



Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe

D5.1 Intermediate demonstrator for pilot 1

| | |
|---------------------------------|--|
| PROJECT ACRONYM | Lynx |
| PROJECT TITLE | Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe |
| GRANT AGREEMENT | H2020-780602 |
| FUNDING SCHEME | ICT-14-2017 - Innovation Action (IA) |
| STARTING DATE (DURATION) | 01/12/2017 (36 months) |
| PROJECT WEBSITE | http://lynx-project.eu |
| COORDINATOR | Elena Montiel-Ponsoda (UPM) |
| RESPONSIBLE AUTHORS | Anna Katharina Strumegger (openlaws), Christian Sageder (openlaws) |
| CONTRIBUTORS | |
| REVIEWERS | Pascual Boil Ballesteros (CC), Pieter Verhoeven (D.V.G.L), Socorro Bernardos (UPM), Víctor Rodríguez-Doncel (UPM), Elena Montiel-Ponsoda (UPM) |
| VERSION STATUS | V1.0 Final |
| NATURE | Demonstrator |
| DISSEMINATION LEVEL | Public |
| DOCUMENT DOI | 10.5281/zenodo.3558269 |
| DATE | 29/11/2019 (M24) |



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780602

| VERSION | MODIFICATION(S) | DATE | AUTHOR(S) |
|---------|--|------------|--|
| 01 | Initial draft | 10/11/2019 | Anna Katharina Strumegger (openlaws), Christian Sageder (openlaws) |
| 02 | First review | 14/11/2019 | Pieter Verhoeven(DNV GI) |
| 03 | Second review | 18/11/2019 | Pascual Boil (CUATRECASAS) |
| 04 | Third review | 22/11/2019 | Socorro Bernados, Víctor Rodríguez-Dondel, Elena Montiel-Ponsoda (UPM) |
| 05 | Updated version based on Input from review | 25/11/2019 | Anna Katharina Strumegger (openlaws), Christian Sageder (openlaws) |
| 1.0 | Final | 29/11/2019 | Christian Sageder (openlaws) |

DISCLAIMER

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. Neither the Lynx consortium as a whole, nor a certain party of the Lynx consortium warrants that the information contained in this document is capable of use, nor that using the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

EXECUTIVE SUMMARY

The main purpose of this document is to give an insight into the current development status of pilot 1 (contract management) which is being developed within the Lynx project.

Contracting is a common activity in companies, but managing contracts systematically is a cumbersome activity only few companies are effective at. Due to a lack of central administration, contracts are physically and electronically distributed across the entire organization. Against this backdrop, we are focusing on automated contract analysis to provide smart contract archiving solutions and compliance services.

In order to achieve this, two approaches have been pursued:

- A pure back-end solution to analysis single and multiple contracts including an archive which can be used by other services / applications.
- A front-end solution where a single user has the possibility to manage contracts.

For the front-end solution selected provisional screenshots and descriptions are provided. In addition, a description of the architecture and newly developed services for the back-end is given.

This deliverable is structured as follows. The first section deals with the relevance of contract management and embeds the efforts into the broader Lynx context. Section 2 describes several use cases we aim at solving with this pilot. In light of this, we are discussing objectives, our working basis in the pilot phase and the visualization of basic modules. Section 3 (proof of concept) and 4 (pilot architecture) contain information about the underlying back-end solutions. Finally, the deliverable closes with an outlook.

TABLE OF CONTENTS

| | | |
|-----|---|----|
| 1 | INTRODUCTION | 6 |
| 2 | USE CASES AND OBJECTIVES | 7 |
| 2.1 | Objectives and working basis in the pilot phase | 7 |
| 2.2 | Visualization of basic modules | 10 |
| 3 | PROOF OF CONCEPT | 15 |
| 3.1 | Contract Analysis | 15 |
| 3.2 | Classification Results | 17 |
| 4 | PILOT ARCHITECTURE | 18 |
| 4.1 | Introduction | 18 |
| 4.2 | Broker Application | 19 |
| 4.3 | Document Service | 20 |
| 4.4 | Annotation Service | 20 |
| 4.5 | RegExFinder Service | 21 |
| 4.6 | Scoring Service | 22 |
| 4.7 | CPIndex Service | 22 |
| 4.8 | Lynx Services | 22 |
| 5 | Outlook | 24 |
| 6 | References | 25 |

TABLE OF FIGURES

| | |
|--|----|
| Figure 1 Solution approaches | 7 |
| Figure 2 Sample of an index clause | 8 |
| Figure 3. Main page for a single contract | 10 |
| Figure 4. Table of Content | 11 |
| Figure 5. Detailed View of an Extracted Information Unit (Seller)..... | 11 |
| Figure 6. Information Query | 12 |
| Figure 7. Manually Add Information (Part 1)..... | 12 |
| Figure 8. Manually Add Information (Part 2)..... | 13 |
| Figure 9. Manually Add Information (Part 3)..... | 13 |
| Figure 10. Further Legal Information..... | 14 |
| Figure 11. Further Legal Information (Detailed View)..... | 14 |
| Figure 12. CSV File and Directories | 15 |
| Figure 13. Information Representation in the Customer’s CRM System | 16 |
| Figure 14. Back-end components | 18 |
| Figure 15. Sequence Diagram | 19 |
| Figure 16. An order-request within neo4j database | 20 |
| Figure 17 Annotation Service | 21 |
| Figure 18 Sample definition for regex pattern | 21 |

LIST OF TABLES

| | |
|--|----|
| Table 1. Relevant Extractable Information out of a contract..... | 9 |
| Table 2. Classification Categories | 16 |
| Table 3. Classification Results of the First Run | 17 |

1 INTRODUCTION

Contracting is a common activity in companies, but managing contracts systematically, which means keeping track of changes or updates, is a cumbersome activity only few companies are effective at. In addition, most companies do not have a database with all the information in their contracts, which prevents them from easily finding information or applying changes.

Let us imagine the following situations in the context of a company:

- a) A contract is needed urgently and no one knows where to find the latest version, because the responsible employee left the company. Moreover, the opposing party confronts you with a signed amendment you've never seen before.
- b) There is a change in law, and you need to know which contracts are effected.
- c) An overview about all obligations with a certain company is needed.

Countless organizations are confronted with similar scenarios, although we are all significantly shaping our legal reality by concluding various contracts. Abstractly, the problem can be summarized as follows: Due to a lack of central administration, contracts are physically and electronically distributed across the entire organization. As a result, often there is no overview, which leads to inconsistent applications, breaches of contracts and (financial) disadvantages.

The implementation of a comprehensive cross-organizational contract management process appears to be the solution. Flitsch (2010) defines contract management as the creation of ideal structures for:

- contract planning
- contract design
- contract negotiations
- implementation of contracts
- contract administration
- contract archiving

In many cases, organizations are lacking these structures. Against this backdrop, we are focusing on automated contract analysis. Building on this, we provide smart contract archiving solutions and compliance services. We expect our application to result in enhanced contract compliance, which will ultimately lead to reduced risks and costs for organisations.

All these efforts are embedded into the broader Lynx context and the associated assumption that building a legal knowledge graph - duly interlinked and integrated - results in reducing possible distances to the applicable law and, thus, in facilitating compliant and diligent actions. To this end, we are channeling our efforts to fulfil what Richard Hamming (1962) so aptly formulated decades ago: "The purpose of (scientific) computing is insight, not numbers."

2 USE CASES AND OBJECTIVES

The starting point and, at the same time, the simplest use case is the analysis of a single contract. However, reality is much more complex: Regularly, a large number of contracts of diverse nature and purpose needs to be analysed and kept track of, taking into account various regulatory frameworks. In order to achieve this, we are currently pursuing two approaches. On the one hand, we are working on pure back-end solutions, and, on the other hand, we are providing a visualisation of the created data space for end users. Figure 1 illustrates the different use cases or rather solution approaches in a general manner:

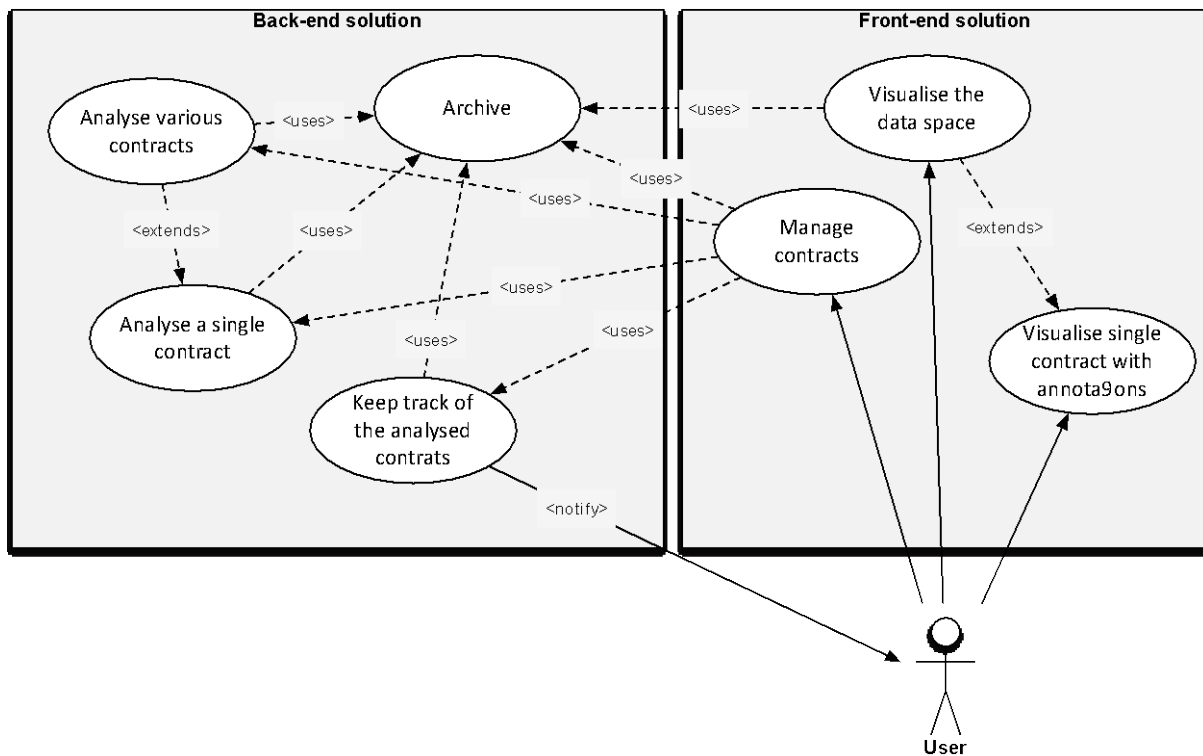


Figure 1 Solution approaches

As core functionality, the back-end solution provides the analysis of single contracts. In addition, multiple contracts can be analysed. Both functionalities are based on the archive where contracts and all the extracted information is stored. These services can be used by other services / applications.

