KNOWLEDGE MINING

The Next Wave of Artificial Intelligence-Led Transformation

The opportunities created by digital transformation are fueling unprecedented change, driven by advances in cloud, data, and AI. Still, many organizations face challenges when it comes to dealing with, and gleaning knowledge from, the ever-increasing amounts of data under their control. These challenges are exacerbated by the wide variety of content types that make up the bulk of this data, such as PDFs, images, and videos. Transforming all this content into insights requires extensive time, resources, and data science expertise.

This has led to a new wave of AI-powered digital transformation with knowledge mining at its heart—empowering organizations to harness this wide variety of content with applied intelligence, dramatically reducing time to market and unlocking new insights that were previously out of reach. Customers investing in knowledge mining today are seeing tremendous impact resulting in increased business performance and profitability. Financial services firms have reduced the time of financial reporting from months to days, health care companies have been able to create holistic views of their patient data, and manufacturing companies have been able to provide engineers witbetter answers faster to foster customer loyalty.

With knowledge mining, organizations can:

- Consume and process all types of content across multiple locations and formats

- Extract new insights and knowledge from the content by using deeply integrated prebuilt AI technologies across vision, speech, and language

- Surface this knowledge through a variety of tools such as search, data visualization, and business applications to make it available to customers or end users in an organization

This report explores the opportunity organizations have to effectively, efficiently, and contextually understand all their content with knowledge mining, enabling them to engage better with customers, transform products and services, optimize operations, and empower employees to do more valuable and impactful work.

JOHN "JG" CHIRAPURATH
GENERAL MANAGER,
AZURE DATA & AI

KNOWLEDGE MINING
The Next Wave of Artificial Intelligence-Led Transformation

Knowledge mining is an emerging category in artificial intelligence (AI), using a combination of AI services to drive content understanding over vast amounts of unstructured, semi-structured, and structured information that allow businesses to deeply understand their information, explore it, uncover insights, and find relationships and patterns at scale. While the first wave of AI involved many narrow applications, such as training a single model over one data source of a certain type for a single problem, knowledge mining is the next wave of AI, generating a dynamic understanding of relationships and patterns in a corpus of information. It is quickly becoming a key part of enterprise digital transformation initiatives that fundamentally change how organizations make sense of real-world information.

While knowledge mining helps uncover latent insights in all types of information, much of the information that's critical to business exists in unstructured formats like PDFs, images, videos, audio files, paper documents, and even handwritten notes. There's key information in this content that isn't readily visible or available to process. Extracting that vital information at scale is a crippling burden for organizations when they try to do it the traditional way, which usually involves someone scanning, interpreting, annotating, cutting, and pasting. Knowledge mining helps uncover latent insights across all types of information, whether structured or unstructured.

More than two-thirds (68%) of respondents to a recent Harvard Business Review Analytic Services survey believe knowledge mining is important to achieving their companies' strategic goals in the next 18 months. And the need for knowledge mining is quickly growing—77% are using manual methods to handle unstructured information, and those methods will quickly be outpaced by the growth of data and potential use cases in which this information could provide great value.

Organizations clearly need a means to discover and gain intelligence from all of their data reliably and consistently. Knowledge mining has the ability to transform the way businesses understand and leverage their content, and as the survey finds, some pioneers are already finding and exploiting the value in vast troves of previously untapped knowledge to create more insights and opportunities.
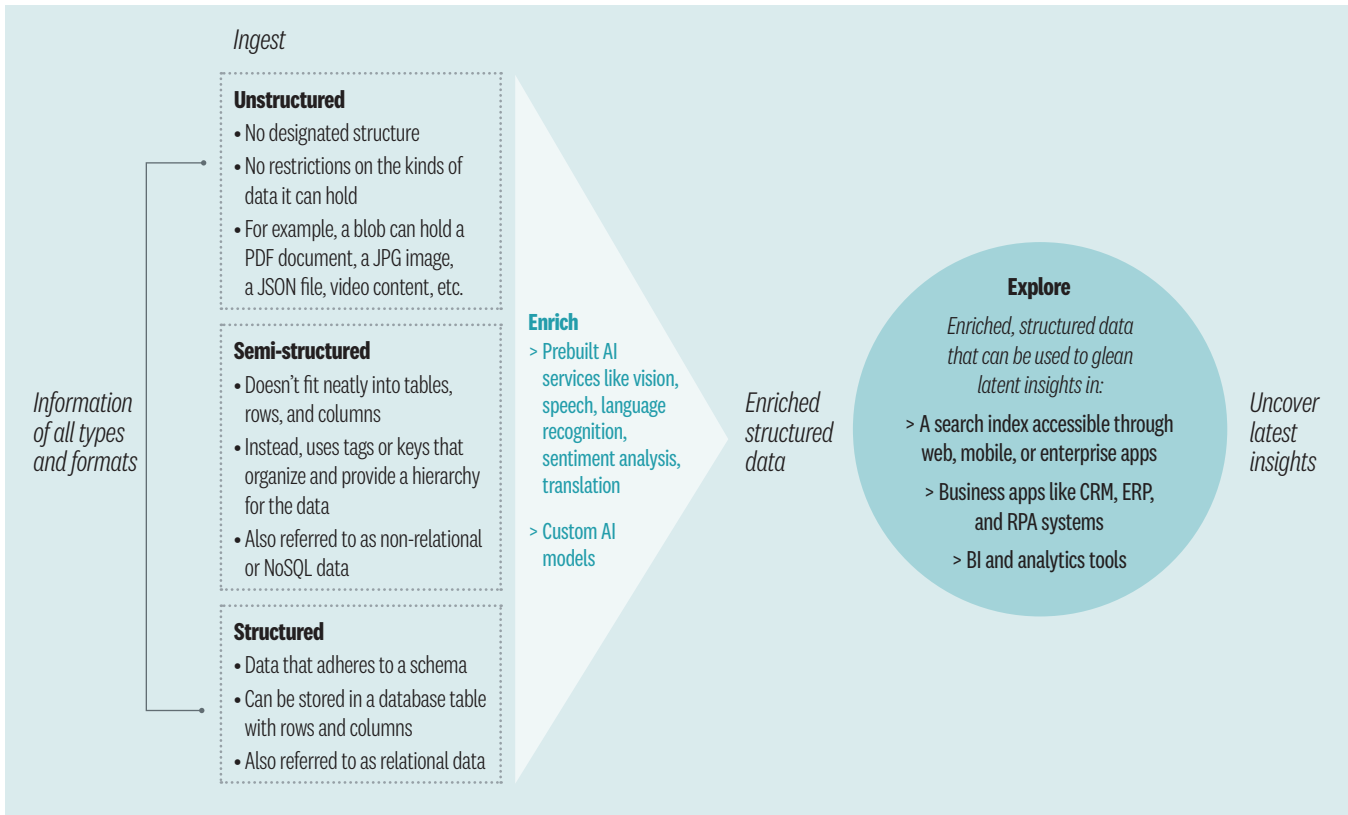
Ingest

