Watson Discovery

Lecture Deck

Contents

- Watson Discovery Overview

- Possible Use Cases

- Using Discovery News

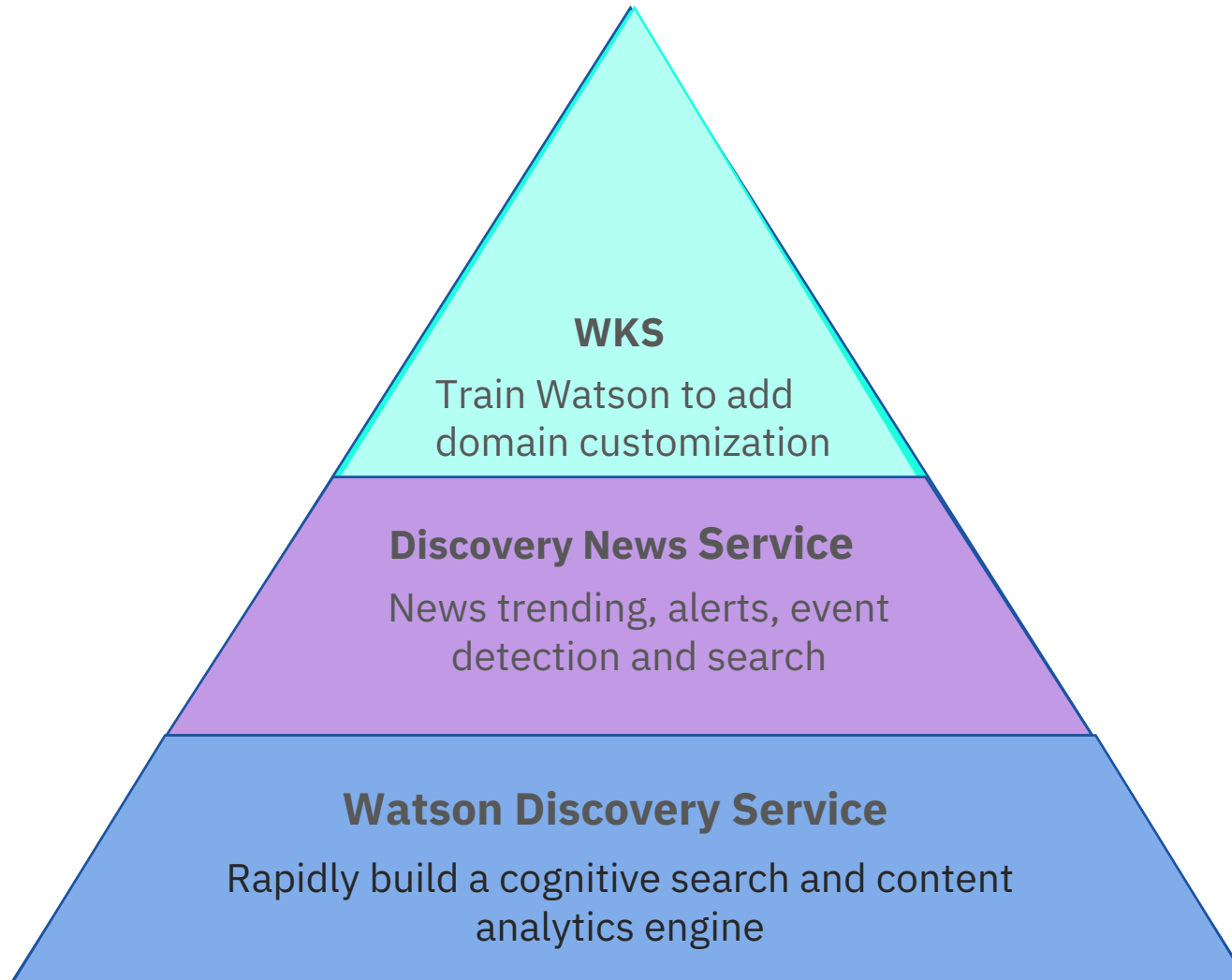- API Fields

- Building Queries

- Links

Watson Discovery Overview

- Bring your own structured or unstructured data – proprietary, public, or third-party data

- Data ingestion, storage, and query in the cloud

- REST API Available

- Discovery news- Pre-enriched news collection with cognitive insights

- Watson knowledge studio - Teach Discovery to apply the knowledge of unique entities and relations in your industry

