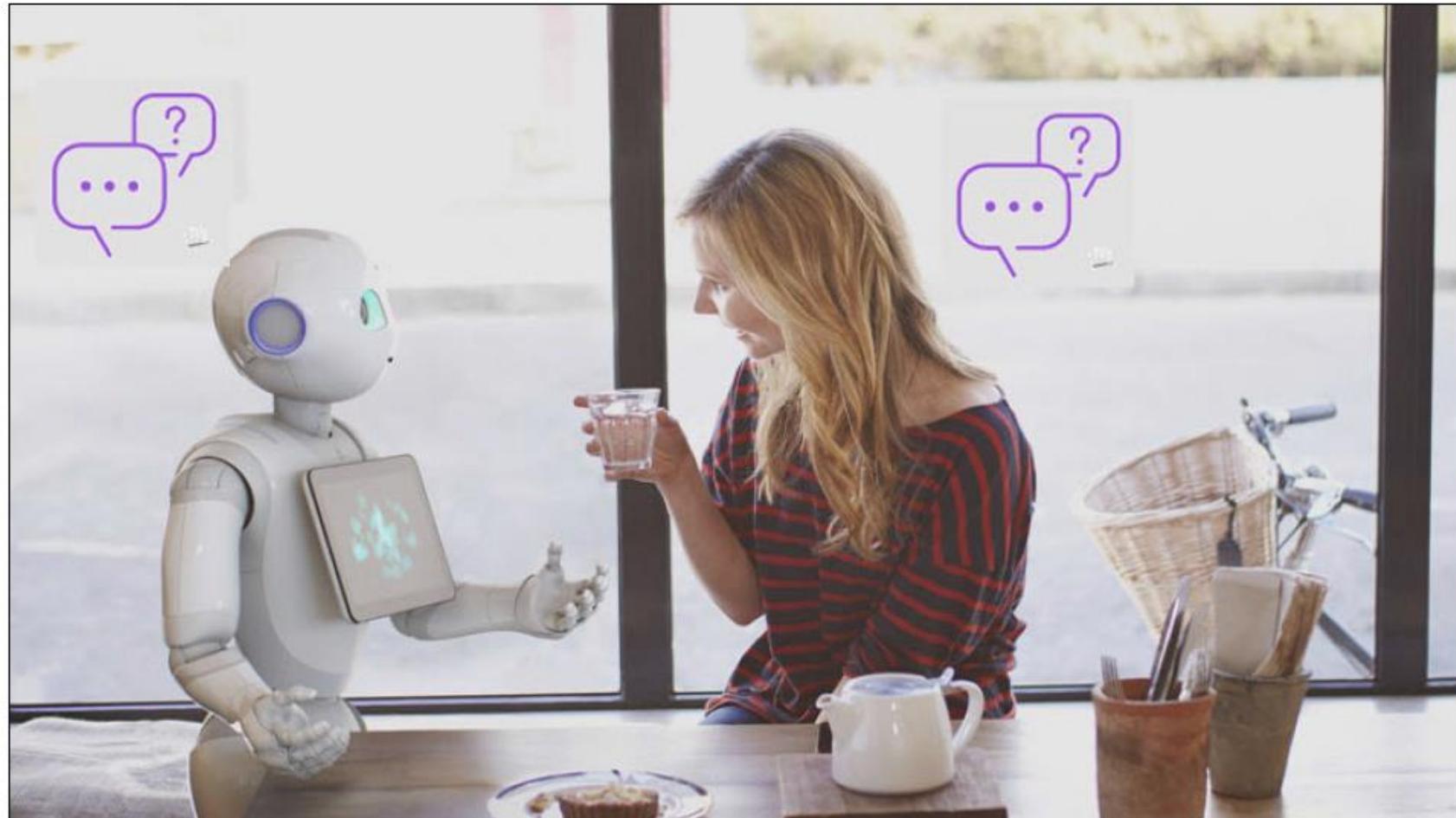


Figure: Engagement and conversation

Cloud development for Watson (1 of 2)

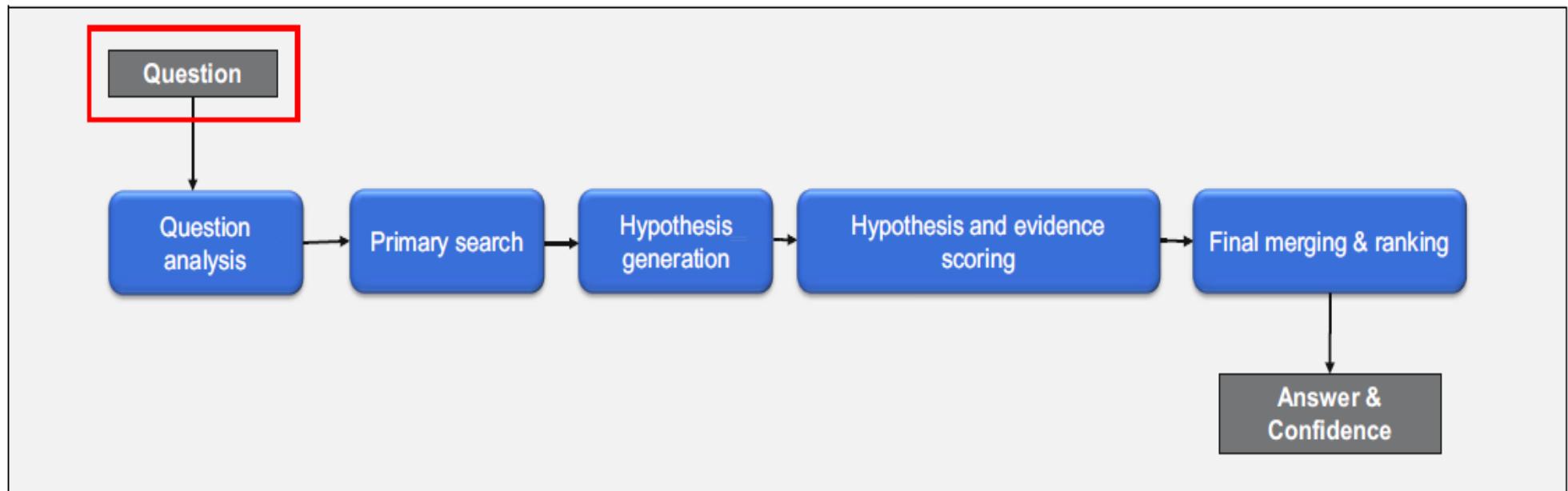
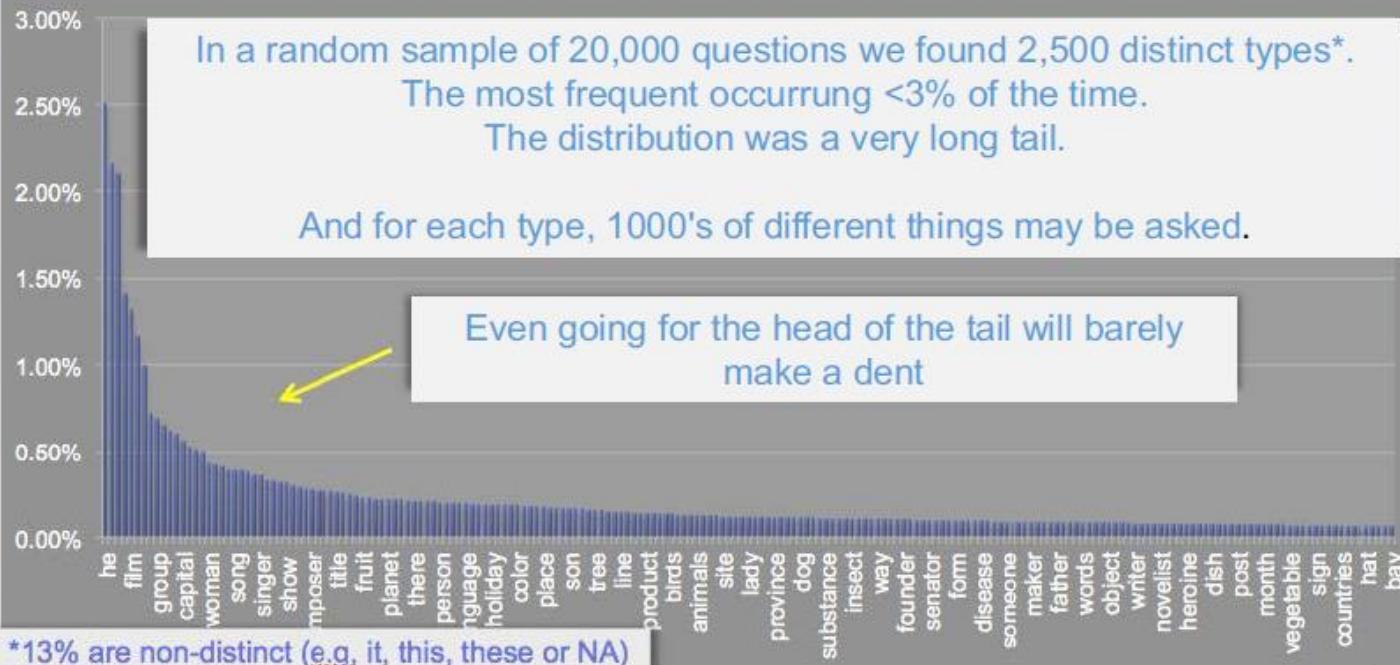


Figure: DeepQA pipeline the question

Cloud development for Watson (2 of 3)

We do NOT attempt to anticipate all questions and build databases.

We do NOT try to build a formal model of the world.



Our focus is on reusable NLP technology for analyzing vast volumes of *as-is* text.
Structured sources (DBs and KBs) provide background knowledge for interpreting the text.

Figure: Broad domain (long tail distribution)

Cloud development for Watson (2 of 2)

- The question distribution from a typical customer service use-case.

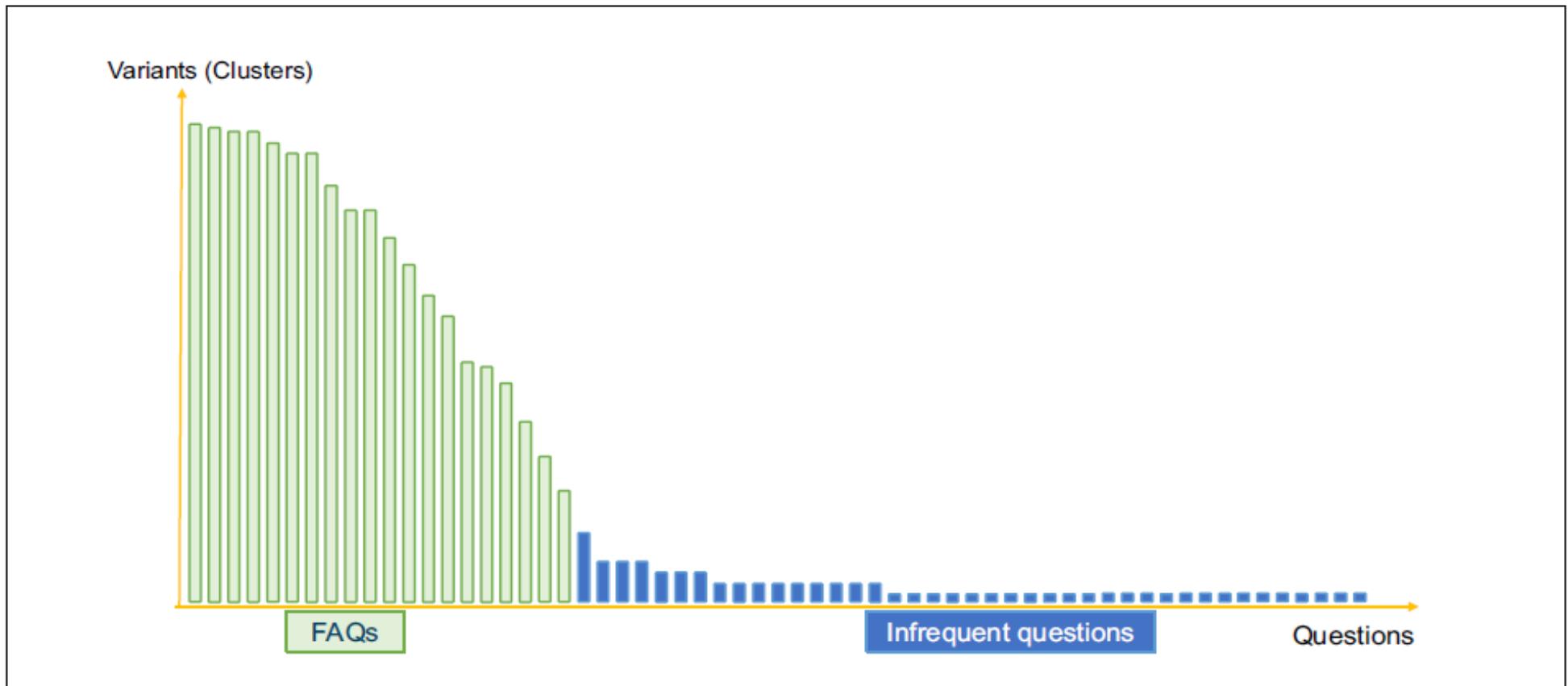


Figure: The question distribution from a typical customer service use-case

Evolution of question analysis (1 of 2)

- Leverage state of the art Deep Learning techniques to derive intent
- Understand varied language and concepts
- Learn over time based on usage
- Handcrafted rules unable to scale and do not benefit from data

“ I forgot my password...
How do I get a new password?
Can't log into your site...
My login isn't working, please help...
Can you reset my password? ”

Intent = Password Reset

Figure: Understand the customer's intent

Evolution of question analysis (2 of 2)

- Detect intent, entities and emotion.

