

Using a Legal Knowledge Graph for Multilingual Compliance Services in Labor Law, Contract Management, and Geothermal Energy



Martin Kaltenboeck, Pascual Boil, Pieter Verhoeven, Christian Sageder, Elena Montiel-Ponsoda, and Pablo Calleja-Ibáñez

Abstract This chapter provides insights about the work done and the results achieved by the Horizon 2020-funded Innovation Action “Lynx—Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe.” The main objective of Lynx is to create an ecosystem of multilingual, smart cloud services to manage compliance based on a Legal Knowledge Graph (LKG), which integrates and links heterogeneous compliance data sources including legislation, case law, regulations, standards, and private contracts. The chapter provides a short introduction in regards to the market needs, gives an overview of the Lynx services available on the Lynx Services Platform (LySP), and provides valuable insights into three real-world compliance solutions developed on top of LySP together with Lynx’s industry partners, namely, (1) Labor Law (Cuatrecasas, Spain), (2) Contract Management (Cybly, Austria), and (3) Geothermal Energy (DNV.GL, the Netherlands).

Keywords Compliance · Legal knowledge graph · Multilingualism · NLP · Labor law · Contract management · Geothermal energy

M. Kaltenboeck (✉)
Semantic Web Company, Vienna, Austria
e-mail: martin.kaltenboeck@semantic-web.com

P. Boil
Cuatrecasas, Barcelona, Spain

P. Verhoeven
DNV, Utrecht, The Netherlands

C. Sageder
Cybly, Salzburg, Austria

E. Montiel-Ponsoda · P. Calleja-Ibáñez
Universidad Politécnica de Madrid, Madrid, Spain

Insights into the industry solutions include the concrete business cases, the problem statements and requirements, the relevant data identified and used, and the LySP AI services combined to realize powerful multilingual compliance solutions in the respective fields. The chapter closes with findings and learnings from the implementation phase and a future outlook for further developments, specifically for the three vertical solutions and LySP.

The chapter relates to the technical priorities of Data Management and Data Analytics of the European Big Data Value Strategic Research and Innovation Agenda [1]. It addresses all challenges of the horizontal concern Data Management and some of the challenges of the horizontal concern Data Analytics of the BDV Technical Reference Model. It addresses the vertical concerns: (a) Big Data Types and Semantics (with a focus on Text data, including Natural Language Processing data and Graph data, Network/Web data and Metadata) as well as (b) Standards (standardization of Big Data technology areas to facilitate data integration, sharing, and interoperability). The chapter relates to the Reasoning and Decision Making cross-sectorial technology enablers of the AI, Data and Robotics Strategic Research, Innovation and Deployment Agenda [16].

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

Currently, European small and medium-sized enterprises (SMEs) and companies operating internationally or wanting to branch out to other countries and markets, *face multiple difficulties to engage in trade abroad and to localize their products and services to other countries*, owing to legal and also to language barriers in Europe. As reported by the European Commission, only 7% of European SMEs sell across borders. SMEs that sell their products and services internationally exhibit 7% job growth and 26% innovation in their offering, compared to 1% and 8% for SMEs that do not go outside their local markets [2]. A key challenge for businesses in Europe is, thus, how to engage with customers effectively across the legal and language barriers.

One of the main problems is the *management of compliance across different countries*. “Compliance is a term generally used to refer to the conformance to a set of laws, regulations, policies, or best practices” [3]. When companies want to sell a product or offer a service in a new market they must comply with the applicable legislation (European, regional, local), implement different standards (e.g., from ISO, AENOR, or DIN [4]) and possibly follow sector-specific best practices. Dealing with *legal and regulatory compliance data* is a cumbersome task usually delegated to legal and consultancy firms that obtain documents from several data sources, published by various institutions according to different criteria and formats.

While data analytics is trying to address the issue of data heterogeneity from a technical viewpoint, the more human side of the data still remains a greenfield, i.e., the inherent incompatibility of multiple natural languages, which not only involve different words but also different syntax and different semantics. Europe is determined to make the most of the linguistic wealth that characterizes the continent. An increasing number of voices are in favor of a stronger commitment towards a multilingual Digital Single Market (DSM) as the key for becoming the most competitive market in the world [5]. As per the former EC vice president Andrus Ansip: “Overcoming language barriers is vital for building the DSM, which is by definition multilingual. It is now time to reduce and remove the language barriers that are holding back its advance, and turn them into competitive advantages” [6].

With the aim of addressing the challenges posed by the European market, currently fragmented into legal silos and split into more than 20 linguistic islands, constituting a competitive disadvantage for SMEs and large companies in general, and in line with other initiatives in Europe that share the same spirit (e.g., Digital Single Market, ISA, CEF.AT [7]), Lynx has created an ecosystem of smart cloud services that exploit a multilingual Legal Knowledge Graph (LKG) of legislation, regulations, policies, and standards from multiple jurisdictions.

This cloud of services integrated in the Lynx Services Platform (LySP) provides mass-customized regulatory information to European businesses. Additionally, it supports the creation of a common legal ICT infrastructure that will contribute to unlocking the potential of a multilingual and truly single digital market.

In order to achieve these objectives, the Lynx platform managed to: (1) create a novel and unique knowledge base related to compliance, integrating information from heterogeneous data and content sources; (2) provide a set of multilingual and smart core services to extract value from the knowledge base, and (3) translate its value into the market in the form of three business-driven pilots, making use of LySP.

