



Bridging the Gap

**More productivity and lower costs through automation of
knowledge-based processes**

October 2021

Executive Summary

In one sentence: Intelligent document analysis using Natural Language Processing (NLP) based on Machine Learning (ML) enables the **automation of knowledge-based processes** by capturing structured data in documents.

Digital transformation is like many other things: Where opportunities arise, there are also risks and associated challenges: More data, more information, more documents also mean more work to evaluate, process and use them for the benefit of the company. In other words: **turning unstructured data into structured benefits** and thus building the bridge between systems that already exist and are in use and new insights that lie dormant in a multitude of documents (contracts, for example).

Traditional workflow and content management systems (for example DMS) are excellent solutions in and of themselves. A special added value is created by linking the systems. While many trivial business processes can be optimised well by means of rule-based algorithms, many documents with process-relevant information lie unused in content management systems and cannot be used as structured data.

NLP is a solution to turn exactly these challenges into a valuable advantage.

In the first step, Optical Character Recognition (OCR) is used to transfer documents into a machine-readable format. In the second step, NLP uses machine learning to turn these unstructured texts into structured assets - in a matter of seconds and with an accuracy that is comparable to human editing...

It depends on the individual business process how and for what NLP and document analysis can be used. This can be to completely automate processes, to identify unrecognised business risks or even to relieve skilled staff through automated preparatory activities.

But in any case, intelligent document analysis can build the so intrinsically important bridge that has emerged as part of the digital transformation between legacy systems and new applications. For **seamless process automation without large-scale architecture projects.**

Challenge: Process-relevant information is often hidden in plain text

There are many industries that have a high volume of documents due to their business - the insurance industry, for example. While some of the documents directly contain structured data for further processing, for example forms, this is not possible with free communication or openly designed texts. Contracts, letters and emails do not follow a predefined pattern, they cannot simply be processed automatically. But it is precisely in these sources that a lot of valuable information is hidden.

In addition, there is the increasing change from synchronous to asynchronous communication as well as the dwindling willingness of customers to act as "data entry helpers" in the course of the customer journey. Convenience is the buzzword of customer-oriented service design. With as little information as possible in an ideally completely free format to success. The future means that the complexity of communication can be solved by the provider itself and no longer shifted to the customer as before.

The same applies to the internal ability to make long documents and large amounts of information in text form quickly available and usable. Qualified employees are already in short supply today, a trend that will become even more critical in the coming years. Customers expect responses in seconds. This cannot be managed with conventional methods.

It is therefore a matter of quickly expanding existing systems and making them fit for the future in order to meet the process requirements of the coming years. If possible without large IT projects. In this context, it is advantageous to build on existing systems and expand them in a meaningful way. OCR as a technology is standard in many companies, as are document management systems (DMS), ERP, CRM and robotic process automation (RPA).

The task is therefore to connect these systems with each other and to realise cost reductions and time savings in addition to more effective customer communication. The goal is always full automation. In times of knowledge work, this can only succeed if systems come up with actual semantic understanding.

In short: machines that can read documents like contracts elevate process automation to a new level.

Case Study: Purchasing optimisation through automated detection of contract deadlines

Challenge:

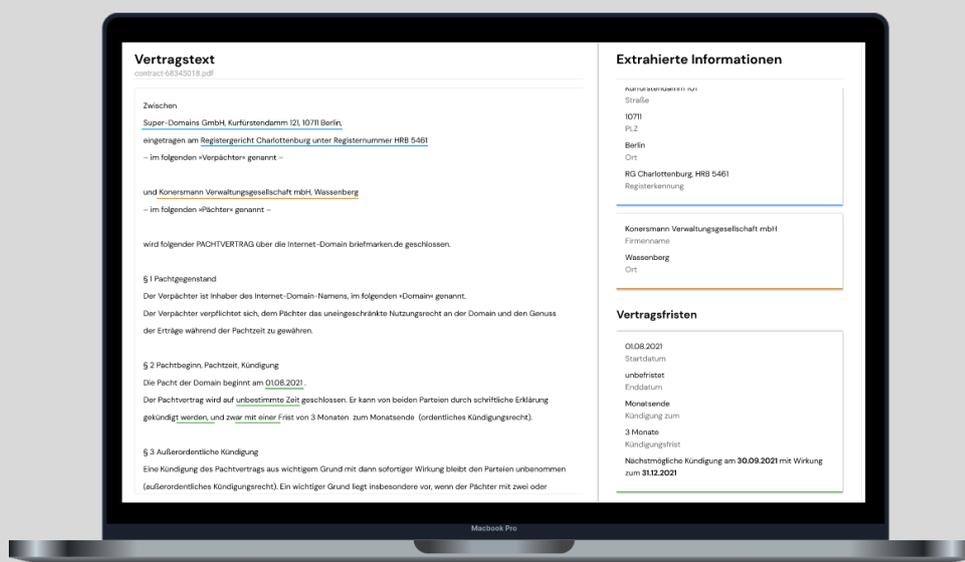
- A hidden champion in the automotive industry had several thousand active contracts with customers and suppliers.
- Termination dates and price escalation clauses could only be tracked sporadically. For old contracts the information was not available at all, for new contracts it was recorded with fluctuating quality.

Solution:

- With the doculect standard module "Contract Master Data", relevant information was automatically extracted from contracts.
- Critical cases were submitted to caseworkers for review – human-in-the-loop.

Result:

- Important contract deadlines were neatly recorded in the DMS, responsible persons were automatically reminded in sufficient time.
- In the following 12 months, savings in the seven-digit range were achieved through contracts that were terminated or renegotiated on time and price escalation clauses that were utilised.



Solution: Automatic extraction of structured information with built-in high quality control loop

Document analysis is a very large and diversified field with a multitude of use cases. The inherent complexity comes from three main sources:

1. Different document types such as contracts, e-mails, letters, forms, notes, historical documents from company archives, but also transcriptions of audio files each have their own characteristics and are available in different formats.
2. Relevant information is not expressed in a uniform way in documents. Effective document analysis must recognise information from different terms and depending on the context.

The diagram illustrates the process of document analysis and information extraction. It shows a document snippet with highlighted sections and a corresponding structured output table.

The document snippet contains the following text:

Orgus Baustoffe GmbH, Glockengießerwall 1168, 27168 Northeim **Stammdaten** .

Mit der Gesellschaft (übernehmender Rechtsträger) ist aufgrund des Verschmelzungsvertrages vom 15.12.2019 mit Nachtrag vom 18.01.2020 die Gesellschaft mit beschränkter Haftung ABC GmbH, Dresden (Amtsgericht Dresden, HRB 12345) verschmolzen. Auf die bei Gericht eingereichten Urkunden wird Bezug genommen **Verschmelzung**

The structured output table is as follows:

Inhalte	
Stammdaten	Verschmelzung
Orgus Baustoffe GmbH Firmenname	Amtsgericht Dresden, HRB 12... Aufgenommene Gesellschaft (Reg.)

Figure: Depending on the target, relevant information is extracted and classified.

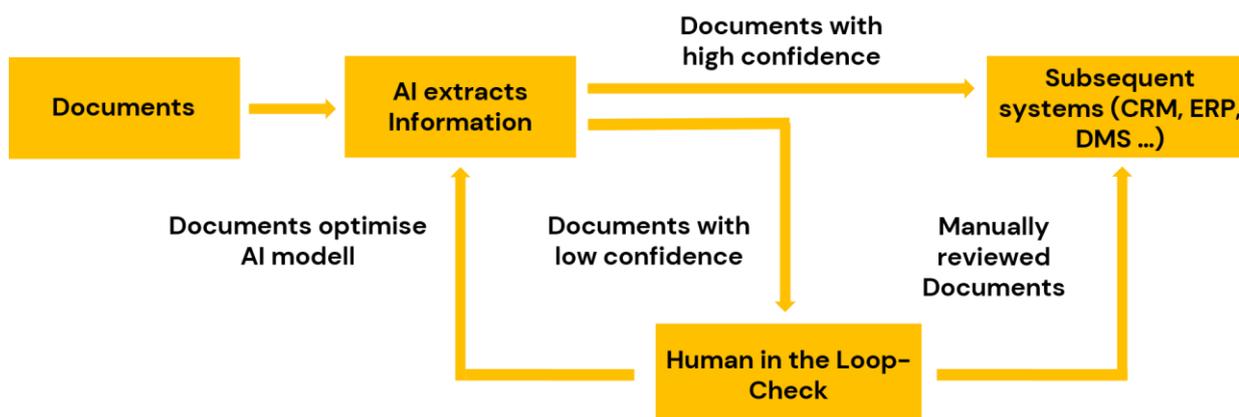
3. Information must be received along a wide range of processes and then processed in a structured way and returned, which means that flexible integration with various systems is required.