“ I’m frustrated, I haven’t been able to log into your online billing system ”	Intent Password Reset
--	-----------------------

Figure: Extract other key information from a question: detect intent

“ I’m frustrated, I haven’t been able to log into your online billing system ”	Intent Password Reset Entities Online Billing System
--	---

Figure: Extract other key information from a question: Entities

“ I’m frustrated, I haven’t been able to log into your online billing system ”	Intent Password Reset Entities Online Billing System Emotional Tone Anger
--	---

Figure: Extract other key information from a question: Emotion

Evolution of question analysis (3 of 3)

- Context:

" I'm frustrated, I haven't been able to log 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
--	---

Figure: Extract other key information from a question: Context

Question	Answer
How do I reset my password?	— Dialog —> Guide the user through a set of steps
Someone has stolen my credit card.	— Deflect —> Transfer to human agent
Where is the nearest store?	— Map —> Application launches map with directions
I need to pay my outstanding invoice.	— App Nav. —> Bring user to pay bill screen
Can I pay my bills using my credit card?	— Info. Retrieval —> Bring back an answer

Figure: Take action: Responses come in different forms

Evolution of the DeepQA question analysis step (1 of 2)



IBM ICE (Innovation Centre for Education)

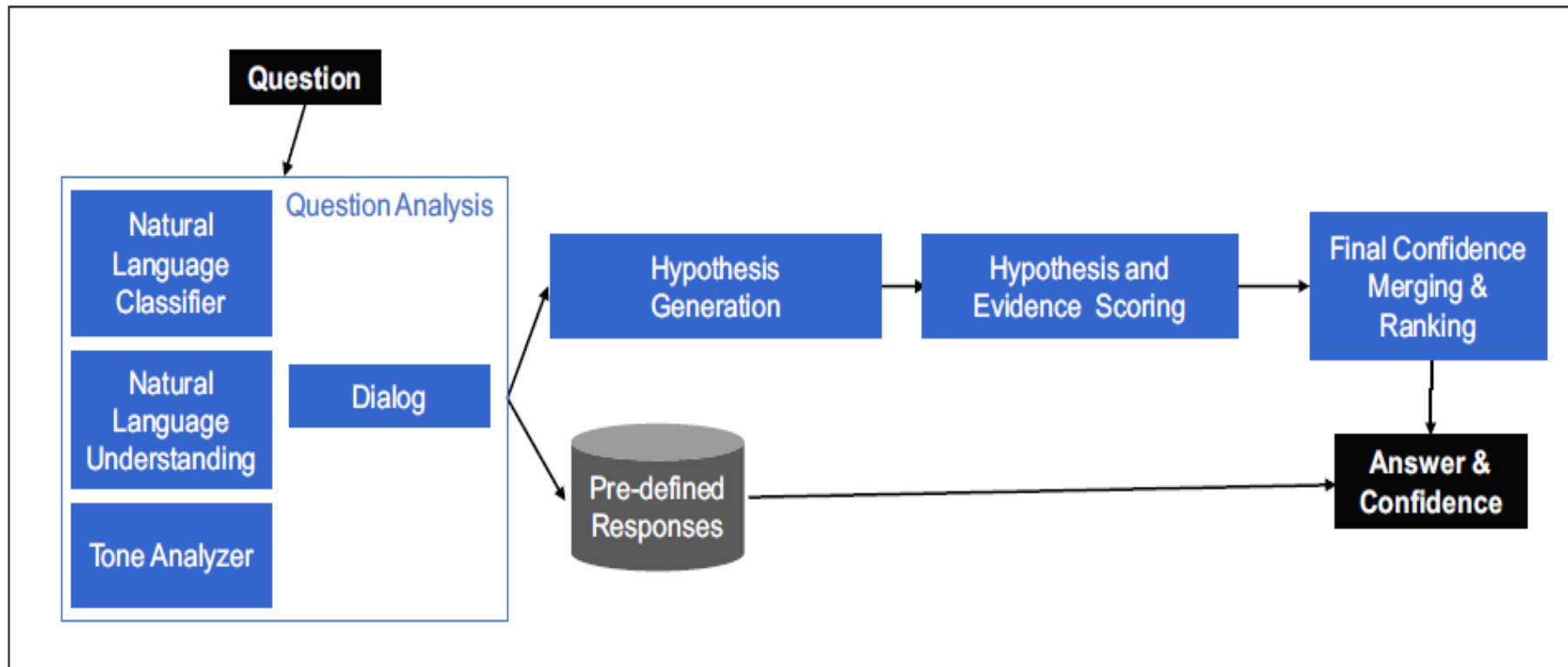


Figure: Evolution® of the DeepQA question analysis step configurable microservices

Evolution of the DeepQA question analysis step (2 of 4)



IBM ICE (Innovation Centre for Education)

"I reset my password but now I'm getting a sporadic error and only when I log in from my desktop app."

Corpus Service desk tickets, product manuals, KBs

Figure: Build a long tail solution for complex question: Corpus

"I reset my password but now I'm getting a sporadic error and only when I log in from my desktop app."

Corpus Service desk tickets, product manuals, KBs

Ingestion Convert documents into relevant answer units

Figure: Build a long tail solution for complex question: Ingestion

Build a long-tail solution for this Question

"I reset my password but now I'm getting a sporadic error and only when I log in from my desktop app."

Corpus Service desk tickets, product manuals, KBs

Ingestion Convert documents into relevant answer units

Search and Score Index answer units and score against input query

Rank Use machine learning to re-rank candidate answers

Figure: Primary search and score and rank

Evolution of the DeepQA question analysis step (2 of 2)



IBM ICE (Innovation Centre for Education)

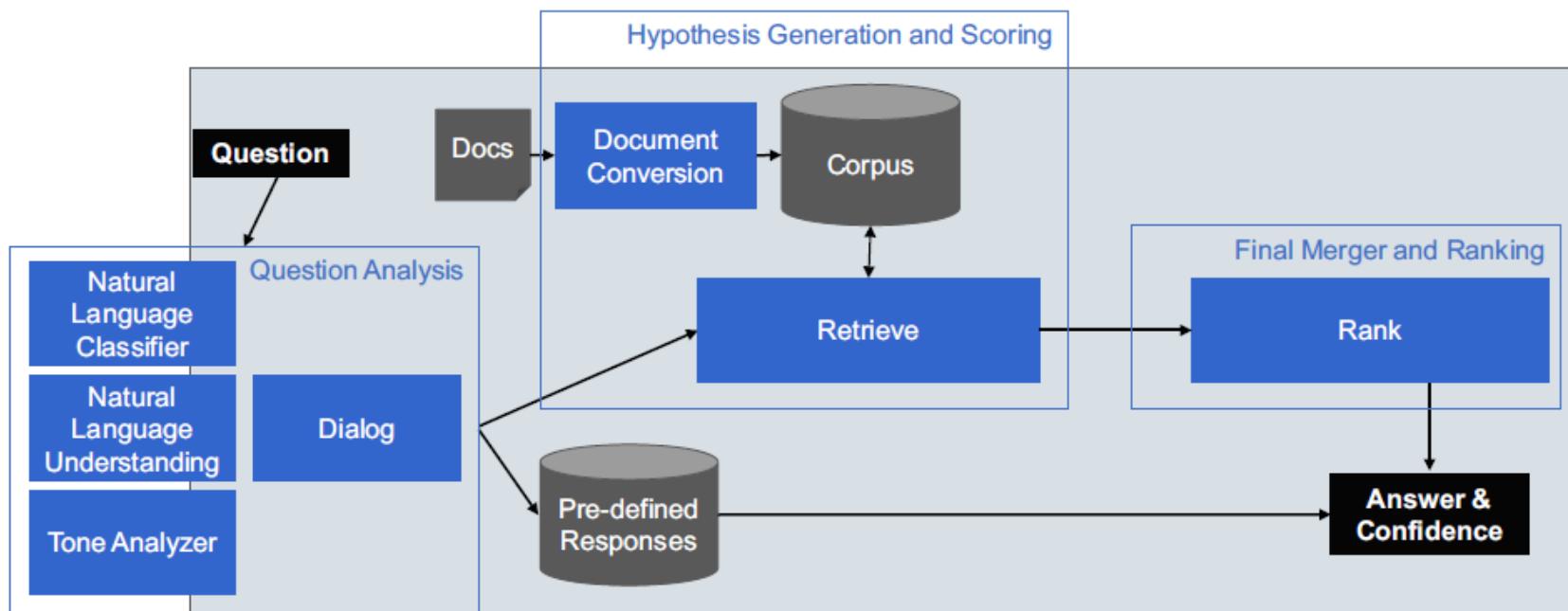


Figure: DeepQA pipeline; evolution to Watson Developer Cloud: Retrieve and Rank

Evolution of the DeepQA question analysis step (4 of 4)



IBM ICE (Innovation Centre for Education)

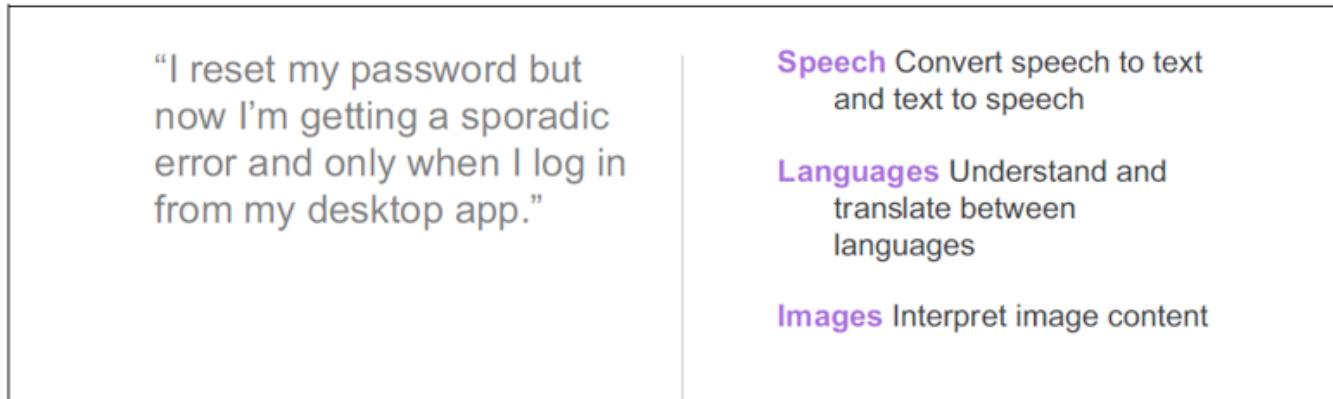


Figure: Deal with multi-modal communication

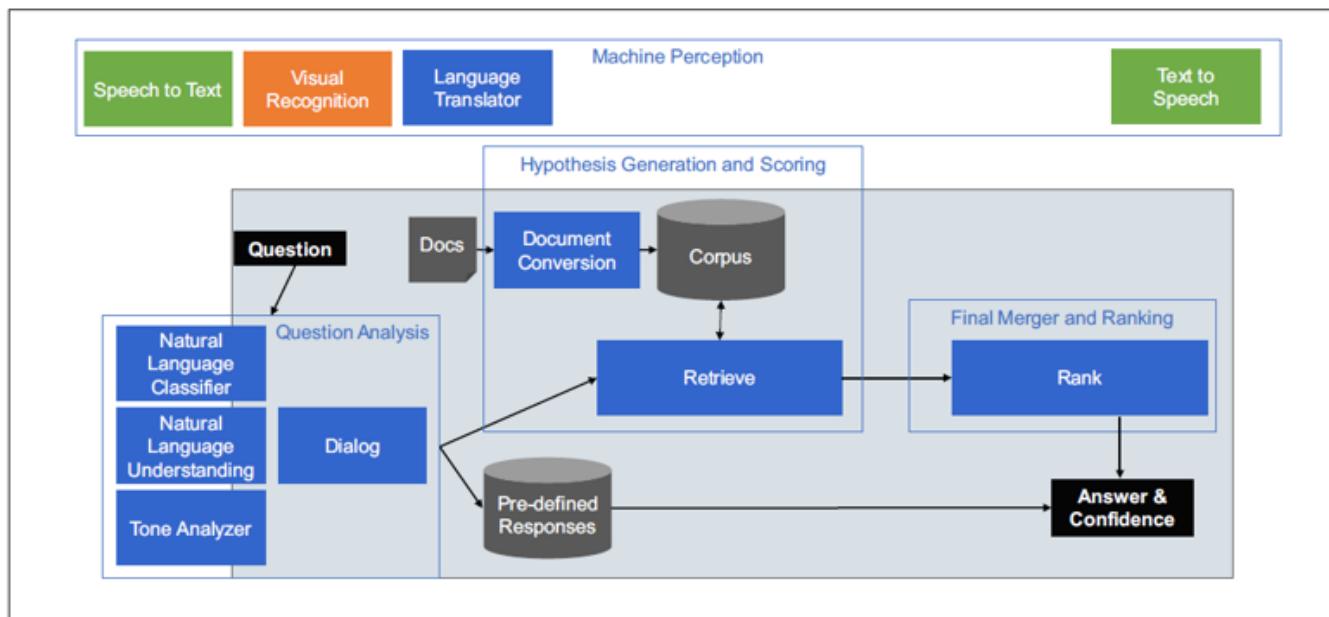
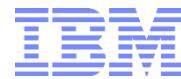


Figure: DeepQA pipeline; evolution to Watson Developer Cloud: Watson services overview

Microservices and robust tooling evolved from DeepQA



IBM ICE (Innovation Centre for Education)

