How Content Intelligence Transforms the Enterprise

Making unstructured content more valuable for the enterprise

Produced for:

ABBYY

Unstructured Content Dominates the Enterprise

Enterprises today have more data than ever before. Between documents, images, emails, online data, and videos, organizations are drowning in data. Up to 90% of the content in the enterprise is in the form of unstructured data -- which is any data that's not organized in a way that computers can easily process the information. Structured information is usually classified and organized in databases and data stores with schema and well-organized and arranged information. Unstructured data may be in digital format, but there's no predictability to the structure or schema of that data. Examples of unstructured data includes emails, documents, images, videos, social media posts, and the wide range of documents in paper and converted forms including invoices, purchase orders, communications, contracts, applications, IDs, meeting notes, and contracts. The rate of unstructured data grows at an astounding 55%-65% per year.

Because of the sheer quantity of unstructured data and the rate of growth in the enterprise, organizations are increasingly looking to make sense of this broad content of data, and gain intelligence from the valuable information. Currently, much of unstructured data goes unanalyzed and valuable information is "lost" and unable to be used because of the complexity of extracting valuable information and making use of it without requiring significant human labor to process and understand.

- **Up to 90%+ of content is unstructured, growing at up to 65% per year.**
- **Complexity of processing unstructured content prevents organizations from gaining value from this significant data asset**

Forward-thinking organizations are taking advantage of the tremendous advancements in computing, storage, and software technology to bring the necessary understanding to content to make intelligent business decisions. Specifically, the development of artificial intelligence (AI) and machine learning (ML) technologies have given rise to the ability to extract more information, and *intelligence* from the wide range of *content* in the enterprise, whether structured or unstructured. Enterprises are increasingly looking not only for technology solutions, but also greater service capabilities, delivered as skills that they can leverage throughout the organization.

> ***"Content Intelligence enables the digital workforce with the necessary skills and understanding to make intelligent business decisions.***

Content intelligence is the idea of applying AI and cognitive technologies to extract useful value and gain intelligence and learning from various forms of content, derive meaning and intent of documents, and add decision-making intelligence. Content intelligence enables organizations to:

- **Handle the full range of data in the organization**
- **Free up human labor to handle processing of data and documents**
- **Give a competitive advantage in document-intensive industries and business processes**
- **Engage customers more proactively to deliver better service and business processes**
- **Engage customers in a more proactive manner and deliver better service**

The Four Levels of Content Intelligence

Applying cognitive technologies to content is not an all-or-nothing thing. Indeed, the greater the value an organization wants to get out of its content, the more cognitive technology it needs to apply. Each increased level of content intelligence providers greater value to the organization, but also requires increasingly more advanced technology to realize those goals. Content intelligence in many ways is not a destination, but a journey - one of increased content value from data to increased application of cognitive technologies.

To get a better idea of the journey an organization needs to take to achieve content intelligence, there are four levels of content intelligence that reflects the incremental value enterprises can realize. This journey starts at Level 0, which is not particularly intelligent from a content perspective, to Level 3 which is the ultimate goal of this content intelligence journey. The idea is to get organizations to the point where systems can on their own generate significant intelligence and value from content without requiring human intervention. The goal is an intelligent system that can process a wide range of content in many forms, and provide the deep understanding that organizations need. Moving up the ladder of cognitive ability results in increasingly greater value to business organizations by tackling increasingly harder business problems of increasingly more strategic value.

Table 1: Levels of Content Intelligence

Level 0: Digitization	**Level 1: Classification, Extraction, and Learning**	**Level 2: Understanding Content Context and Details**	**Level 3: Semantic Understanding, Meaning, and Relationships**
➢ Turn typed or handwritten text into digitized machine-readable text	➢ Classify content based on the layout and text associated within it ➢ Extract and identify text in the relevant portions of content ➢ Application of machine learning models	➢ Applying Natural Language Processing (NLP) to understand content ➢ Determining patterns in content learned from other documents	➢ Understanding the meaning of content ➢ Connecting documents and content together into longer threads and conversations (Content Context)

Source: Cognilytica

The more AI-enabled an intelligent content solution is, the more it can handle tasks that have previously been handled by employees. In addition, there are many times when information is incomplete, requires additional enhancement, or a combination with multiple sources to complete a particular task. For example, vendor information on an invoice would need to be verified against other systems.. The goal of content intelligence is to apply cognitive technologies to automatically understand and create meaning of unstructured content, transforming into structured data that can be connected into various systems and processes.. In this way, content intelligent systems can help eliminate many of the situations that require human handling, eliminating error and speeding up processes.