Through the front-end solution a single user has the possibility to manage (add, delete, update, group, search, etc.) contracts. The user can view a single contract and related annotations or get a broader view of the corresponding data space, e.g. legislation, similar contracts, other contracts with the same partner, etc. In addition, the user is notified when legislation changes with effects on a given contract. All of this supports companies in achieving compliance.

2.1 OBJECTIVES AND WORKING BASIS IN THE PILOT PHASE

As outlined above, the following three elements are crucial when it comes to defining the objective of our application:

- contract analysis
- contract archiving
- compliance services creation

In the course of the pilot phase, our goal is to develop reasonable strategies for automated contract analysis and contract archiving. In a first step, we **concentrate on the back-end** of the application and **visualise basic modules** to provide all necessary services and evaluate possible solutions.

Our proof of concept is based on a rich dataset of 100 000 documents (of different type, e.g. contracts, communication and format, e.g. MS-word, email) provided by a real customer of openlaws, the Lynx partner leading the pilot development. Previously, the pilot was trained with publicly available contract templates, which proved to be inefficient, as general templates did not provide the variety needed for our queries. In addition to the data set mentioned above, we use a testing data set of approximately 10 000 documents, which provides a collection of different contract types, e.g. non-disclosure agreements, rental contracts, decisions of civil authorities, etc. This dataset offers a wide variety, as the contracts originate from different sources. This data set will be used at a later stage of the project.

With the primary training dataset, we strive for a highly **practice-oriented goal**: We aim at identifying **index clauses**, that is, those clauses that refer to the price being tied to a consumer price index to ensure value stability (Figure 2). **Moreover**, additional information, such as document dates, customer IDs, offer numbers, etc. should also be located. We assume that the ability to determine index clauses, will enable the identification of **other contractual clauses, as the process/procedure can be reused**.

“Index Assurance: Prices are subject to annual value adjustment in accordance with the Consumer Price Index (CPI) 2010 or the index replacing it. The starting point for the calculation of value assurance is the index number published for the month in which the offer was placed.”

Figure 2 Sample of an index clause

Moreover, we want to **classify** these documents according to content categories based on the retrieved information. Classification plays an important role here: On the one hand, it can provide various important stimuli for future work and, on the other hand, it is a crucial component of another envisioned service, contract clause evaluation. In the latter, our aim is to **generate useful evaluations**, for instance discover which clause is better, equally good or worse than a defined standard clause.

Index clauses are very common in long-term contracts between companies and between companies and consumers as well. These clauses allow companies to adapt prices according to inflation rates or other agreed price indices. To translate these clauses into action, companies need to identify such clauses and apply them in the relevant contracts. Sometimes it is even the case that there are various index clauses in use in the same contract, or there are some contracts that contain such clauses and others that lack them.

Obviously, it is necessary to actively extract information from contracts and other documents to achieve the above-mentioned objectives. Table 1 lists examples for relevant extractable information.

| Hard facts that entail major legal implications | | Documents related to contracts | Important background information |
|--|---|---|---|
| | Information available in standardised form | | |
| <ul style="list-style-type: none"> ● type of contract ● name of contracting partner (and parent company) ● contract signature date ● contract duration | <ul style="list-style-type: none"> ● index clauses ● termination options ● restrictions (such as exclusivity agreements) ● options and rights of first refusal ● change-of-ownership clauses ● minimum purchase requirements ● warranty periods ● securities (such as deposits, bank guarantees, etc.) and deadlines for recovery ● insurance obligations ● post contractual obligations (such as presentation of annual financial statements, reporting obligations) ● accounting standards and due dates | <ul style="list-style-type: none"> ● linked contracts, e.g. rental and sublease agreements, energy supply and cleaning contracts for rental properties ● documentation of contract fulfilment ● e-mails (from various employees) | <ul style="list-style-type: none"> ● contract owner (person responsible for the contract within the organization) ● contact person ● place of storage of the original document |

Table 1. Relevant Extractable Information out of a contract

Working with this real data set is very beneficial in terms of testing the capability and the performance of the application. However, this means working with highly confidential information. We agreed with our customer on performing our queries within the customer's environment, where we deploy parts of the openlaws and the Lynx System. Therefore, processing the data in the current Lynx environment is unfortunately not possible. For those services provided within the publicly available Lynx environment, it is indicated in the component description. When we present data segments for illustration purposes, we use anonymised data.

2.2 VISUALIZATION OF BASIC MODULES

As mentioned above, we are currently concentrating on back-end solutions. Nevertheless, we are also working on visualizing basic modules to better illustrate the functionalities of the pilot. In that sense, we present selected provisional screenshots of our **visualization of a single contract**, specifically of a purchase contract (“Kaufvertrag zwischen Unternehmern über bewegliche Sachen aus Sicht des Verkäufers”).

Figure 3 shows the main page for a single contract. The top bar contains the contract title as well as relevant metadata, e.g. document title, document date and language. Underneath, the contract is presented on the left-hand side and the headings of extracted information units (in this case about the seller (“Verkäufer”) and the buyer (“Käufer”)) on the right-hand side. Further information units are added based on the contract type.

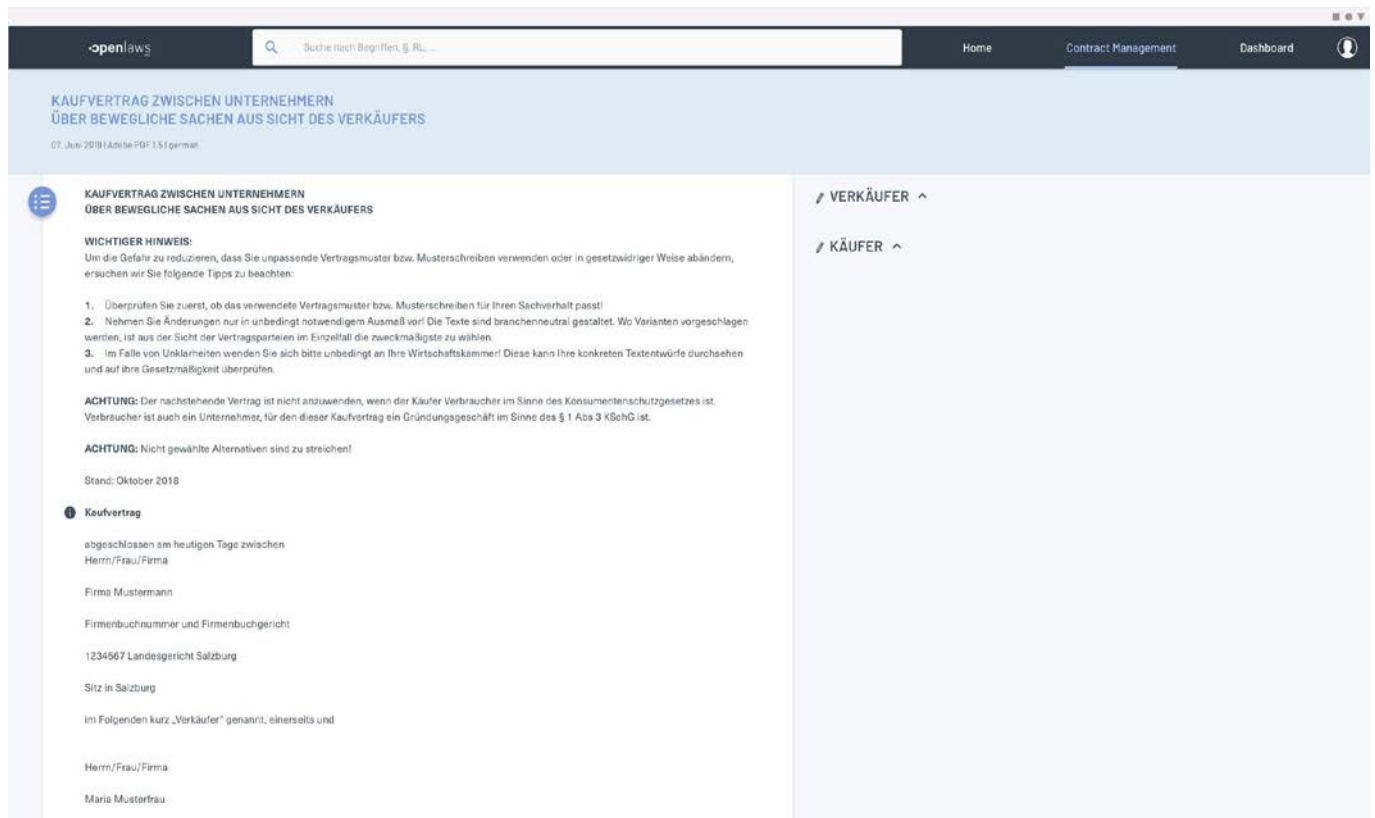


Figure 3. Main page for a single contract

Figure 4 shows the table of contents which is hidden behind the blue button on the left side. The user can click on a certain heading in order to instantly jump to the desired position.