Unstructured
- No designated structure
- No restrictions on the kinds of data it can hold
- For example, a blob can hold a PDF document, a JPG image, a JSON file, video content, etc.

Semi-structured
- Doesn't fit neatly into tables, rows, and columns
- Instead, uses tags or keys that organize and provide a hierarchy for the data
- Also referred to as non-relational or NoSQL data

Structured
- Data that adheres to a schema
- Can be stored in a database table with rows and columns
- Also referred to as relational data

Information of all types and formats

Enrich
> Prebuilt AI services like vision, speech, language recognition, sentiment analysis, translation

> Custom AI models

Enriched structured data

Explore

Enriched, structured data that can be used to glean latent insights in:

> A search index accessible through web, mobile, or enterprise apps

> Business apps like CRM, ERP, and RPA systems

> BI and analytics tools

Uncover latest insights

THE BASIS FOR SOME OF THE GREATEST FUTURE INNOVATIONS MAY BE LOCKED, UNSEEN, IN SPREADSHEETS, PDFS, PRESENTATIONS, AND DOCUMENTS.

Emerging Need for Disruption in How Content Is Explored

Most areas of the workplace have seen breathtaking innovation in recent years, but there is one area that remains bound in tradition: the way unstructured information is understood and used. At the same time, the growing volume of unstructured information—information that is critical to business functions—is overwhelming current practices.

Many organizations are struggling to make better use of all of their available information. The great majority of surveyed executives (82%) say that exploring, understanding, and using unstructured information in a timely way is a challenge to their organization.

The survey highlights six key factors that illustrate the need for a better way to explore, understand, and use unstructured information.

1. The footprint of unstructured information in the workplace is growing. A majority of respondents (54%) note that at least half of

the available information in their organization is unstructured. There's key information in this content that isn't readily visible or available to process. FIGURE 1

2. The value from unstructured information remains trapped in traditional formats. The basis for some of the greatest future innovations may be locked, unseen, in the spreadsheets, PDFs, presentations, and documents that are the most common forms of semi-structured and unstructured information. Often, the employees involved are highly trained, highly paid professionals such as risk management specialists and engineers—those with the experience to know what to look for and why it may be relevant. Industry-specific information, such as computer-assisted design (CAD) files, affects one in five of the organizations surveyed. FIGURE 2

3. Critical business functions frequently depend on unstructured information. While difficult to use, unstructured information is needed for a range of business-critical functions.

Companies have identified numerous ways they use their unstructured information, according to the survey. More than half of respondents are using unstructured data today for business performance analysis, knowledge management, and routine business processes and workflows as well as to find patterns and trends and to project outcomes. A significant number (45%) also use it for end-customer experiences and for auditing and compliance. FIGURE 3

In practice, every use case might make multiple uses of unstructured information—if it is accessible and consumable. Because of these dependencies, the accuracy, speed, and completeness of the transformation of raw unstructured information into usable formats have a huge impact on business health and growth.

4. Methods to extract and understand unstructured information are outdated and insufficient. Despite the importance of their unstructured information, most companies are reverting to manual methods to understand it—which usually involves someone reading, interpreting, annotating, cutting, and pasting. The survey shows that this laborious process is a problem for many. Only 28% of respondents think their organization has met its goals for gaining insights from unstructured information. FIGURE 4

5. Manual methods are time-intensive, error-prone—and expensive. Most organizations still process nearly all of this information manually, with workers searching for information and then copying, pasting, or reentering it into another form. The problem is not just the time wasted but the quality and accuracy of what is produced. "Think of all the errors that can happen when someone is moving all that data around, typing and retyping," says Claudia Imhoff, president and founder of Boulder BI Brain Trust and Intelligent Solutions. There is also the problem of reading, interpreting, and understanding vast troves of content. It is far too easy to miss something important when employees need to do this manually. Nearly every employee must perform one or more of these tasks as part of their regular work. At