- Relevance training tool available

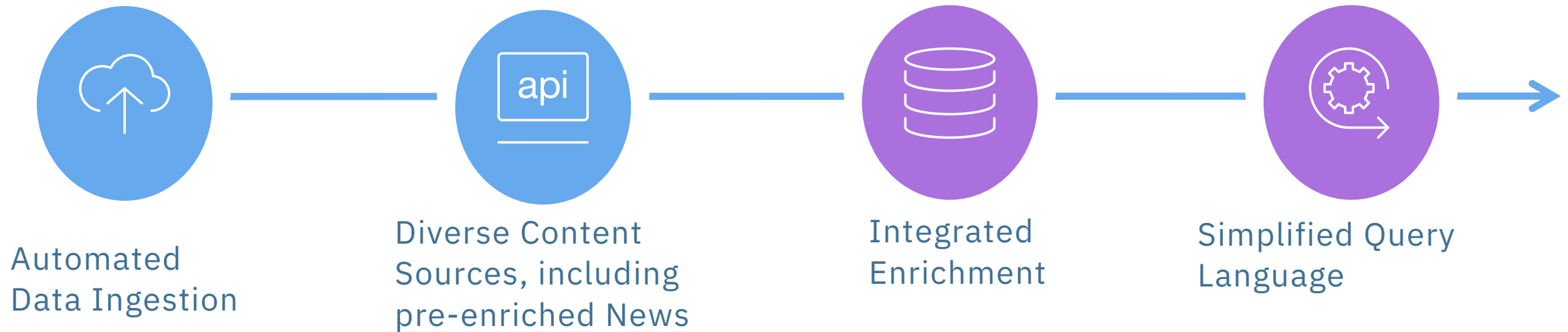- AI powered queries, passage retrieval, relationship graphs and anomaly

Possible Use Cases For Watson Discovery

- **Voice of Customer Analysis -** Look through reviews and surveys and identify patterns and trends in customer opinions.

- **Document Discovery** - Search through a large data set of documents (e.g. white papers, abstracts, patents) for content relating to your field of study

Watson Discovery Service extends functionality to allow for cognitive search and content analytics capabilities to be combined into new and existing applications
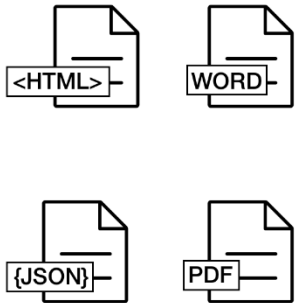
WKS
Train Watson to add domain customization

Discovery News Service
News trending, alerts, event detection and search

Watson Discovery Service
Rapidly build a cognitive search and content analytics engine

With Watson Discovery Service, go from data to insight rapidly and with less effort

Automated
Data Ingestion

Diverse Content
Sources, including
pre-enriched News

Integrated
Enrichment

Simplified Query
Language

Architecture of a complete Discovery service solution

Data

Private data

Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

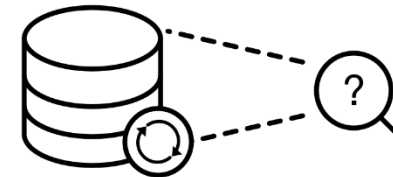**Clean and normalize** through an automated processing of NLP results, improving data quality

Storage

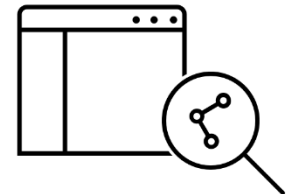Normalized data is **indexed into a collection** as part of your **environment** in the cloud

Query

Understand data faster, create better hypothesis and deliver better outcomes

Output

Actionable insights into your app

Adding Content

- Content to be queried can be added via:

  - Watson Discovery API
    Useful if integrating the upload of content with an existing app or creating a custom upload mechanism

  - Watson Discovery Service Tooling
    Useful if uploading locally accessible files

  - Watson Discovery Data Crawler
    Command line tool that will take documents from the repositories where they reside (for example: file shares, databases, Microsoft SharePoint®) and push them to the Discovery Service. Useful for a managed upload of a significant number of files from a remote system, or to extract content from a supported repository (such as a relational database).

- Added content is run through Watson Natural Language Understanding and enhanced with metadata that contains entities, keywords, concepts, semantic roles, sentiment, concepts etc

How content is organized

- Each document added to Watson Discovery must be associated with a previously defined collection

  - A collection is a grouping of content. At least one collection must be created to be able to upload content.

  - Each query must target a specific collection
    - Single queries against multiple collections are not allowed

Queries

- 2 types of queries

  - Structured – based on Watson Natural Language Understanding enhancement of uploaded content. Allows queries that retrieve doc fragments with specific entities, keywords, concepts, semantic roles, sentiment, concepts etc

  - Natural language – e.g. "Why is the sky blue ?"
    - Uses machine learning and training data to match these queries to the most relevant docs
      - Training data must be uploaded and consists of sample questions and the uploaded docs that are most relevant