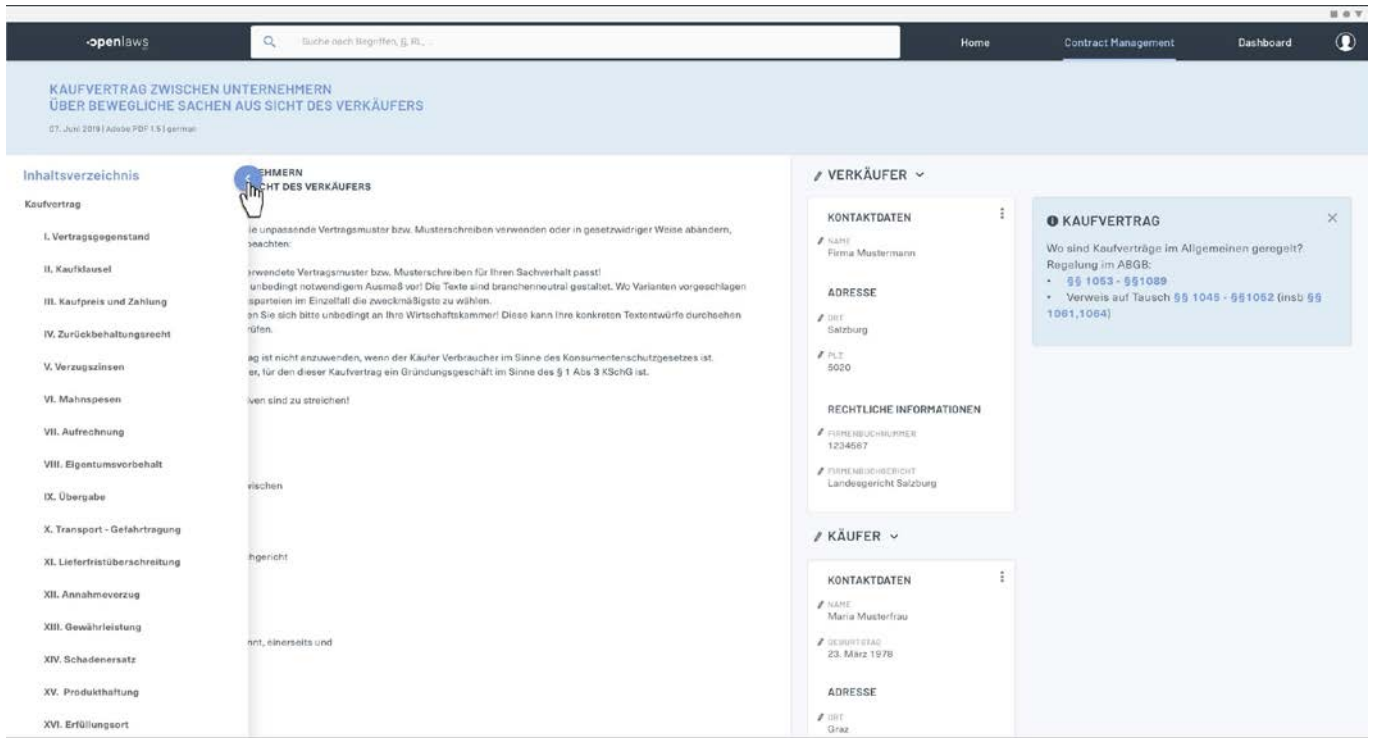


Figure 4. Table of Content

Figure 5 displays the detailed view of an extracted information unit. In this case it displays information about the seller ("Verkäufer"). The text passages from which we obtain the relevant information are highlighted in a light blue colour. These annotations are only visible when the respective detailed view is open.

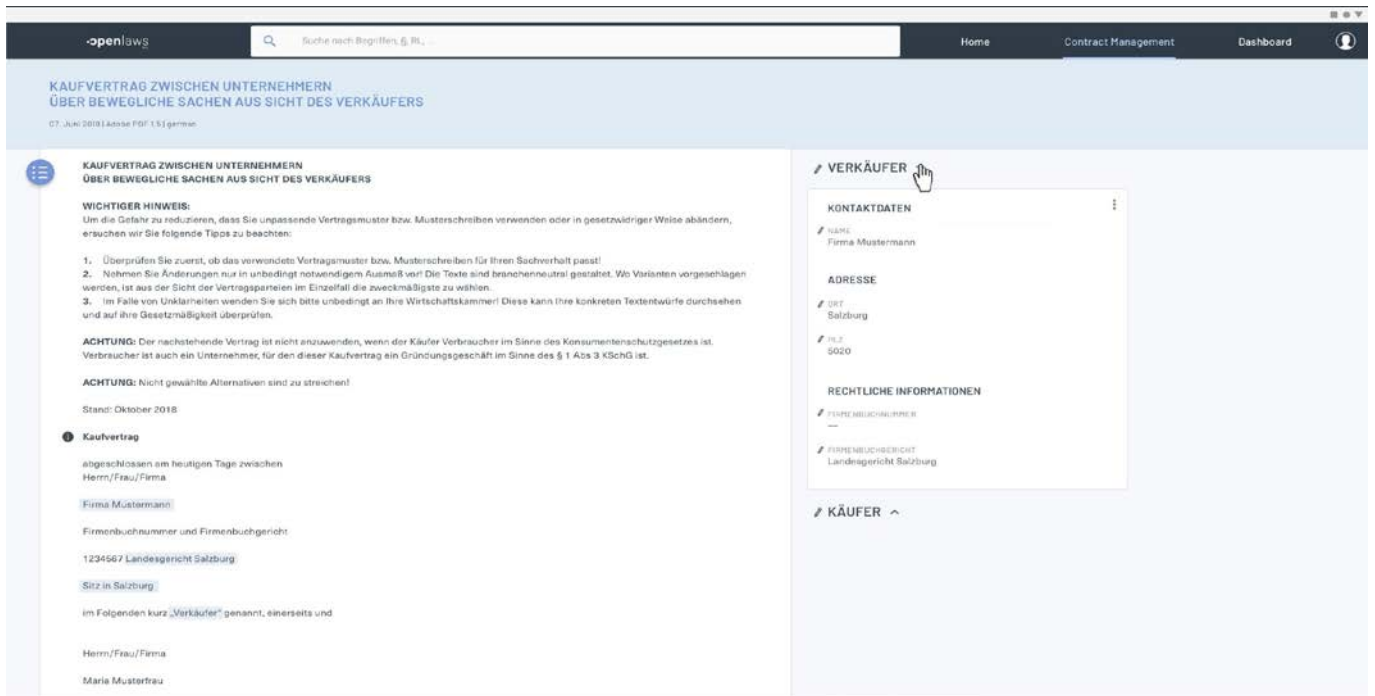


Figure 5. Detailed View of an Extracted Information Unit (Seller)

The user can select any of the annotations on the right-hand side to mark the position in the document. It helps users to find certain information in the document without reading the whole contract again. Figure 6 shows an example of this information query.