Technically, a distinction must also be made between implementation as real-time and mass processing of the documents ("documents" here and in the following as a generic term for all data that are available digitally).

Economically, it should be noted that for recurring real-time processing, there are often existing AI models that can be individualised and adapted to corresponding goals. For one-time mass processing of documents, on the other hand, it is regularly necessary to create dedicated models. It is therefore important to check whether the individual case justifies the effort and whether a regular benefit can be identified for the model once it has been trained.

Mass processing:

Central results are collected in an archive that is continuously enriched with all digitised documents as well as intermediate results from the workflows. These findings are used to train AI models. The human-in-the-loop is essential here:



The monitoring and evaluation of the analysis results is carried out by specialists in order to intervene and control if necessary. These models, developed through mass processing, are then later used in a supportive way in real-time processing.

Unsupervised learning creates special added value here, for example through the possibility of topic analyses that semantically summarise information about the contents of a document archive. In this way, changes in customer behaviour or an accumulation of certain processes can be identified. Specific pattern recognition ensures that conspicuous features are detected and that the data stock remains of high quality. Questionable documents are indexed in this way and identified separately for manual examination and transferred to the human-in-the-loop.

Real-time processing:

Optimising processes/workflows by creating structured data is the goal of real-time processing in document analysis. Continuously learning AI modules are used for this purpose. Put simply, the value creation takes place in four steps: pre-processing and loading of the documents, extraction of the desired information, if necessary, transfer to the human-in-the-loop or to further defined redundancy processes and finally return of the desired information in the defined data format for further processing.

Along this value chain, different AI modules are used - depending on the requirements and data source and data quality.

For example, a classification model can be used to automatically assign incoming documents to the correct processor, subsequent system or archive folder. With an extraction model, relevant information such as customer or transaction number, contracting parties or details of the transaction (for example, damage details in an insurance notification) can be automatically recognised and processed. A clustering model can detect irregularities in customer communication or on social media channels in real time, enabling timely intervention.

Interim summary:

Natural Language Processing based on Machine Learning is the booster for intelligent document analysis and enables next level process automation.

Solution: Increased process automation and a strategic knowledge advantage

Intelligent document analysis bridges established application systems and enables the automation of knowledge work.

Patterns enable potentials - unleash the potential

Patterns can be searched for and recognised in structured data. These regularly hold enormous potential for the optimisation of products, services and processes - which could not be raised manually in any commercially justifiable cost-benefit ratio. Intelligent document analysis, which

connects inventory systems and evaluates unstructured data, makes it possible for the first time to raise this potential in a scalable way and to systematically realise the associated business successes.

Automate and optimise business processes

Processes can be accelerated through document analysis - shortened response times increase conversion and customer satisfaction in equal measure. Internally, processes are simplified, accelerated and continuously adapted to changing situations. Certain process steps are eliminated altogether, so that professionals individually and companies as a whole can concentrate more on their core competencies.

Identify unrecognised business risks

In many documents and the processes associated with them, not only potentials but also risks lie dormant. Recognising these efficiently is a great business advantage. With intelligent document analysis, companies can react in a targeted manner - and act proactively to prevent such risks.

