Watson Discovery

Explore the news and gather insights

What is Watson Discovery?

A suite of APIs to ingest, normalize, enrich, index and search data

Input

- Unstructured data from local files, such as email and chat logs
- Unstructured, semi-structured and structured content from different types of data sources
- Structured content from databases and content management systems
- Structured and unstructured content from Cloud sources

What is Watson Discovery?

Output

- Insights extracted via queries, patterns, and trends from your private content augmented with pre-enriched content from IBM
- Insights extracted from news and other third party and public content
- Augmented response options that chatbots and other applications can offer to complex questions

How it works

1. Create Collection via Online Tools, API calls, Data Crawler*
    a.   Or use existing Collection of Newspaper data
    b.   Currently only for US english
2. IBM Watson adds Enrichments to the Collection
    a.   keywords
    b.   docSentiment
    c.   entities
3. Query the data via Online Tools or API calls

*The data crawler lets you automate the upload of content (for example: file shares, databases, Microsoft SharePoint®)

Queries

- **filter:** A cacheable query that excludes any documents that don't mention the query content.
- **query:** A query search returns all documents in your data set with full enrichments and full text.
- **aggregation:** Aggregation queries return a count of documents matching a set of data values.
- **natural_language_query:** A natural language query enables you to perform queries expressed in natural language (beta).

State

- General Available
- Predefined Collections available
- Trainable (in Beta state)
- A natural language query (in Beta state)

Applications

- Rocket Fuel, a predictive marketing platform
    - Uses Discovery to advise customers on digital marketing, in real time
    - Scan the top 10,000 sources every 7 minutes to provide immediate trend analysis
- Johnson & Johnson, automatic assessment of scientific papers
    - Synthesizing the information directly from the medical literature

Hypotheticals

- Legal research teams can quickly extract relevant details from dense documents
- Customer service departments can find insights on customer pain points
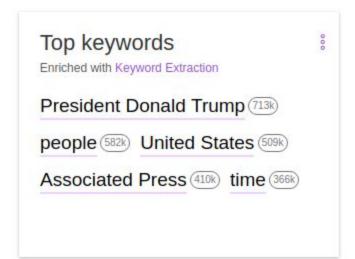
Limits

- No detailed information on the data extraction
- No detailed information on the data analysis or the cognitive intuition
- No detailed information on the stored data or the searching algorithms
- No information on the level of performance
- Only supports US English

Example

Demo Time

Enter a query or keyword

```
spacex
```

Number of results to return (Count)

```
10
```

Default: 10

Narrow your query results (Filter)

```
language:german
```

Group query results (Aggregation)

```
term(entities.text)
```

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
PHILOSOPHISCHE FAKULTÄT
UND FACHBEREICH THEOLOGIE

Links

- https://www.ibm.com/developerworks/library/cc-watson-discovery-service-bluemix-explore/index.html
- https://www.youtube.com/watch?v=9ks-cEG6KPs - Overview
- https://www.ibm.com/watson/developercloud/discovery/api/v1/
- https://www.ibm.com/blogs/bluemix/2016/12/introducing-watson-discovery-service/
- https://www.ibm.com/watson/developercloud/doc/discovery/release-notes.html