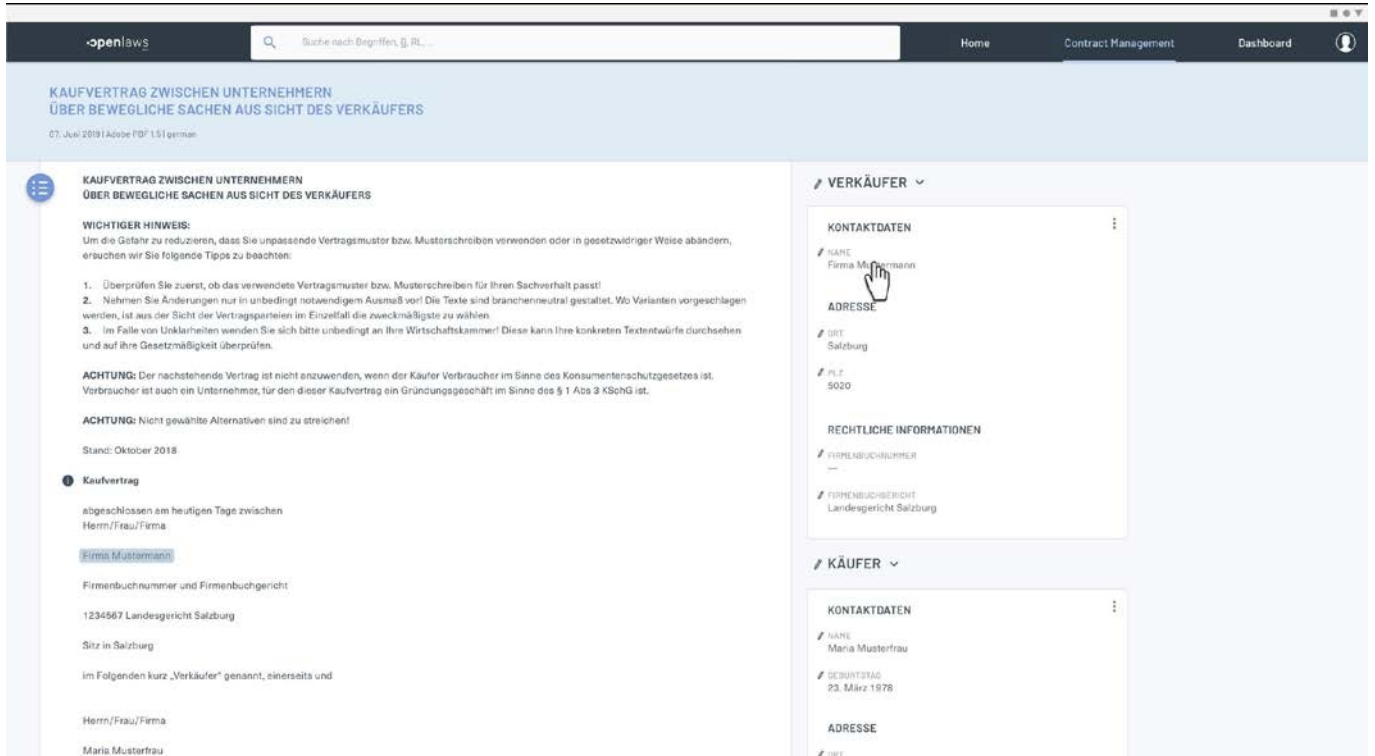


Figure 6. Information Query

If the system has not recognized the important information completely, the user can add it manually, preferably, by highlighting the relevant text passages. To do so, the user has to click on the pencil next to the title of the respective information unit. A mouse cursor to annotate the new information will appear. To get a new tag, the user has to mark the position in the document and drop the cursor. Afterwards, the new annotation appears on the right-hand side at the appropriate position. Figure 7, 8 and 9 illustrate this process.

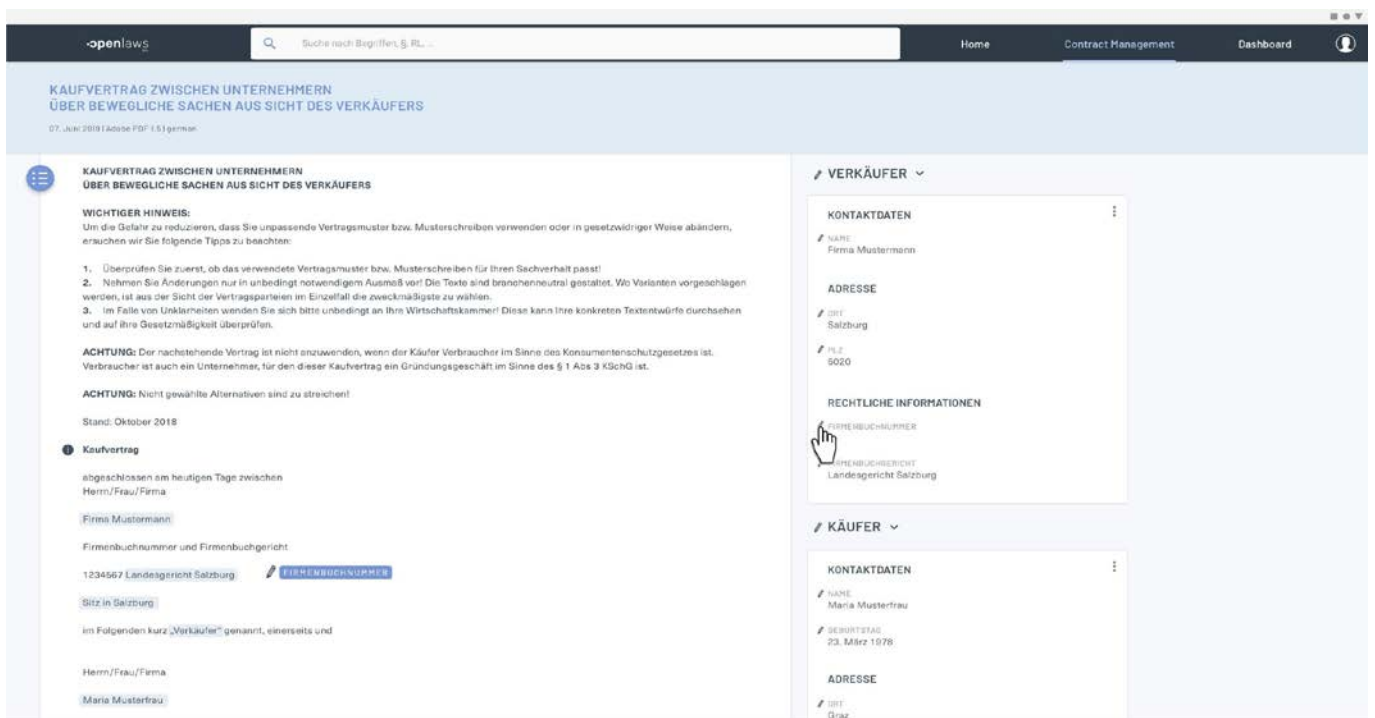


Figure 7. Manually Add Information (Part 1)

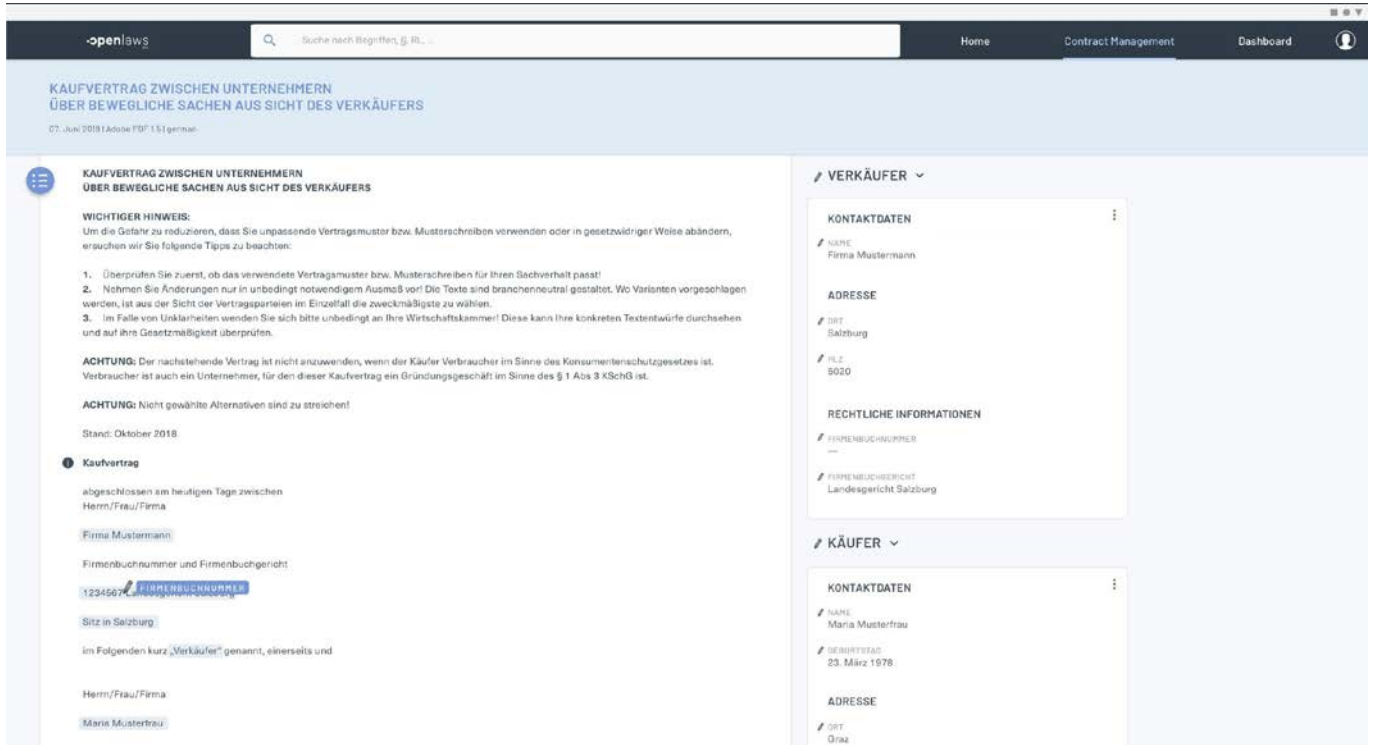


Figure 8. Manually Add Information (Part 2)

