

# Knowledge Mining





As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know.

# 2020 *This Is What Happens In An Internet Minute*

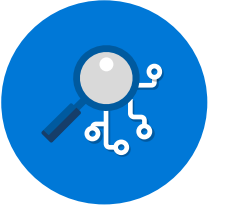


- 1.7MB of data every second by every person
- In the last two years alone, the astonishing 90% of the world's data has been created.
- 2.5 quintillion bytes of data / day are produced
- 95 million photos and videos are shared every day on Instagram.
- By the end of 2020, 44 zettabytes will make up the entire digital universe.
- Every day, 306.4 billion emails are sent, and 500 million Tweets are made.
- 80% of this business data is estimated to be unstructured
- \$14,209 lost per year per employee from information search



Understanding the latent value in all content

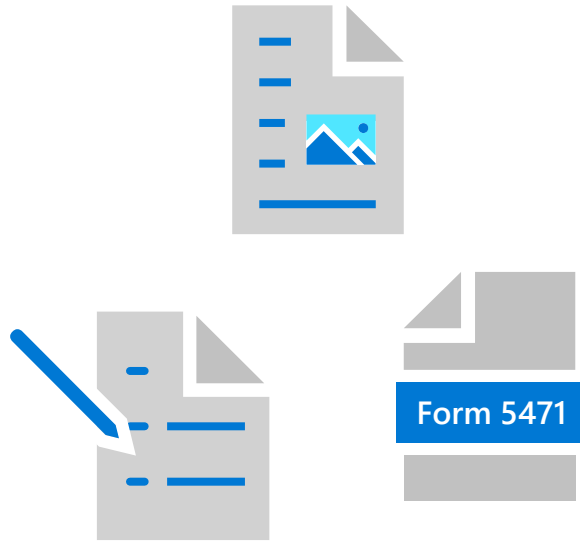
# What if you could...



Turn your **DATA** into **KNOWLEDGE**  
with AI powered search?

The use of advanced data science to understand relationships across multiple sources of structured & unstructured data in a human-like fashion in order to uncover critical business insights and deliver transformational outcomes

# We call this Knowledge Mining

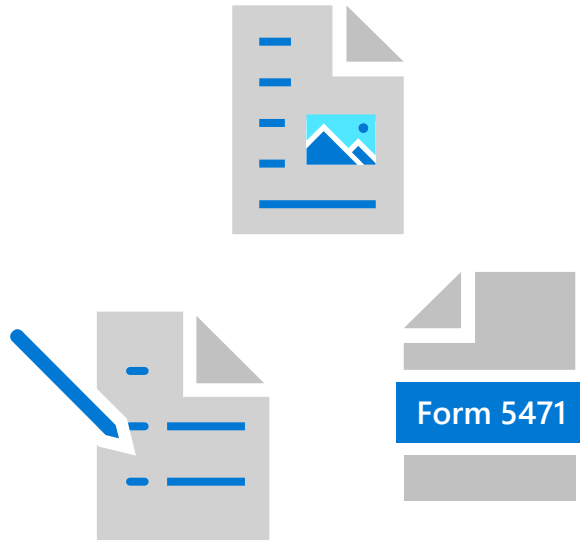


Your documents



Unlock valuable information lying latent in all your content in order to perform an action or make decisions