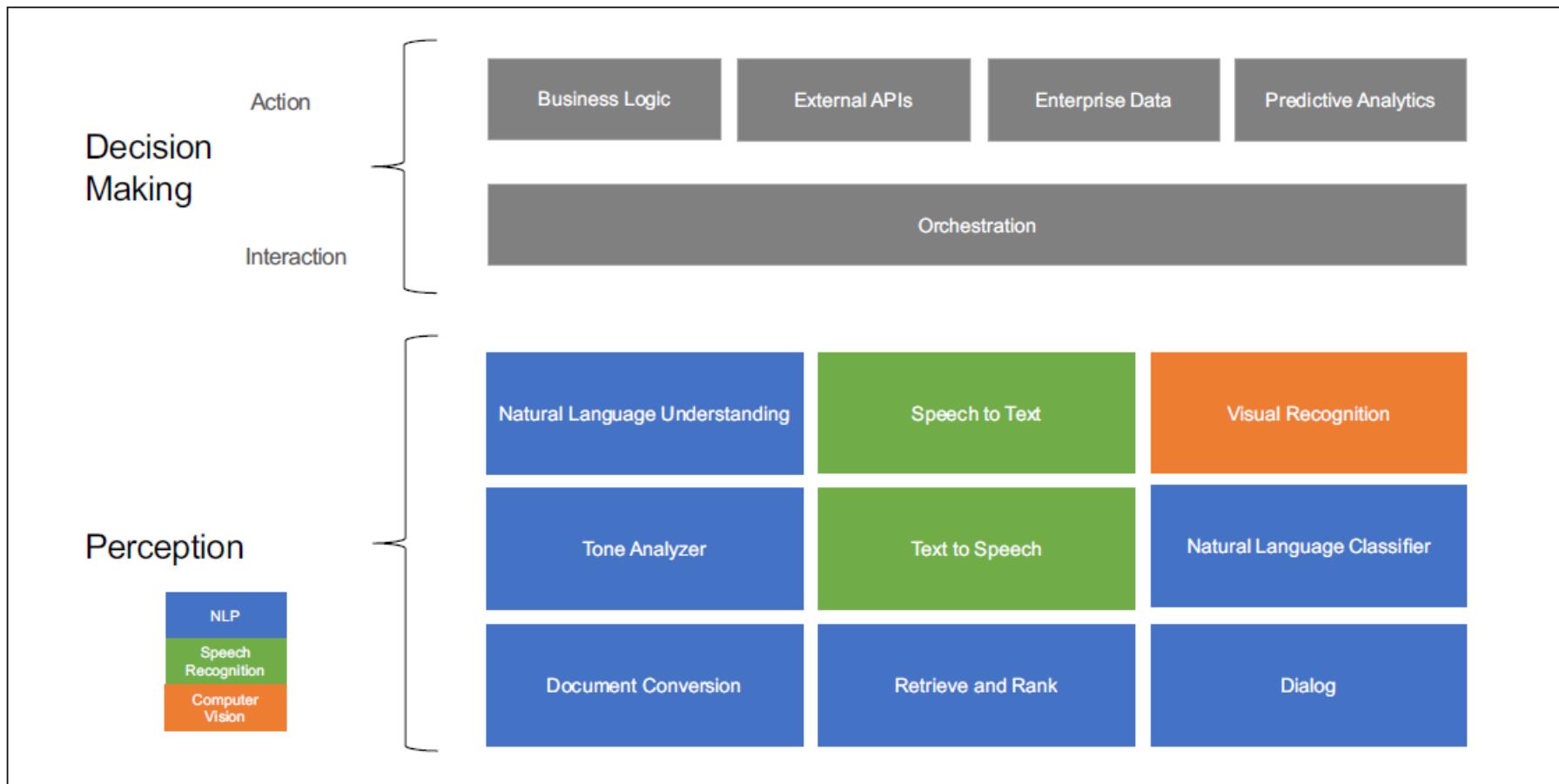


Figure: Watson Developer Cloud in approximately 2019

Service of Watson conversation

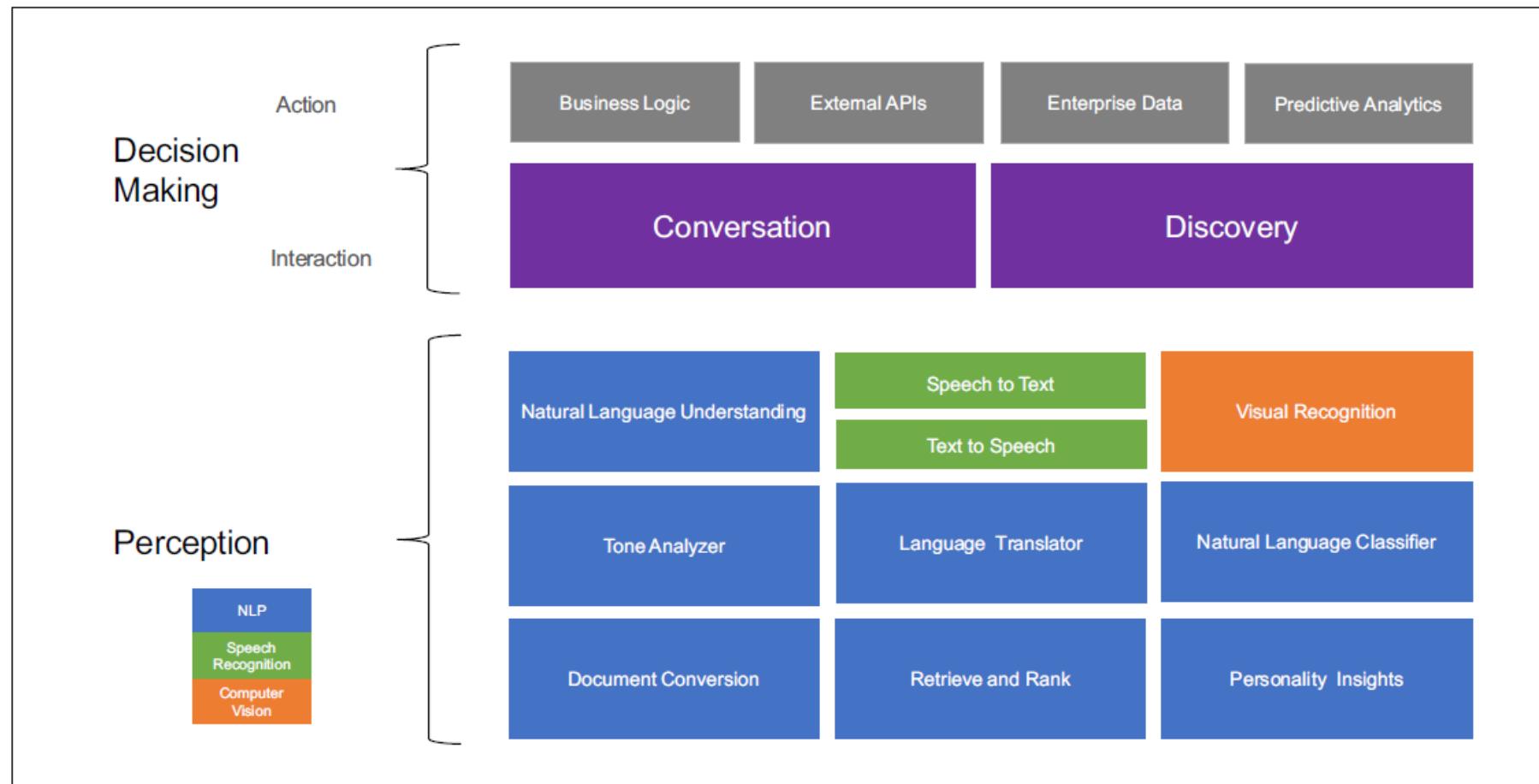


Figure: Conversation and Discovery services: Watson Developer Cloud today (2020)

Watson discovery service (1 of 2)

The screenshot shows the Watson Discovery Service interface. On the left, there's a sidebar with icons for file management and search. The main area is titled "Query your data > Build a query". It includes fields for "Enter your query" (with a note to run it in JSON), "Query:" (empty), "Aggregation:" (set to "term(host_id)"), "Filter:" (set to "comments_enriched.docSentiment.type:negative"), "Fields to return:" (radio button selected for "All"), and a "Run Query" button. To the right, the "Collection: AirbnbReviews" is selected. The "Results" section shows a JSON response with aggregations:

```
aggregations": [ { "type": "term", "field": "host_id", "results": [ { "key": "127772", "matching_results": 5 }, { "key": "19331457", "matching_results": 5 }, { "key": "6294856", "matching_results": 3 }, { "key": "2495797", "matching_results": 2 }, { "key": "101172", "matching_results": 1 }, { "key": "1408973", "matching_results": 1 } ] }
```

Figure: Discovery service-User interface

Watson discovery service (2 of 2)

- Evolution summary.

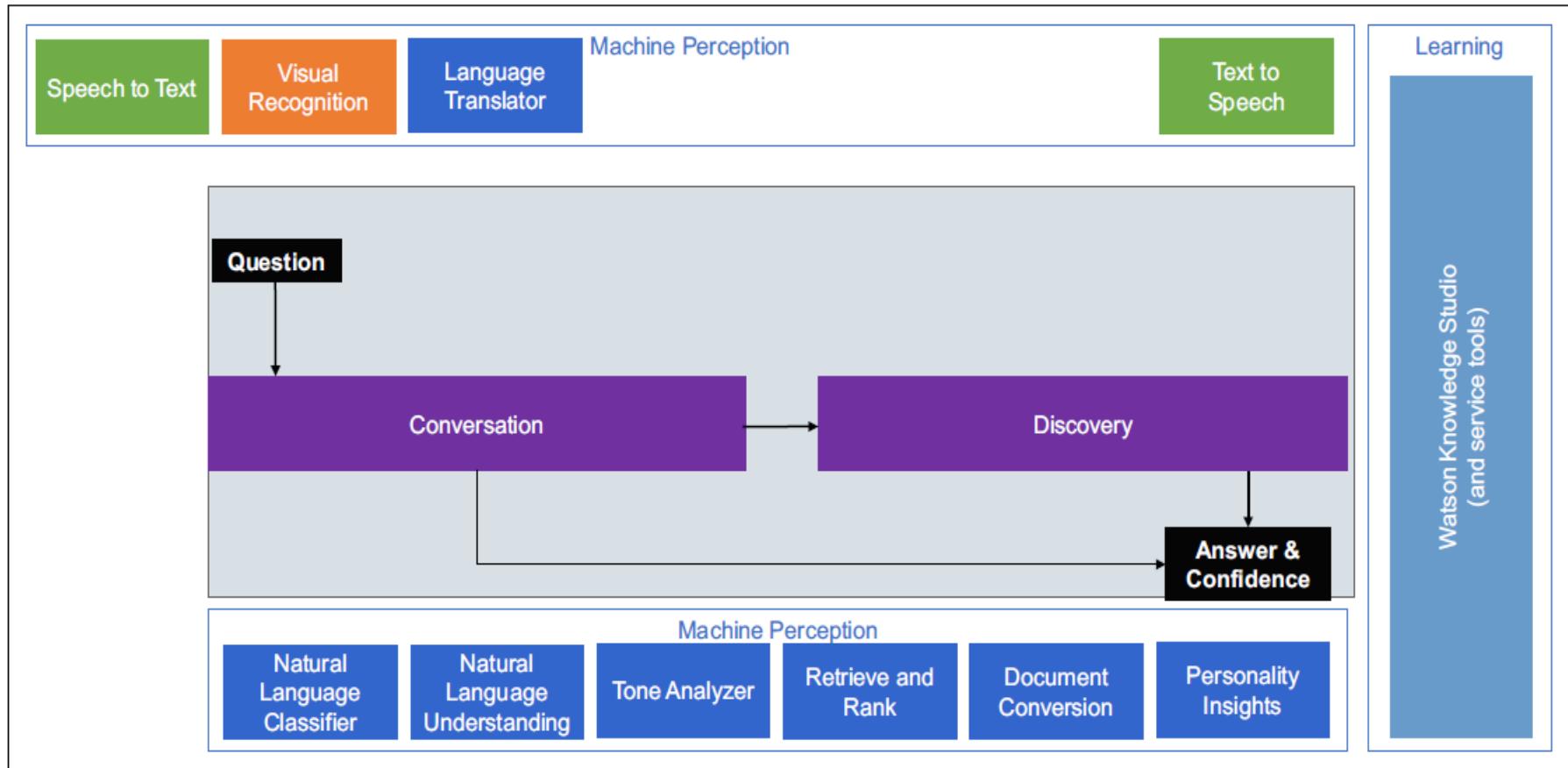


Figure: DeepQA pipeline, evolution to Watson Developer Cloud Summary

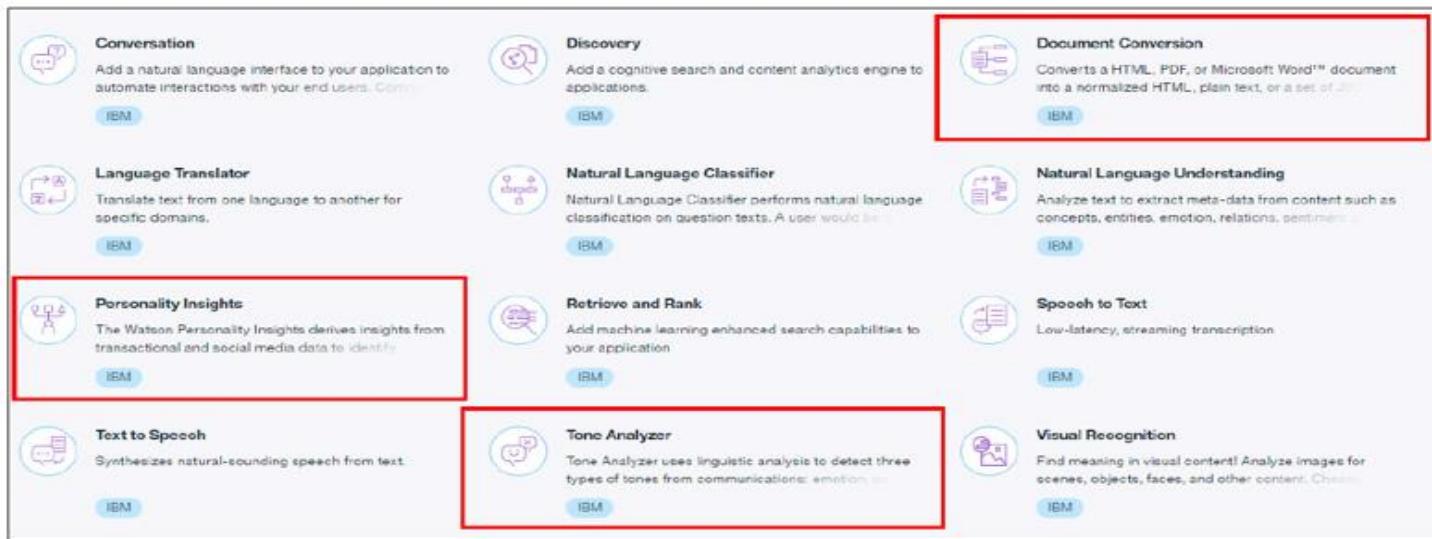


Figure: Watson services that cannot be trained by users

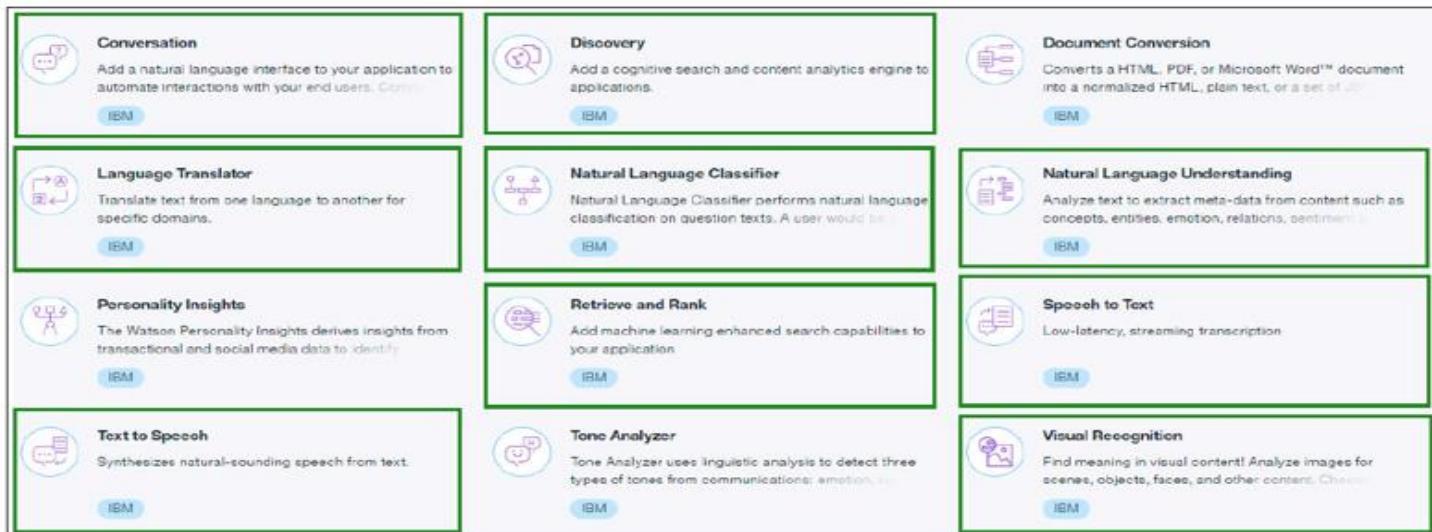


Figure: Watson services that can be trained by users

Watson conversation

- #intents -The purpose of a user's input; what the user wants to achieve.
- @entity-A term or object that is relevant to the intent. Provides the context.
- Dialog-Enables the service to respond to users; based on intents and entities.