UNSTRUCTURED INFORMATION ABOUNDS
Percentage of data that is unstructured in respondents' organizations
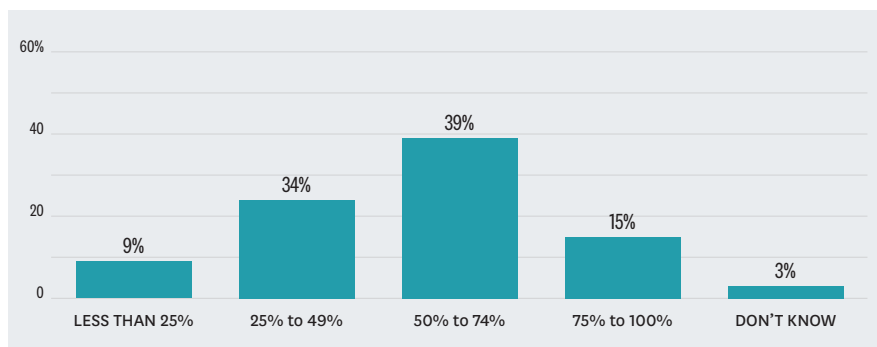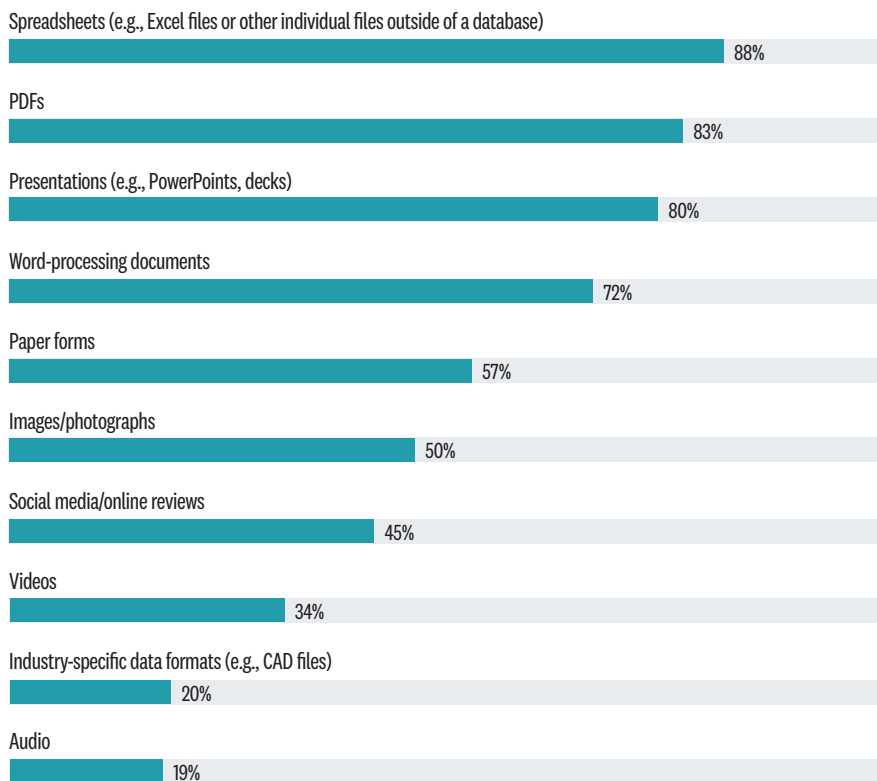

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES, JULY 2019

FIGURE 2

THE MANY FORMS OF UNSTRUCTURED INFORMATION
Types of unstructured information respondents' organizations explore, understand, and analyze manually today
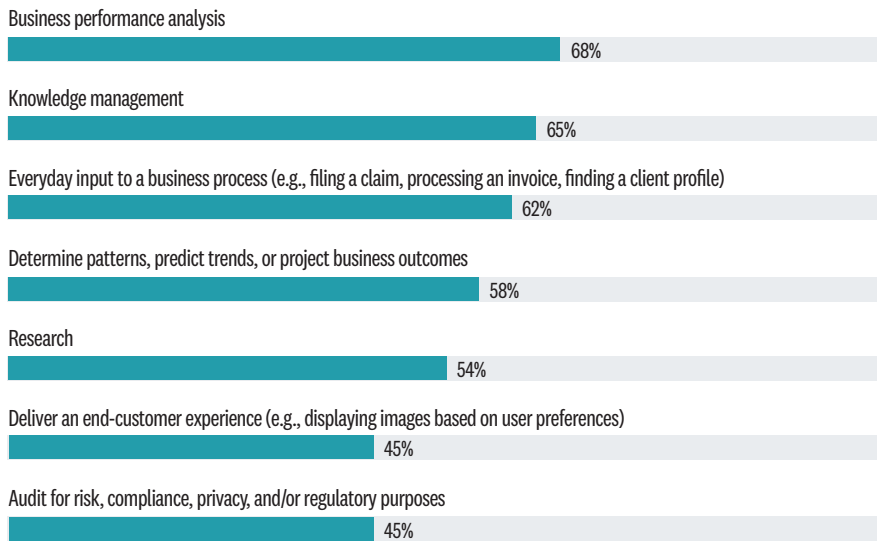

Spreadsheets (e.g., Excel files or other individual files outside of a database) 88%
PDFs 83%
Presentations (e.g., PowerPoints, decks) 80%
Word-processing documents 72%
Paper forms 57%
Images/photographs 50%
Social media/online reviews 45%
Videos 34%
Industry-specific data formats (e.g., CAD files) 20%
Audio 19%

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES, JULY 2019

FIGURE 3

UNSTRUCTURED INFORMATION NEEDS TO BE USED WIDELY

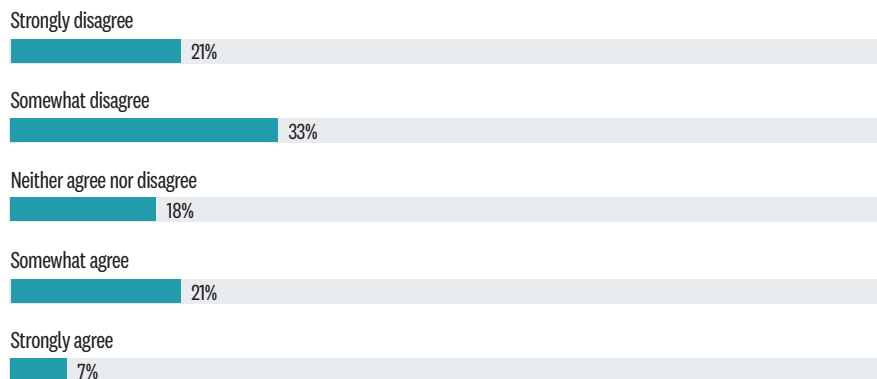How unstructured information is applied in organizations