# We call this Knowledge Mining



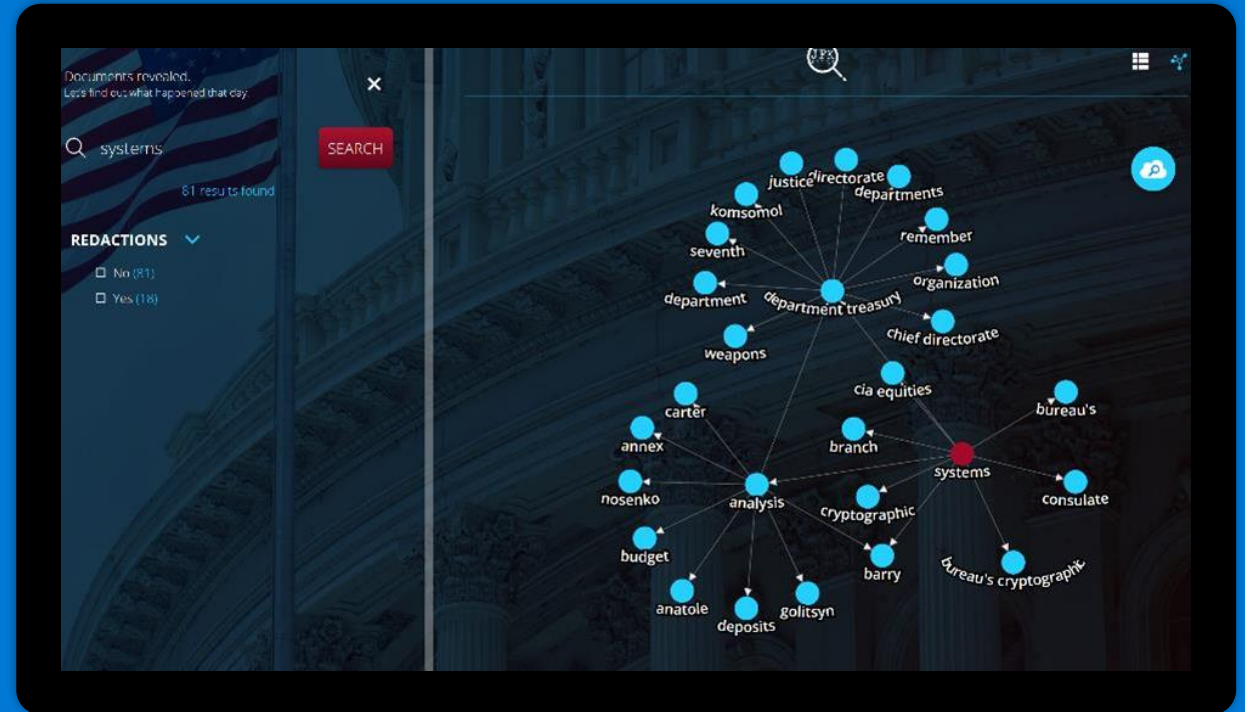
Your documents



Unlock valuable information lying latent in all your content in order to perform an action or make decisions

# Knowledge mining Demo

JFK Files





# At a high level...

INGEST

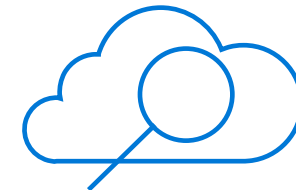
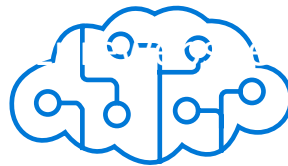
ENRICH

