

INTELLIGENT DOCUMENT PROCESSING



HOW TO CREATE
ACTIONABLE
INFORMATION
FROM COMPLEX &
UNSTRUCTURED
DATA

Grooper®

ON ONE SIDE OF EVERY DOCUMENT

processing project is a vast quantity of information locked in physical or digital documents, and at the other, critical business workflows that demand accurate and timely information.

Connecting the two requires a flexible, transparent, and unified fabric of technologies that balances ease-of-use with powerful business outcomes.

This whitepaper highlights the most relevant algorithms, data science tools, and intelligent automation techniques included in a software platform called Grooper that are required to turn document-based data into trusted and actionable information.

We will dispel some myths around artificial intelligence and provide the reader with insight into the complexity of extracting information from complex and unstructured data.

To be fair, this isn't an easy thing to do, and the good news is that we have enjoyed immense success with customers who understand the journey they are undertaking.

And even if we do all the heavy lifting for you, you need to understand the amount of effort and the steps required to provide a successful outcome.

THE TRADITIONAL APPROACH TO DOCUMENT PROCESSING

The traditional approach to document processing is a combination of document imaging, capture technology, and manual human review.

There is a great chasm of time and innovation separating traditional optical character recognition (OCR)-based document capture and modern intelligent document processing.

Traditional approaches are afforded the luxury of time and the patience of subject matter experts who spend more time looking for information than acting on it. In today's fast-paced and rapidly evolving ecosystem of business, we no longer have these luxuries.

What is needed is human-level understanding baked into a software platform that is accurate, dependable, and useful for multiple document-based use-cases (not just invoices, email, structured forms, etc.).

BECAUSE IF THE SOFTWARE IS “INTELLIGENT,” SHOULDN'T IT WORK ON ANY DOCUMENT?

CAN'T A.I. JUST UNDERSTAND WHAT'S ON MY DOCUMENTS?

A WORD OF CAUTION

As artificial intelligence, machine learning, and natural language processing continue to gain momentum, it is easy to have a fallacy in our thinking: "Can't A.I. just understand what's on my documents?"

A.I. should only be viewed from the context of using computers (machines) to do something really difficult. With this idea in mind, we must understand that we've been bombarded with a lot of expectation-shaping when it comes to A.I., and especially deep learning, and neural networks.

Steer clear of the implications that A.I.-based "intelligence" should, will, or can, work like human intelligence. For example, neural networks do not, in fact "copy" the structure of the human brain, as is often claimed.

A.I. doesn't, and won't ever, work like that. There is absolutely zero evidence that a trained neural network, or any other A.I. system, has, or can have anything like generalized intelligence. Absolutely none.

A.I.-based tools are incredibly helpful, but they're only ever going to be part of a full solution. Intelligent document processing isn't a narrow problem, and it's never going to be. That's why A.I. won't save you!

BUT, A.I. IS THE FUTURE... SO READ ON...

USING COMPUTERS TO DO SOMETHING REALLY DIFFICULT

Ultimately, in the realm of document-based data integration, we need technology to create an awareness of:

- **The original intent of the document**
 - **What specific information is represented**
 - **Where specific information is located**
 - **When certain words mean something else in relation to the intent of the document**
 - **Known relationships between document information and external data**
-

The difficulty in working with unstructured natural language documents like contracts and leases is as complex as language itself. Similar words have different meanings and intent. When words are combined, they create an idea that is separate from the individual words themselves.

A great example of this is an address. Humans immediately see an address for what it is, but to a computer reading characters on the page, it needs an awareness that when certain “features” are combined it means something more than just the sum of the characters or words.

And the difficulty in working with complex semi-structured forms like explanation of benefits and industry-specific forms / reports is that the overall document structure varies from vendor to vendor. Humans quickly grasp how the layout of fields and labels represent information, but it is much more difficult to get software to do the same thing reliably and at scale.

To create this kind of document / information awareness while preserving data integrity is extremely difficult and requires a methodical approach using several core technological approaches that at first, seem to be taking steps backwards. These are necessary; you’ll see why.

THE TECHNOLOGY POWERING INTELLIGENT DOCUMENT PROCESSING

Documents (digital, text-only, paper, microform, etc.) are artifacts that represent an event that has taken place, or are designed to communicate an idea composed from multiple sources. They are a work product, and as such are not source data.

If all software applications communicated perfectly in sync, we wouldn’t need documents at all. That isn’t the world we live in, so we’ve amassed incredible hoards of document-based information and there are no signs of slowing.

TO CREATE AN AWARENESS OF WHAT INFORMATION A DOCUMENT REPRESENTS, AND TO FILL IN THE GAPS USING A.I. AND ADVANCED PROCESSING TECHNIQUES, SEVERAL KEY TECHNOLOGIES ARE REQUIRED:

DOCUMENT INGESTION / SCANNING

Understanding how to get documents into Grooper is important because there are many options (here are the most popular):

Scanners – Grooper will directly control ISIS or TWAIN scanner hardware, including Mekel branded microform scanners

CMIS – Connect to your CMIS-compliant electronic content management system and either move documents into Grooper, or process them where they are stored without moving them

File system – Grooper “watches” an NTFS file store and processes documents as soon as new ones are discovered, or point Grooper at an existing folder for processing

API – Use the Grooper document ingestion API to move documents into Grooper from your software application

Database – Integrate with any SQL or ODBC-compliant database

Cloud storage – Integrate with Box, SharePoint, OneDrive, etc.