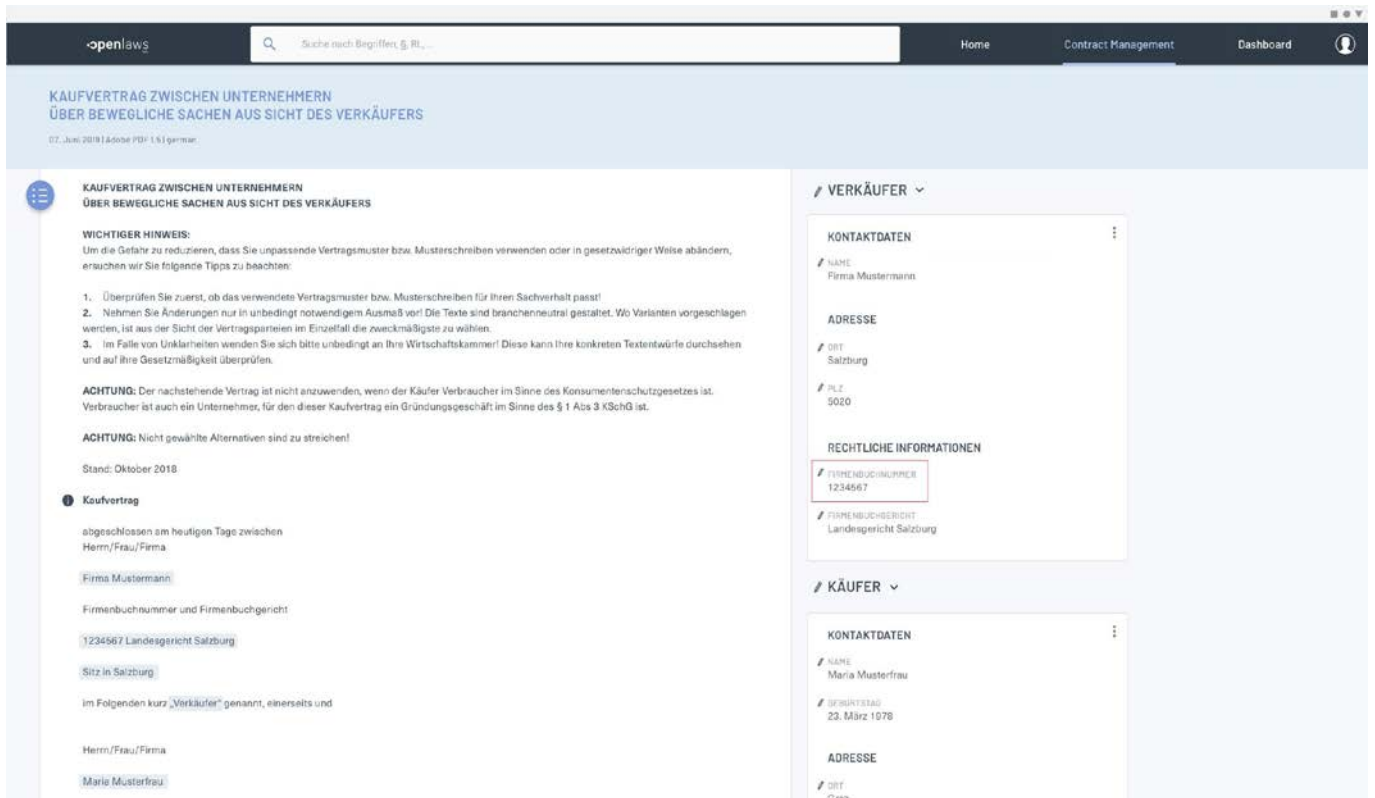


Figure 9. Manually Add Information (Part 3)

The user can also get further legal information about the contract. An info icon is shown next to every identified headline of the document. If the user presses the info icon, a box will appear on the far right (Figure 10). In the box the user gets information where e.g. the purchase agreement (“Kaufvertrag”) is defined in the legislation and further references. This information is provided by the existing openlaws system based on the annotations of the document.

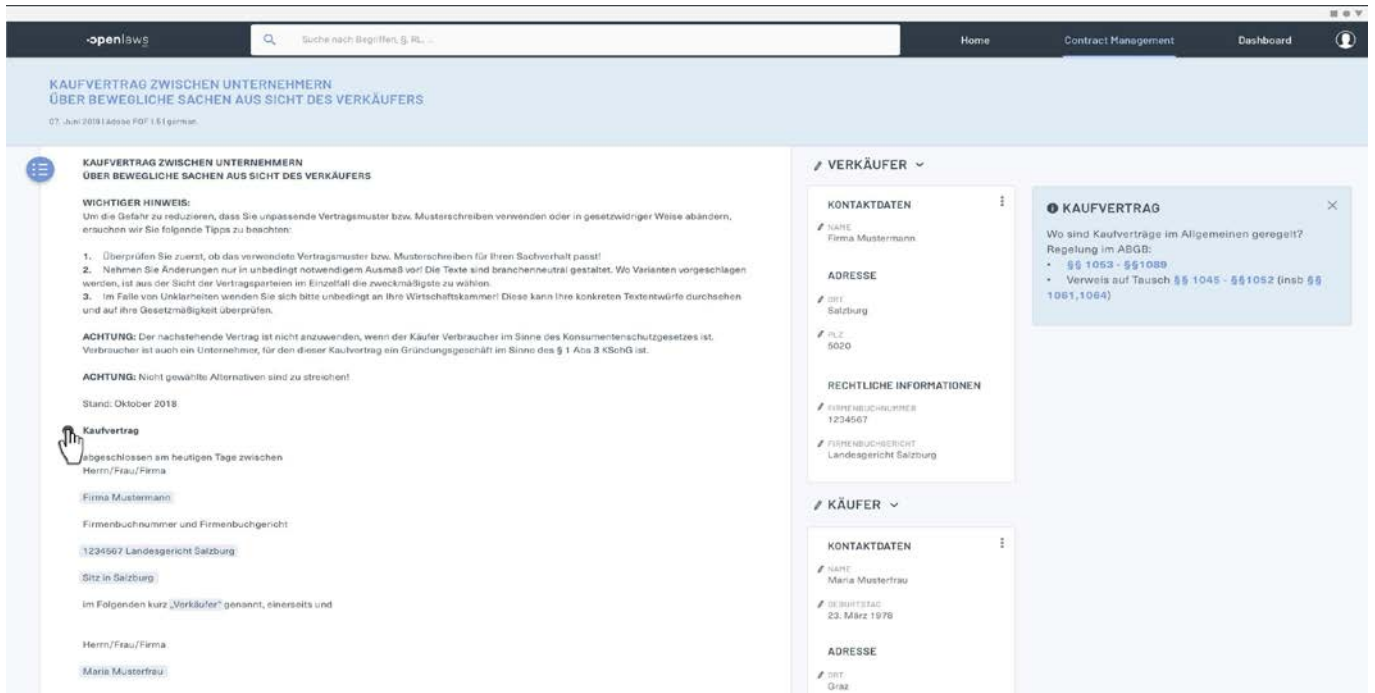


Figure 10. Further Legal Information

The user can click on the related regulatory information, e.g. §§1053 - §§1089 to find out more about it within the legislation without leaving the page (Figure 11).

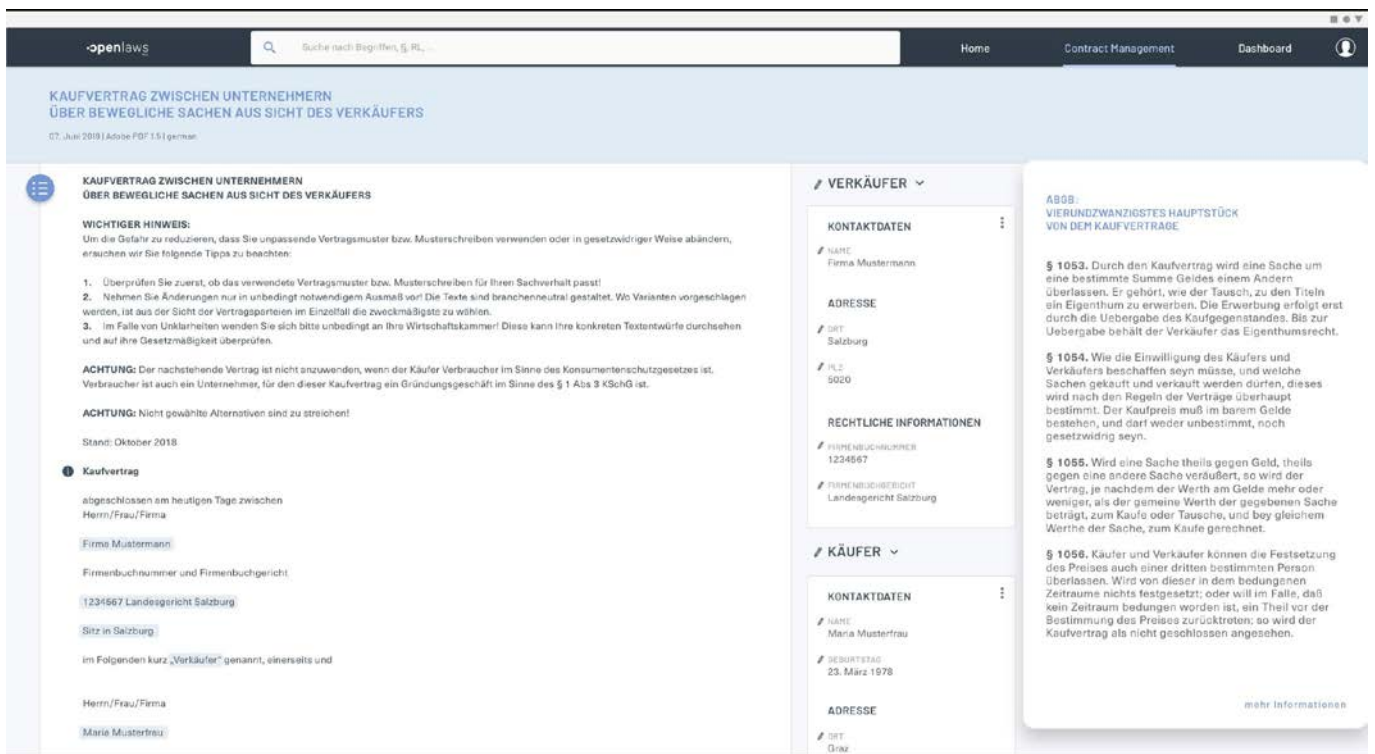


Figure 11. Further Legal Information (Detailed View)

3 PROOF OF CONCEPT

After having reviewed possible visualizations, this Section discusses the underlying back-end solutions.