In the first step, the LySP acquired—and continuously maintains—data and documents related to compliance from multiple jurisdictions in different languages, as well as interlinked terminologies and language resources, open standards, and sectorial best practice guidelines. This collection of structured data and unstructured documents, obtained from open sources, was the base of the LKG [8].

In the second step, a set of core domain-agnostic services was put in place to analyze and process the documents and data in order to integrate them into the LKG. Existing multilingual terminologies, semantic tools, and machine learning mechanisms were adapted and customized to the legal domain and used to annotate, structure, and interlink the LKG contents. Iteratively and incrementally, the LKG has been developed and augmented by linking to external databases and corpora, by discovering topics and entities linked implicitly, as well as by using translation services to translate documents not previously available in certain languages.

In the third and final step, these services have been configured in three real-world pilots according to the industry needs represented by the Lynx business cases. These

vertical solutions exploit the knowledge available in the LKG and have been driven forward and evaluated by companies with an existing customer base.

The main objective of Lynx is to *facilitate compliance for companies in internationalization processes*, by leveraging the existing European *legal and regulatory open data* seamlessly *interlinked* and offered through a set of *cross-sectorial, cross-lingual smart services in the Lynx Services Platform: LySP*. SMEs and other organizations can benefit from LySP through: (1) companies directly making use of the LySP Services and (2) companies in the portfolio of law firms and consultancy companies making use of LySP, both through using LySP Services either standalone (as a self-service) or integrated into existing IT systems.

2 The Lynx Services Platform: LySP

LySP is a cloud of smart and multilingual services working on top of the LKG and acting as a basis for training and operation of end user services. As illustrated in Fig. 1, the LKG contains law and legal information, directives, regulations, and other

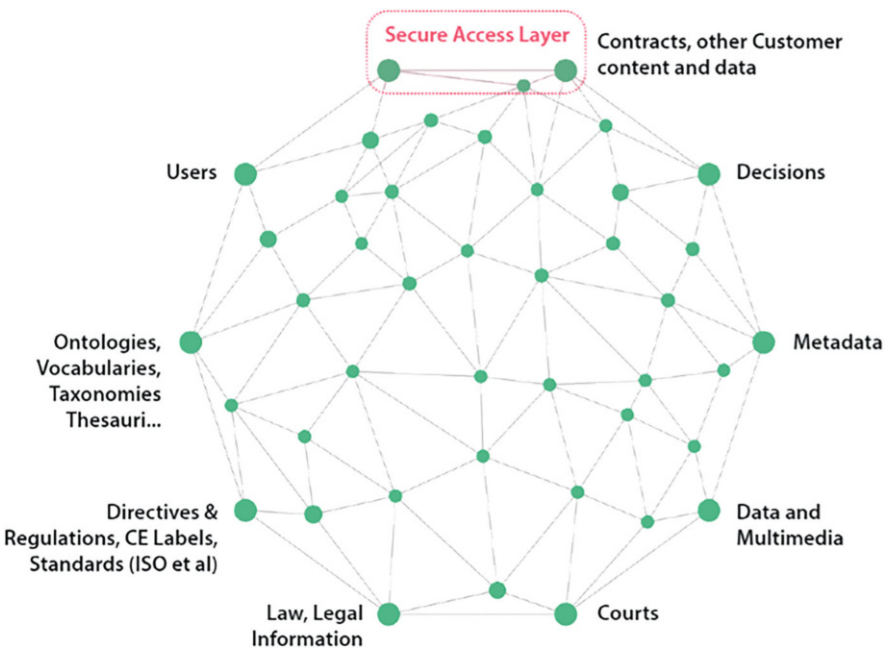


Fig. 1 Lynx legal knowledge graph

relevant data and information harvested from public sources, integrated and enriched by making use of the Lynx data model [15]. Additionally, the LKG has been expanded—in a secure layer—by data and information of the pilot applications.

LySP Services have been developed (1) from scratch by the Lynx partners: Universidad Politécnica de Madrid (Spain), Semantic Web Company (Austria), Cybly GmbH (Austria), Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany), Alpenite (Italy), or (2) through the adaptation of existing software components, namely: Tilde Translator (<https://tilde.com/products-and-services/machine-translation>) by Tilde, Lexicala (<https://www.lexicala.com/>) by KDictionaries, and PoolParty Semantic Suite (<https://www.poolparty.biz>) by Semantic Web Company, all of them docked onto and trained by the LKG to ensure the full value of LySP Services in the field of LegalTech and compliance. Through the orchestration of these services, a broad portfolio of real-world use cases can be created [9]. In the framework of the project, three pilots have been developed, as further explained in Sect. 3.

The key principles according to which LySP is being built are summarized in the following [10]:

- Token-based OAuth2 protocol for authorization together with the centralized access control and authorization rules management based on Keycloak.
- An established LynxDocument schema according to the LKG ontology.
- Containerized deployment in an orchestrated application platform making use of Red Hat OpenShift.
- Workflow Manager based on Camunda.
- LinkedDataPlatform-inspired Document Manager.
- Common rules for the development of web APIs: REST + API gateway patterns, including OpenAPI 3 description.