DOCUMENT IMAGE CLEANUP

Because many of the documents we store are not digitally born, but are scanned documents, they will need to be re-scanned with modern software. This critical step of re-scanning documents ensures accurate OCR. This re-scanning isn't printing and then physically scanning; rather, a re-ingestion into document image processing software.

Grooper contains all the built-in connectors and APIs you need to process documents from virtually any source.

While this may seem like a step backwards, it is critical for accuracy. You need a basic understanding of two key technologies involved in modern scanning: Image Processing and Computer Vision.

IMAGE PROCESSING & COMPUTER VISION

Image Processing (IP) and Computer Vision (CV) are applied in three phases to improve OCR:

Phase 1

Enhance page images to create the best visual representation of the original page

Phase 2

Create intermediate images to ensure the best possible OCR and data extraction results (these intermediate images are a vital resource to guide Grooper architects' decisions later in document processing)

Phase 3

Apply fully automated IP and CV techniques as needed to automate the collection of various data elements such as information stored in tables, boxes, and other bound regions (the goal is to get the intelligent document processing platform to “read” the document the same way humans do)

MODERN IP AND CV ALGORITHMS PERFORM THE FOLLOWING IMPORTANT TASKS

Intelligent Comb Removal:

Combs were problematic in line removal because they are often shorter than lines used to draw characters. Comb removal automatically detects and removes combs which intersect detected lines (think of forms that have little lines separating where each character is typed or written in).

Advanced Line Detection: This command supplements basic line detection with a Hough Transform. A Hough Transform is a feature extraction technique used in CV to detect imperfect instances of geometric shapes such as lines, circles, and ellipses. This advanced function is very useful on low-quality images where lines are broken and/or significantly skewed.

Negative Region Removal: This command detects rectangular negated text regions and inverts or removes them to help create perfectly clean character images.

Photo Removal: Photos are temporarily removed prior to OCR to improve OCR performance and reduce junk characters.

Speck Removal: This is a hardware-accelerated speck removal algorithm that doesn't remove important punctuation marks.

Border Cleanup Commands:

Border detection uses image analysis to determine the boundaries of any borders. Users define how sensitive this feature should be by how thick they want to trim the borders.

Bounded Region Detection: A bound region is a page selection which is bound on all sides by lines. This command can be used to execute other IP and data extraction commands directly into these regions.

Mask-Based Dropout: This algorithm generates a mask indicating areas to be temporarily removed before OCR. The masked pixels are removed from black and white, color, and greyscale images.

Shape Detection: This command identifies the locations of shapes such as stamps and seals and aids in the extraction of information from them.

Color Dropout: Drop out any color from any image and any region of an image.

Inpainting: A digital restoration technique that prevents OCR inaccuracies by restoring damaged or unknown parts of a document image using information from surrounding pixels.

Barcode Removal: Temporarily remove barcode structures that will negatively impact OCR results.

Blob Detection: A "blob" is a set of connected black pixels on a binarized image. These must be detected, filtered, and removed. **Halftone Removal:** A halftone pattern is a matrix of small dots used to simulate shades of gray on black and white documents.

Hole Punch Removal: Hole punches are located by eroding the image until only hole punches remain. This process erodes away thin features like lines and characters, leaving only "thick" features like hole punches behind. The holes are then digitally removed.

Line Removal: This is a new approach to an old problem. A line removal algorithm causes complete form removal, including small vertical bars, and combs. Characters connected to lines are trimmed so that no edge artifacts from lines are left behind.

OPTICAL CHARACTER RECOGNITION

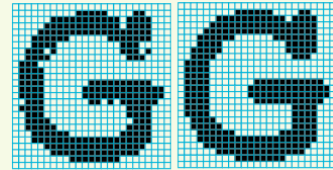
OCR is the most crucial part of intelligent document processing. OCR is responsible for transforming pixels on a page that represent the text we read with our eyes to text that software will work with to understand what is written.



Traditional OCR uses a "matrix matching" algorithm to identify characters using pattern recognition. The character on the document image may look like this.



It is compared to a stored example that looks like this.



By comparing a matrix of pixels between the character on the image and the stored example, the software determines the character is the letter "G."



This may appear to be a good approach, but there are pitfalls. Because the software is comparing text to stored examples pixel by pixel, the text must be very similar and easy to read. Even if there are hundreds of examples stored for a single character, problems often arise when attempting to match text on poor quality images or using uncommon fonts.

The solution to high accuracy OCR is a method for decomposing characters into their component features rather than by comparing pixels to known examples. Instead of pixels, features:



Features matching how the character is drawn are often easier for software to understand since the margin of error is less. Feature detection is also less susceptible from errors caused by random pixelization.

Grooper OCR works by combining multiple OCR engines, and both traditional and feature-based OCR techniques (the best of both worlds!). The results of both algorithms are combined to produce the best-matching result. Each character is given a "confidence score," which corresponds to how closely the character pixels, or features match, or a combination of the two.

As an example, what is this character below? Is it a "G," a "C," a "0," or is it even a character at all?

OCR must decide, and it may not make sense within the context of the word or sentence. And if a human can't read the character, then OCR will certainly have trouble.

BIS has two patents that make OCR processes within Grooper especially powerful:

Patent 10,740,638 – Data element profiles and overrides for dynamic OCR-based data extraction

Patent 10,679,089 – Systems and methods for optical character recognition

OCR POST-PROCESSING TO THE RESCUE

Without additional context, it makes sense that there are character recognition errors. Even if the character isn't discernable, a human knows "ballboy" refers to a "ball boy," or an indie band from Scotland, but that "bollboy" is just gibberish. The most common

post-processing done by Grooper OCR engines is basic spell correction. Often, errors from poor recognition result in minor spelling mistakes. Grooper OCR compares words with a lexicon of common words and attempts to make logical replacements.

ballboy ← **word**
bollboy ← **nonsense**

BUT WHAT ABOUT PROPER NOUNS & OTHER IMPORTANT WORDS THAT AREN'T IN A LEXICON OF COMMON WORDS?