Support professionals

Automated processes that perform intelligent document analyses relieve valuable employees who are otherwise tied up in monotonous routine processes. This relief frees up capacities for higher-value tasks. As a side effect, this also increases employee satisfaction.

Quintessence:

As soon as documents are digitised and intelligently evaluated, there are many opportunities to optimise the value chain. Previously unknown patterns and correlations become visible, potentials are recognised, risks are avoided and knowledge about processes is extracted.

Case Study: Identifying cyber risks in commercial contracts

Identifying explicit and implicit inclusions of cyber risks in commercial insurance contracts in wholesale individual business.

Situation:

- Cyber is explicitly and implicitly included in insurance contracts.
- These contracts are not standardised and can therefore usually not be documented in a structured form.
- There is no clarity on the overall cyber risk in the portfolio.

Challenge:

- The phrasing in contracts is heterogeneous.
- The contract portfolio is inconsistent due to indexation and different tariff generations.
- Underwriters are too valuable as a resource to be tasked with a manual evaluation.
- Conventionally, large amounts of data are needed to train an NLP model.

Solution:

- Identification of typical wordings on the basis of a few sample documents.
- Training of an NLP model with corresponding synthetically generated training data.
- Evaluation of the existing database with the trained model.

Extract unstructured information from contracts and claims notices using **Natural Language Processing (NLP)**.

Direct operative benefits:

- **More turnover** through better conversion of offers thanks to significantly faster processing.
- **Lower costs** in claims processing thanks to automatically generated summaries of the damage reports.

Results:

- More turnover thanks to faster evaluation and processing of offers.
- Structured database of all historical cyber covers in the portfolio.
- New contracts are automatically evaluated and added.
- Real-time risk and profitability assessment for cyber across the entire portfolio.
- Accumulation analysis for cyber can be partially automated.

Roadmap: From proof of concept to enterprise platform

It is recommended to implement intelligent document analysis in a structured process. A promising and small-scale proof of concept (PoC) should first be envisaged before the ultimate goal of a self-sufficiently usable NLP platform for the entire company is conceived and implemented.

This ensures that the necessary technical and professional expertise can be built up successively and that all employees recognise and learn to appreciate the value of the technology. Only in this way can a solution be developed and deployed, and a precise roadmap with corresponding KPIs be drawn up. Roughly summarised:

- In a kick-off workshop, the current situation is evaluated, the problem is elicited and the goal is defined.
- In the first phase, the pilot solution, the problem is penetrated, cast into assumptions and solved by algorithms and models. The result is a proof of concept. Duration approximately: 3-6 weeks.
- In the second phase, implementation and integration, the solution is prepared for productive use and iteratively integrated into relevant systems and processes. Duration: 1-3 months.
- In the third phase, the performance of the document analysis is evaluated and monitored - the solution is continuously adapted and optimised. This phase is de facto an ongoing process from implementation and integration onwards.

Summary

Extracting structured data from documents and automating knowledge-based processes can already be implemented quickly and pragmatically today - without having to set up or establish lengthy projects with many different departments.

Technical implementation is also less of an issue, as an intelligent document analysis platform can be used independently of existing system landscapes. This applies to the case when a new document management system is to be integrated or an existing system is to be adapted, as well as in the case of robotic process automation.

Recommendation

If you already have data teams in place, if you have already set up RPA, if you have an up-to-date real-time DMS, then consult on how you can quickly and pragmatically generate valuable assets from unstructured data with simple means and solutions. Because intelligent document analysis can build the bridge that leads you to your goal on your way to business challenges.

About tetrel

At tetrel, autonomous data science teams work for companies and government organisations on turnkey document analysis solutions to extract structured data from complex documents. These teams are assembled individually according to requirements.

tetrel is an independent and self-funded company. Together with our partners, our mission is to identify strategic use cases for Artificial Intelligence, evaluate them and turn them into productive solutions. In doing so, we understand AI as a tool to support highly qualified employees and a lever to efficiently increase productivity.

Your contact



Johannes Humbert
+49 176 83 33 51 46
johannes.humbert@tetrel.ai