Business performance analysis

68%

Knowledge management

65%

Everyday input to a business process (e.g., filing a claim, processing an invoice, finding a client profile)

62%

Determine patterns, predict trends, or project business outcomes

58%

Research

54%

Deliver an end-customer experience (e.g., displaying images based on user preferences)

45%

Audit for risk, compliance, privacy, and/or regulatory purposes

45%

FIGURE 4

UNTAPPED POTENTIAL

Insights from unstructured information elude the vast majority of organizations

My organization has met all of its goals for gaining insights from unstructured information

Strongly disagree

21%

Somewhat disagree

33%

Neither agree nor disagree

18%

Somewhat agree

21%

Strongly agree

7%

36% of companies surveyed, more than half of employees are involved in extracting data from unstructured information. FIGURE 5

6. Growing volume of unstructured content creates scalability challenges. Scale is a key factor now and will remain so when it comes to unstructured information. Roughly a third of organizations surveyed are already trying to make use of terabytes of unstructured information—including some that are measuring it in petabytes. FIGURE 6 At this volume, manual processing could never keep up, yet the survey finds that 77% of organizations handle most or all of their unstructured resources manually. Only 6% say that processing is mostly automated.

Knowledge Mining in Action

Nearly all enterprises say they see a clear need to automate the understanding of data from their unstructured information. Thirty percent are already automating the extraction of data from their unstructured information, and another 35% are investigating or piloting automation, according to the survey. More telling, however, are the 24% who say that they have no current automation projects even though they see the need to develop these capabilities. FIGURE 7

Most organizations want to use their unstructured information for analytics (80%), to make it searchable (53%), or to visualize it in a dashboard (49%). One in five organizations is using it to set up knowledge graphs. The ability to transform information into a comprehensible structure can expand visibility into real-world business practices, streamline cumbersome processes, and create new knowledge and actionable insights. FIGURE 8

Knowledge Mining Trailblazers

By using a variety of AI technologies together, knowledge mining can help solve three of the key challenges of handling unstructured information: time, scale, and insights.

Knowledge mining can provide the ability to automatically categorize and curate vast streams of content that would be far too labor intensive to handle manually. Knowledge mining can extract metadata from scanned documents that was never available in a searchable format before. However, the real power of knowledge mining comes when trapped data can be surfaced, analyzed, and illuminated with intelligent methods to solve a real business problem. It can draw out connections in related information, aid in real-time anomaly detection, and index data for more accurate simulations and forecasting.

As these four case studies show, the ability to access and mine unstructured information can be transformational.

CASE 1: CREATING INSTANT INSIGHTS OUT OF TIME-SENSITIVE FINANCIAL INFORMATION

"Across the financial sector, unstructured data is either not extracted at all, which is mostly the case, or it is done manually," says Subhra Bose, founder of Financial Fabric, a financial technology service company. "Even the largest companies still have huge operations where they are manually entering the data from forms and statements into some form of structured repository, as well as teams of developers writing the code necessary to extract data from different sources," he says.
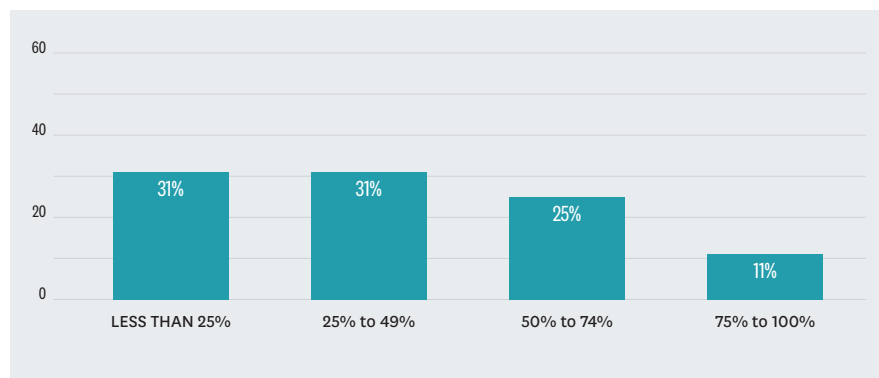
By the time that manual process completes, much of the information is useful only for record keeping or historical analysis. And there is much more relevant information that is never analyzed because it is locked within PDF documents, presentations, printed documents, or spreadsheets. The average investment fund management team, for example, might receive thousands of pages of information every day, far too many to process manually. Bose estimates that fund managers see only 10% to 40% of the information that could be useful to them while it is still relevant.

As the former CTO of a large bank's asset management division, Bose saw

EXTRACTING KNOWLEDGE IS LABORIOUS AND TIME CONSUMING

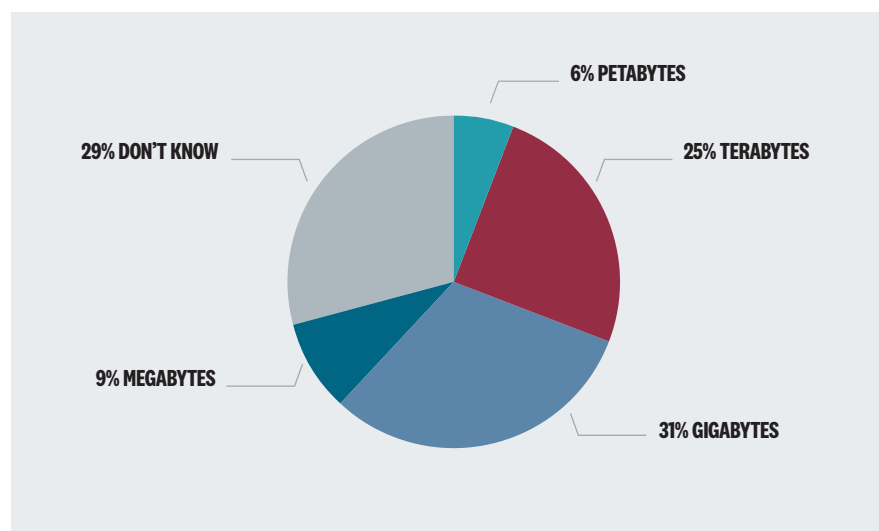Percentage of full-time employees involved in extracting data from unstructured information