There are two methods to identify incorrect characters:

Lexicons

Import custom lexicons for words and terms related to known industry jargon.

Fuzzy Matching

This is a method of providing weighted thresholds to characters and allowing the software to substitute characters based on likely good replacements. For example, the software would be allowed to try an "o" when a "0" provides a bad result. Same for an "l" instead of a "1," etc.

5 FEATURES OF ADVANCED OCR

WHAT ABOUT HANDWRITING?

We have discovered the best handwriting recognition engine is powered by Azure Cognitive Services. This service is available through an application programming interface (API) call in Grooper (why reinvent the wheel?).

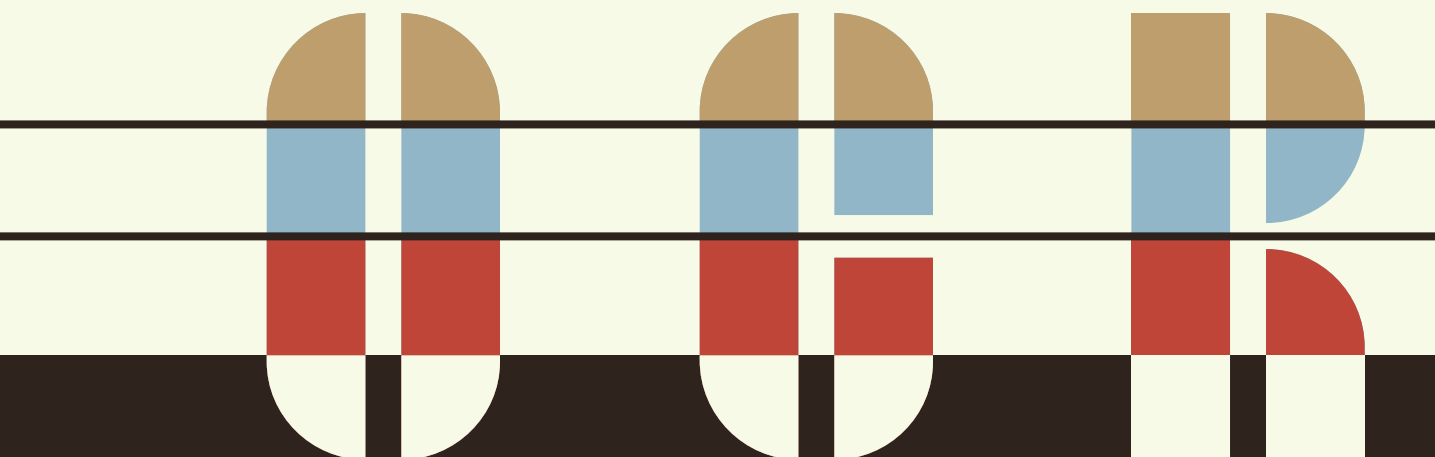
1. ITERATIVE OCR – Captures text missed on the first “pass” of OCR by running OCR multiple times. On each pass, successfully recognized text is digitally dropped out of the document image. The OCR engines find missing text easier without the distraction of nearby text.

2. CELLULAR VALIDATION – Multi-column layouts of text present a challenge for OCR, especially when they are offset, or have different fonts, or font sizes. Cellular Validation splits an image into a grid, and each area is processed independently.

3. BOUND REGION DETECTION – A bound region is a page section which is bound on all sides by lines. During bound region processing, text inside boxes is processed first. This ensures that content outside of the text boxes isn’t confused with other text on the page. Because the location of each box is understood, all text is intelligently joined back together for more powerful contextual understanding of document data.

4. SEGMENT REPROCESSING – Because Grooper understands the layout of text on a page, groups of data are viewed together as segments – the same way a human would read. OCR engines independently re-run recognition on low-accuracy segments of text until the best accuracy is achieved.

5. LAYERED OCR – A document with multiple fonts makes accurate character recognition tricky. Grooper overcomes this by using a different OCR engine for each font type and then synthetically combines the results. Some OCR engines do a better job on certain fonts (including handwriting) than others, so it makes sense to use the right tool for the job.



MACHINE LEARNING CLASSIFICATION SEPARATION

ML algorithms are generally divided into two types: supervised and unsupervised. Within Grooper, all classification, document separation, and data extraction incorporate ML algorithms. There are four methods for classification within Grooper:

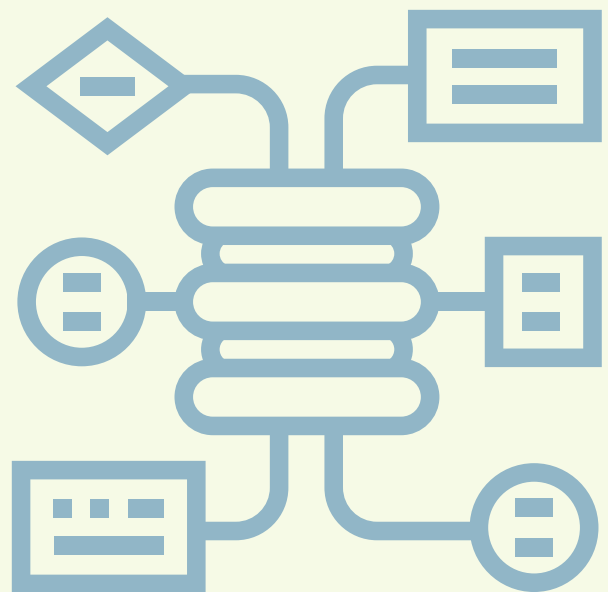
1. **Lexical** – This is an NLP and ML training-based approach that uses document examples for training based on document context
2. **Rules-Based** – Uses unique key words or features that identify a document
3. **Visual** – Computer vision is used to analyze the visual structure of a document without need of OCR
4. **Smart label classification** – Uses hierarchical header-level information to determine document type