3.1 CONTRACT ANALYSIS

Initially, our objective requires a rule-based classification of the documents in question. This classification is done on top of the annotations which are created by the different Lynx Services (Section 4.8). There are 4 classes (Class A to Class D). A document is assigned to a class depending on the availability of certain information (Table 2).

To achieve this, it was necessary to extract the 100 000 documents from our customer's CRM (Customer-Relationship-Management) system. These documents are grouped by orders and exported in the following way:

- Metadata of an order is stored as a single line in a defined csv-file.
E.g., order number, opportunity number, reference to a frame contract, quote number, offer date.
- Documents are attached to an order in a separate directory, one for each order.
The format of the attached documents differs in the following ways:
 - e-mail (outlook), partially with attachments (Often, these attachments were in returned e-mails.)
 - MS-office documents (word, excel)
 - pdf (either searchable or only as a scan)
 - others, e.g. tiff

For illustration purposes, Figure 12 shows how orders are archived/stored, namely:

- the CSV file containing all orders which should be processed
- the folder with all orders, and
- an exemplary folder with all documents of a single order.

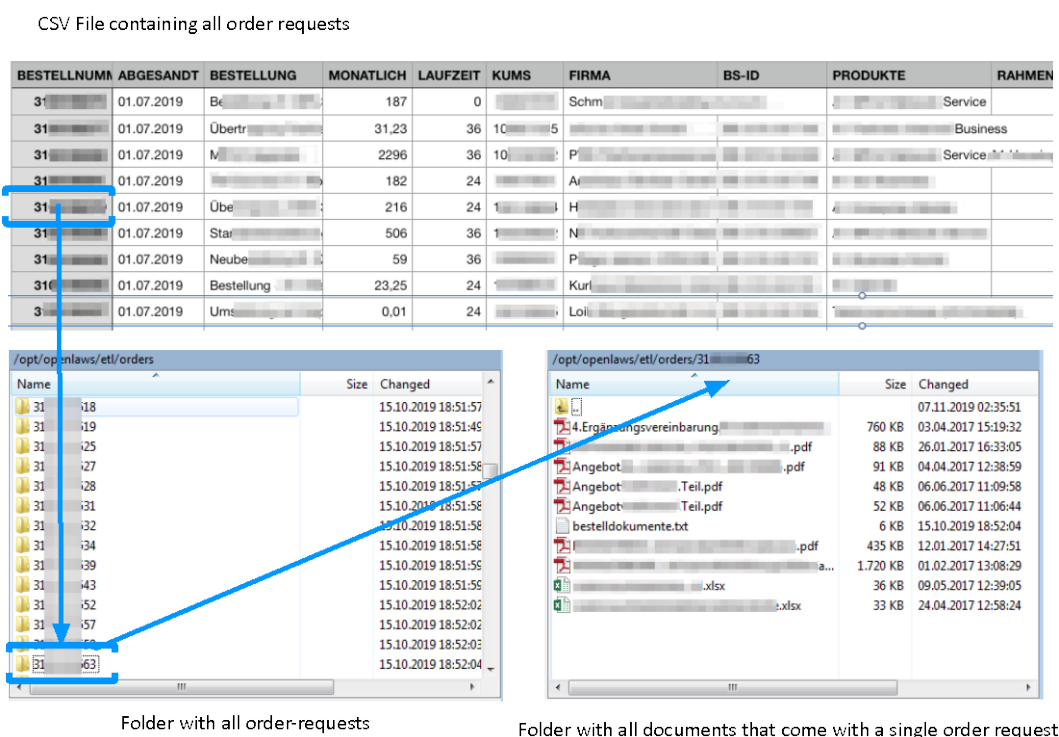


Figure 12. CSV File and Directories

Each of these orders is based on an offer which has been sent to the customer and was accepted by the customer. This is important because the classification of the documents builds upon offer data. Such a classification process results in four categories:

| Class | Requested Information | | | Further Remarks |
|---------|-----------------------------------|---|--|---|
| | Is the document an offer ? | Does the offer contain an index clause? | Does the index clause refer to a certain index , e. g. VPI (from Statistics Austria)? | |
| Class A | Yes | Yes | Yes | Additional automatic calculation of the index increase. |
| Class B | Yes | Yes | No, there is no index disclosed. | |
| Class C | C.1 | Yes | No, it only refers to the terms and conditions. | |
| | C.2 | | No, and it does not refer to terms and conditions | |
| Class D | No | --- | --- | Document cannot be assigned to class A, B or C. Either additional information and/or documents are missing, or the document is not an offer at all. |

Table 2. Classification Categories

Subsequently, the results are passed back to the CRM system through a SOAP-Call. In particular, the following information is passed back:

- Classification: A, B, C or D
- Explanation why a certain class is chosen, with additional information about the data source (e.g. file name, position of the index clause within the document, consumer price index reference)
- Additionally, for class A the price adjustment in percentage is calculated based on the difference of the original consumer price index when the order has been made and the current date.

Figure 13 illustrates how the information is presented in the current stage within the customer's CRM system.

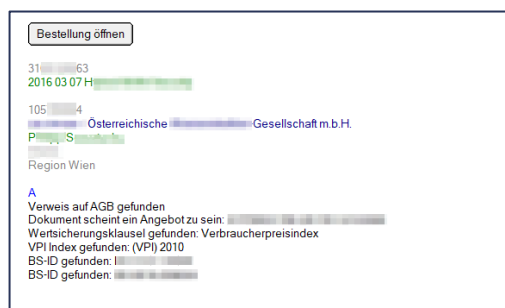


Figure 13. Information Representation in the Customer's CRM System

3.2 CLASSIFICATION RESULTS

We have performed now a first run over the complete dataset. In Table 3 we list the classification results of the first run:

| Class | Percentage Share | Further Remarks |
|---------|------------------|--|
| Class A | 20% | Documents in this Class A are mainly correctly identified. |
| Class B | 20% | In Class B we have a high number of false positives when it comes to the question: Does the offer contain an index clause? One of the root causes is that the rule for the definition of index clauses is too broad. Therefore, our application identifies sentences with specific keywords also as an index clause. |
| Class C | 10% | Documents in class C are mainly correctly identified. To support our customer, additional information should be extracted from the documents. |
| Class D | 50% | If a document is assigned to class D, there is a need for action in terms of assigning it to class A, B or C. On the one hand, this can be achieved, if our customer provides further information on the basis of our first results (e.g. frame contract). On the other hand, we are going to reduce this 50%-quota by optimizing our annotation and rule service. |

Table 3. Classification Results of the First Run

At the current stage we are analysing the results, mainly for Class B and D, to improve the different services.

4 PILOT ARCHITECTURE

4.1 INTRODUCTION

The front is developed with React¹, a JavaScript² library, and uses Material-UI³ as a component library. It communicates through REST⁴ calls with the openlaws⁵ system. The openlaws system provides all needed functionalities and includes user management, search, querying individual documents (legislation, case law, summaries and contracts) and the attached meta information including references to other documents.

The back-end system of the pilot is implemented in Java Spring⁶ and follows a microservice architecture like the rest of the openlaws and Lynx Platform. All services are individual applications providing REST interfaces.

As mentioned above, our main efforts have focused on the back-end components so far. Figure 14 shows only new components developed for this pilot. We will go into detail in the following sections.