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES SURVEY, JULY 2019

A MEASURE OF WHAT'S TRAPPED

The amount of unstructured data being employed can be measured in gigabytes and terabytes
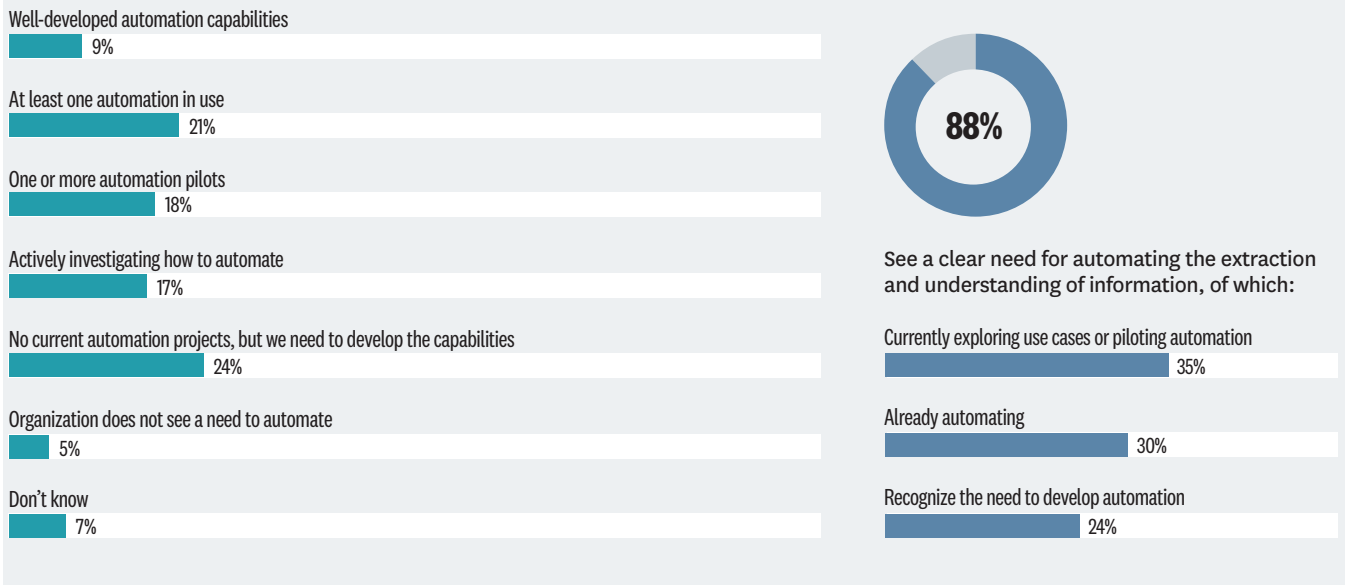

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES SURVEY, JULY 2019

FIGURE 7

ORGANIZATIONS ARE READY FOR KNOWLEDGE MINING

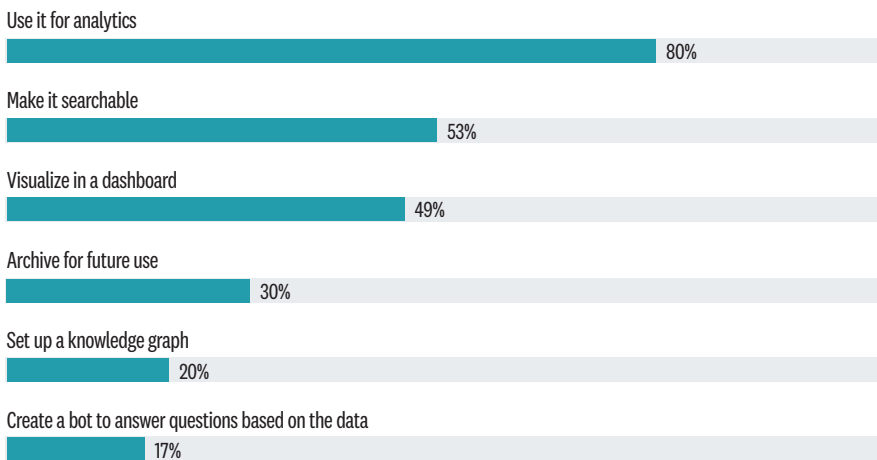Stage of organization when it comes to automating the extraction of data from its unstructured information

Well-developed automation capabilities
9%

At least one automation in use
21%

One or more automation pilots
18%

Actively investigating how to automate
17%

No current automation projects, but we need to develop the capabilities
24%

Organization does not see a need to automate
5%

Don't know
7%

88%

See a clear need for automating the extraction and understanding of information, of which:

Currently exploring use cases or piloting automation
35%

Already automating
30%

Recognize the need to develop automation
24%

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES SURVEY, JUNE 2016

FIGURE 8

DATA SURFACED HAS MULTIPLE USES

Outcomes most important when it comes to extracting data from unstructured information

Use it for analytics
80%

Make it searchable
53%

Visualize in a dashboard
49%

Archive for future use
30%

Set up a knowledge graph
20%

Create a bot to answer questions based on the data
17%

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES, JULY 2019

the opportunity to transform the huge stores of information that remain dark at most banks and normalize them for analytic insights. The firm was able to get ahead of others in the industry by writing custom code to automate the extraction of key data from semi-structured and unstructured sources such as statements, invoices, and investment documents. But in the three years since Financial Fabric was founded, the technology has advanced so dramatically that there is no longer a need for custom coding for a large set of scenarios.