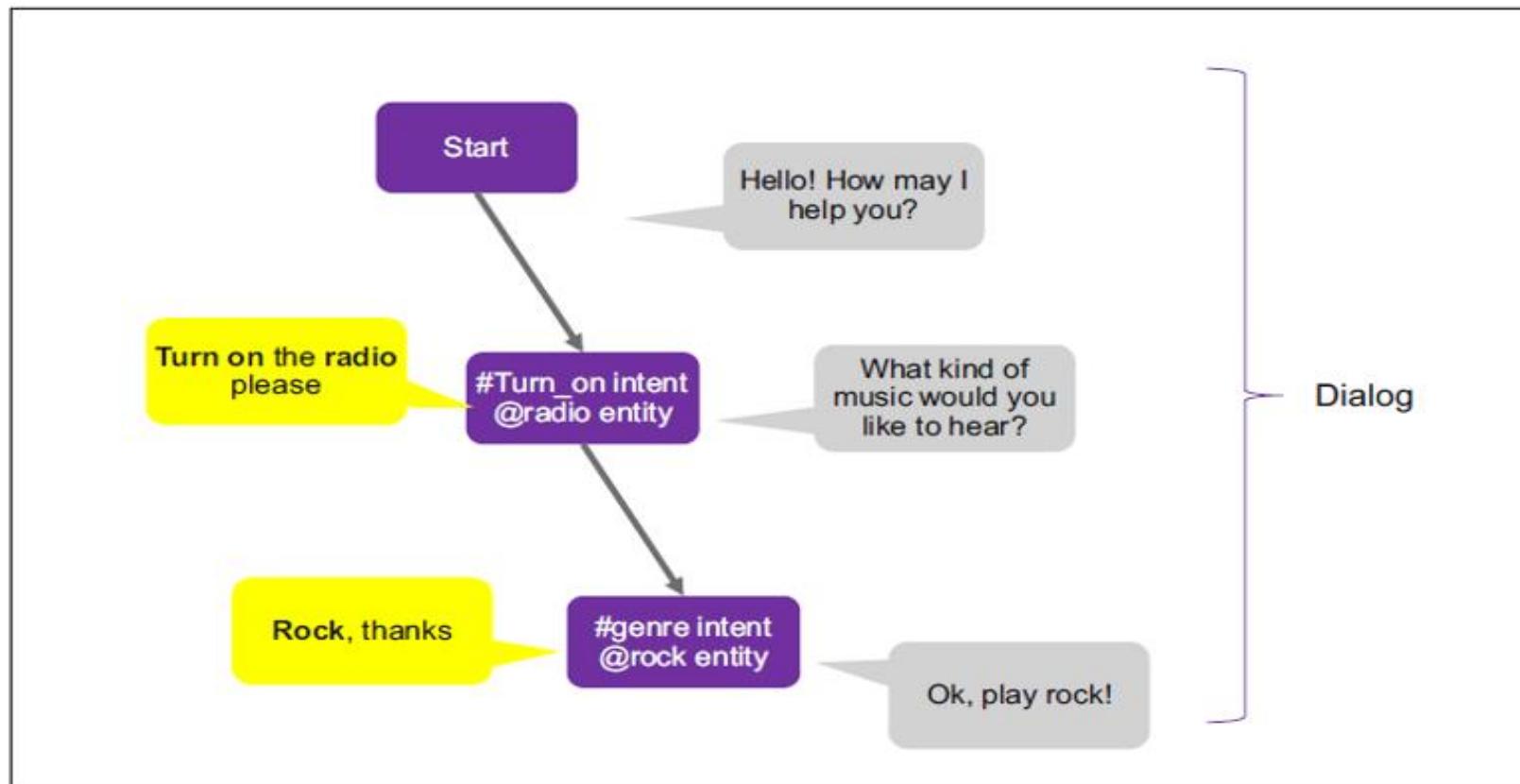


Figure: Conversation service: Showing intent, entity, dialog

Watson language translator

- The Watson language translation software provides the following language models:
 - News: Targeting news and transcript articles. Arabic, French, German, Italian, Japanese and Spanish, Brazilian Portuguese. Spanish to and from French can also be translated.
 - Conversational: The conversational symposiums were targeted. Translate Arabic, Portuguese, French, Italian and Spanish, to and from. Translate English.
 - Patents: Technical and legal terminology targeted. Translate Portuguese, Chinese, and English from Brazilian to Spanish.

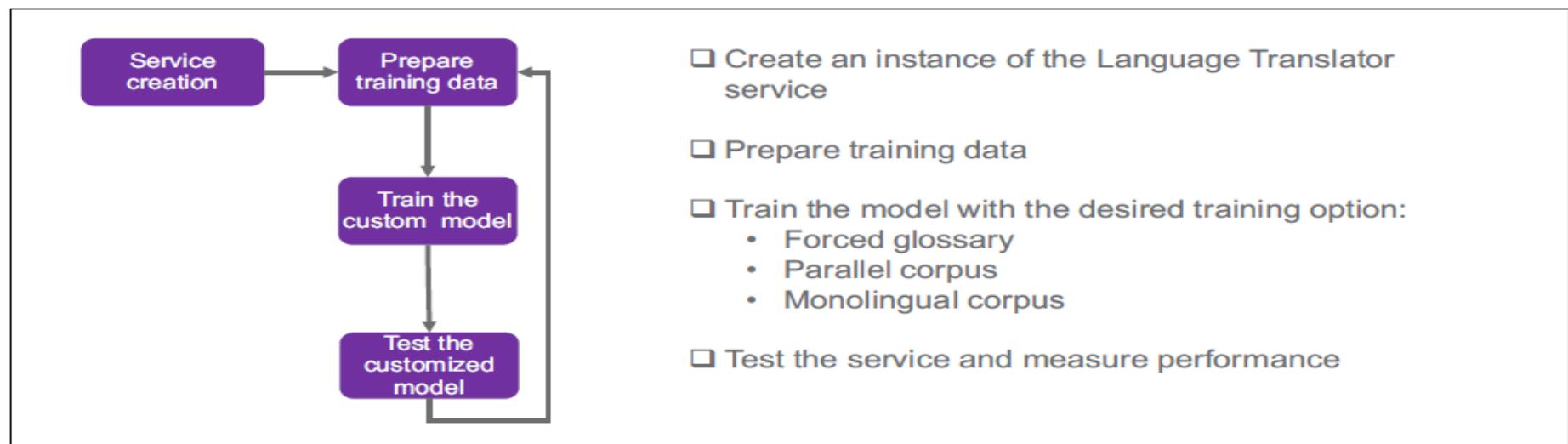


Figure: Language Translator service adaptation

Watson natural language classifier (1 of 2)



IBM ICE (Innovation Centre for Education)

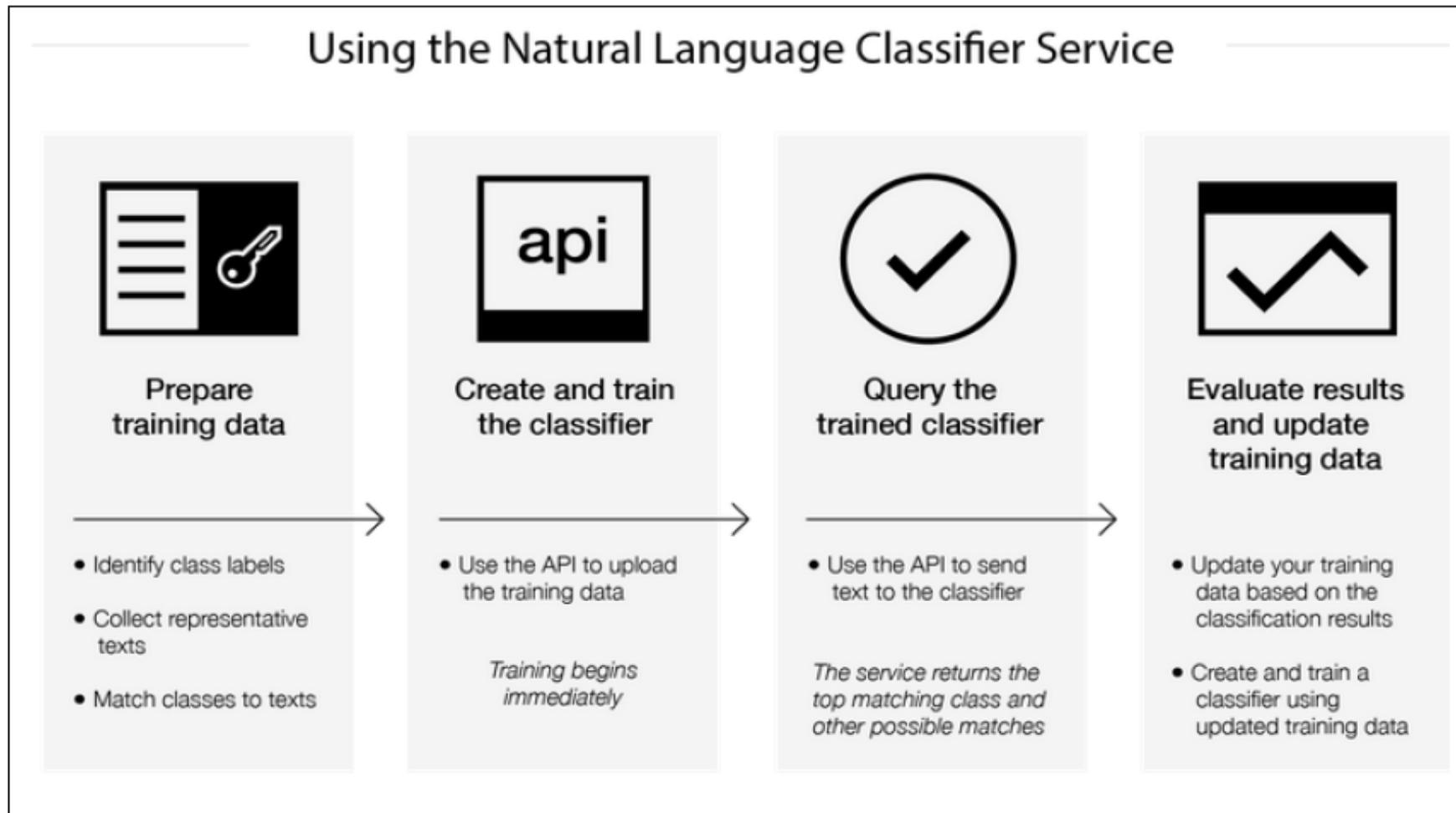


Figure: Using the Natural Language Classifier service-Process steps

Watson natural language classifier

(2 of 2)



IBM ICE (Innovation Centre for Education)

- The aim of this step is to improve the results of the classification process:
 - Detect wrong or weak confidence cases for user input text.
 - Change or restructure user's phrases into generic representative text.
 - Match text to their corresponding class label.
 - Add new text to the original training data and create a new classifier.
 - Repeat this cycle when quality of classification drops to a certain lower limit

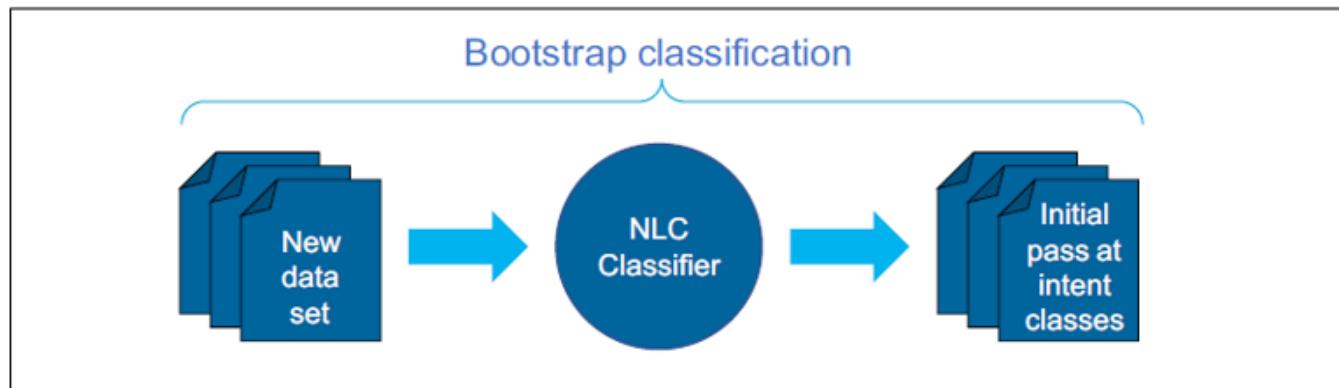


Figure: Bootstrap classification

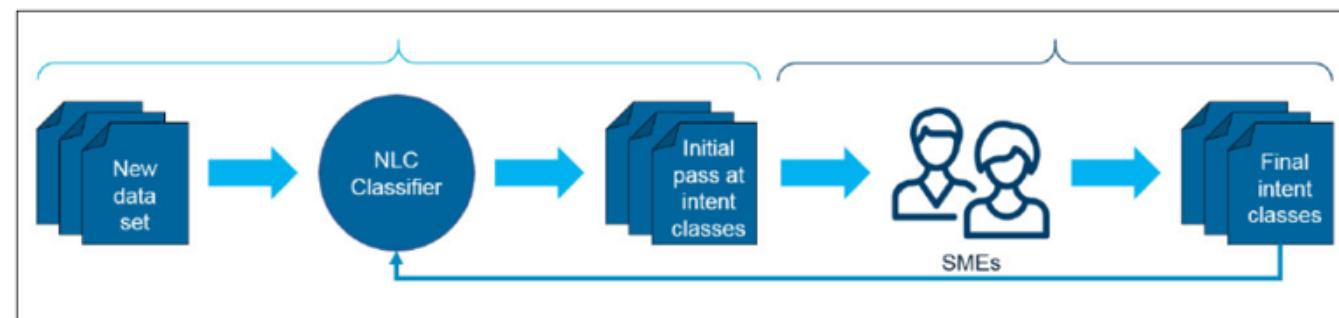


Figure: Manual validation of classifier

Watson retrieve and rank (1 of 2)

Using the Retrieve and Rank Service

Collect and load content



- Collect content
- Modify and upload Solr configuration files
- Upload content

Train the machine learning rank model



- Collect queries and relevant answers to leverage as training data
- Create and upload training data

Query service and evaluate results



- Send runtime queries to trained model
- Evaluate results and improve model

Figure: Using the Retrieve and Rank service

Watson retrieve and rank (2 of 2)

- Retrieve and rank service.