In its current status, LySP provides 16 services that accomplish different purposes, as listed below:

LySP Enrichment Services

LySP Annotation Services

1. Temporal Expression Recognition (TimEx): finds temporal expressions in documents.
2. Named Entity Recognition (NER): finds named entities using state-of-the-art methods.
3. Geographical NER (Geo): finds geographical entities in documents.
4. Relation Extraction (RelEx): extracts relations between entities.
5. Entity Linking (EL): identifies and links entities, provides annotations, including word sense disambiguation.

LySP Conversion Services

1. Machine Translation: translates documents.
2. Summarization: summarizes the content of a document.

LySP Search and Information Retrieval Services

1. Question Answering (QADoc): retrieves the most relevant answer for a given question.
2. Cross Lingual Search (Sear): searches a text string in documents across different languages.
3. Semantic Similarity (SeSim): calculates similarity between any two documents.
4. Terminology Query (TermQ): obtains information about a certain term with examples and notes of use.

LySP Vocabulary Services

1. Dictionary Services (DA): queries domain-independent dictionaries from SPARQL endpoint.
2. Terminology Extraction (TermEx): extracts terminology from document corpus.

LySP Platform Services

1. Workflow Manager (WM): manages workflows defined in BPMN.
2. Document Manager (DCM): manages documents and annotations in the LKG.
3. Authentication and Identity Management (APIM): provides Lynx identity, OAuth2 flows, and social login.

At the moment, most of the services are available for the languages English, German, Spanish, and Dutch [11]. The current version of the LySP Architecture can be seen in Fig. 2, where arrows and colors illustrate the principal workflows. Broadly speaking, a collection of documents (corpus) is ingested into the platform where TermEx performs a terminology extraction process (red arrow). Next, documents are annotated by means of LySP Enrichment Services. Some services depend on the annotations produced by specific services, whereas others can run in parallel. To efficiently orchestrate the different services, a dedicated Workflow Manager based on Camunda is used. The result of this process is what we call an Enriched Lynx Document. The service in charge of efficiently storing, updating, and retrieving documents is the Document Manager. Enriched documents are then stored for subsequent retrieval by LySP Search and Information Retrieval Services (Storage and Information Retrieval box in Fig. 2). For more details on how the services are orchestrated in LySP, we refer the interested reader to [9]. All three Lynx compliance solutions described in detail in the next section are built on top of LySP and the workflows explained above.

3 Lynx Compliance Solutions

The purpose of this section is to describe the three real-world compliance solutions that have been developed on top of LySP together with Lynx's industry partners, namely: (1) Labor Law (Cuatrecasas, Spain), (2) Contract Management (Cybly, Austria), and (3) Geothermal Energy (DNV.GL, the Netherlands).

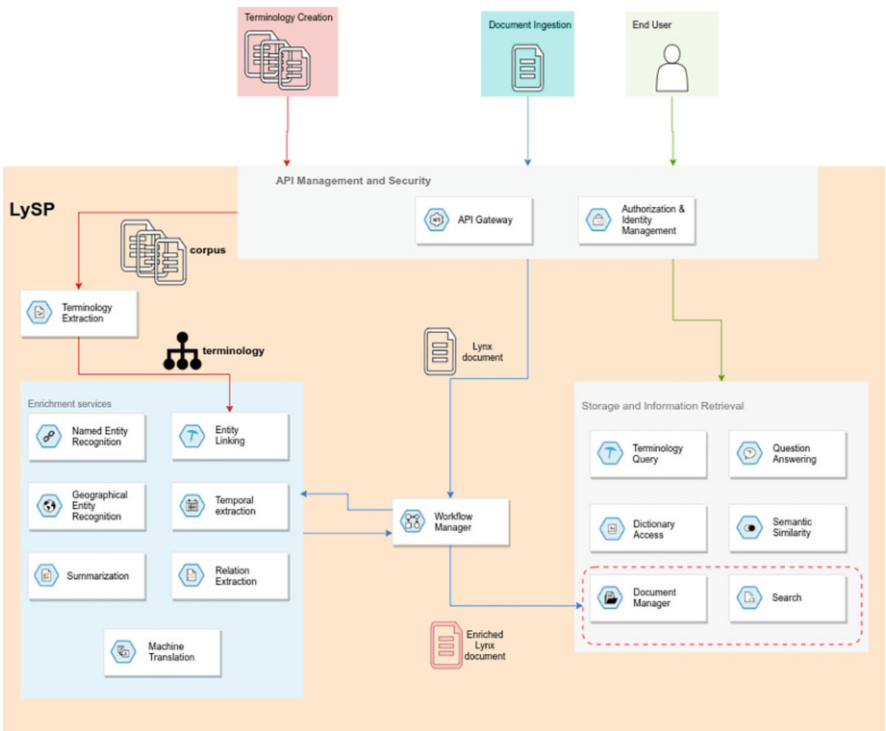


Fig. 2 Architecture of LySP

Each solution is described according to the same structure. First, the industry partner involved in the solution is introduced, to provide the context for the needs and requirements of each business case in what we have called “Problem Statement and Business Case.” Next, the solution is spelled out and, finally, details of the Lynx services involved are provided.

3.1 Compliance Services in Labor Law (Cuatrecasas, Spain)

About Cuatrecasas

Cuatrecasas (www.cuatrecasas.com) is an international law firm with headquarters in Barcelona, Madrid, and Lisbon. The firm is specialized in all areas of business law, applying a sectoral approach and covering all types of business. It represents several of the largest international companies, advising them on their investments in the major markets in which they operate. Cuatrecasas is present in the main financial centers of Europe, America, Africa, and Asia through international offices; European Network teams in Germany, France, and Italy; and international desks

covering over 20 regions. Thanks to its international presence, it has expert knowledge of various industries and regions, and is well aware of the challenges posed to companies in internationalization processes.