Structuring a Basic Query

Queries return results ranked by relevance.
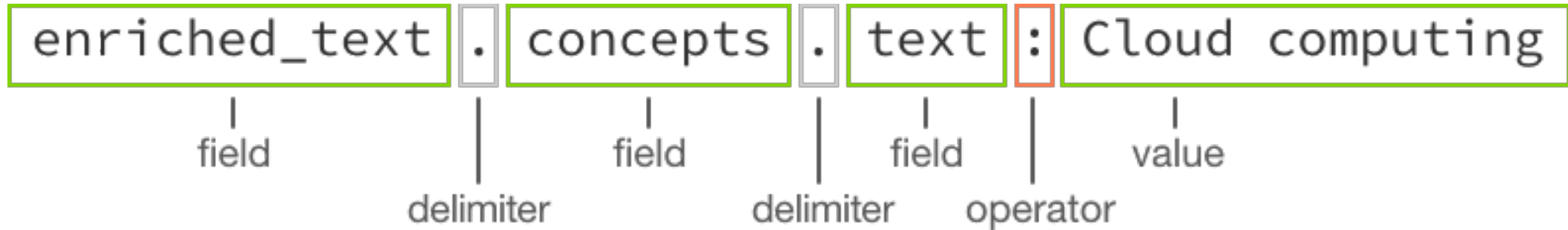
enriched_text.entities.text:IBM

More Complex Structured Queries – Aggregations

Aggregation queries return a set of data values.
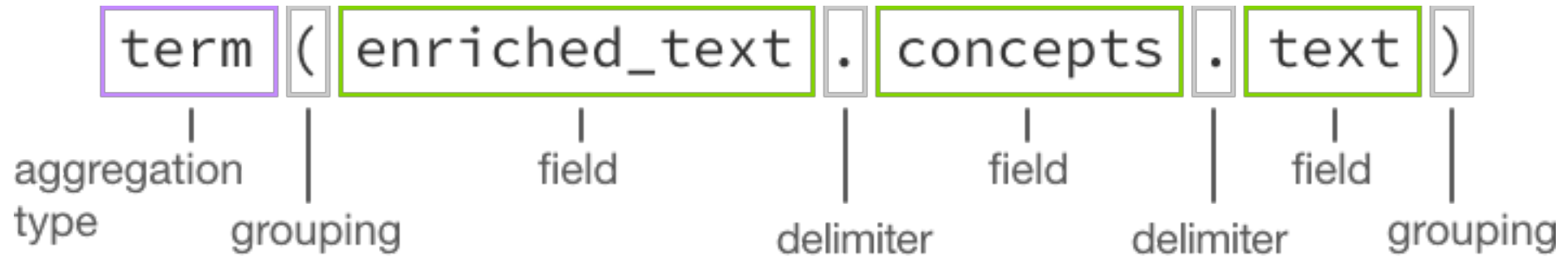
term(enriched_text.keywords.text,count:5)

Fields that can be queried using structured queries

enriched_text.relations.arguments.text
enriched_text.relations.score
enriched_text.relations.sentence
enriched_text.relations.type
enriched_title.relations
enriched_title.relations.arguments
enriched_title.relations.arguments.entities
enriched_title.relations.arguments.entities.text
enriched_title.relations.arguments.entities.type
enriched_title.relations.arguments.text
enriched_title.relations.score
enriched_title.relations.sentence

enriched_title.relations.type
external_links
extracted_metadata
extracted_metadata.file_type
extracted_metadata.filename
extracted_metadata.sha1
forum_title
main_image_url

List of all API fields: https://watson-developer-cloud.github.io/doc-tutorial-downloads/discovery/News_migration_v_1.01.xlsx

AI Powered Queries – Anomaly detection

Anomaly detection using timeslice

- Anomaly detection is currently available only on top-level timeslice aggregations. It is not available in lower-level (nested) aggregations.
- The maximum number of points that can be processed by anomaly detection in any given timeslice aggregation is 1500. The maximum number of top-level timeslice aggregations that can be processed by anomaly detection is 20

timeslice(field:<date>,interval:<interval>,anomaly: true)

grouping

aggregation type

grouping field

data interval

anomaly flag true/false
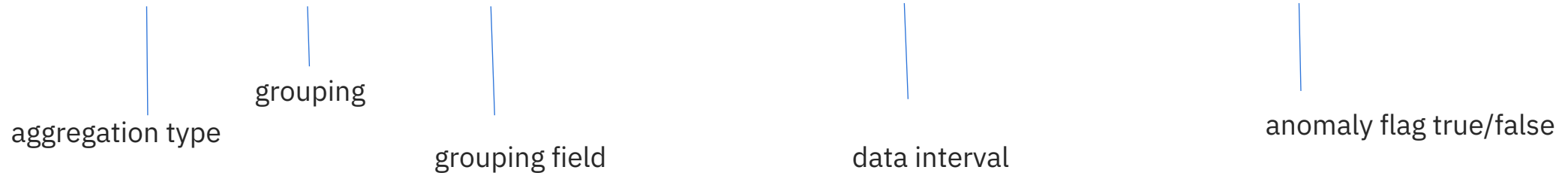
Example

timeslice(field:review_date,interval:1day,anomaly:true)

14
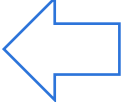
AI Powered Queries – deduplication
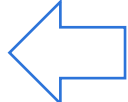
- Deduplication option available only for discovery news.