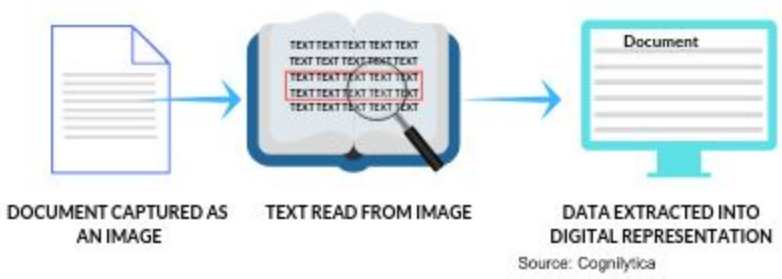
Content Intelligence Levels Explained

It's important to realize that gaining intelligence from content requires more than simply digitizing information into a form that computers understand. To truly extract more value requires applying machine learning and cognitive technologies that provide increasingly more capabilities. The levels mentioned above are further explained below. These are referred to as "levels" in much the same way that the automotive industry as defined "levels of automation" for autonomous vehicles and other industries have adopted the concept of levels of intelligence for their various AI and machine learning applications to industry solutions.

Level 0: Digitization

At this level, the primary objective is to get information that computers can't process at all into a format they can handle. The goal here is one of simply digitizing information. At this level, organizations are simply encoding their paper-based documents or images into machine readable text. Known as "Optical Character Recognition" (OCR) or document / data capture, this technology has been around for decades and is in prevalent use today. This step is incredibly important and necessary for organizations that deal heavily with document or image data, and it is a required step before moving to greater, more intelligent levels of content intelligence.

Level 0 Content Intelligence: Digitization and Data Capture



Source: Cognilytica

However, the journey doesn't stop here. In fact, it's just the foundation to the digital journey. While useful, digitization of content does not provide any intelligence on the content in itself. The objective of tools and technologies at this level is to simply take documents that originated in paper form (handwritten invoices, checks, and various forms of paper documents) or images captured from various different devices and sources and store them for use by a computer system.

Level 1: Classification, Extraction, and Learning

In the business value chain, Content Intelligence platforms are able to digitize content, identify the structure and elements (fields, tables, boxes) of a document, locate and extract data, and perform business rule validations to ensure a high level of accuracy around the data. We see "entity" extraction as part of the level. Level 1 content intelligence systems enable processes and the new digital workforce to become smarter and faster with handling content. For example, level 1 content intelligent platforms can accurately classify a supporting loan document like a W-2 or paystub, and route to the appropriate loan officer. Or, it can determine that a document is an invoice based on the layout of the page.

In addition to document classification, data extraction is crucial for moving up the ladder of content intelligence. Level 1 content intelligence platforms can automatically identify and extract data such as names, organizations, locations, dates, quantities, monetary value, payment terms, and other important information from orders, invoices, contracts and other documents. It can recognize patterns of data that fit criteria and use these matched patterns to extract information that might be valuable or relevant for further processing.

Level 1 Content Intelligence: Entity Extraction, Classification, and Structure

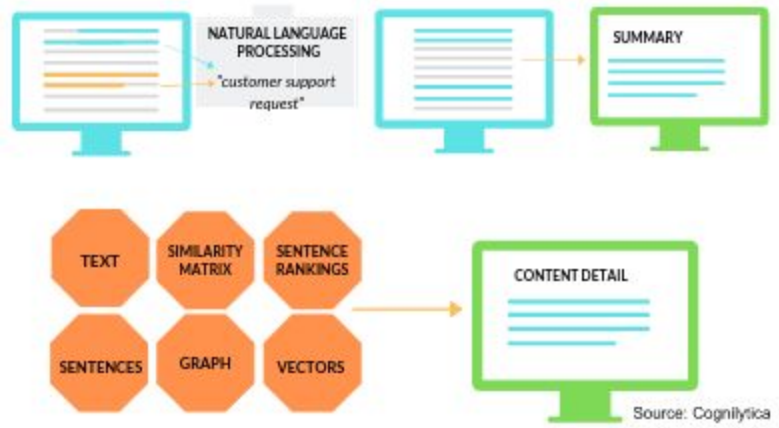
Source: Cognilytica

At this level of content intelligence, we're now applying machine learning. The machine learning systems are classifying a document based on the structure and text contained with a document. The system starts with a "single truth" document for each type and are being trained on variations of the document so that over time they get more intelligent and better able to handle the wide range of document types in the organization. The key to these systems is not requiring thousands of documents to begin learning which enables organizations to realize quicker results in the form of operational efficiency. Overtime, the system creates new learning models based on expanded sample set and user input eventually getting to a point where little human intervention is required.

Level 2: Understanding Content Context and Details

Until this level, we've been primarily concerned with getting the document into a format where machines can process the data in various structured ways. However, to gain more value, systems need to be able to actually understand what's in the content itself. At level 2 Content Intelligence, the system applies natural language processing (NLP), content summarization and other technologies to actually understand what the content is actually about and the context for what the content is regarding. For example, a level 2 content intelligence system can accurately examine and understand contracts and leases, identifying the terms, clauses and commitments within the document, comparing and analyzing contract language.