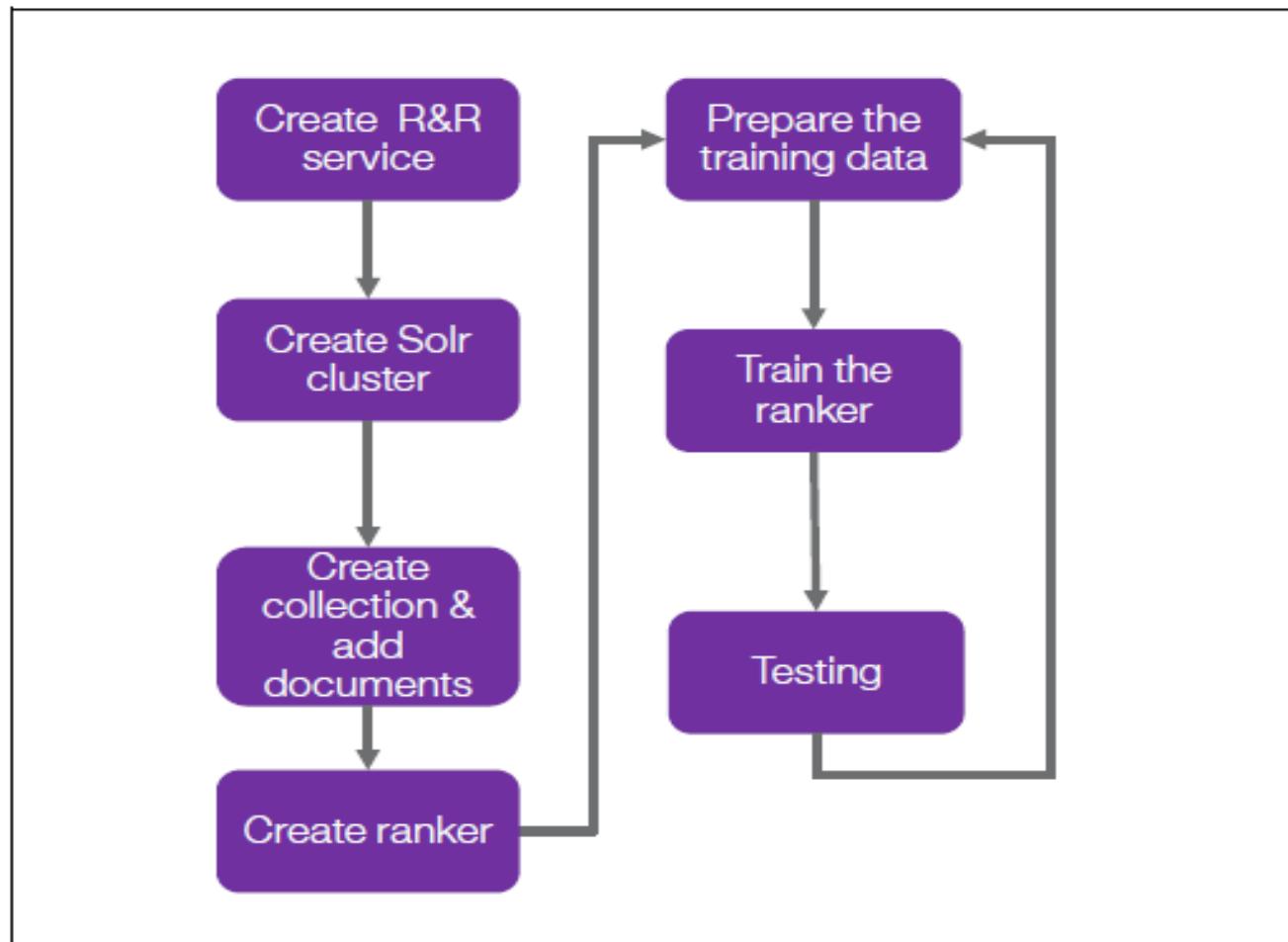


Figure: Adaptation of the Retrieve and Rank service

Image identification of Watson (1 of 2)

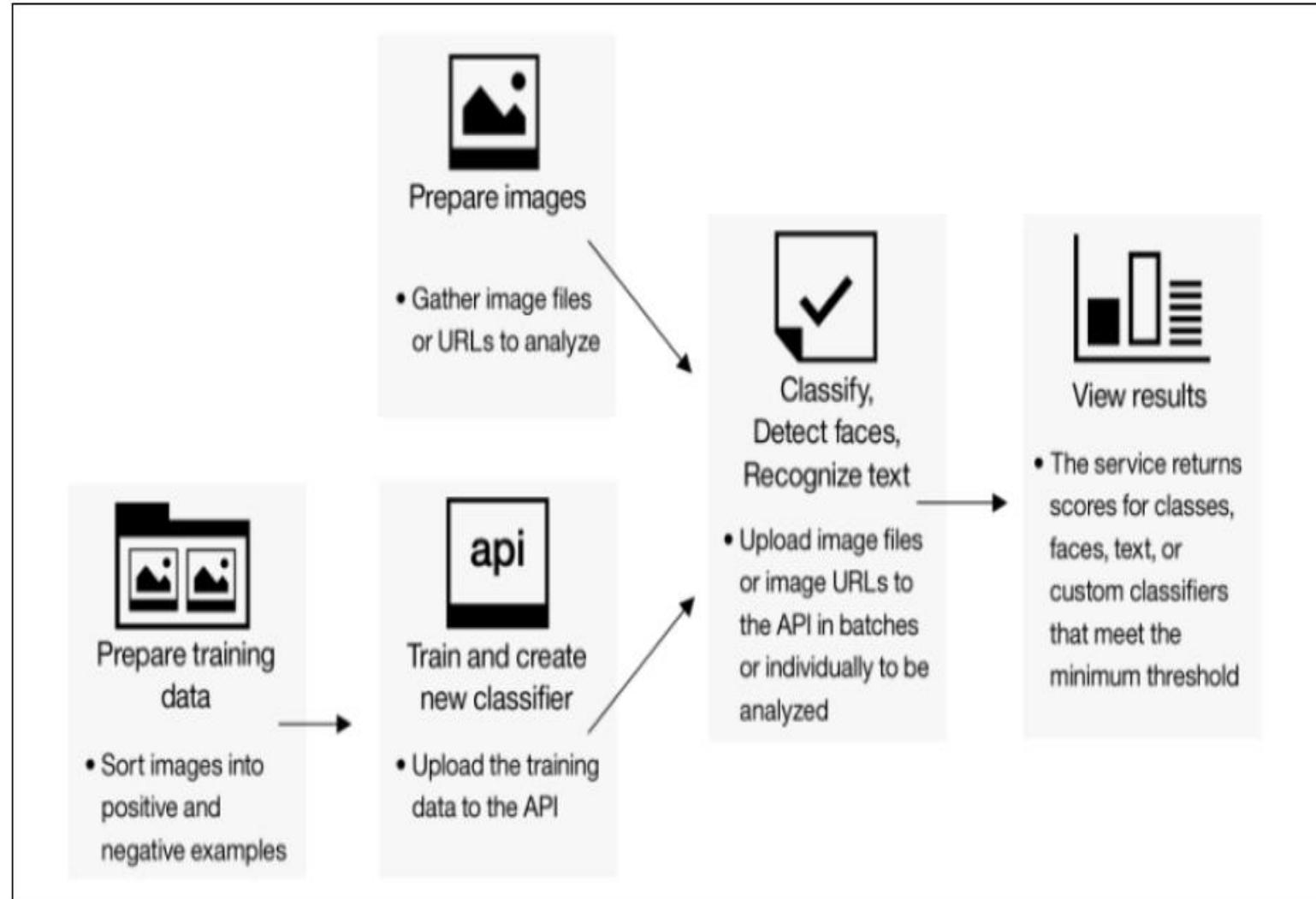


Figure: Visual Recognition process with custom classifier

Image identification of Watson (2 of 3)

- Visual recognition training images.

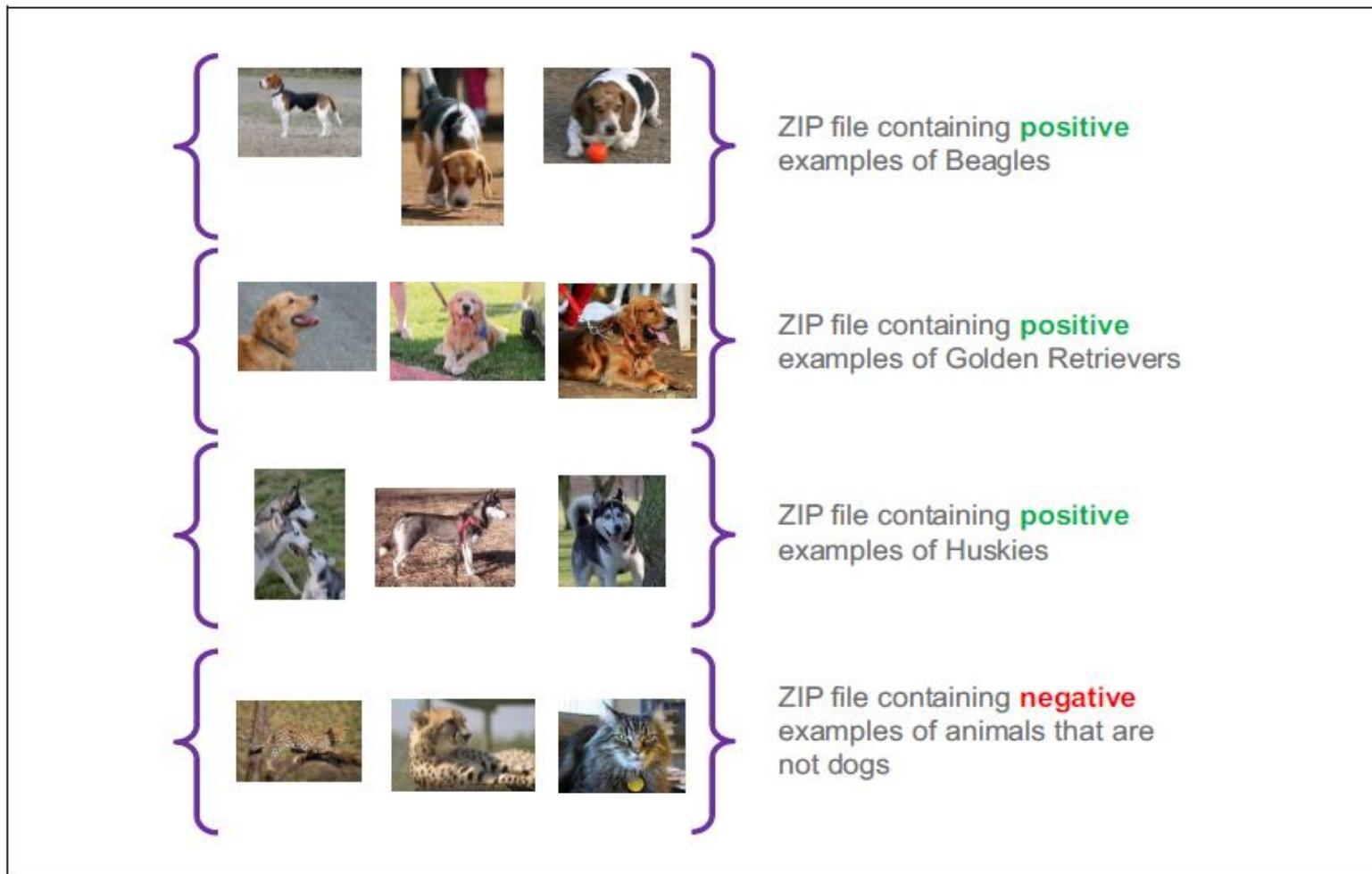


Figure: Example of Visual Recognition training images

Image identification of Watson (2 of 2)

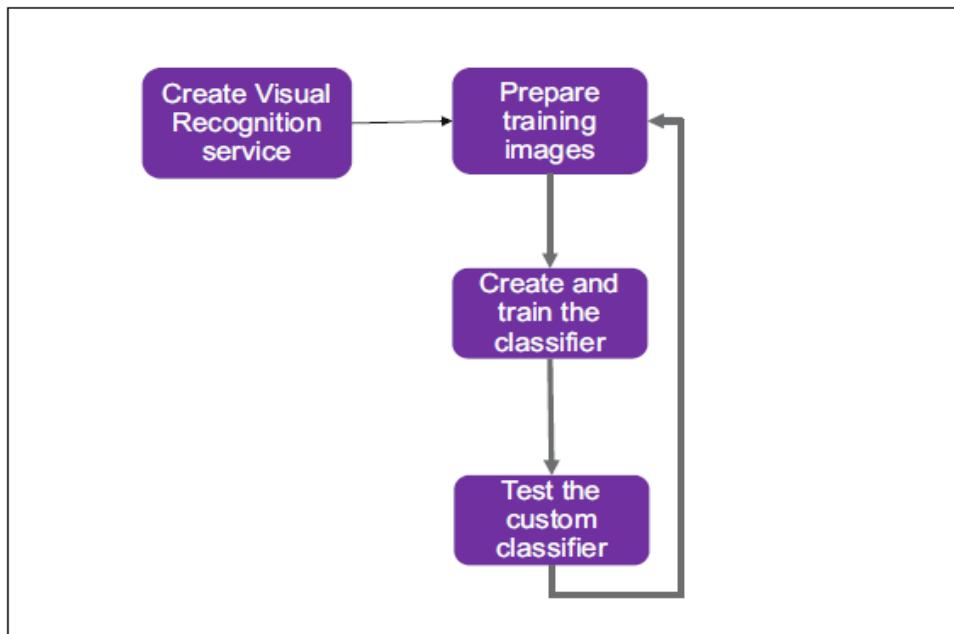


Figure: Adaptation of the Visual Recognition service

The following steps are shown above:

1. Prepare training data.

Gather image files to use as positive and negative example training data.

2. Create and train the classifier.

Specify the location of the training images and call the Visual Recognition API to create the custom classifier.

3. Test the custom classifier.

Classify images with the new custom classifier and measure the classifier performance.

Figure: Visual Recognition process with custom classifier

Speech to text from Watson (1 of 2)

To add a corpus to a specific model, use the below method.

POST /v1/customizations/{customization_id}/corpora/{corpus_name}

Below example shows an abbreviated corpus for the healthcare domain. A corpus file is typically much longer.

How Is Coronary Microvascular Disease Treated?

If you're diagnosed with coronary MVD and also have anemia, you may benefit from treatment for that condition.

Anemia is thought to slow the growth of cells needed to repair damaged blood vessels.

What causes autoimmune hepatitis?

A combination of autoimmunity, environmental triggers, and a genetic predisposition can lead to autoimmune hepatitis.

What research is being done for Spinal Cord Injury?

The National Institute of Neurological Disorders and Stroke NINDS conducts spinal cord research in its laboratories at the National Institutes of Health NIH.

NINDS also supports additional research through grants to major research institutions across the country.

What is Osteogenesis imperfecta OI?

Osteogenesis imperfecta OI is a rare genetic disorder that, like juvenile osteoporosis, is characterized by bones that break easily, often from little or no apparent cause.