- Deduplication is enabled, query results excludes the identical and nearly identical documents.

- To enable Add '&deduplicate=true' to your query parameter

Training for Natural language queries

Sample training data for Natural language queries

```
{
    "natural_language_query": "who is keyser soze",          ⇐  **Question**
    "filter": "text:criminology",
    "examples": [                                            **Documents**
        { "document_id": "ada…", "relevance": 2 },       ⇐  **w/best**
        { "document_id": "639…", "relevance": 4 }            **answers**
    ]
}
```

New API Fields

enriched_text.relations.arguments.text
enriched_text.relations.score
enriched_text.relations.sentence
enriched_text.relations.type
enriched_title.relations
enriched_title.relations.arguments
enriched_title.relations.arguments.entities
enriched_title.relations.arguments.entities.text
enriched_title.relations.arguments.entities.type
enriched_title.relations.arguments.text
enriched_title.relations.score
enriched_title.relations.sentence

enriched_title.relations.type
external_links
extracted_metadata
extracted_metadata.file_type
extracted_metadata.filename
extracted_metadata.sha1
forum_title
main_image_url

List of all API fields: https://watson-developer-cloud.github.io/doc-tutorial-downloads/discovery/News_migration_v_1.01.xlsx

Using Discovery

Manage Discovery instance with online tool
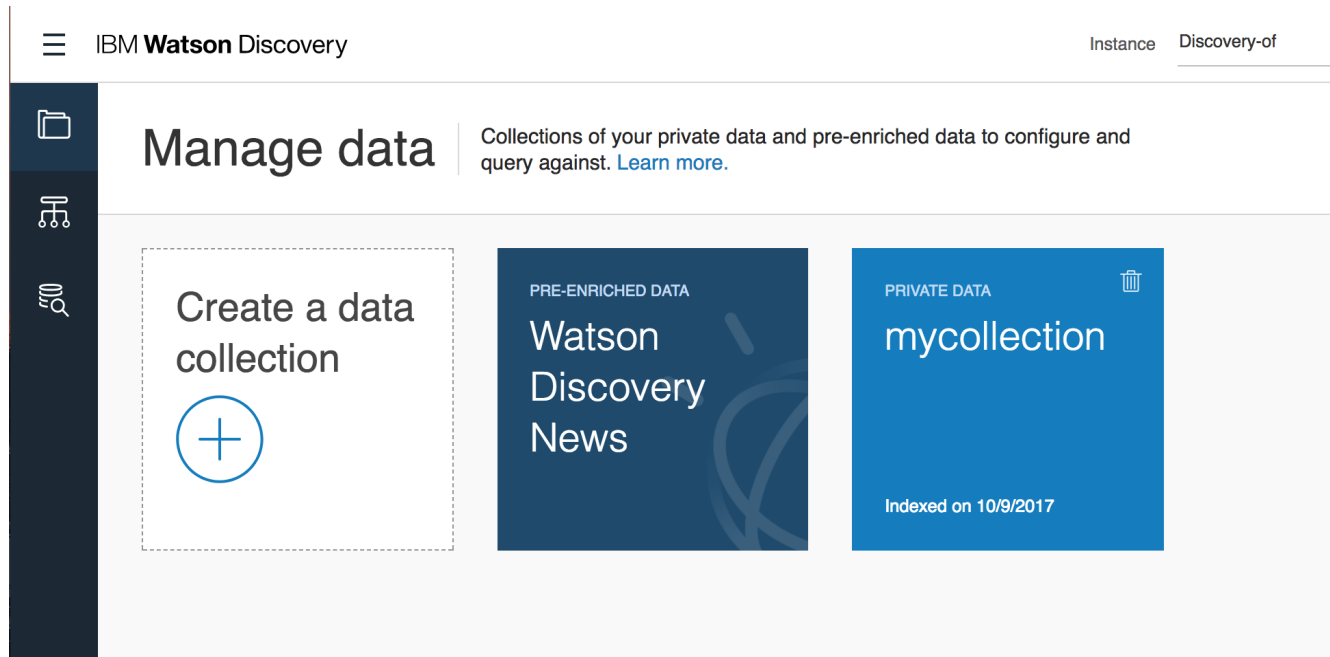
With the tool you can:

- Create/explore/modify/delete collections
- Run queries

Links

- [Short introductory videos](#)
- [Service Documentation](#)
- [Demo application](#)
- [Building a basic query](#)
- [Developer Journey using Watson Discovery](#)
- [Watson Starter Kits using Discovery](#)