Problem Statement and Business Case

Labor law is generally the most common regulation companies have to deal with on a daily basis. In any company acquisition—Mergers and Acquisitions (M&A) operations or prior Due Diligence analysis—or when supporting international business expansion, the “local labor legislation” has crucial implications. Due to the relevance of labor law, Cuatrecasas has highly specialized lawyers in Spanish and Portuguese labor law and dedicates a constant effort to be updated on legislative changes and binding precedents. However, when crossing the Iberian borders, coverage decreases and the firm has to rely on associated firms (similar to International Legal Networks).

The main objective of this business case is to provide a reliable service that helps companies (starting with Cuatrecasas itself and ending with any company and, of course, the majority of Cuatrecasas’ clients) to solve typical issues related to labor law, which are commonly regulated by each country with significant discrepancies. More commonly than not, these differences are crucial to making strategic business decisions in an overseas expansion strategy. The typical Cuatrecasas client operates internationally. Around 30% of the current customer base has international problems of one kind or another, and at least half of them have implications for labor law.

In this context, Cuatrecasas has created two distinct but complementary business cases in the context of the Lynx project:

Internal Usage. A tool for Cuatrecasas’ lawyers

The first business case is based on an internal approach that would result in time and cost savings for Cuatrecasas’ legal assessment related to country-specific labor law. The firm, typically as part of an M&A (Mergers and Acquisitions) operation or Due Diligence, normally sends out questionnaires on labor law with frequently asked questions (one questionnaire has between 10 and 50 questions) to an average of six to ten partner firms from different countries (jurisdictions). Usually, the completion of these questionnaires is subcontracted to local partner law firms. The partial cost savings estimation gives a basis for the ROI justification. The pilot developed during the Lynx project lifetime covers the four countries/languages of the Lynx project (ES, IT, DE, EN) due to available resources. However, the numbers are also interesting when other countries (languages and jurisdictions) of interest for Cuatrecasas clients (e.g., Russia, China, Mexico, Brazil) are included. In the second scenario (EU and non-EU countries), the expected cumulative cost reduction benefits would be close to the 3 MM € in 5 years.

External Usage. A tool for Cuatrecasas customers

The second business case is based on a SaaS (Software as a Service) approach (new line of business income), putting the solution directly in the hands of Cuatrecasas’ big customers with a high level of internationalization. This scenario is not a very aggressive one regarding pricing, since it could be considered rather a loyalty system than a software product itself. The cumulative figures of this second

use case are quite similar to the first one. Estimating 3MM € in a 5-year plan is a conservative projection, given that a minimum of 25 existing Cuatrecasas' clients could use the Lynx platform.

The Solution

In Lynx we focused on the internal use case in the Lynx project. The idea was to test, improve, and evaluate accuracy, and show internal value before presenting the solution to customers. As a support tool for Cuatrecasas' lawyers this application should be executed internally (inside the corporate Cuatrecasas' network).

Cuatrecasas provides a range of services to clients, including (1) specific operations, which typically involve a project with a limited scope and period; and (2) general legal advice, which is usually categorized by practice area (e.g., labor, tax, and corporate) due to the different legal specializations required. Moreover, lawyers at Cuatrecasas may work on more than one matter and with multiple clients at a given time. For this reason, the system must provide lawyers with the tools they need to organize and optimize their tasks, enabling them to configure and save their favorite options (more common/default): personal or client/company.

Although Cuatrecasas has offices in multiple countries, the firm's official languages are Spanish, English, and Portuguese. Despite being specialized in Spanish and Portuguese jurisdictions, the firm offers global international coverage to its clients, with a focus on Latin America. Typical clients of Cuatrecasas include large (Spanish and Portuguese) companies with businesses around the world. The main problem that nonlocal lawyers usually face is accessing and understanding foreign local laws and regulations that are often unavailable in other languages.

For this internal use case, users are assumed to be legal experts. Often, they are junior lawyers who are tasked to investigate external regulations. Currently, these lawyers have to contact the internal Knowledge and Innovation Team to find out about (1) the legal particularities of a specific country/jurisdiction, (2) the legal sources available and (3) whether they can count on local lawyers from partnering institutions that can be contacted, if necessary. These lawyers are accustomed to use legal databases and other information resources (e.g., the ones provided by LexisNexis, Thomson Reuters or vLex). Moreover, they usually have a good command of the legal terminology in their own language and in English, but limited knowledge of the legal terminology in other languages.

To fulfill the requirements of this business case, an application has been developed in which the user formulates a complete query in natural language (Spanish, English, German, Dutch) about labor law and workers' regulations, specifying one or more jurisdictions (Spain, Austria, Netherlands). Then, the system returns the most relevant results based on the direct texts of the law, and translated to the language previously selected by the user, including the following:

- The most precise answer possible (when the question is specific, asking for a value, data and name).
- The paragraph(s) related with the topic/question, where the possible answer appears as part of the text (ideally highlighted).

- The context by showing the article (and section) from which the paragraph(s) is extracted, showing the number and title, and allowing the user to view the full text of the article and law, which the user should be able to access and download.