Machine learning (ML) is a way to build systems that display artificial intelligence (remember, solving hard problems). Generally speaking, ML is a strategy that builds systems that:

1. Are not specifically programmed to perform their task until given training data
2. Get better (to a point) at performing their task when exposed to more data or more inputs

There are eight document separation providers:

1. Control sheets
2. Event-based
3. Pattern-based
4. Change-in-value
5. Embedded page information
6. ESP auto separation
7. Multi-separation (uses multiple separation providers)
8. Undo separation



Here's how machine learning is performed in Grooper

A Grooper architect determines a discrete set of categories, and gives an initial set of properly labeled training for the algorithm to begin making predictions. This is all done within an easy-to-use point-and-click user interface.

Grooper is designed this way, because human-supervised ML works better, especially for the sorts of complicated problems that arise in document data integration. **Unsupervised machine learning should not be used in document-based A.I. projects because it:**

1. Requires a very large data set to work
2. Doesn't work very well in general

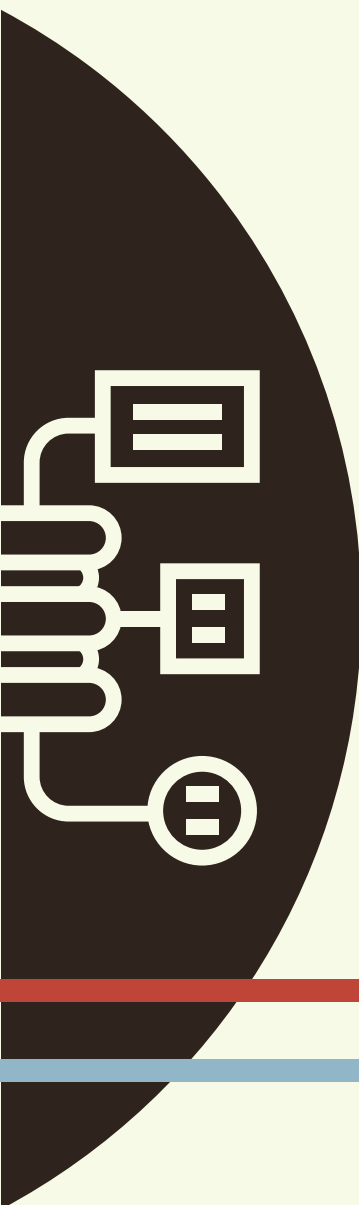
If people ask about or promote unsupervised ML, they are talking about one of two different things:

1. Will the A.I. get better at classifying (or extracting) on its own?
2. Will A.I. save me from having to understand my own documents or data?

The answer to both of these questions is "No!" Document A.I. doesn't get better on its own, because a human-centric design of document processing systems achieves better outcomes than unsupervised ML in almost every use-case.

Nothing – not even Amazon, or Azure, or Watson – will save you from having to understand your own documents and data.

If someone tells you otherwise, they're lying; and if someone you're talking to believes that Grooper (or any other document processing system) will save you from your own document or data problems, they're setting you up for failure.



THE TF-IDF MACHINE LEARNING ALGORITHM

TF-IDF is the core A.I. and ML algorithm used in the Grooper platform. TF-IDF stands for “term frequency / inverse document frequency,” and classifies documents (or document collections) by comparing how frequently words are seen on the target document type versus how frequently they occur in the sample set as a whole.

TF-IDF works by identifying words (or inputs) that are unique (or more common) in a particular type of document compared to the document set overall. It’s a deceptively simple way of classifying documents, and it generally does so in a similar way to how humans do it: by looking at the individual words on the document.