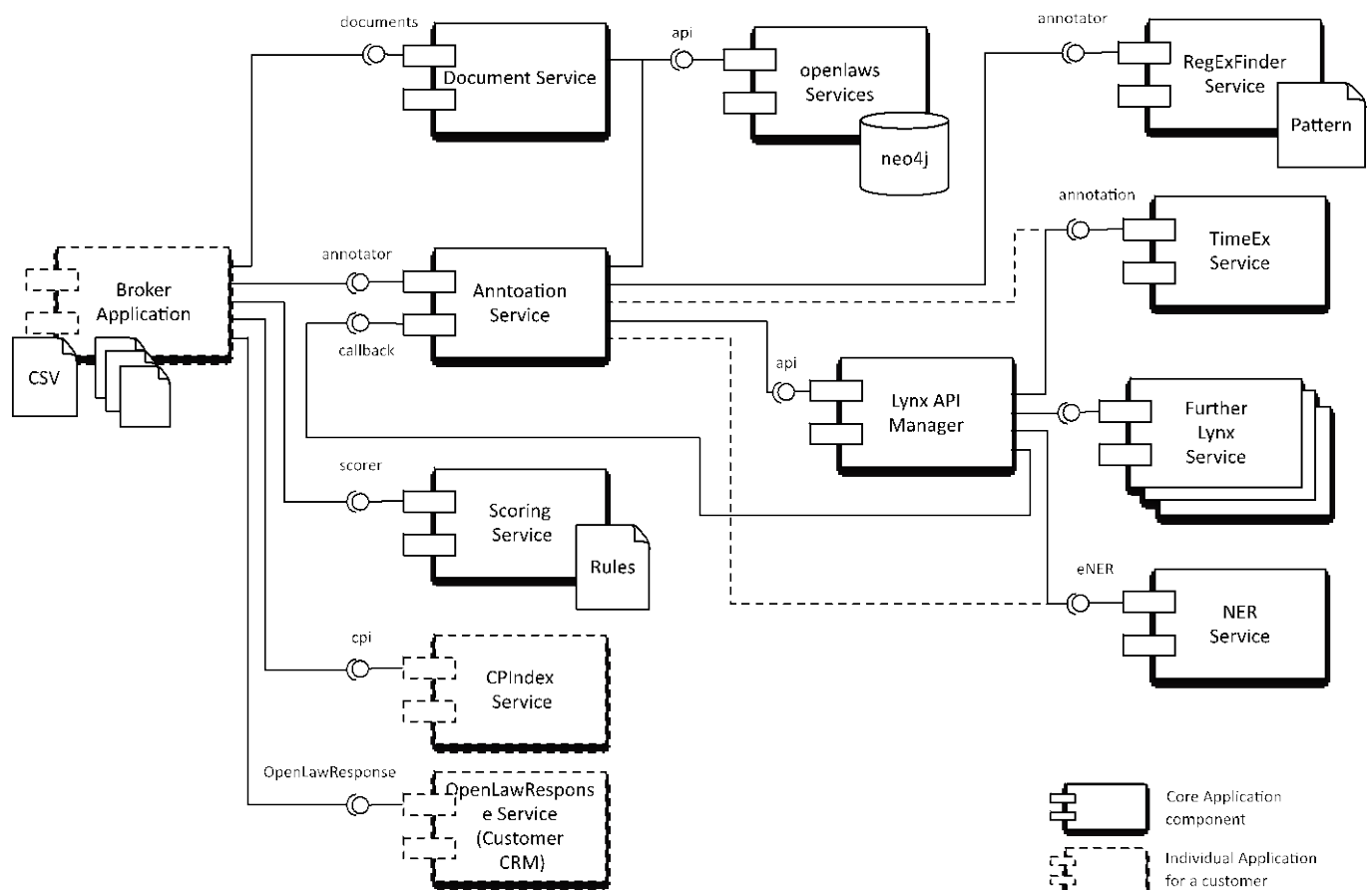


Figure 14. Back-end components

¹ <https://reactjs.org>

² <https://en.wikipedia.org/wiki/JavaScript>

³ <https://material-ui.com>

⁴ https://en.wikipedia.org/wiki/Representational_state_transfer

⁵ <https://openlaws.com>

⁶ https://en.wikipedia.org/wiki/Spring_Framework

4.2 BROKER APPLICATION

The Broker Application is the central component. It orchestrates the call to the different services and consists of a Java Spring Boot application which applies Spring Batch.

The application checks periodically if there are changes to the CSV⁷ file and, if so, the newly added lines are processed. First, a single line of the CSV file is read (see Figure 15). Based on the order number, it checks if there are documents attached to this order and passes this information to the Document Service (Section 4.3 where the text is extracted and converted into a Lynx Document for further processing with Lynx Services. Once the information is extracted, the Annotation Service (Chapter 4.4) is called. Based on the annotations, the Scoring Service (Section 4.6) scores the individual document and those scores are used further to classify the complete order. Optionally, an index calculation through the CPIIndex Service (Section 4.7) is done. The result is passed back to the Customer CRM System (OpenlawsResponse Service).

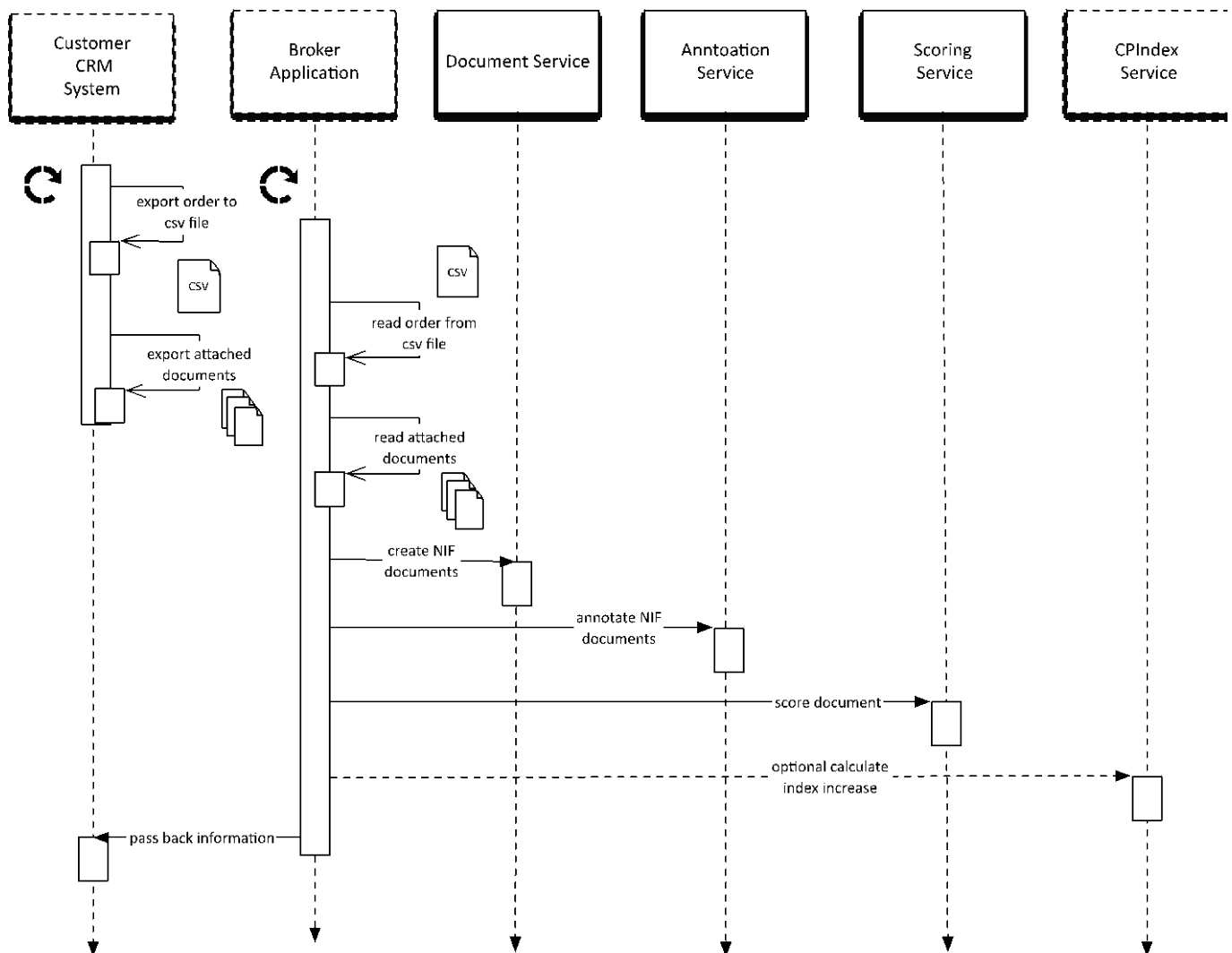


Figure 15. Sequence Diagram

The main parts of this service will be transformed into a new service for monitoring directories in the file system. In case such changes occur, the service will pass them to the openlaws system.

⁷ https://en.wikipedia.org/wiki/Comma-separated_values

4.3 DOCUMENT SERVICE

The Document Service is a Java Spring Boot application. It provides a REST API to extract text content and meta information of different document formats, e.g. e-mail, office documents and pdf files. Afterwards, this information is returned in the form of a Lynx Document enriched with metadata⁸ (in Json-LD format).

In case a document is an e-mail, the attachments are also extracted in a recursion. Finally, a list of Lynx Documents including all e-mails and related attachments is returned. For pdf files an OCR process is performed.

In addition, the documents are stored into the local openlaws documents store, which is a neo4j database (see Figure 16).

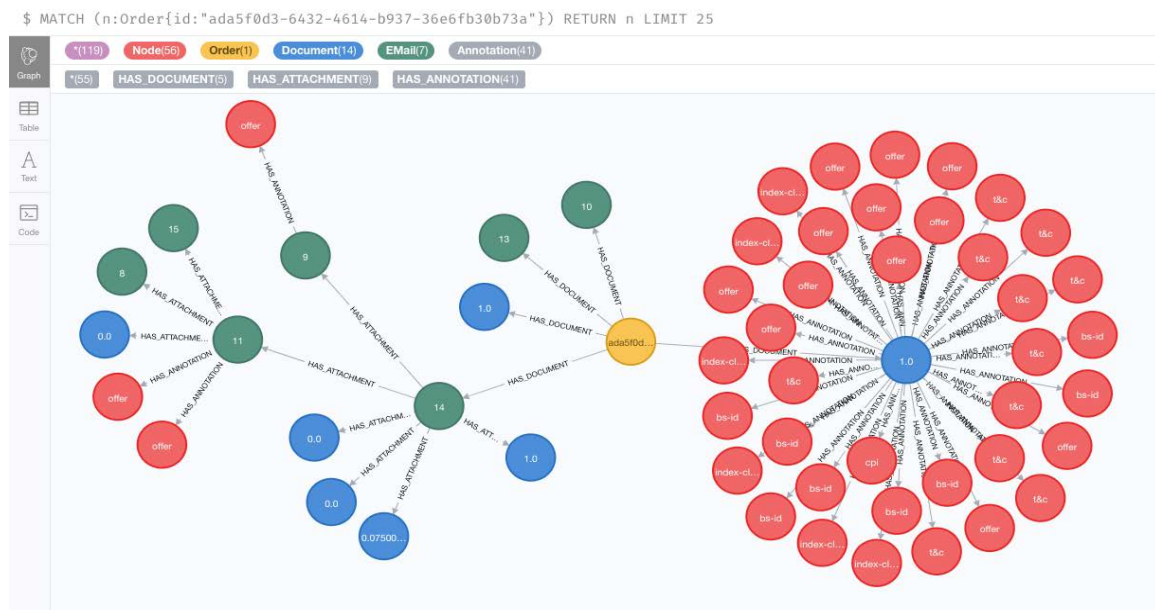


Figure 16. An order-request within neo4j database

The OpenAPI specification for the service can be found online⁹. In addition, a small demo application which returns a Lynx Document for different file formats is available online¹⁰. This demo application does not store any data in the database.