Example: Abbreviated sample corpus for health care domain

Speech to text from Watson (2 of 2)

```
curl -X POST -u {username}:{password}
--header "Content-Type: application/json"
--data "{\"words\": [
  {\"word\": \"HHonors\", \"sounds_like\": [\"hilton honors\", \"h honors\"],
  \"display_as\": \"HHonors\"},
  {\"word\": \"IEEE\", \"sounds_like\": [\"i triple e\"]}]}"
"https://stream.watsonplatform.net/speech-to-text/api/v1/customizations/74f4807
e-b5ff-4866-824e-6bba1a84fe96/words"
```

Figure: Adding multiple words to a custom language model

```
curl -X PUT -u {username}:{password}
--header "Content-Type:application/json"
--data "{\"translation\":\"I triple E\"}"
"https://stream.watsonplatform.net/text-to-speech/api/v1/customizations/{custom
ization_id}/words/IEEE"
```

Figure: Adding the word IEEE to a custom voice model by using the sounds-like method

```
curl -X GET -u {username}:{password}
--header "Accept: audio/wav"
--output ieee-new.wav
"https://stream.watsonplatform.net/text-to-speech/api/v1/synthesize?text=IEEE&c
ustomization_id={customization_id}"
```

Example: Using the custom model

Watson text to speech

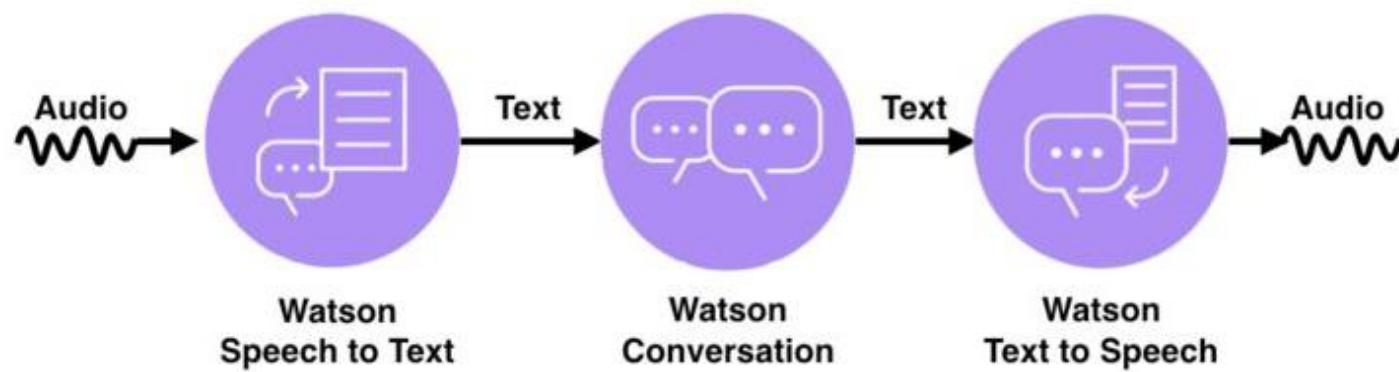


Figure: Watson Text to Speech

Source: <https://images.app.goo.gl/7xKJ7tTxCvoS8YcH7>

Watson natural language understanding

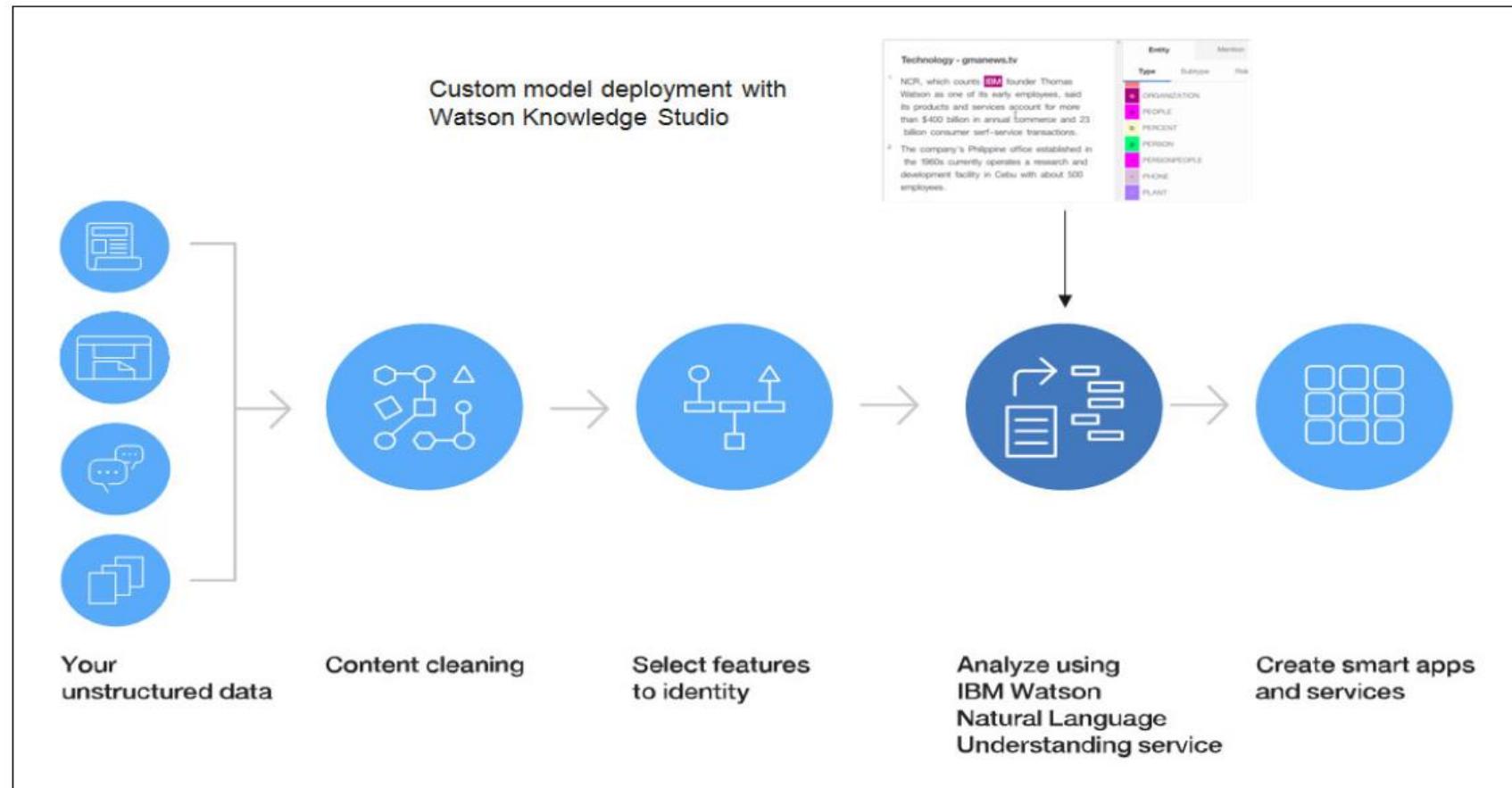


Figure: Analyzing unstructured data with Watson Natural Language Understanding

Watson discovery

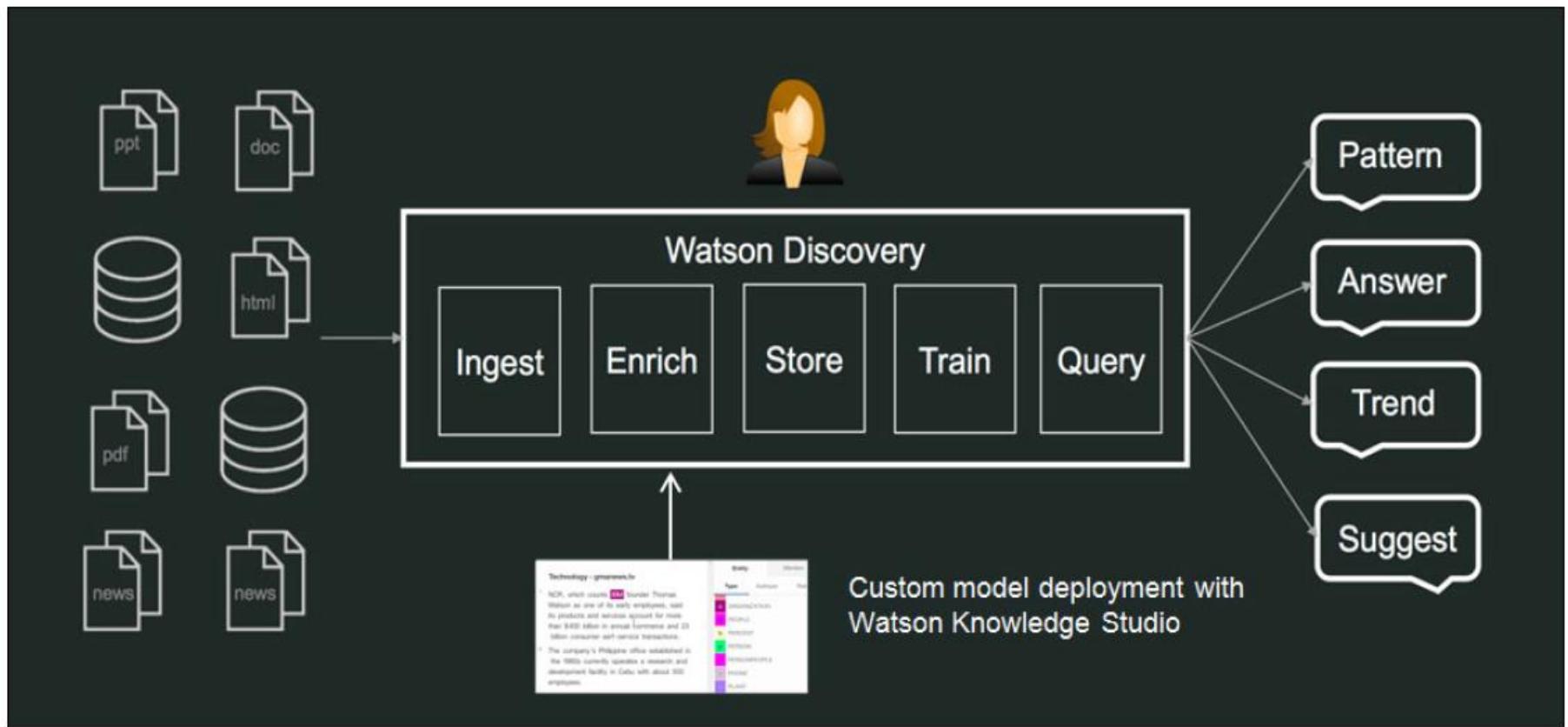


Figure: Ingest, enrich, and query unstructured data with Watson Discovery service

Watson knowledge studio (1 of 2)

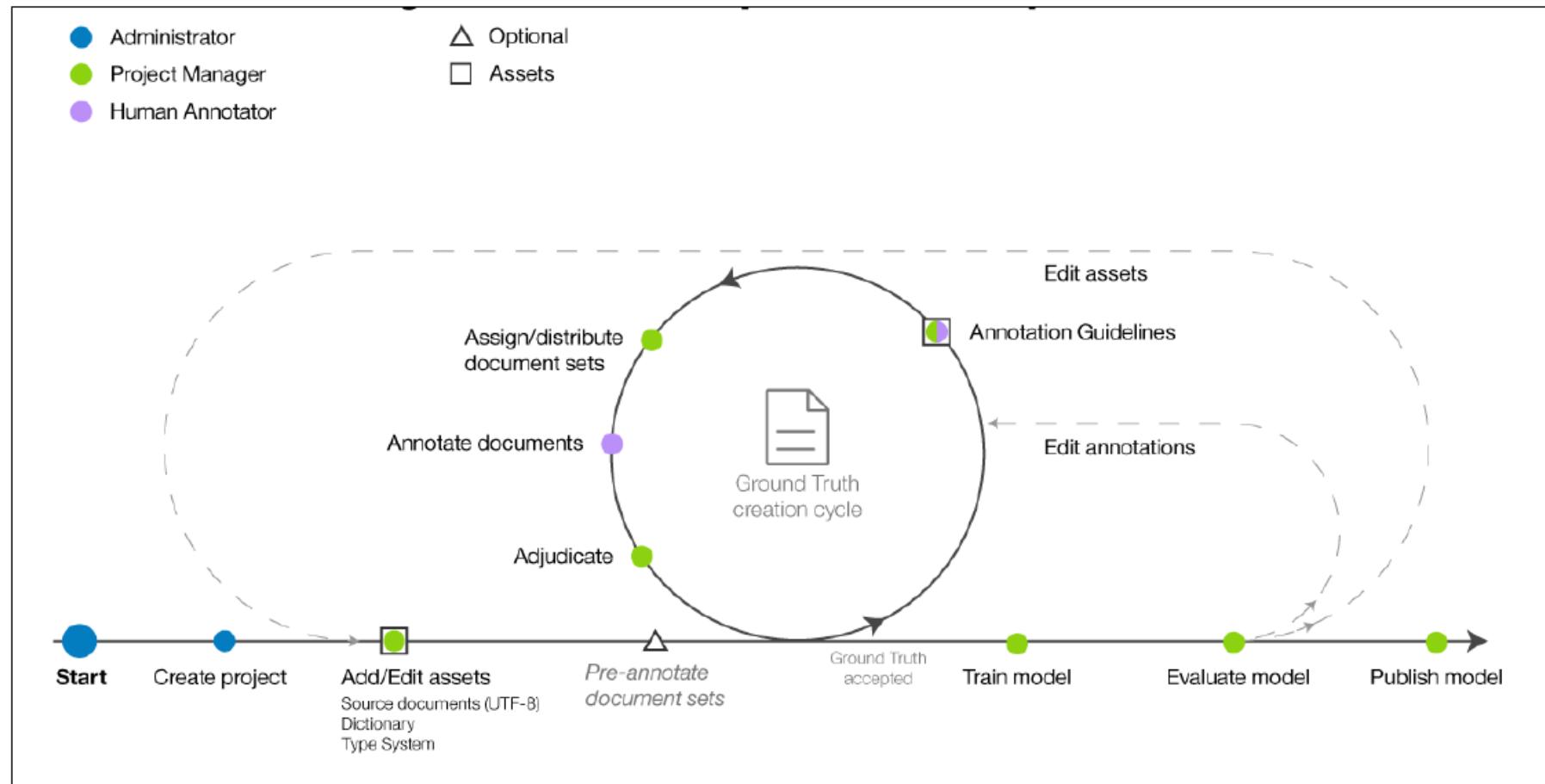


Figure: Machine-learning annotator component development workflow

Watson knowledge studio (2 of 2)

Step	Description
Assign user roles	An annotator component requires input from subject matter experts, project managers, and users who can understand and interpret statistical models. A user account must be created for each user who needs to log in to Watson Knowledge Studio
Create a project	A project contains the resources that are used to create the annotator component, including these: <ul style="list-style-type: none">▶ Type system▶ Source documents▶ Dictionaries
Optional: Pre-annotate documents	Pre-annotate documents according to the terms in the project dictionaries or based on rules that you define
Annotate documents	The project manager assigns annotation tasks to human annotators, configures the inter-annotator agreement threshold, and provides annotation guidelines for the human annotators to follow Human annotators use the Ground Truth Editor to manually annotate documents

Step	Description
Adjudicate and promote documents	Accept or reject the ground truth that was generated by human annotators and adjudicate any annotation differences to resolve conflicts. Accepted documents are promoted to ground truth
Train the model	Create a machine learning annotator component
Evaluate the model	Evaluate the accuracy of the annotator component
Publish the model	Export or deploy the model

Table: Creating a machine learning model with Watson Knowledge Studio overview

Example: Creating a machine learning model



IBM ICE (Innovation Centre for Education)

- Assign user roles.
- Create a project.
- Create a type system.
- Add a dictionary.
- Add documents for annotation.
- Create and assign annotation sets.
- Optional: Pre-annotate with a dictionary-based annotator.
- Create an annotation task.
- Annotate documents.
- Adjudicate conflicts and promote documents to ground truth.
- Create a machine-learning annotator.