Complex legal questions are almost impossible to answer by only highlighting parts of the law. Context and additional information are often needed. This additional information is sometimes difficult to incorporate into a question and these context words are not always easy to find directly mentioned in laws. To palliate this issue, the system is designed to be used as an intelligent search tool, providing legal guidance to lawyers, to help substitute or minimize some of their less-valued work.

The Use of LySP Services

The Cuatrecasas Lynx Pilot is comprised of four main parts/components (see Fig. 3): a Front-end Application, with the presentation layer responsible for the user experience; a Back-end Application and business logic layer, to encapsulate the defined modules (login, configuration, and Q&A modules) and provide all the required application functionalities; an Application database, to ensure data persistency; and, finally, the Cuatrecasas-Lynx API, a middleware component to encapsulate and centralize interaction with Lynx Services.

The LySP Services used in the Cuatrecasas Lynx Pilot are described below:

- *WM, DCM, and LKG*

The Cuatrecasas Lynx Pilot makes use of the LySP Services, as defined in Sect. 2, namely, the Workflow Manager (WM), responsible for the effective orchestration of the LySP Services, and the Document Manager (DCM) service, where documents are stored and maintained. As already mentioned, the basic

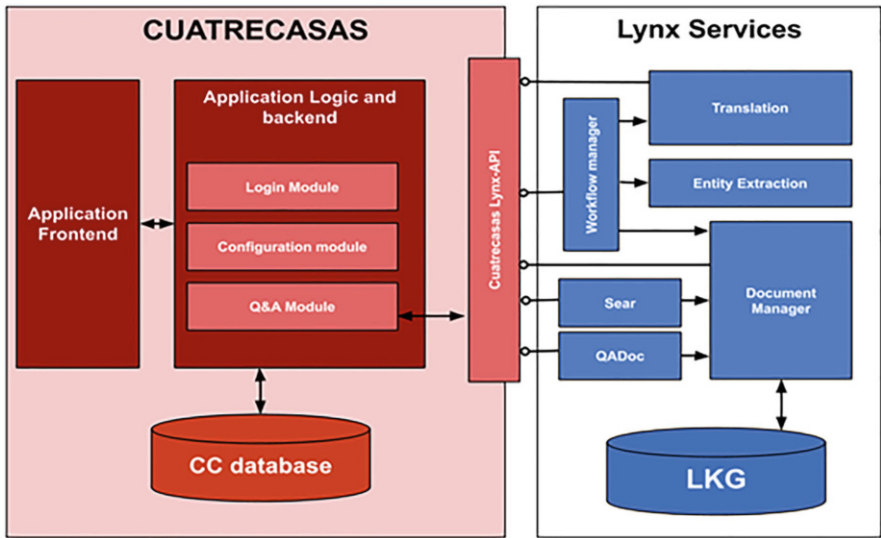


Fig. 3 Pilot modules and components schema of the Cuatrecasas Lynx Pilot

functionality of the DCM includes storing documents and their annotations, particularly in regards to the support of their synchronization, providing read and write access, as well as updates of documents and annotations. The DCM can be queried in terms of annotations (e.g., “which documents mention this entity?”), as well as in terms of documents (e.g., “what are the contents/annotations of document X?”). The interface includes a set of APIs to manage the following resources within LySP: collections, documents, and annotations. DCM is responsible for storing the LKG (Legal Knowledge Graph) and the documents once they have been processed through the different workflows.

Additionally to the LySP Platform services, this application relies on some LySP Annotation and LySP Search and Information Retrieval Services, as specified below:

- *SEAR Service*

The Cross-lingual search service is used to retrieve documents from collections previously defined by end users. Documents are retrieved from the Document Manager based on metadata and content filters. The service generates a first list of ranked candidate answers (previously broken down into paragraphs), and highlights the text segment that is responsible for the selection. The SEAR service relies on the document enrichment processes performed by the LySP Enrichment Services to allow filtering out searches and to score the results based on the query. Additionally, this service uses Query Expansion (QE) mechanisms to improve search precision and cover the main use case requirements.

- *QADoc Service*

The QADoc service receives a query posed in natural language and a source text to find a precise answer within it. Only when the service returns a result with a high level of confidence, the application will show this result to the user.

- *TimEx Service*

Temporal expressions are very relevant in any legal document. For example, expressions for deadlines or regulated procedures are common in the labor context, such as “something has to be done 10 days after the contract is signed,” “the probationary period does not exceed six months,” or “the cost of dismissing an employee is 20 days per worked year.” The pilot makes use of this service to identify time expressions that may contain the answer to a question.

- *Machine Translation Service*

The translation service provides automated machine translation by using the Tilde MT cloud platform. Currently, the translation service provides support for a runtime scenario and an endpoint for the Lynx platform’s asynchronous process in the background. Neural Machine Translation (NMT) systems were trained for the project languages. In the domain of labor law, specific legal and business data was gathered and processed before training the NMT systems on a mix of broad-domain and in-domain data that is able to translate both in-domain and out-of-domain texts.