In the future this service will also be available through the Lynx Platform. In this case it will not store any document. The persistent will be done through the Lynx Workflow.

4.4 ANNOTATION SERVICE

The Annotation Service is a Java Spring Boot application which orchestrates the calls to different Lynx Services (Figure 17 Annotation Service) (See also Section 4.8 Lynx Services), Lynx Workflows¹¹, and openlaws services, and persists the result within the openlaws document store.

This service provides also the callback endpoint for the Lynx Workflow Manager for asynchronous annotation requests.

⁸ <http://lynx-project.eu/data2/data-models>

⁹ <http://document-service-88-contracts.cloud.itandtel.at/api/swagger-ui.html>

¹⁰ <http://document-service-88-contracts.cloud.itandtel.at/api/view/index.html>

¹¹ D4.3 Final version of Workflow definition (<https://doi.org/10.5281/zenodo.3235767>)

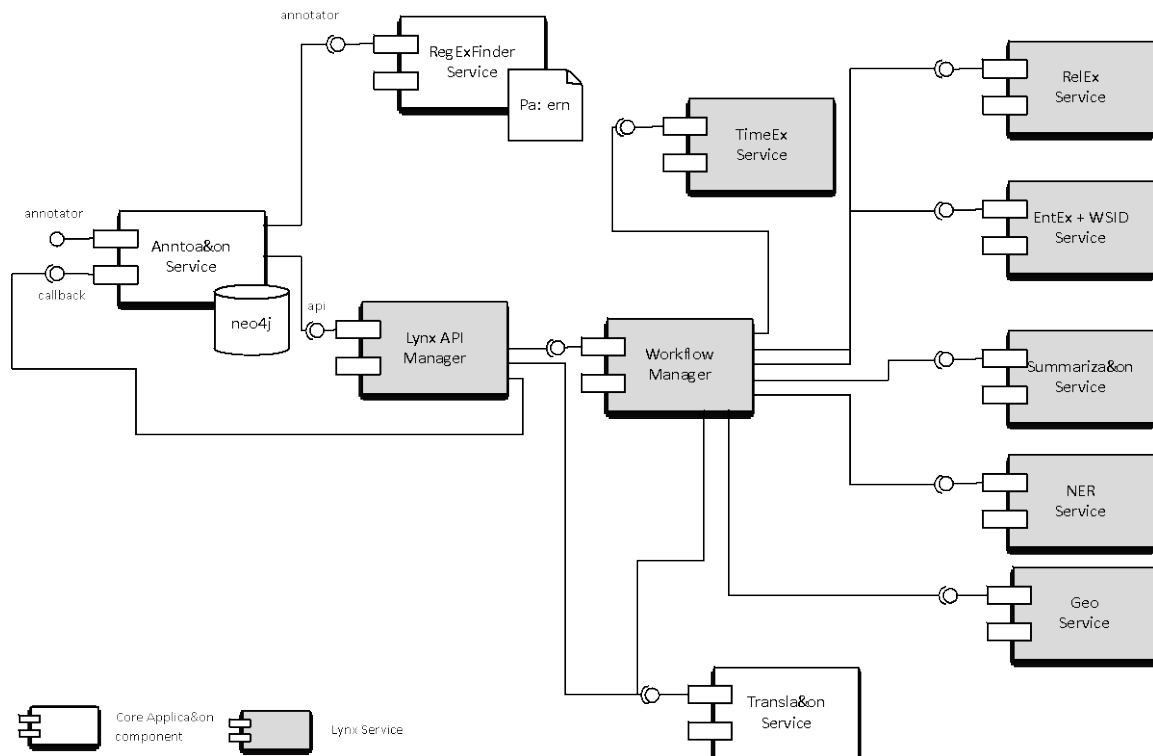


Figure 17 Annotation Service

4.5 REGEXFINDER SERVICE

The RegExFinder Service searches for entities based on Regular Expressions¹² (Patterns) and passes them back as NIF annotations. This service is used to extract entities like offer number, customer number, quote number, and consumer price index number of the text corpus.

The regular expression can be defined in a file which is loaded at service start. Figure 18 shows a sample for such a file.

```
{
  "rules": [
    {
      "taClassRef": "ols:cpi",
      "pattern": [
        "\\((?VPI\\)??\\s?[1|2]\\d\\d\\d\\d\\d\\)"
      ]
    },
    {
      "taClassRef": "ols:offer-number",
      "pattern": [
        "\\b20[1|0]\\dA\\d\\d\\d\\d\\d\\d\\d\\b",
        "\\bAN\\d{9}\\b"
      ]
    }
  ]
}
```

Figure 18 Sample definition for regex pattern

¹² https://en.wikipedia.org/wiki/Regular_expression

The OpenAPI spec for the service can be found at the Lynx Service Catalogue¹³. In addition, a small demo application is available online¹⁴. This demo application searches for 4-digit numbers in a pdf file and returns a pdf where the numbers are annotated as pdf annotations.

In the future this service will also be available through the Lynx Platform. To do so, additional functionalities to upload rules must be provided.

4.6 SCORING SERVICE

The Scoring Service is Java Spring Boot application which scores an individual document, based on a defined set of rules. It accepts a Lynx Document as input and returns a score for the provided document.

The Scoring Service uses individual annotations and combines them to a higher-level annotation, e.g. for index clauses.

The paragraph “*Index Assurance: Prices are subject to **annual value adjustment** in accordance with the **Consumer Price Index (CPI) 2010** or the index replacing it. The starting point for the calculation of value assurance is the **index number** published for the month in which the offer was placed.*” includes the following individual annotations:

- Index Assurance
- Annual value adjustment
- Consumer Price Index
- (CPI) 2010
- Index number

Only if a majority of these annotations are included in a paragraph, the clause will be identified as an index clause.

Currently we investigate how the main functionality of this service could also be made available through the Lynx Platform, e.g. to get higher-level annotations (e.g. a complete clause, here the index clause) based on low-level annotations (e.g. annotations from NER, TimeEx, ...). To do so different rule engines are evaluated to provide this functionality.

4.7 CPINDEX SERVICE

The CPIIndex Service is a Java Spring Boot application which provides a REST API to calculate the consumer price index increase within a given period. The service is loading the current CPI information from the Statistic Austria Portal¹⁵. This increase is used to calculate the value increase for the offered product.

4.8 LYNX SERVICES

The Lynx Services¹⁶ used in this pilot (as of the time of editing this document or in a near future) are the following:

- **TimEx** - Temporal Expression Extraction. Temporal expressions in documents are detected, e.g., the date when the offer has been made.
- **NER** - Named Entity Recognition – Is used to identify named entities such as persons and organisations (companies).

¹³ <http://regex-extractor-service-88-contracts.cloud.itandtel.at/api/swagger-ui.html>

¹⁴ <http://regex-extractor-service-88-contracts.cloud.itandtel.at/api/view/index.html>

¹⁵ https://data.statistik.gv.at/data/OGD_vpi10_VPI_2010_1.csv

¹⁶ <http://lynx-project.eu/doc/api/>

- **APIM** - Api Manager – It exposes RESTful API to first party clients, end-users and administrators. It will represent the main entry point to the Lynx Services in the future.
- **WM** - Workflow Manager. This service is used to implement the necessary annotation workflows, within Lynx. For the pilot a workflow is defined to perform all annotations, translations and summarizations with a single call to the Lynx Platform and return the information once the result is available.
- **ReIEx** - Relation Extraction between entities within a single document will be used to find e.g. cause-effect relation: Example: The agreement ends by 20.1.2020
- **EntEx+WSID** - Entity Extraction and Word Sens Disambiguation Service. The service produces the necessary annotations of the documents enriching the documents with entities from the defined vocabulary.
- **TRANS** - Machine Translation is used to translate non-English document to English which is then used for the summarization service.
- **Summarization** – Is used to create a short textual summary of the contract. This result is used to create for each contract a single page with all necessary information.
- **Geo** – to find geographical expressions in documents, mainly addresses.

5 OUTLOOK

To conclude this deliverable, the subsequent time-plan has been developed for the next steps:

November / December 2019:

Improvement of the back-end services

- Document Service:
 - Extracting Structure Information
- Annotation Service:
 - Further integration of additional Lynx Services / Workflows
- Scoring Service
 - Improve Rule set, Generalize Service for higher level annotations

February 2020:

- First working UI prototype
- First version of the enhanced controlled vocabulary with:
 - Links to the legislation
 - Explanation

March – October 2020:

- Evaluation and continuous improvement of the prototype with openlaws customers
- Integration with openlaws system, e.g. Actively inform companies, persons in charge (directors, managers, data protection officers, etc.) or the company's lawyer(s) about whenever there is a change in relevant legislation, case law, or contractual obligations.

6 REFERENCES

- Flitsch, Martina (2010) Verträge und Vertragsmanagement in Unternehmen.
- Hamming, Richard W. (1962). Numerical Methods for Scientists and Engineers. New York: McGraw-Hill.; second edition 1973