Self evaluation: Exercise 20

- To continue with the training, after learning the various steps involved in cognitive analytics and watson machine learning, it is instructed to utilize the concepts of cognitive machine learning algorithms to perform the following activity.
- You are instructed to write the following activities using Python code.
- Exercise 20: Simple Top-Down Parsing for cognitive computing

Cognitive device system of the industry

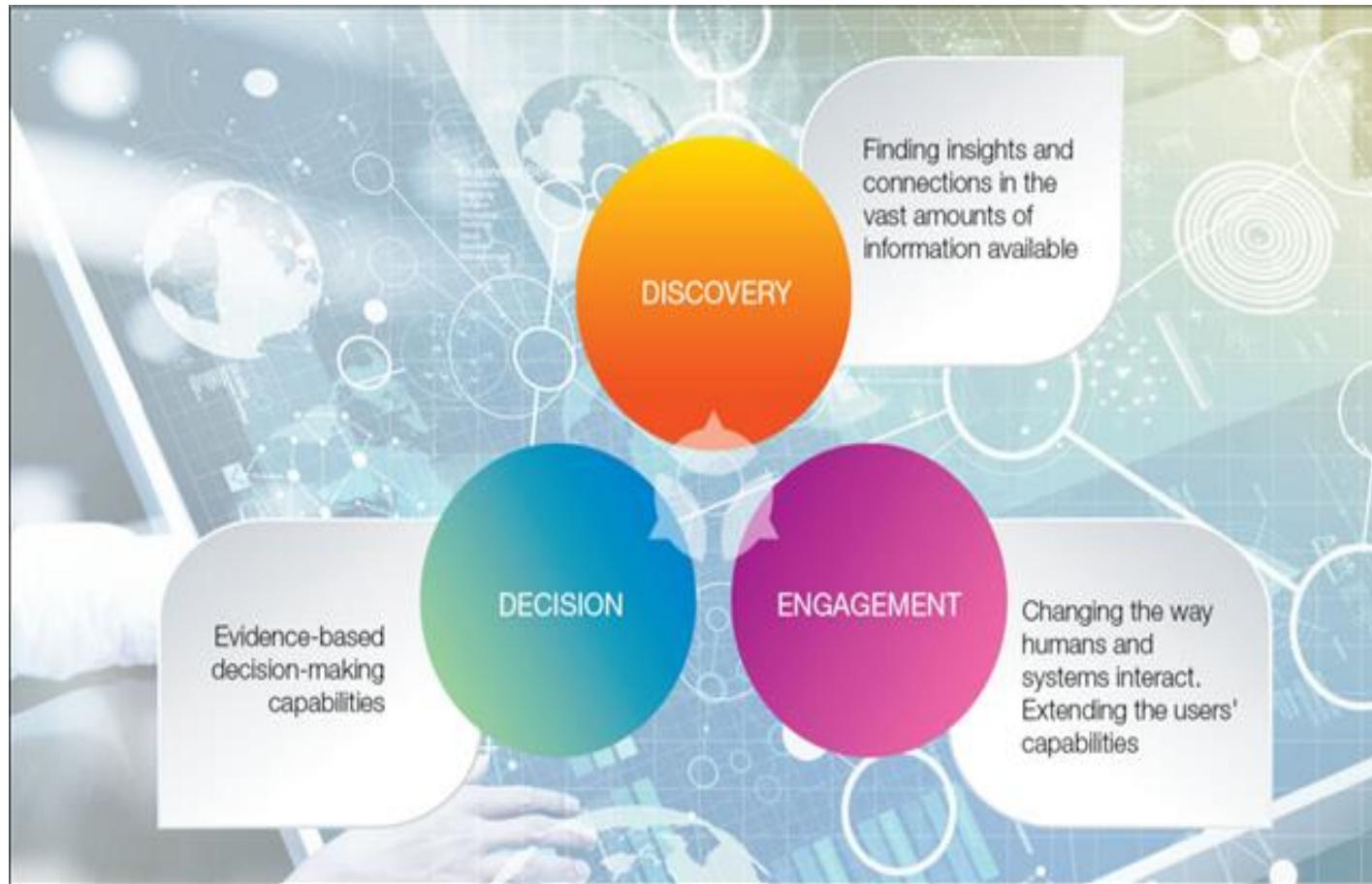


Figure: Broad capability areas of cognitive systems

Market consumer: Cognitive computing services



IBM ICE (Innovation Centre for Education)

Examples of Cognitive Computing

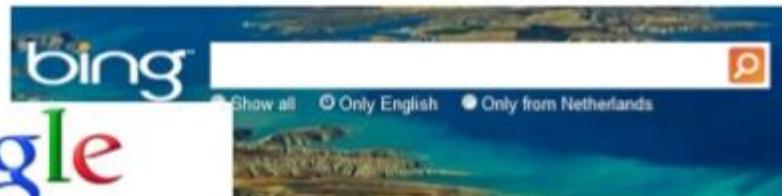
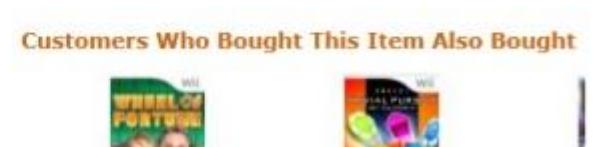
- Web Search 
- Image Search  
- Event Search 
- Social Computing 
- Natural Language Processing 

Figure: Market consumer: cognitive computing services

Source: <https://images.app.goo.gl/zQGKP8TbN2wgCND79>

Cognitive services delivery: Open source and cloud projects



IBM ICE (Innovation Centre for Education)

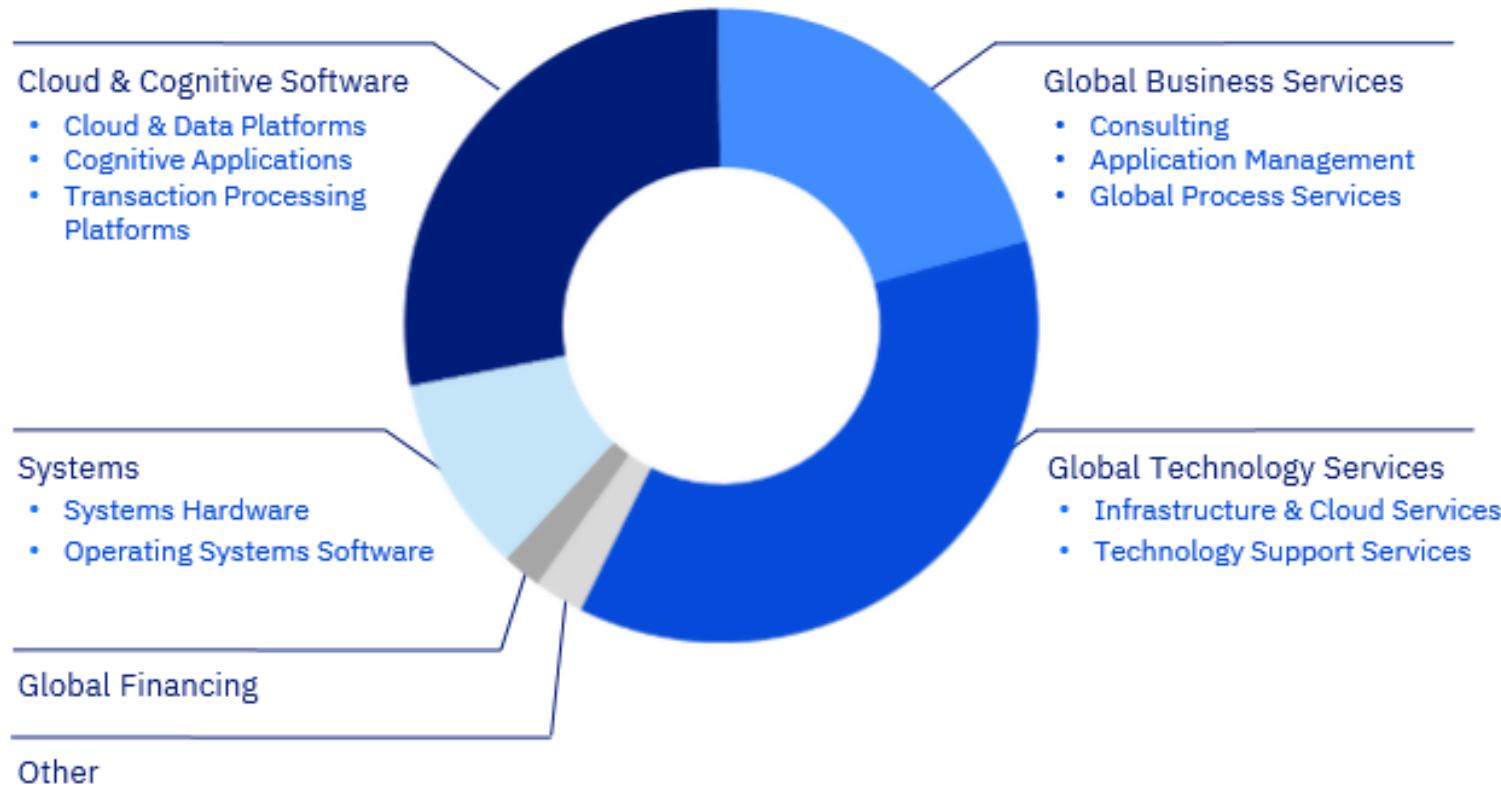


Figure: Future of cognitive services delivery

Source: <https://images.app.goo.gl/EbD25eNWhwfueMpd9>

What is new in IBM Watson content analytics?



IBM ICE (Innovation Centre for Education)

- Content mining:
 - Discovery's powerful new text analytics capability gives you the ability to enrich and explore your data. By combining AI-powered search, Natural Language Processing (NLP) and content analytics, users can rely on Content Mining to find the information they need, using Discovery's highly-sophisticated depth of context, relationship, and pattern understanding. Doesn't matter if the data is structured (text documents) or unstructured (tables, images, graphs) users are now able to work more efficiently and find the right information the first time.
- Content intelligence:
 - Now you can take advantage of pre-built NLP models. With this out of the box content, users can skip document training and configure Discovery projects using contracts, invoices, and purchase orders. This helpful feature is uniquely suited for organizations working in the general contract governance space, the procurement space, and to identify changes to contracts necessary for the LIBOR transition. The IBM Watson team has specifically built specialized LIBOR NLP models to assist with the 2021 transition for this December 4, 2019 release.

New features for the content analytics administrator



IBM ICE (Innovation Centre for Education)

Social analytics and social search

New social media crawler	If you have a BoardReader license, you can configure a BoardReader crawler to collect content from blogs, message board sites and forums, news sites, reviews, and videos. Because all information on social media sites is presumed to be public, secure search is not supported.
Enhanced social media crawlers	The FileNet P8 crawler, SharePoint crawler, and Seed list crawler for IBM Connections can collect social data. The enhancements allow users to explore relationships between people, recommendations, tags that are associated with documents added to a collection by these crawlers.
Social search	If social search is enabled for a collection, the system can aggregate information from various social networking sources and extract relationships between documents, people, and tags. For example, users can discover people who are relevant to a document; see recommendations for other documents and people that might be of interest; and drill down through a tag cloud to explore related information.

User experience

Redesigned user interfaces	The content analytics miner and enterprise search applications were redesigned to provide a common user experience that matches other IBM products. The redesign includes performance enhancements, functional enhancements, and usability enhancements.
Layout customization	You can easily change the appearance and widgets in the application interfaces by selecting a predefined layout option. For example, in the enterprise search application you can select a faceted search layout that includes time series analysis and correlation analysis. You can also create your own layouts based on templates such as a three-column page design or a two-row page design.

Figure: The administration console interface

Source: https://www.ibm.com/support/knowledgecenter/es/SS5RWK_3.5.0/com.ibm.discovery.es.nav.doc/iysawhatsnew.htm

Self evaluation: Exercise 21

- To continue with the training, after learning the various steps involved in cognitive analytics and watson machine learning, it is instructed to utilize the concepts of cognitive machine learning algorithms to perform the following activity.
- You are instructed to write the following activities using Python code.
- Exercise 21: Latent Semantic Analysis (LSA)

Content analytics architecture (1 of 2)

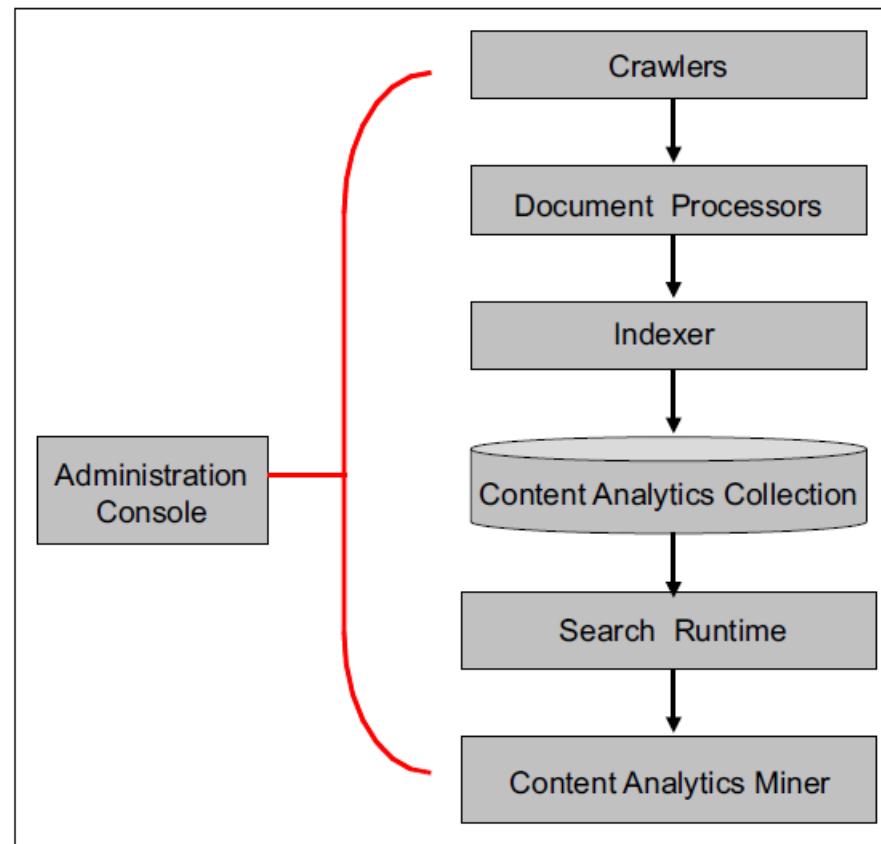


Figure: Component architecture of Content Analytics

Content analytics architecture (2 of 2)

- Document processors.