3.2 *Smart Contract Management (Cybly, Austria)*

About Cybly

Cybly (www.cybly.tech) is a legal tech company based in Salzburg and Vienna, Austria. It combines two brands or product lines under one roof—"LawThek" and "Legalnetics." "LawThek" is a legal database with content from EUR-Lex (directive, regulation, and decisions), Austria (federal and state laws, decisions), and Germany (federal laws), offering cross-platform access to standardized and interlinked sources of law. In addition to the desktop version, the RIS:App (Right Information System) is distributed. This enables mobile access and is available for free download in the Apple and Google app stores. "LawThek" is complemented by the high-end products and services of "Legalnetics." The range of services offered by "Legalnetics" includes process-oriented, integrated IT solutions in the areas of law, finance, and compliance, as well as all other areas with a legal or legal information background.

Problem Statement and Business Case

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. Many SMEs (small and medium enterprises) do not have a database with all the information of their contracts, which prevents them from easily finding information or applying changes.

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

1. There is a change in law, and you need to know which contracts are affected.
2. An overview on all obligations with a certain company is needed.
3. 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 or a subsidiary agreement you've never seen before.

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: Contracts and contract-relevant documents are physically and electronically distributed across the entire organization and tools, e.g., file server, emails, physical documents. As a result, there is often 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 [14] defines contract management as the creation of ideal structures for: contract planning, contract design, contract negotiations, implementation of contracts, contract administration, and contract archiving. In many cases, organizations are lacking these structures.

When it comes to contracts, there are very few tools being used: a word processor to create the contracts, email to communicate with the client or the other party,

a file storage in a defined directory structure, and/or a legal software to store the documents. To keep history recorded, they often add the different versions of the document and individual mail communication to this software or put them on the file system. With this in mind, we are focusing on automated contract administration and archiving.

The Solution

The aim of our solution is not to change the existing workflow, which is well-established in most companies and law firms, but to provide an integrated solution to the existing toolset and workflows.

The starting point and, at the same time, the simplest use case is the analysis of a single contract/document. However, the reality is much more complex: Regularly, a large number of contracts of diverse nature and purposes need to be analyzed and kept track of, taking into account various regulatory frameworks. In order to achieve this, we have two approaches. On the one hand, we have a pure back-end solution, and, on the other hand, we provide a visualization of the created data space for end users.

In the course of the Contracts Management Lynx Pilot, our goal is to develop reasonable strategies for automated contract analysis and contract archiving. Contract administration and management are crucial when it comes to defining the application:

To harvest documents, a command line tool has been implemented to provide the following two main functionalities: Recursively send all documents of a directory and its subdirectory to the Document Service for processing them. Monitoring a given directory and its subdirectory and send notifications when the contents of the specified files or directories are modified. With this tool it is possible to ingest a large set of documents to the system and also to monitor this set for any subsequent changes. Other external systems can use the REST interface of the Document Service to ingest contract-related documents too.

To make use of the harvested document, the Conversion Service converts documents in different data formats into a Lynx-Document that includes metadata and document structure where possible. The following main document formats are currently supported: Microsoft Office document formats, Open Document Format, iWorks, HTML, PDF, Images, Outlook Messages (*.msg), MIME Messages. The newly created Lynx Document is then annotated by the Annotation Service which orchestrates the calls to the different LySP Annotations Services and also to others. The document and its extracted information are stored within the LawThek Document Store. To do so, LawThek has been extended with the possibility to store Lynx Documents with the annotations beside the original document.

Through the front-end solution a single user has the possibility of managing (add, delete, update, group, search, etc.) contracts/documents. 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.

The search builds on top of the Lynx SEAR service. It is possible to search for documents by full text, document type, annotations, metadata, e.g., document

date (facets) and any combination. It is therefore possible to search for newly added/modified documents or for certain document names, etc.

The Use of LySP Services

To ingest documents into the system, the following tasks are performed: (1) convert files into Lynx Documents with the Converter Service; (2) store, update, delete the file in Customers' local LawThek Document Store, which is a neo4j database for metadata and relationships in combination with file storage to persist the original file; and (3) store, update, delete the document in the search index.

The Lynx Services used in this pilot (at the time of writing or in a near future) are specified in the following:

- *TimEx*—Temporal Expression Recognition—used to detect temporal expressions in documents, e.g., the date when the offer was made.
- *NER*—Named Entity Recognition—used to identify named entities such as persons and organizations (companies).
- *RelEx*—Relation Extraction between entities within a single document—used to find, e.g., cause-effect relationships, such as “The agreement ends by 20.1.2020.”
- *EntEx + WSID*—Entity Extraction and Word Sense Disambiguation Service—used to enrich documents with entities from a previously defined vocabulary.
- *Geo*—Geographical NER—used to find geographical expressions in documents, mainly addresses.
- *SEAR*—Cross Lingual Search—provides the ability to search for documents by full text, document type, annotations, metadata, e.g., document date (facets) and any combinations. It is therefore possible to search for newly added/modified documents, or for certain document names, for example.
- *APIM*—Authentication and Identity Management—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.

3.3 Compliance Solution for Geothermal Energy (DNV GL, the Netherlands)

About DNV GL

DNV GL (www.dnvgl.com) is the independent expert in risk management and assurance, operating in more than 100 countries. Through its broad experience and deep expertise, DNV GL advances safety and sustainable performance, sets industry benchmarks, and inspires and invents solutions.

Whether assessing a new ship design, optimizing the performance of a wind farm, analyzing sensor data from a gas pipeline, or certifying a food company's supply chain, DNV GL enables its customers and their stakeholders to make critical decisions with confidence.