Level 2 Content Intelligence: Understanding Content Context and Details



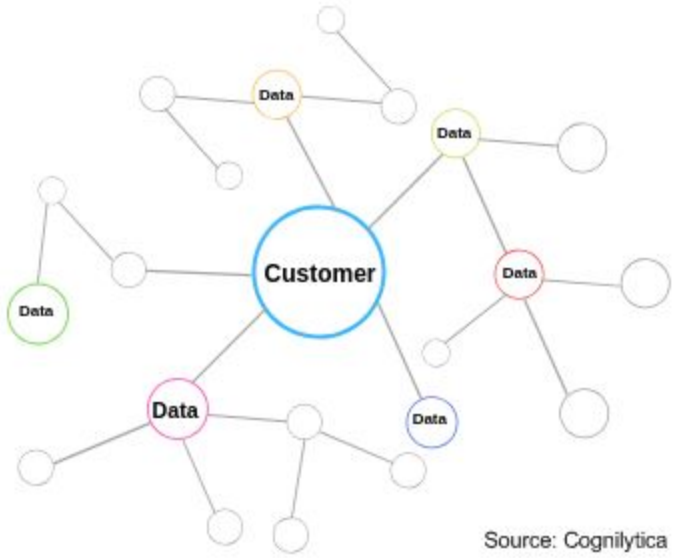
At level 2, the system also has Natural Language Processing (NLP) capabilities. NLP helps to read and analyze text, infer meaning in context and identify important parts of documents by analyzing text relationships and ontologies. NLP systems can actually understand the intent of specific portions of the document and allow users to query documents using natural language and conversational forms of interaction. In this way you're not just searching for keywords but actually "talking" to the system to get access to content that may not even contain the specific keywords that you'd otherwise use. This level of document understanding begins to truly transform and differentiate these systems from non-intelligent counterparts.

Level 3: Semantic Understanding, Meaning, and Relationships

This is the ultimate journey of content intelligence - gaining real understanding and intelligence from data, regardless of its source or format. At level 3 content intelligence, the system can automatically understand what the data and documents mean, and also how they might be related to other documents. The intelligent system can understand that a particular document is part of a larger interaction, thread, or conversation and know how concepts in one data source or document relate to another.

For example, a level 3 content intelligence system in a document setting is where intelligent content systems can not only identify and classify documents such as contracts and invoices, but intelligently understand how these contracts and documents relate to each other and additional data in the organization. These systems can connect the dots and have a full understanding of the content. The system uses interconnected pieces of information and knowledge graphs so that it truly gains intelligence for what the data means and how the data fits in to the overall process and organization, in much the same way as their current human counterparts. For example, if a customer signs a contract, the system is able to understand that the contract is connected to that particular customer and that any purchase orders sent are relevant to that contract. This helps take the human out of the loop and have the system automatically identify and tie all these documents together for the correct customer.

Level 3 Content Intelligence: Semantic Understanding, Meaning, and Relationships



Very few companies have realized this level of content intelligence as it is the edge of what is possible with content intelligence technologies. However, this is the goal. Much of what we have humans do when they are processing documents is not just to understand the letters or even the words on the page, but to understand what the document is all about, why is it important, and how does it relate to the other documents and data in the organization. It's the ability to gain this understanding, the interrelationships with other data, and the meaning behind the information that is what makes level 3 Content Intelligence so powerful and the ultimate goal of applying cognitive technology to the range of information in the enterprise.

Use Cases for Content Intelligence

Increasingly, a wide range of industries are turning to various levels of content intelligence to glean more value from all their various forms of unstructured content. As they make their way up the scale of cognitive intelligence, they continue to gain value. The use cases for content intelligence are broad and span many industries with content intelligence increasingly being leveraged within various automation platforms including robotic process automation (RPA) or business process management (BPM). Applying content intelligence to say RPA makes "digital worker" smarter taking the place of a human worker by read documents, extracting and inputting into systems, determining if a contract is compliant, processing documentation as part of an onboarding process, or initiating an electronic signature as part of a transaction. A few use cases are outlined below but this is by no means a comprehensive list.

Use Case - Accounts Receivable and Accounts Payable

Accounts receivable and accounts payable is one area where content intelligence can have a huge impact on organizations across a wide range of industries. Currently companies make purchases or receive payments that are still paper based. Companies many have hundreds of suppliers they work with, each sending multiple invoices for processing which can quickly add up to several thousands of invoices. However, applying the appropriate intelligence to accounts

payable and receivable process is doing more of the heavy lifting in helping companies create more efficient processes, better accuracy, and doing so at a lower costs.