With knowledge mining, it is now possible to train a system to recognize the key data to extract from a statement—whether it is in a PDF, a scanned document, or a spreadsheet format—and to do it consistently. The same is true for more complex processes, such as allocating invoices to the right account or pulling data from investment documents, which can vary in their presentation, and using that data to validate investment terms. Financial Fabric is now uploading documents that are thousands of pages

long and testing new ways to surface insights while they still have value.

For example, a large corporate treasury with many derivatives trades was traditionally able to reconcile data quarterly or monthly at best. The process involves manually entering data from counter-party bank PDF statements. With knowledge mining, reconciliation of the derivatives portfolio can happen on a daily basis, identifying trade breaks, non-captured trades, and other errors. "It is not only reducing risk but also providing other insights, which create different portfolio opportunities," says Bose. "Having a fully reconciled, daily view of their hedge portfolios enables treasury groups to trade more nimbly and with greater accuracy while reducing the potential for errors."

CASE 2: STREAMLINING THE CONTRACTING PROCESS

At Howden, most of the fans, compressors, steam turbines, and other equipment made by the 164-year-old engineering business predate the digital revolution, as does the knowledge to create and service them.

New projects still draw on that traditional knowledge, often stored in paper files, as well as on a range of twenty-first century digital and visual resources. Bidding on large projects involves reviewing and evaluating complex technical standards and material requirements under a tight timeline. Engineers must sort through documents and technical drawings, usually printed on stacks of paper, searching for anomalies and identifying potential risks. "Even our most experienced and skilled engineers can sometimes miss important requirements, which will ultimately result in delays or quality issues that we will need to address later on in the process," says Chris van Ravenswaay, global business solutions manager for Howden. Mistakes can cost up to millions of dollars.

"Traditionally, most information is locked away in paper or electronic archives. Retrieving information required experience and knowledge

"Unlocking this potential will **significantly change the way we do business** with our customers and how we service their equipment," says Chris van Ravenswaay, global business solutions manager for Howden.

of how the archives are structured, how information is organized in documents, the ability to interpret technical and company-specific jargon, and historical naming conventions," he says. "We wondered how we might optimize our consumption of customer requirements and started exploring ways to use knowledge mining to mimic the human brain through intelligent search algorithms."

Howden worked with partners to develop a knowledge mining solution to upload relevant documents that can then search through technical jargon, automatically identify potential risks, specify international standards, and red-flag conflicting specifications. Engineers can annotate directly in the uploaded document, making their comments available to anyone else who uses it.

"In our industry, most of the life cycle of equipment is stored in unstructured data," says van Ravenswaay. "Unlocking this potential will significantly change the way we do business with our customers and how we service their equipment."

CASE 3: SOLVING THE MEDICAL CODING PUZZLE

BUDDI Health is, a deep-learning company focused on contextualizing medical records and claims data to enable business applications across both the financial and the clinical sides for hospitals, clinics, payers, and pharmaceutical companies. The company is building deep-learning models for billing and coding, using knowledge mining to bring a holistic

view to health care providers while also helping to sort through what really matters at that moment. "A hospital stay could easily produce up to 600 pages of documentation, with notes on diagnosis, procedures, laboratory data, patient observations, and drugs prescribed," says CEO Ram Swaminathan. Plus, there are medical histories; images from, say, radiology; and financial records. Between 60% and 80% of the data generated is unstructured, and much of it can only be interpreted manually by an expert.

On the other side of the sheer volume of documentation and data generated by the health care industry is the medical coding system in the United States. In the latest version of the medical coding system, medical practitioners have more than 68,000 possible diagnosis codes from which to choose—five times more than five years ago. Without proper coding, insurers will not pay; 40% of all claim denials are a result of erroneous or missed coding. Keeping track of it all is a crippling burden for the medical profession.

To understand and prioritize content in vastly different formats, BUDDI Health applies knowledge mining to bring together a variety of techniques, including character recognition, natural language processing, and class-imbalance algorithms to help cut through the noise common in faxed documents such as streaks and black spots, and predictive algorithms to sort through anomalies like misspellings that might lead to missed documentation.

Knowledge mining offers a way for organizations to compile, understand, and see the relationship between disparate sources and formats of information as a structured story line in the form of a graph. For example,

knowledge mining can take any element in the available content, such as an MRI image or a diagnosis and its corresponding radiologist report with technique, impressions, and findings text, and show it in a web of relevance with every other piece of information about the patient or the procedure in a proprietary format. This type of interface is far more intuitive and informative than sifting through paper documents or scrolling through a long list of elements on a screen. Adding information in any particular category (a new image or symptom, for example) will also trigger suggestions and predictions about other conditions or procedures that are commonly found in similar situations.

For BUDDI Health, the ultimate goal of knowledge mining is to make sure the right medical codes are assigned to every diagnosis and procedure so medical practitioners are paid accurately for the work that they do.

CASE 4: TURNING STATIC DOCUMENTS INTO A DYNAMIC SOURCE OF INTELLIGENCE

The average large enterprise has thousands of different contracts—millions in some cases—covering everything from the global supply chain to ordering coffee for the break room. But most organizations have no way to view all their contracts in one place—or the collective risk and compliance issues they may be signing on to with each new contract. The language describing certain risks and obligations may be contained in a few words that are buried among hundreds of pages of a contract. With knowledge mining, it is possible to surface those elements and focus attention where it is most needed.

"Contracts are the foundation of commerce. If we can make that more

Knowledge mining offers a way for organizations to compile, understand, and **see the relationship** between disparate sources and formats of information.

efficient, we can change the way business works in fundamental ways," says Monish Darda, CTO and cofounder of Icertis, a contract management company. "For the first time in history, contracts are being digitized. Applying AI to this new pool of digitized data transforms them from static documents into live contracts that can interact with people, surrounding systems, and ultimately even other contracts."

Contracts are inherently heterogeneous and can be very complex. "Sometimes they have images or design documents and they also have metadata, such as supplier data or cost center information," explains Darda. "Sometimes people write notes on top of the contract during the negotiating process, or there are email exchanges that need to be preserved." All of that unstructured data can have value and meaning beyond the contract document alone.