Driven by its purpose, to safeguard life, property, and the environment, DNV GL helps tackle the challenges and global transformations facing its customers and the world today and is a trusted voice for many of the world's most successful and forward-thinking companies.

Problem Statement and Business Case

Geothermal Energy is an emerging source of sustainable energy. Its application is expected to show accelerated growth driven by the need for the global energy transition. To achieve sustainable and controlled growth, modernization of legislation and regulations, as well as industry standards and best practices, is required. Stakeholders—such as project developers, regulators, and engineers—typically struggle to find this information, resulting in delayed application and imposing additional risks. The pilot developed during the Lynx project aims to demonstrate how the structuring of documents in the Legal Knowledge Graphs can help users to find and select relevant regulatory documents (e.g., permits) and recommended reading on safety and environmental risks. For this pilot, it was essential to find and correctly link entities from the regulations to taxonomies (in the enrichment phase) and to quickly and reliably estimate the semantic similarity between the user's document and the previously collected documents. Moreover, to improve the accessibility and discoverability of data, it was essential to translate the documents automatically.

Governments play a crucial role in legislating and assuring compliance to mitigate safety and environmental risks, in all sectors, and in the energy industry in particular, due to the transition which is currently undergoing. With the expected growth in sustainable energy alternatives, continuous standardization of technology to bring down costs and risks can be expected. Most countries will develop policies and laws individually or together with other countries. Governments will seek balance in the use of subsidy schemes to accelerate growth and develop regulation or legislation to mitigate safety and environmental risks to guide the sustainable growth of technologies and markets. Companies active in these supply chains are likely to seek cross-border growth in order to develop economies of scale and bring costs down. If cross-border growth is envisioned, keeping up with the latest legal and regulatory rules is likely to become a challenge as country-specific clauses and local languages complicate when trying to gain an overview.

In the DNV GL Lynx Pilot, this specific context and challenge is explored for the geothermal energy domain as a proxy of the wider renewable energy domain. Geothermal energy is heat generated in the sub-surface of the earth. A geothermal fluid or steam carries the geothermal energy to the earth's surface. Geothermal energy operators drill a production and an injection well (also known as a doublet) to a certain depth (between 100 m and 4000 m) to circulate fluid to produce "heat." Depending on the temperatures, this fluid can be used to produce clean electricity, or as a baseload for municipal district or industry heating or cooling. Geothermal energy is seen as a promising sustainable energy alternative, and the industry (supply) and its users (demand) are at the dawn of accelerated growth.

To prove the value of LySP, two business cases were designed to explore the typical problems and challenges in this domain:

1. National actors in the geothermal energy supply chain facing regulatory risks, missing potential opportunities, are taking poor decisions due to compliance information being fragmented over multiple information sources. The first geothermal energy challenge is better expressed by the following question: “Can value be generated by connecting machine-readable regulatory information resources for geothermal energy?”
2. International actors in the geothermal energy supply chain struggle with a lack of understanding of country-specific regulatory frameworks (which is a competitive disadvantage), thus limiting international competition and the potential benefits of economies of scale as well as standardization. The second geothermal energy challenge is: “Can internationalization be stimulated by providing the same level of access to relevant compliance information for, and from, different EU countries?”

The Solution: The Use of LySP Services

To address these two challenges, a web application “Recommender” (see Fig. 4) was developed on top of LySP. It facilitates searching for relevant documents in multilingual corpora. The Recommender accepts plain text and PDF documents. The documents are preprocessed and plain text is extracted. The plain text is then annotated by the *Entity Linking* (EL) service. The annotated documents are processed by the *Semantic Similarity* (SeSim) service (see Fig. 4). On the left the original document title and content are displayed (A) with highlighted entities from the LKG, identified through the EL service. The SeSim service not only creates similarity scores, but it is also the reasoning behind these scores visualized as a

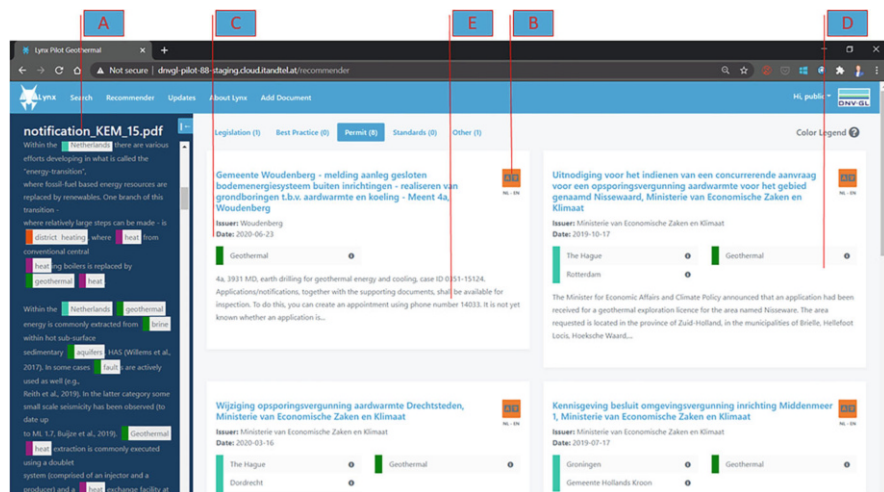


Fig. 4 Geothermal use case: recommender results page

table (D) and relevant metadata (C). The documents are translated (B) using the *MachineTranslation* (MT) service and presented to the user in the user's language (E).