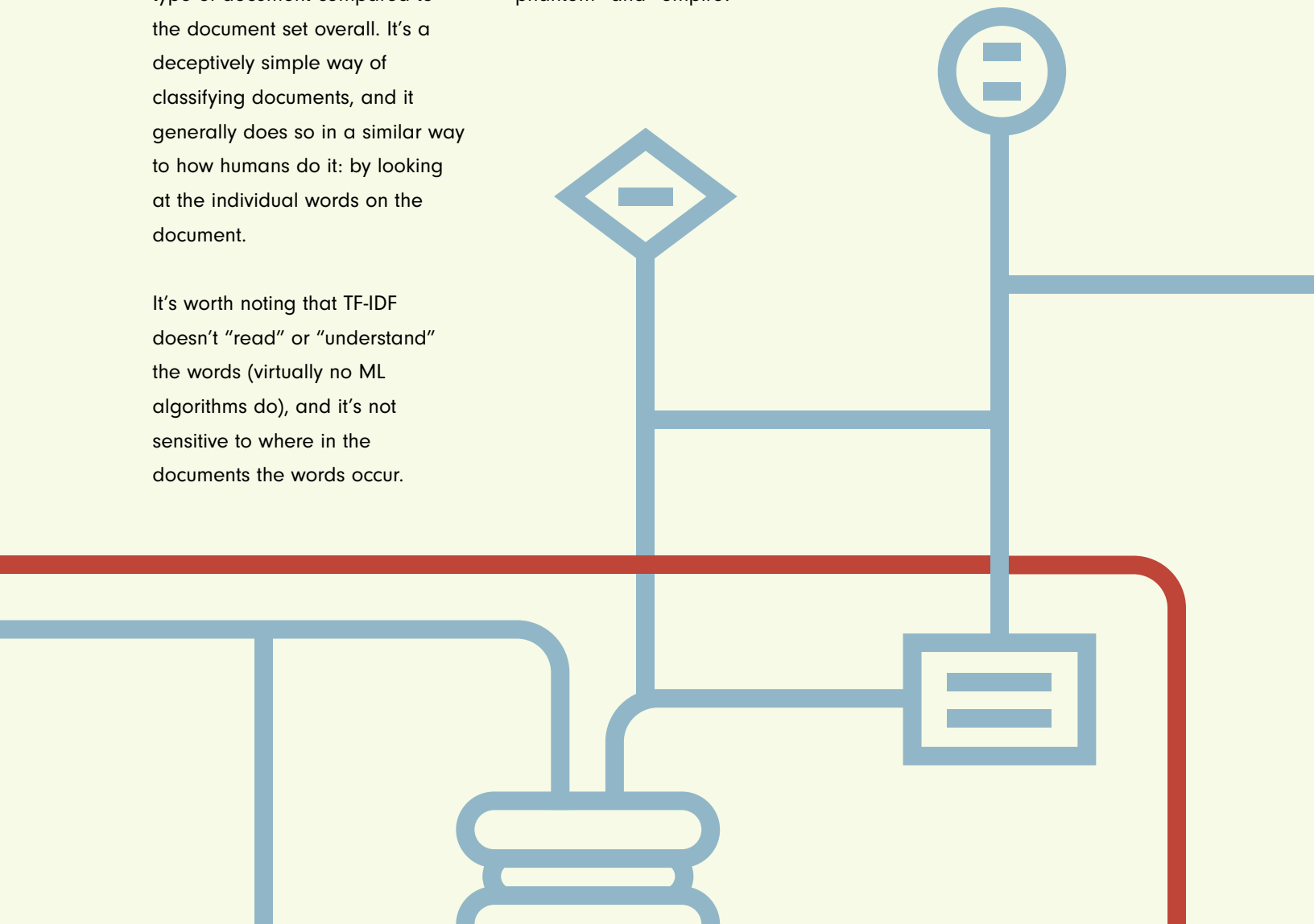
It’s worth noting that TF-IDF doesn’t “read” or “understand” the words (virtually no ML algorithms do), and it’s not sensitive to where in the documents the words occur.



It’s just counting words (or features, like the presence of addresses, phone numbers, percent values, etc., if you aren’t using words).

Grooper does a few things that make TF-IDF work better than other ML algorithms. First, and most importantly, it lets you feed the algorithm anything.

Do you want the algorithm to look at words? Two-word pairs? Names of mammals? The presence of dates or names? The words “phantom” and “empire?”

Good, great, fantastic, even! You can write extractors to feed that into the algorithm. This is crucial, because it lets a human determine which types of features are most likely to matter for a given document – and there are some types of documents where the content of the words won’t ever tell you what the document is. This is when you turn to features for document classification.





Our TF-IDF implementation also lets architects inspect the machine learning assigned weightings (transparent A.I.), which lets them directly, and completely understand how it's making decisions. This is not possible using neural nets! And generally not available in other systems, even those implementing TF-IDF, or similar algorithms.

TF-IDF also requires a relatively small number of training samples (generally less than 100, although not always so) to reach optimum decision making, which makes it much quicker to train than most machine learning algorithms.

As mentioned, TF-IDF is used in data extraction, document separation, and classification within Grooper. In separation and classification, it lets you work with unstructured documents.

In data extraction it lets you pick out the correct instance of a value type on a page (e.g. tell date of birth from date of service) by looking at features surrounding the detected value.

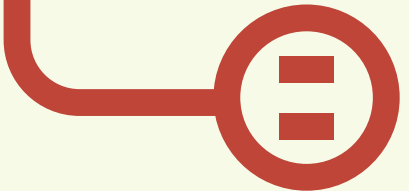
Natural Language Processing

Natural language processing (NLP) finds paragraphs, sentences, or other language elements that convey specific meaning. NLP is responsible for context-based data capture which is important in creating an understanding of what specific dates, names, etc. on a document represent.

NLP is also responsible for distinguishing between commonly

used descriptions that differ, but mean the same thing, like "SW," "Southwest," and "SW ¼ of the NW ¼," and "SW ¼ and the NW ¼." These are difficult problems to solve without built-in NLP.

Many possible combinations of common data could be provided based on known document data and provided to Grooper for more flexible data discovery.



Grooper natively processes text as n-grams and via porter stemming in addition to supporting configurations that implement more complex NLP methods like:

- Sentiment analysis
- Part-of-speech tagging
- Named entity tagging
- Feature-based tagging

The main difference between a standard NLP library (like the Stanford Library), and Grooper is the use of NLP throughout the product, not just as an add-on.

NLP also provides a practical understanding of things humans take for granted when looking at a document:

- Paragraph Isolation – Indentations, double spacing, bullets, key phrases, line length, and many other factors are analyzed to determine where each paragraph starts and stops

- Lexical Analysis – Collect features from within paragraphs, like n-grams, entries from a lexicon, or a non-value feature count grouped by data types, like address, phone number, given name, etc.