The Icertis platform uses knowledge mining to make sense of what is in a contract and then apply different applications to build a business flow. Users can apply any parameters they need to look for valuable information, such as termination clauses or liability limitations that would have required hours of communication and research using traditional methods. "So, we can see the contract by its attributes and its clauses as well as its obligations to answer the question 'What are you actually agreeing to do as part of this contract?'" says Darda.

At the same time, the platform can bring contracts, together with all of their metadata, into a single repository. This is no small feat for a large, global enterprise. An industrial company might have millions of contracts worldwide in thousands of different formats, says Darda. Each global enterprise might consist of hundreds of different entities, each with its own customers, suppliers, and in-house systems to create and store contracts between them.

"It is a huge task to bring it all under one roof and then unlock the value there," says Darda. Knowledge mining allows Icertis to take unstructured data,

"A reliable knowledge mining platform can drive roughly a third of the costs out of the medical claims process," says Ram Swaminathan, CEO at BUDDI Health.

use a variety of AI services together such as optical character recognition and named entity recognition and then apply custom cognitive skills to make contracts readable in a number of common languages. For example, it is possible to search for a clause in Spanish and find all the contracts that contain a similar clause, even if they are written in English or Japanese.

Knowledge mining is already changing the way contracts are negotiated, executed, and enforced. "Many of our customers might take six weeks to go through the process of executing the contract. With this technology, we have brought that time down to four or five days," says Darda. But the real power of knowledge mining is in the way it will interact with other emerging technologies to drive action from insight. "AI tells you what is inside the contract. It also tells you what the relationship of the contract is with the outside world," he explains.

New Benefits for Business

There are rich opportunities for automating the understanding of content and creation of knowledge to meet a clear business need in many fields—defect identification in manufacturing, fraud detection in financial services, and smarter use of resources in logistics, for example.

Other areas where knowledge mining has potential include:

Reduced costs. For information that is needed in downstream processes, knowledge mining can bring both immediate and recurring cost savings. In the medical industry, a reliable knowledge mining platform can drive roughly a third of the costs

out of the medical claims process, says BUDDI Health's Swaminathan. Knowledge mining can help generate relevant codes based on all available data while a practitioner is with a patient or completing his or her chart. An elevated white blood count together with a fever, for example, might suggest a test or treatment for a sepsis infection. If related actions are not recorded in the moment, the average physician will have forgotten about them by the time the insurance denial arrives.

Competitive advantage. As knowledge mining evolves, more organizations will be able to differentiate themselves, says Bose. "We see opportunities in our domain using external and internal data and automation in workflows." The investment management business model is already facing tremendous pressures from new regulations, reduced margins, and the explosion of information available to investors, which raises the bar for financial service professionals to prove how they add value to their clients. "Those who don't take advantage of these technologies may even go out of business because of higher operational costs and their inability to process the data that is available to them," he predicts.

The return on investment (ROI) for knowledge mining at a small fund with one or two analysts is 30% to 58%, compared with the current industry practice of manually capturing the data or outsourcing to a provider, says Bose. For medium-sized funds with five or six analysts, the ROI is between 130% to 150%, and for much larger funds, with 50 or more analysts, it is over 500%. But the greater value in a

competitive field is the elimination of latency. "With automation, we are able to capture the data in near real time, versus hours or days with the manual process," he says.

More efficient work processes. Most analysts, engineers, architects, and other professionals would rather not spend any part of their workday combing through documents to find answers. Knowledge mining can help uncover important details and meaningful insights buried in the data. Putting more knowledge at the fingertips of highly paid professionals means they can do what is most valuable: analyzing insights or updating models, for example, instead of searching through CAD files or spreadsheets. Since 67% of organizations say that more than a quarter of their full-time employees spend part of their day extracting information from unstructured information, knowledge mining has the potential to give employees some of their time back.

Improved compliance. The supply chain is governed by agreements to buy, sell, deliver, and pay for goods and services, all of which come in a variety of electronic and paper formats that are tracked through a variety of accounting and inventory management systems. Matching orders and deliveries can be a time-consuming and inaccurate process, says Thomas Davenport, the president's distinguished professor of information technology and management at Babson College. "Extracting the information from those documents

and systems and comparing it to what is actually delivered is the sort of thing AI does very well," he says. By surfacing details when they are most needed, knowledge mining can help companies determine whether the goods and services delivered match what was ordered and invoiced. Darda at Icertis estimates that knowledge mining reduces the cost of compliance by 60%.

More effective decision making. At Howden, preparing a bid for a new project requires a high level of expertise. Engineers must review customer specifications and material requirements as well as maintain compliance with a host of regulations and guidelines. Knowledge mining can surface more of that relevant information and highlight risks and anomalies that might otherwise be lost in a stack of documents and technical drawings.

Best Practices to Ensure Successful Implementation

Going forward, how can leaders realize the full value of knowledge and insights from the breadth of their organization's information? Knowledge mining is an emerging suite of capabilities. Early adopters have discovered some best practices to make the most of any knowledge mining project.

Start with a use case, then design the system. Howden chose to focus first on the bidding process in the compressor market by creating a portal for its engineers and salespeople to improve collaboration, put more technical information at their fingertips, and automatically flag potential risks. This is a key step in the firm's transformation from a supplier of large capital equipment to a service-oriented model. "AI-driven technology has allowed Howden's engineers to increase accuracy, decrease response time, and release locked potential that is hidden in our idle documents and make it available to all of our engineering and aftermarket teams, increasing the size and scale of our knowledge bank," explains van Ravenswaay.

"AI TELLS YOU WHAT IS INSIDE THE CONTRACT. IT ALSO TELLS YOU WHAT THE RELATIONSHIP OF THE CONTRACT IS WITH THE OUTSIDE WORLD."
MONISH DARDA, CTO AND COFOUNDER OF ICERTIS

Putting more knowledge at the fingertips of highly paid professionals means **they can do what is most valuable**: analyzing insights or updating models.