4 Key Findings, Challenges, and Outlook: LySP—The Lynx Services Platform

To summarize, the Lynx Services Platform (LySP) provides a total of 16 smart services that can be used either standalone or orchestrated in specific combinations to provide powerful solutions for multilingual compliance-related applications. The services have been developed and trained on top of the Legal Knowledge Graph (LKG) to ensure domain-specific solutions with high precision, whereby the LKG consists of both: open data from public sources as well as solution-specific data and information inside a secure layer that can only be used by the respective vertical solution.

LySP Services include: 7 Enrichment Services (5 Annotation Services, 2 Conversion Services), 4 Search and Information Retrieval Services, 2 Vocabulary Services, 3 Platform Services, and are available in 4 languages—English, German, Spanish, and Dutch at the moment—and trained for legal, regulatory, and compliance use cases. However, it is worth noting that LySP Services are developed for a generic use, meaning that LySP Services can be trained for other domains (e.g., health information of financial industry) and for other languages (e.g., French and Portuguese) to allow for future scalability and the exploitation of LySP.

As a result, LySP Services can become an integral part of the European Digital Single Market [12] to be used and provided via the continuously growing number of European Data Markets and Data Spaces. Services can be trained on specific domains and languages and can thereby be used either as a core service of a Data Market and/or offered as a service for customers of Data Markets and Data Spaces. A first exercise into this direction is in progress with the European Language Grid, ELG [13] and ongoing discussions in regards to industrial Data Spaces and Markets are currently taking place.

The biggest challenges in the realization of LySP and the three industry business cases can be summarized as below:

1. The specification and the development of the LKG in regards to the harvesting of available law and regulations and other relevant information as well as the regular update of these, as such legal information is only partly available as open data, and is available in different formats and through different access paths.
2. The training of LySP Services for the three different industry use cases, as of the requirement to make use of additional data and information and as of the training effort required for specific areas of law. Again in regards to identification, specification and harvesting of data, but also the training of some services by domain experts.

3. The development of a LySP pricing model that takes into account dynamic infrastructure costs, as well as continuous maintenance costs of services and the Legal Knowledge Graph, to establish a stable and sustainable pricing model in a complex market of services available on the internet.

At the time of writing, the Lynx consortium is working on the exploitation strategy of LySP to bring LySP to the market in 2021. Besides the commercial offering of LySP, the pilot partners are going to use LySP Services internally and for their businesses, and the Lynx technology partners are integrating LySP Services into their own product and professional service offerings.

References

1. Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European big data value strategic research & innovation agenda*. Big Data Value Association.
2. Press Release, Brussels 06 July 2010: Internationally active SMEs yield better results. https://ec.europa.eu/commission/presscorner/detail/en/IP_10_895, link opened 01/2021.
3. Silveira, P., Rodríguez, C., Casati, F., Daniel, F., D'Andrea, V., Worledge, C., & Taheri, Z.. (2010). On the design of compliance governance dashboards for effective compliance and audit management. In *Service-oriented computing*. ICSOC/ServiceWave 2009 (pp. 208–217). Berlin: Springer.
4. ISO is the International Standards Office, AENOR is Asociación Española de Normalización y Certificación, DIN is Deutsches Institut für Normung.
5. See the “META-NET Strategic Research Agenda for Multilingual Europe 2020”, and the more recent “Strategic Research and Innovation Agenda for the Multilingual Digital Single Market” reports.
6. <https://medium.com/@JochenHummel/how-multilingual-is-europes-digital-single-market-9b8d908fce6c>, link opened 12/2020.
7. Connecting Europe Facility, Automated Translation: machine translation for selected Europe’s public digital services.
8. Even if the name LKG refers to ‘legal’, it must be understood in a broad sense, comprising also regulatory data and the terminologies to facilitate cross-lingual analysis.
9. Orchestrating legal NLP services for a portfolio of use cases, medium, 19/10/2020, Artem Revenko. <https://medium.com/semantic-tech-hotspot/lynx-service-platform-architecture-ac8d88c754f6>
10. Details of the used and listed technologies: Keycloak: <https://www.keycloak.org/>, LKG ontology: <https://www.lynx-project.eu/doc/lkg/>, Red Hat OpenShift: <https://www.openshift.com/>, Camunda: <https://camunda.com/>, and OpenAPI3: <https://swagger.io/specification/>
11. Overview of LySP Services plus links to documentation, videos, services websites and the respective service API: <https://lynx-project.eu/doc/api/>
12. <https://ec.europa.eu/digital-single-market/en/shaping-digital-single-market>, link opened 12/2020.
13. The ELG, the European Language Grid: <https://www.european-language-grid.eu/>, link opened 12/2020.
14. Flitsch, M. (2010). *Verträge und Vertragsmanagement in Unternehmen*. Linde Verlag.
15. LKG Ontology. Accessed 01/2021, from <https://www.lynx-project.eu/doc/lkg/>
16. Zillner, S., Bisset, D., Milano, M., Curry, E., García Robles, A., Hahn, T., Irgens, M., Lafrenz, R., Liepert, B., O’Sullivan, B., & Smeulders, A. (Eds.). (2020). *Strategic research, innovation and deployment agenda – AI, data and robotics partnership*. Third Release. September 2020, Brussels. BDVA, euRobotics, ELLIS, EurAI and CLAIRE.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