EXPLORE

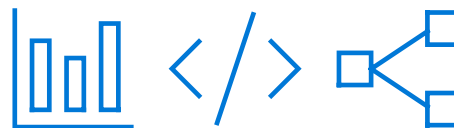
101010  
010101  
101010

Data in any  
format, any  
Azure store

Cognitive skills

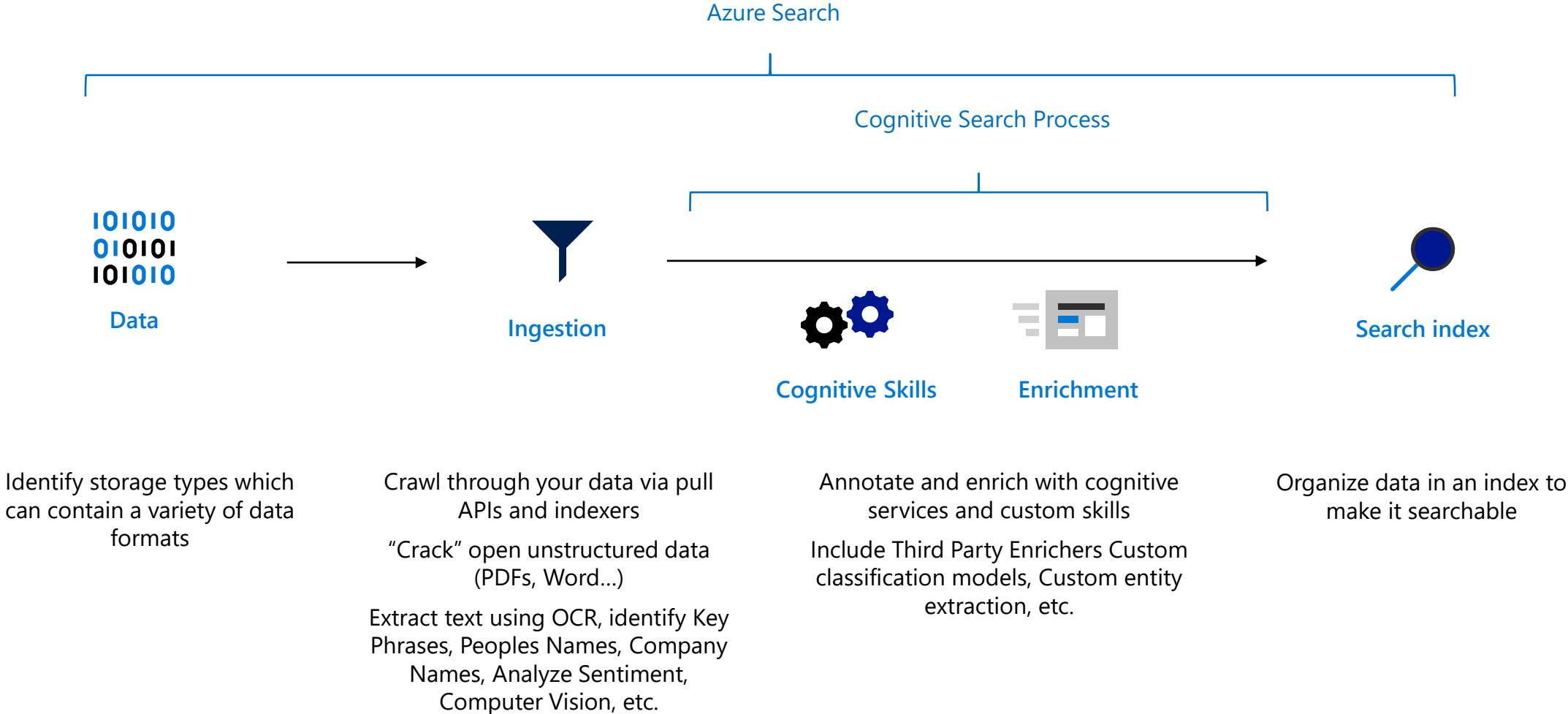


Search



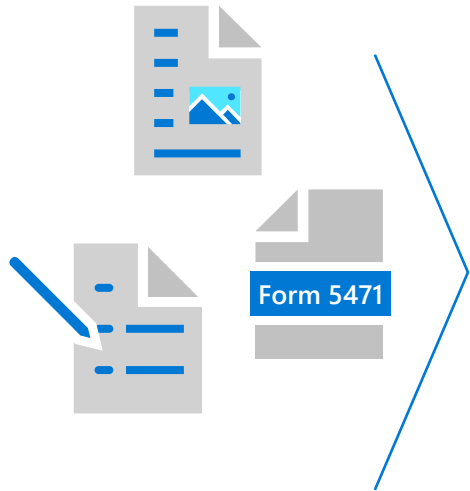
Annotations

# Knowledge Mining – Azure & Cognitive Search

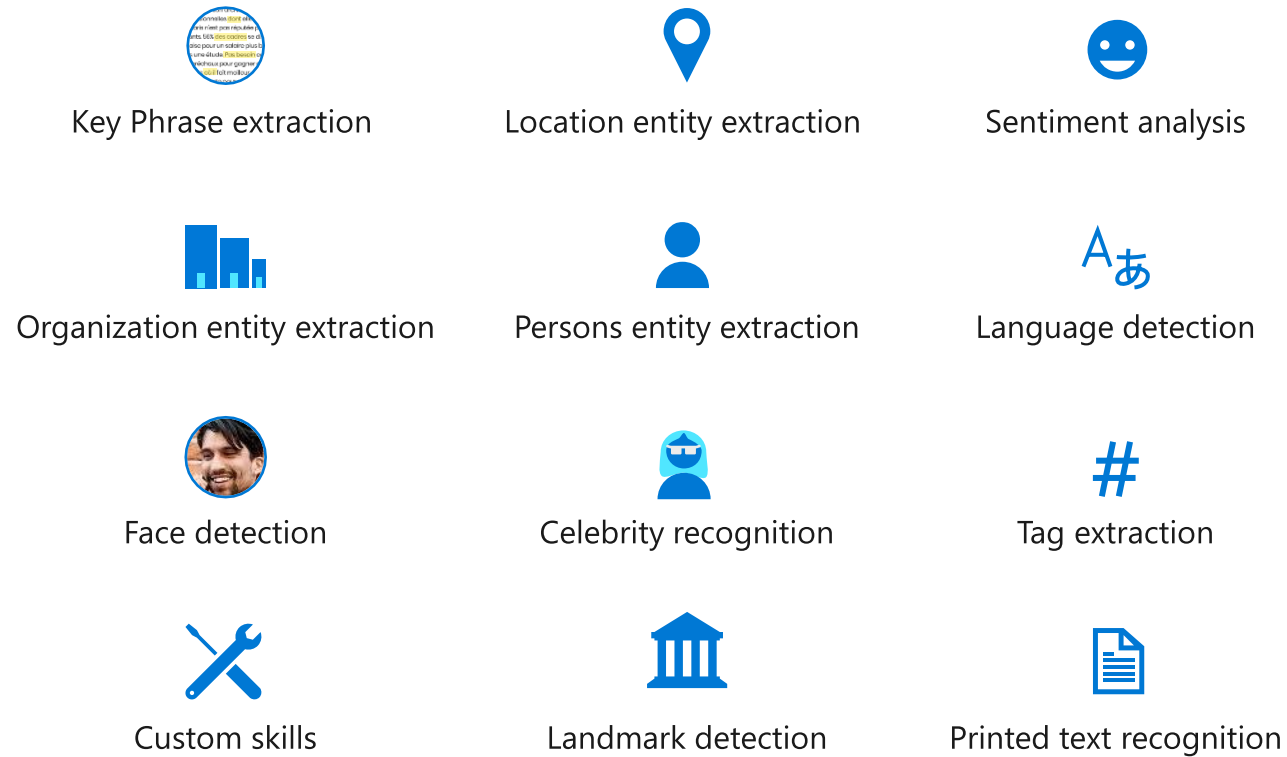


# Knowledge mining with Azure Search

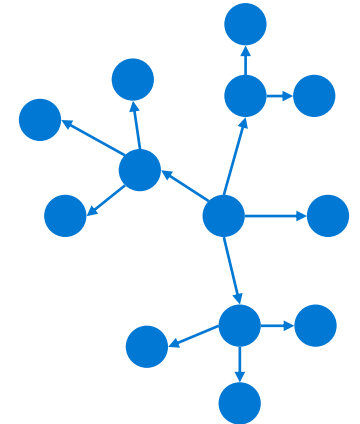
## Information



## Cutting-edge AI



## Fully text-searchable rich index



# Top Scenarios & Industries



## Customer Facing Search

*"I want to surface relevant product details or content to my customer"*

**Retail  
Telecommunication  
Media & Entertainment**



## Internal LOB Search

*"I want to manage my business content and have a one-stop-shop to search everything"*

**Cross-Industry  
Application Integration, Finance,  
Legal, Sales, Support, HR**



## Entity Linking and Document Matching

*"I want to match, group, and merge related material across my data sources"*

**Cross-Industry  
Knowledge Discovery  
Legal, Compliance, Media, Gov...**



## Contract Management & Custom Alerts

*"I want to deeply understand my data in order to filter and drill down on specific signals"*

**Cross-Industry  
Knowledge Discovery  
FSI, Retail, Healthcare, ...**

# Introducing Form Recognizer Cognitive Service (public preview)

Easily extract text, keys, and tables from your documents

The image shows a document with a header, contact information, a table, and a sidebar. The sidebar on the right lists extracted key-value pairs from the document.

**Contoso**

Bill To: Contoso, Ltd      Phone: 432-555-0189      Invoice #: 3-456-2  
Fax: 432-555-0123      Invoice Date: 4/14/2019  
Email: contoso@example.com

Invoice For: Project 2

Item #	Description	Qty	Unit Price	Discount	Price