- Data Merge – Paragraphs are recognized that span multiple pages and then joined together in a machine understanding of the document

- Label / Value Pairs – In structured documents, most data are recorded in label / value pairs; a value will have a corresponding label somewhere on the page that communicates the meaning of the value to the software

Data Parsing, Extraction, Verification, and Integration

There are many methods for extracting information in Grooper that rely on optical character recognition, image processing, computer vision, natural language processing, and machine learning. Here are some of the more popular data extraction methods:

- Understand checkboxes with optical mark recognition
- Table data
- Header data
- Key value pairs
- Pattern matching
- Fuzzy matching
- Zonal extraction
- Barcodes

Data Verification

For a data extraction and integration system to be trusted, it must produce 100% accurate data. This is no easy feat, and absolutely requires a human-in-the-loop design approach. Grooper is configured to go and find information,

and when possible, validate it using either known information from external databases, custom / built-in lexicons, or built-in validations using character or word-level accuracy thresholds, mathematical validation, business rules, etc.

Any data extractor may be configured for human review of collected information based on validation pass / fails. In this method, the time of human operators is maximized as they are only looking at document data that has been flagged for review.

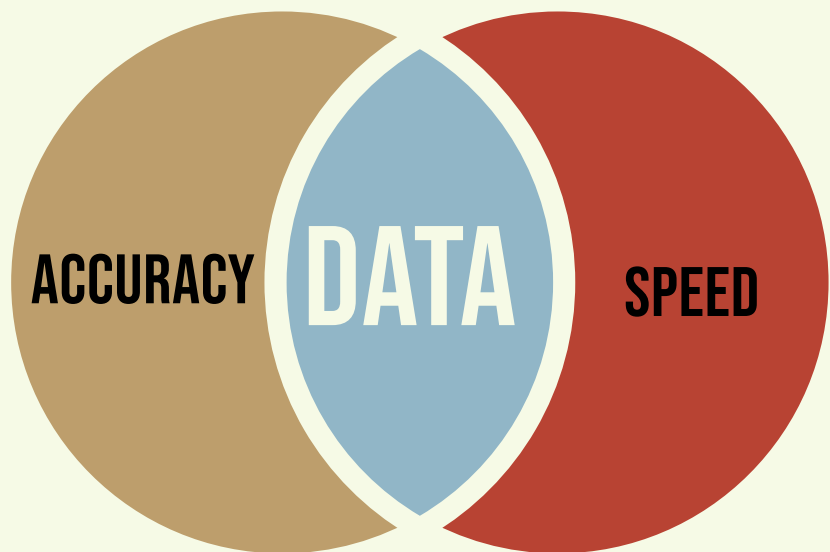
The majority of Grooper users spend their time in the Grooper Attended Client.

The Attended Client is an intuitive interface that enables human verification of extracted data before integration. The console presents extracted information as a virtual document on one side of the screen and the actual document image on the other. When a user clicks on an extracted field, it is highlighted on the document for quick verification.

Grooper architects will set field-level accuracy thresholds and / or business rules on extracted data to ensure accuracy. Fields that have an extraction accuracy

threshold below a certain percentage are automatically flagged for human review.

This ensures operators only spend time looking at data which needs correction or oversight. Missing or inaccurate data is either filled in manually, or by Grooper Field Sense which automatically supplies certain document data (like dates) for quick entry.



Data Integration

After data has been validated, it will be integrated using the file / format of your choice into virtually any content management system, robotic process automation tool, or downstream line of business application / database for further processing. Data may be integrated directly, added to a staging database, included as metadata, formatted as CSV, JSON, XML, etc.

Here are some of the most commonly used data integration tools within Grooper:

- **CMIS** – Use CMIS to integrate with external document storage platforms for import and export operations
- **Pre-built connectors** – Alfresco, Box, Documentum, FileBound, M-Files, Microsoft, ApplicationXtender, OnBase, Laserfiche, etc.
- **File share** – NTFS, FTP, SFTP, etc.
- **Databases** – SQL, or any ODBC-compliant destination
- **Smart PDFs** – Integrate data within documents as bookmarks, annotations, or metadata

Data Accuracy & Speed

We're often asked about accuracy and speed. "What's your OCR accuracy? How accurate is the data? How fast can we process x amount of pages / data? How much more efficient will our processes be?" And nobody likes the answer "It depends." You also don't want to be lied to.

But we recognize the importance of OCR accuracy. Anything less than 100% accuracy creates massive error rates. Here's why: Say you're getting 95% accuracy on an invoice and you need to extract ten independent fields. The overall per-field accuracy is actually going to be 60%, not 95% because .95 to the power of 10 is .60, or 60%.

Imagine, just 4% better OCR accuracy boosts overall accuracy to 90%. This is the math we live by.

Accuracy and Speed? It Depends...

We've worked on enough documents and data sets to know that there is no such thing as a one-size-fits-all data accuracy or speed answer. Here are the factors that affect accuracy (and the reason image processing and OCR processes are so important):

- Poorly scanned documents
- Damaged or very old original paper
- Scratched microfilm / microfiche
- Handprinted text
- The amount of data to be extracted

We are confident that with high quality image processing and advanced OCR processes, a very high level of accuracy will ensure as little human review needed as possible. But after human review? 100% accuracy on field extractions.