As organizations apply content intelligence to finance documents like invoices, purchase orders, or payment remittance efficiencies and savings are quickly recognized. These solutions can recognize and extract customer names, addresses, invoice terms, and other necessary information located in invoices. Additionally, these tools can spot anomalies and identify relevant data in invoices, assigning them to various people for approvals or applying internal codes to pass them along to the next step in the payment process.

At higher levels of intelligence, these systems further augment the human workforce. For example someone in accounts payable who is looking for payment information from a specific organization can see the range of interactions with that company and help filter based on the specific relationship or document type they are looking for by using semantic searching for information contained in the documents. Teams are able to locate content that is related to whatever their current need is without having to go through hoops to figure out the right terms to search or hope the content is appropriated tagged and filed.

- ***By applying content intelligence to current workflows TRN Group boosted customer satisfaction with 35% reduction in processing time for creditors' PDF and paper-based invoices.***
- ***Hendre, a large UK housing association was able to automate invoice processing which improved efficiency, reduced time and costs by more than 50% with their content intelligence approach.***
- ***Leveraging content intelligence solutions, PAPCO reduced invoice-processing times by 6 man-hours a day and streamlined approval process, resulting in savings of $30,000 annually.***

Use Case - Shipping and Logistics

Shipping and logistics companies struggle under heavy administrative burdens. These firms produce large volumes of data including various shipping records, bills of lading, customs forms, and other documentation that pertains to the movement of goods on roads, planes, ships, and across borders.

By applying content intelligence, shipping and logistics firms can handle their document-intensive workloads without adding staff to their already overburdened budgets. Starting at level 0, these documents are digitized, then at level 1, can be further analyzed to reveal critical information. The system can automatically classify and validate data from Bill of Lading, Certificates of Origin, waybills and more, providing transparency and awareness of when goods are delivered, lowering operational costs, and initiating billing processes.

Use Case - Financial Services

Banking and financial institutions regularly look to technologies to more effectively serve clients, reduce costs, increase profits, prevent fraud, and meet regulatory compliance issues. Moving

towards digital transformation where more and more banking process are moving online and automating solutions to replace paper with digital documents that are easily searched, accessed, managed, and processed is a goal for most all banking and financial institutions. However merely moving documents online doesn't go far enough. Many banking processes like onboarding a customer, processing a loan, and verifying an individual or business identity involves the processing of content. These institutions need to move up the ladder of cognitive intelligence to better ensure they comply with Know Your Customer (KYC) requirements. When banks do not properly vet a business or individual they expose themselves to being out of compliance and can face significant fines.

Content intelligence benefits for financial services institutions:

- **Reduced cost of compliance for regulatory processes**
- **Increased auditability of systems, data, and processes**
- **Accelerated delivery of mandatory compliance and regulatory requirements**

Content Intelligence tools can automatically read and complete compliance or regulatory policies, fill and file necessary paperwork, and make sure rules and KYC regulations are not being violated. When a bank is doing business with a corporation or individual there are many steps that must be performed to ensure compliance that become part of a KYC file, and much of these steps involve content from various documents. For banks that face large penalties for non-compliance or have other large potential risks, moving up the ladder of cognitive intelligence adds a lot of value.

Taking the Next Steps on Content Intelligence

Enterprises are striving to achieve the goals of Digital Transformation - combining digitized process, data, collaboration, mobility and intelligence to help companies take advantage of the rapidly evolving digital economy. Without content intelligence digital transformation is just not possible. When combined with process automation, content intelligence provides the AI and cognitive machine-learning capabilities and helps automate content-centric processes.

Picking the right vendor to help your organization navigate the journey of digital transformation and content intelligence is important. Efficiency. Speed. Quality. Cost savings. Productivity. These are all things companies need to consider. This whitepaper has explored just a few industries that can benefit from content intelligence. But this is by no means a comprehensive list. Engage further to learn how your organization can benefit from adopting content intelligence as part of your digital transformation journey.

ABBYY Quick Overview

ABBYY is a global leader in content IQ technologies and solutions. ABBYY offers a complete range of AI-based technologies and solutions transforming business documents and content into business value. By providing digital transformation solutions to financial services, insurance, transportation, healthcare, and other industries, the company helps organizations achieve the next wave of growth by understanding customers and delivering responsive real-time intelligent systems.

The flexibility of ABBYY AI solutions enables customers to utilize a diverse range of advanced technologies, platforms and solutions for classification, text analytics, data and entity extraction, and data validation via any communication channel and in any format.

Find out how ABBYY can help you on your content IQ journey

ABBYY Content IQ is a class of enabling technologies that help digital workforces understand and create meaning from enterprise content. Content IQ delivers cognitive skills that the new digital workforce can harness to turn your unstructured content into structured, actionable information making your digital workers smarter and your processes run more efficient.

Reach out today to schedule a personalized meeting to find out how ABBYY can meet the exact needs of your business.