Z4567	Invoice 3-456-2 Data 1	39	\$ 5.00	\$ -	\$ 195.00
Z4568	Invoice 3-456-2 Data 2	40	4.00	5.00	155.00
Z4569	Invoice 3-456-2 Data 3	30	6.00	7.00	173.00
Z4570	Invoice 3-456-2 Data 4	40	7.00	-	280.00
Z4571	Invoice 3-456-2 Data 5	10	4.00	-	40.00
Z4572	Invoice 3-456-2 Data 6	5	8.00	-	40.00

**Key-value pairs**

-----

Bill To: Contoso, Ltd  
Phone: 432-555-0189  
Invoice #: 3-456-2  
Fax: 432-555-0123  
Invoice Date: 4/14/2019  
Email: contoso@example.com  
Invoice For: Project 2

# Why Azure for Knowledge Mining?



Fully managed search experience

**AI-optimized** and **industry leading** ingestion, annotation, and indexing capabilities, out-of-the-box

**Fully managed service** that you can configure rather than code

Use **any Azure storage solution** and easily schedule to update on a regular time interval

Document cracking + OCR ingests both structured and **unstructured data**

[Azure Search Compete FAQ](#)



One-click integration of AI services

Choose from a variety of Cognitive Skills to derive **more understanding from your data**

**Vision**



**Language**



**Speech**



**Search Services**



**Knowledge**



Customize to your specific business needs

Orchestrate **custom AI models** and run updates only on the section of the solution you need

**Custom skills to train your data**

- Classifications models
- Entity extraction
- Scoring models
- Custom analyzers
- Complex type shaping



Secure and reliable

Built for the **scale and performance** needs of enterprises, with **advanced security and privacy features**

SLA of **99.9% availability** with industry leading privacy and security

- ✓ FedRamp
- ✓ ISO 27001:2013
- ✓ SOC 2Type 2
- ✓ GxP (21 CFR Part 11)
- ✓ HIPAA and the HITECH act
- ✓ HITRUST
- ✓ PCI DSS Level 1
- ✓ Australia IRAP Unclassified