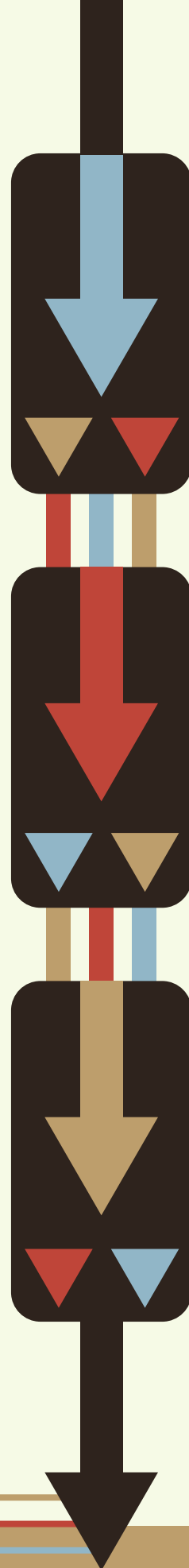
Overall accuracy on full text search is a little easier to answer – it is typically in the 97% – 99% range, but again, it depends on your documents.

Questions around speed are mainly answered by a question back:

"How much technical compute resources are you willing to provide?"

Grooper operates in memory and takes advantage of parallel processing, and all available processing cores. And even speed is dependent on the amount of work required to clean up poor quality images.

Grooper is optimized for speed and is so important to us that we provide built-in reporting for every operation performed in Grooper down to the millisecond to help architects optimize performance. We have individual customers with production environments processing millions of pages a year or pure data integration of over a billion extracts a day.

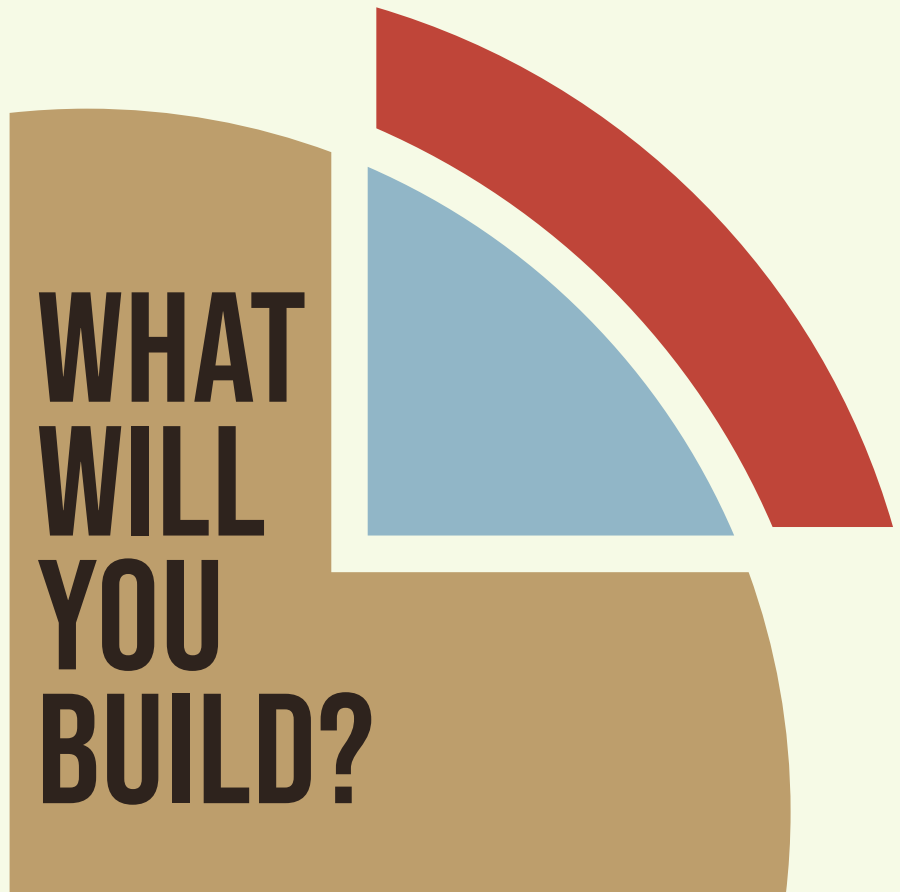


Common Intelligent Document Processing Use-Cases

If you are considering using intelligent document processing to enhance processes, increase revenue, decrease manual labor, or bring new products to market, you aren't alone. Below are some of the more common use-cases we've seen.

Common Use-Cases:

- Accounts payable automation
- B2B data / EDI
- Case files
- Checks
- Contracts / leases
- Custom business forms
- Data mapping
- Digital mailroom
- Email
- Engineering documents / blueprints
- Explanation of benefits
- Financial statements
- Government forms
- Human resources
- Invoices
- Legal descriptions / right of way
- Materials test reports
- Medical claims
- Metadata
- Mineral ownership reports
- Mortgage documents
- PII identification / redaction
- Pipeline integrity
- Reports
- Royalty statements
- Smart / embedded PDFs
- Student transcripts
- Titles
- Transactional data
- Well logs



ABOUT GROOPER

Grooper was built from the ground up by BIS, a company with 35 years of continuous experience developing and delivering new technology. Grooper is an intelligent document processing and digital data integration solution that empowers organizations to extract meaningful information from paper / electronic documents and other forms of unstructured data.

The platform combines patented and sophisticated image processing, capture technology, machine learning, natural language processing, and optical character recognition to enrich and embed human comprehension into data. By tackling tough challenges that other systems cannot resolve, Grooper has become the foundation for many industry-first solutions, financial services, oil and gas, education, and government.

WHY CHOOSE GROOPER?

We have spent 35+ years in business evolving the industry and setting the standard for intelligent document processing with patented tech and deep understanding of the complexity and true potential of unstructured data. We stay laser focused on rapidly discovering and delivering meaningful value.

Grooper was built to succeed where others have failed, taking on tough challenges. The platform processes and integrates massive amounts of difficult data from complex documents and architecture. If we say our system can solve a problem, it can. And we'll show you how.

We provide data fulfillment as a "done-for-you" service or collaborate on a "done-with-you" approach to teach you our tech. Either way, your data is always yours, always secure, and now more useful than ever.