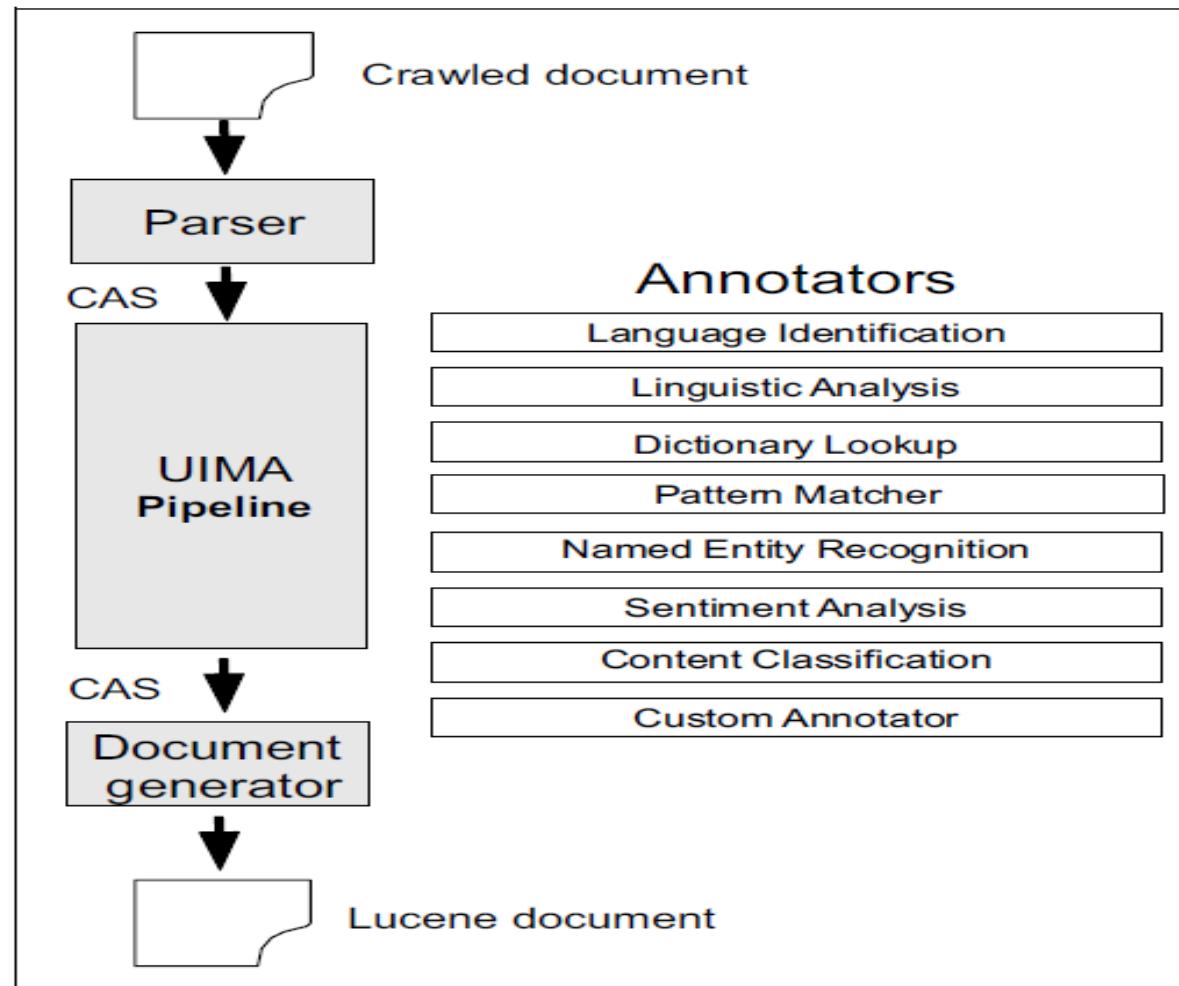


Figure: Document processor architecture in Content Analytics

Content analytics architecture (3 of 3)

- Export points in the data flow.

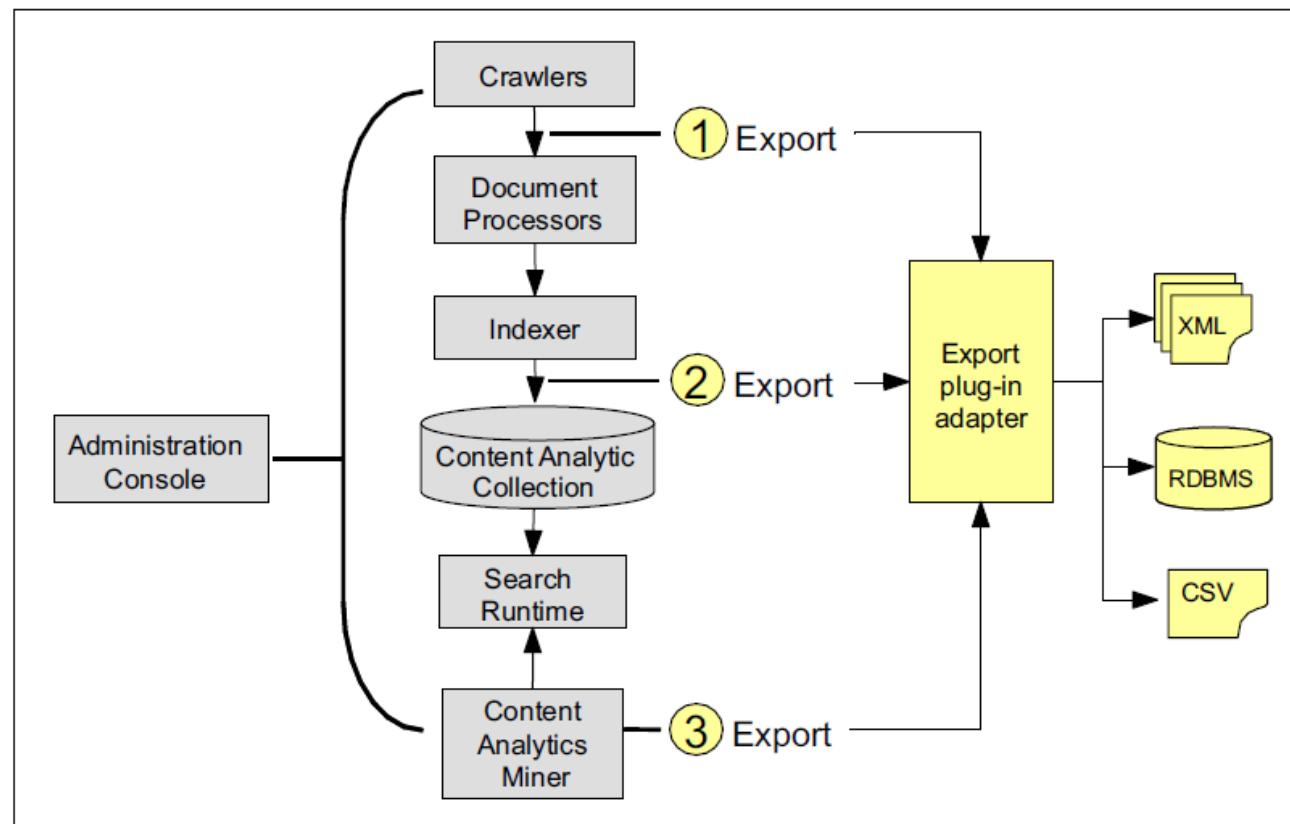
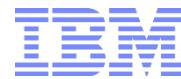


Figure: Export points in Content Analytics

Use case scenarios

- Customer insights.
- Call center.
- Quality assurance.
- Enforcement of legislation and public security.
- Investigation management.
- Insurance Fraud.
- Fraud Detection.
 - Banking.
 - Energy and utilities.
 - Taxation.
 - Healthcare.
 - Insurance.
 - Warranties.
 - Worker's Compensation.
 - Travel and more.

Watson corpus building: Answer and evidence



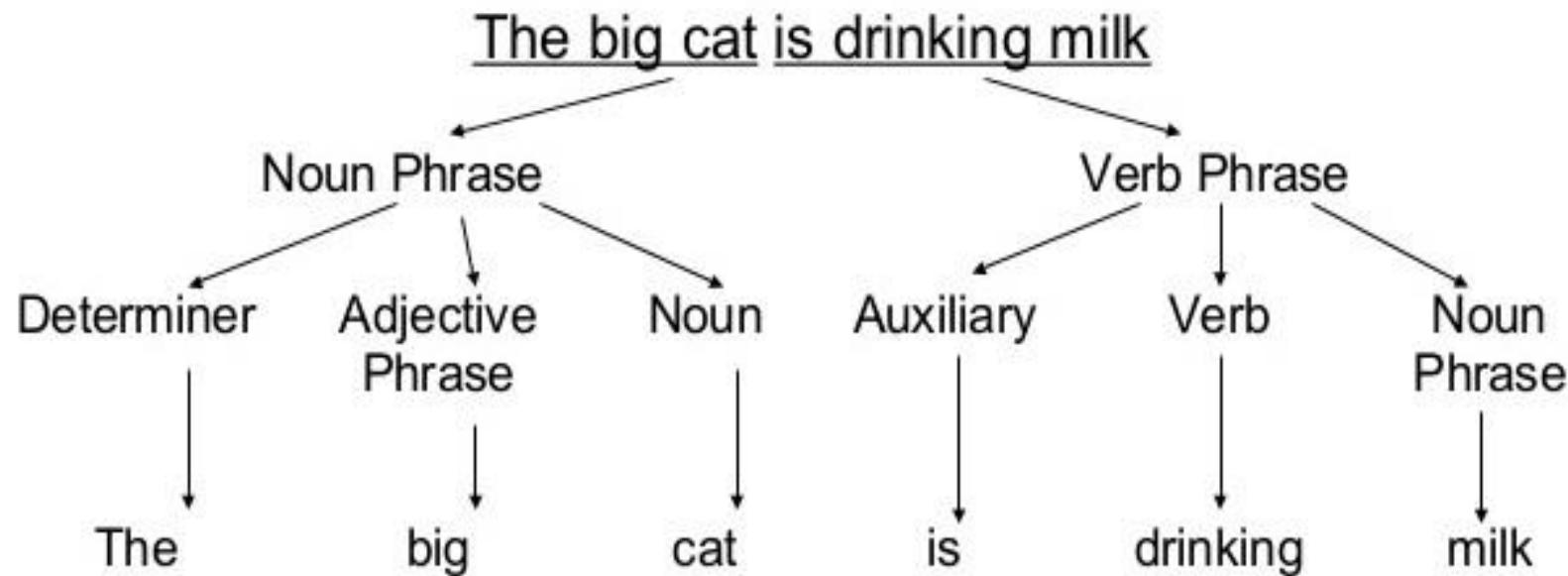
IBM ICE (Innovation Centre for Education)

- Focus: The emphasis is on that part of the problem that is the response. You need to grasp the purpose of the query to be able to answer correctly. Focus identification relies on the detection of target style patterns.
- LAT (Lexical Answer Type): In order to decide what type of response, Watson uses the LAT. Watson, for example, is looking for a film, place or person's name?
- Question classification: Watson uses Query Labeling to discourage the sort of problem that should be asked.
- Question dection: There are fragments of questions that need to be answered with a particular approach.
 - Source acquisition— Identify the appropriate collection of tools for this unique task.
 - Source transformation— Optimize the textual information framework to effectively search.
 - Source expansion and updates— Expansion algorithms are used to avoid more expertise that will help to fill gaps and add depth to the source of information from Watson corpus.

Grammatical slot parser and textual research modules



IBM ICE (Innovation Centre for Education)



www.decideo.fr/bruley

Figure: Example - Naming Syntactic Roles or Phrases subject

Source: <https://images.app.goo.gl/BZRmnthUfWaTXJaF7>

Question classification

Type of Clue	Example	How You Answer the Clue
You need to know the facts.	HEAD NORTH: Two states you could be re-entering if you are crossing Florida's northern border. Answer: Georgia and Alabama	You answer the question based on factual information about one or more entities. Understand what is being asked and which elements of the clue will help you get the answer.
You need to decompose the clue.	DIPLOMATIC RELATIONS: Of the four countries in the world that the United States does not have diplomatic relations with, the one that's farthest North. Answer: North Korea	One subclue is nested in the outer clue. After you replace the subclue with its answer, it become easier to answer the outer clue. In this example: The inner subclue is "the four countries in the world that the United States does not have diplomatic relations with." The answer to the subclue is Bhutan, Cuba, Iran, and North Korea. After replacing the subclue with the answer, the new question reads as follows: Of Bhutan, Cuba, Iran, and North Korea, the one that's farthest North.
You need to solve a puzzle.	BEFORE and AFTER: 13th Century Venetian traveler who's a Ralph Lauren short sleeve top with a collar. Answer: Marco Polo	Two subclues have answers that overlap.

Table: Answering Different Types of Jeopardy! Clues

Self evaluation: Exercise 22

- To continue with the training, after learning the various steps involved in cognitive analytics and watson machine learning, it is instructed to utilize the concepts of cognitive machine learning algorithms to perform the following activity.
- You are instructed to write the following activities using Python code.
- Exercise 22: Question Classification

Checkpoint (1 of 2)

Multiple choice questions:

1. IBM cloud is from which category?
 - a) Public cloud
 - b) Private cloud
 - c) Hybrid cloud
 - d) Community cloud

2. Which of the following is essential concept related to cloud?
 - a) Reliability
 - b) Productivity
 - c) Abstraction
 - d) All of the mentioned

3. Which of the following is cloud platform by IBM?
 - a) Azure
 - b) AWS
 - c) Cloudera
 - d) Bluemix

Checkpoint solutions (1 of 2)

Multiple choice questions:

1. IBM cloud is from which category
 - a) **Public cloud**
 - b) Private cloud
 - c) Hybrid cloud
 - d) Community cloud

2. Which of the following is essential concept related to cloud?
 - a) Reliability
 - b) Productivity
 - c) **Abstraction**
 - d) All of the mentioned

3. Which of the following is cloud platform by IBM?
 - a) Azure
 - b) AWS
 - c) Cloudera
 - d) **Bluemix**

Checkpoint (2 of 2)

Fill in the blanks:

1. You cannot count on a cloud provider maintaining your _____ in the face of government actions.
2. All cloud computing applications suffer from the inherent _____ that is intrinsic in their WAN connectivity.
3. IBM _____ provide cognitive services.
4. _____ connection are required to run IBM cloud.

True or False:

1. IBM Watson is the part of IBM cloud. True/False
2. IBM cloud required Internet connection for operations. True/False
3. Any preloaded software required to run Watson. True/False

Checkpoint solutions (2 of 2)

Fill in the blanks:

1. You cannot count on a cloud provider maintaining your privacy in the face of government actions.
2. All cloud computing applications suffer from the inherent latency that is intrinsic in their WAN connectivity.
3. IBM Watson provide cognitive services.
4. Internet and data connection are required to run IBM cloud.

True or False:

1. IBM Watson is the part of IBM cloud. **True**
2. IBM cloud required Internet connection for operations. **True**
3. Any preloaded software required to run Watson. **False**

Question bank

Two mark question:

1. What is Watson analytics?
2. Mention the group which would be benefitted from the Watson analytics.
3. What else can be accomplished by the business through Watson analytics?
4. Role of cloud in the Watson analytics.

Four mark question:

1. Mention one of the most important feature in the journey of Watson analytics.
2. Name some of the organizations which are using IBM Watson analytics.
3. Give examples of cognitive technologies?
4. What are some of the advantages of IBM Watson analytics?

Eight mark question:

1. What sectors can IBM Watson be applied?
2. What are certain barriers for the adoption of IBM Watson analytics?

Unit summary

Having completed this unit, you should be able to:

- Understand the concepts of IBM Watson with other search engines
- Learn about preparing Watson for commercial applications
- Gain knowledge on Watson's software architecture
- Gain an insight into components of DeepQA architecture
- Learn about Building the Watson Corpus
- Understand the concepts of question analysis, semantic analysis
- Gain knowledge on question classification, hypothesis generation
- Understand scoring and confidence estimation