Ensure top-down alignment and avoid siloed implementation. Knowledge mining is not just a data science or a technology problem, it also has tremendous business implications. "We see the need coming from all levels, from the C-suite to professionals to business lines to operations people," says Bose. "Everyone is in need of extracting insights from unstructured information and setting up either their own or their group's day-to-day workflow more efficiently, armed with the best insights possible." In order to successfully implement knowledge mining, business leaders should lean in, bringing in stakeholders from senior leadership as well as from across departments to plan for an integrated knowledge mining strategy.
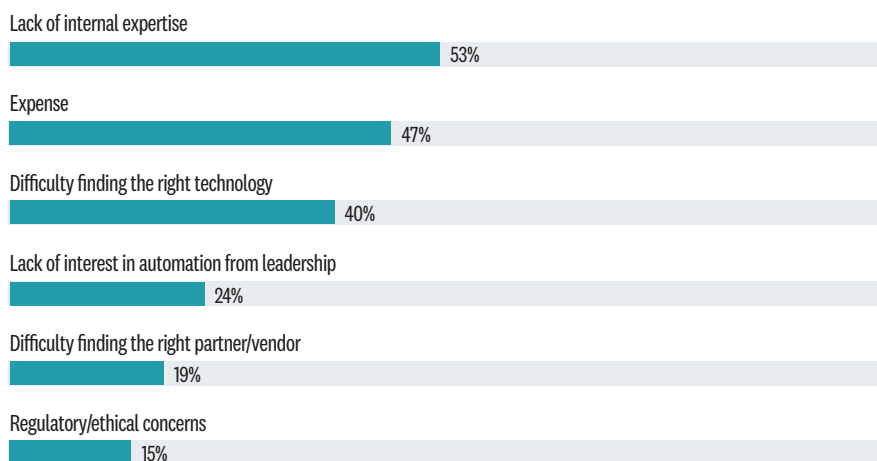
Find the right partner for your knowledge mining project. The knowledge mining toolbox consists of technologies that can be customized to each organization without the need to re-create each capability. Knowledge mining eliminates the need to custom stitch AI services together, which helps keep down development costs and the need for data science expertise. Instead, it is possible to gain the benefits of decades of development by collaborating with knowledge mining specialists.

Partnering offers a way to jump-start any knowledge mining initiative by integrating into an organization's existing data sources. Yet, eight out of 10 organizations (79%) say that managing the extraction of data from unstructured information is done mostly or completely in-house. Over half (53%) name a lack of internal expertise as the most significant barrier to greater automation when it comes to pulling out data from unstructured

BARRIERS TO AUTOMATING CONTENT UNDERSTANDING

Most significant barriers to greater automation of the extraction of data from unstructured information

Lack of internal expertise
53%

Expense
47%

Difficulty finding the right technology
40%

Lack of interest in automation from leadership
24%

Difficulty finding the right partner/vendor
19%

Regulatory/ethical concerns
15%

SOURCE: HARVARD BUSINESS REVIEW ANALYTIC SERVICES, JULY 2019

content, followed closely by expense (47%) and difficulty finding the right technology (40%). FIGURE 9

Conclusion

Ensuring that information is an asset and not a burden requires a means to harvest the knowledge organizations need from the data they already possess. More than half of the executives surveyed (54%) believe that when it comes to creating value for their organization, automating the understanding of unstructured information is important—including 19% who believe it is extremely important.

"The idea of knowledge mining has been around for decades but, just as

with AI, it wasn't possible to make a business impact with it until now," says Bose at Financial Fabric. In fact, understanding, surfacing, and acting on information has only become possible with the remarkable growth in machine learning and cognitive capabilities. Many of the visual, language, and other tools necessary to automate information extraction and understand context are available now to any business or developer, if they know how to use them.

METHODOLOGY AND PARTICIPANT PROFILE

A total of 579 respondents drawn from the HBR audience of readers (magazine/enewsletter readers, customers, HBR.org users) completed the survey.

SIZE OF ORGANIZATION

15%	**21%**	**12%**	**20%**	**8%**	**24%**
FEWER THAN 100 EMPLOYEES	100-499 EMPLOYEES	500-999 EMPLOYEES	1,000-4,999 EMPLOYEES	5,000-9,999 EMPLOYEES	10,000 OR MORE EMPLOYEES

SENIORITY

25%	**44%**	**20%**	**11%**
EXECUTIVE MANAGEMENT	SENIOR MANAGEMENT	MIDDLE MANAGEMENT	OTHER

KEY INDUSTRY SECTORS

ALL OTHER INDUSTRIES ARE UNDER 7% EACH.

13%	**10%**	**10%**	**10%**	**9%**	**8%**	**7%**
MANUFACTURING	TECHNOLOGY	EDUCATION	GOVERNMENT/NOT FOR PROFIT	FINANCIAL SERVICES	BUSINESS/ PROFESSIONAL SERVICES	HEALTH CARE

JOB FUNCTION

ALL OTHER FUNCTIONS ARE UNDER 5% EACH.

21%	**9%**	**8%**	**7%**	**6%**	**6%**	**5%**	**5%**
GENERAL/ EXECUTIVE MANAGEMENT	SALES/BUSINESS DEVELOPMENT/ CUSTOMER SERVICE	FINANCE/RISK	CONSULTING	ADMINISTRATION	STRATEGIC PLANNING	MARKETING/PR/ COMMUNICATIONS	OPERATIONS/ PRODUCTION/ MANUFACTURING

REGIONS

38%	**23%**	**21%**	**7%**	**9%**	**2%**
NORTH AMERICA	EUROPE	ASIA/PACIFIC	MIDDLE EAST/ AFRICA	LATIN AMERICA	OTHER

Figures may not add up to 100% due to rounding.