OUR VISION

A world where unstructured data is easily accessible, understandable, and useful for driving business efficiency and innovation.

OUR MISSION

Helping organizations transform their information and workflows to become industry leaders.

OUR COMPANY VALUES

The outcome is the most important part of the journey. Following guiding principles that keep the goal in focus and filtering out the "noise" keeps projects on track to deliver value.

A project should only be pursued if it's supported by a well-reasoned business case.

Transparency from all parties is essential to be effective. We work in a way that makes it safe for everyone to understand and tell the truth. When we say we can do something, we make it easy to see how.

People are the most important part of our business. We value long-term relationships with employees and with clients. We have spent over three decades cultivating trust.

Coming to grips with the complexities of data is hard. Being willing to have fun while getting things done makes the process more enjoyable and sustainable.

Elegance and simplicity are valuable principles in the production of creative, efficient, and sincere work.



HOW OUR CUSTOMERS WIN WITH GROOPER

- Make data more useful by embedding comprehension into it
- Automate information workflows effectively to streamline operations and capture innovation opportunities
- Create meaningful data that informs better to make faster business decisions and provides operational agility
- Faster time to value from digital transformation with less risk of project failure
- Make IT simpler and more agile, aligning IT with business objectives for revenue growth, customer experience, strategic partnerships, compliance, and more
- Transfer data and knowledge more easily within and across organizations
- Redact at-risk data, improving compliance, and safeguarding private information

HEAR WHAT CUSTOMERS SAY

"Grooper has saved OSU hundreds of thousands of dollars and the ROI was seen in less than six months after going live. This product has taken data processing, document scanning, and import automation to a whole new level. It's now in virtually every department including our president's office."

-Erin Girtan, Database Administrator/Content Management & Capture Administrator, Oklahoma State University

"Grooper allows my staff to process ten times more volume than we could with our previous image capture solution; our office thinks Grooper is worth its weight in gold."

-Marie Ramsey-Hirst, Court Clerk, Canadian County

"Grooper will give us the access to more contract data than ever before by quickly extracting the data across thousands of lengthy contracts allowing employees to spend time on value adding data analysis rather than extraction."

-Glena Brauer, Supervisor - Marketing Contracts & Compliance Chesapeake Energy

"In acquisitions and divestitures, there's millions of dollars at risk for us in knowing precisely what lease documents actually say versus what cover sheets say or the information being represented to us. With Grooper, you get precise information that has an impact on defects and revenue realization within the 45-day buying window. Without Grooper, you'd just be guessing."

-Clay Chamberlain, General Counsel and Director of Legal Operations, Corterra Energy

"Grooper has cut down on our indexing time by 70 percent."

-Ryan Freeman-Smith, Manager, Oklahoma Health Care Authority

"There's infinitely more information in documents that we can draw from to make better business decisions in the future. If we can do things safer, better, faster, and cheaper because of the information we have, from our previous mistakes and successes, then we ought to take advantage of it."

-Gary Ridley, former Oklahoma Secretary of Transportation, former Director of ODOT, former Director of OTA, former President of WASHTO, former Board Member of AASHTO

"Grooper gave Change Healthcare the ability to process huge diverse client print files and extract the data into a normalized data stream for processing. One of the key features of Grooper is the ability to identify, interpret, enhance, and extract data from a large number of media types (TIF, JPG, PDF, Word, Excel, ect...) Another key decision point was there wasn't a requirement for a software developer to write code, a business analyst can use a GUI base application to configure 98% of the extract jobs."

-Todd S, Change Healthcare

"My initial experience with Grooper was to improve our OCR processing. Using the Grooper platform we were able to implement four major projects. We used the CMIS connectivity capabilities to extract images from our ECM system as well as locate and identify documents containing PCI data for compliance regulations from the billions of archived documents. We have been able to purge large amounts of documents and automated the process of getting document from our core system to our repository has been great."

-Gerald D.



"One of our biggest challenges was how to bring in our DocuSign documents and get them into the OpenTXT platform without intervention. This is the ticket! This software is incredibly flexible. The capabilities seem to be endless, I am amazed at how much it can do. We use this software in conjunction with our automation system to bring in documents from DocuSign. However; we are in the process of expanding it for use by our HR department."

-William M.

"We are integrating this software into our processes to augment our analysts capabilities. So far it has been a success and we expect a 6x increase in efficiency for selected activities. It gives me and the other users control. It allows us to create models in a structured environment without programming, but still with precision control. The team behind it is great too."

-Kyle S.

"I personally have been doing hundreds of tax statements for accounting and it takes hours to do, Grooper will save half and even more than half of time when it comes to data entry, Lease Analyst work and more. I can't wait to experience this program more fully."

-Kayl E.

"We have many standardized contracts in our business and it will save time to pick out the things we need to give our attention to rather than go through pages and pages of words. It will save time once you get it set up."

-Frances B.

"Endless possibilities as far as collecting data for output. Grooper is very advantageous in acquisitions and divestitures; collecting, assignments, leases, and contracts."

-Lauri S.

"We are ecstatic about how well Grooper works on our billing system. We have a customer that submitted a bill with 12,000 lines. These lines totaled over \$6,000,000 in fees that needed to be audited. It took just over 1 minute to process every line, and match up against the total. This would have taken close to 2 hours before."

-D, Healthcare SaaS Payment Integrity Firm

IMAGINE

WITH

US

WWW.GROOPER.COM
INFO@BISOK.COM

1-800-662-0777
13900 N. HARVEY AVE.
EDMOND, OK 73013

