

Information extraction

Transforms unstructured content into structured fields.

Output example

```
{  
  "customer": "Maria Chen",  
  "product": "TrailMaster",  
  "issue": "damaged screen"  
}
```

Inputs can include

- Documents and forms
- Images
- Audio
- Video

Key message

The modality can change. The goal stays the same: produce usable data.

Extract fields from documents

Populate predefined fields from documents and forms.

Example transformation · “PDF invoice → structured invoice record”

Possible fields

- Vendor
- Invoice number
- Date
- Total
- Line items
- Contract parties

Common inputs

- Invoices
- Receipts
- Contracts
- Tax forms
- Applications
- Support documents

Exam signal

Populate predefined fields from documents.

Extract data from images

Images may contain text, identifiers, objects, tables, or visual evidence.

Example · “Photograph of a shipping label”

Structured fields

```
tracking_number: "1Z8042"  
destination: "Seattle"
```

Images may contain

- Printed or handwritten text
- Labels and identifiers
- Objects and attributes
- Tables or forms
- Visual evidence

Key distinction

Read all visible characters → OCR. Return specific business fields → information extraction.

Extract information from audio

A recording can yield defined fields—not only a transcript.

Example · “Recorded support call”

Structured fields

```
intent: replacement_request  
product: TrailMaster  
customer_sentiment: negative
```

Possible structured outputs

- Speaker names
- Product mentioned
- Customer intent
- Issue category
- Dates and commitments
- Action items

Key distinction

Produce a transcript → speech recognition. Derive defined fields from the recording → information extraction.

Extract across time and modalities

Video can combine visual scenes, spoken language, on-screen text, and events over time.

Example · “Training video → topic, presenter, equipment shown, and chapter timestamps”

Video can combine

- Visual scenes
- Spoken language
- On-screen text
- Events over time

Possible outputs

- Participants
- Topics
- Products shown
- Key events
- Timestamps
- Action items

Exam signal

Extract fields across frames, audio, text, and time—not from one still image.

Define the output before processing the content

The schema is the contract for what extraction must return.

Example schema

Field	Type
Customer name	String
Product	String
Issue category	Category
Damage visible	Boolean
Requested action	String

Why schemas matter

- Predictable output
- Easier validation
- Consistent processing
- Direct application integration

Key message

Extraction is not merely “tell me about this.” It is “return these fields.”

Azure Content Understanding

A Foundry capability for analyzing multimodal content and producing structured results.

Conceptual flow · “Content → Analyzer → Defined schema → Structured result”

Designed for

- Documents and forms
- Images
- Audio
- Video
- Repeatable extraction workflows

Exam connection

- AI-901: Associates information-extraction solutions with Azure Content Understanding in Foundry Tools.

Study guide

learn.microsoft.com/credentials/certifications/resources/study-guides/ai-901

Content Understanding versus multimodal prompting

Content Understanding	General-purpose multimodal prompt
Schema-first	Question-first
Repeatable field extraction	Flexible conversation and reasoning
Consistent structured results	Usually open-ended responses
Process many similar content items	Explore or interpret individual inputs
Application and workflow integration	Interactive assistance

Content Understanding

“Extract customer, product, damage_type, and requested_action.”

Multimodal prompt

“Look at this photograph and explain what may have happened.”

Rule of thumb

Known fields at scale → Content Understanding. Open-ended question → multimodal prompting.

A unified customer record

“Combined output: one structured case for the support agent.”

Input	Extracted information
Claim form	Customer and order number
Damage photograph	Product and damage type
Recorded call	Intent and requested resolution
Unboxing video	Event and timestamp of damage

Key message

Different media can contribute fields to the same business object.

An insurer processes thousands of claim photographs, forms, and recorded statements. Each claim must produce the same predefined fields for a downstream workflow.

Which approach is most appropriate?

A. General-purpose multimodal chat

B. Image classification only

C. Azure Content Understanding

D. Speech synthesis

Answer reveal

C. Azure Content Understanding

The solution must process multiple modalities repeatedly and return a consistent structured schema.