D4.1

Users’ Needs and Usage Scenarios for Future Legal Applications

Grant Agreement nº: 690974
Project Acronym: MIREL
Project Title: MIning and REasoning with Legal texts
Website: http://www.mirelproject.eu/
Contractual delivery date: 31/12/2016
Actual delivery date: 16/01/2016
Contributing WP: WP4
Dissemination level: Confidential
Deliverable leader: INRIA
Contributors: UNITO, UNILU, UNIBO, NICTA
Document History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Author</th>
<th>Partner</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>30/12/2016</td>
<td>Serena Villata</td>
<td>INRIA</td>
<td>Initial draft</td>
</tr>
<tr>
<td>0.2</td>
<td>11/01/2017</td>
<td>Serena Villata</td>
<td>INRIA</td>
<td>Second draft</td>
</tr>
<tr>
<td>1.0</td>
<td>16/01/2017</td>
<td>Serena Villata</td>
<td>INRIA</td>
<td>Final Version</td>
</tr>
</tbody>
</table>

Contributors

<table>
<thead>
<tr>
<th>Partner</th>
<th>Name</th>
<th>Role</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>INRIA</td>
<td>Serena Villata</td>
<td>Editor and WP leader</td>
<td>Definition of the questionnaire, study of the results, description of the three usage scenarios</td>
</tr>
<tr>
<td>UNITO</td>
<td>Sabrina Praduroux</td>
<td>Reviewer</td>
<td>Questionnaire: legal point of view</td>
</tr>
<tr>
<td>UNILU</td>
<td>Livio Robaldo</td>
<td>Reviewer</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>UNIBO</td>
<td>Monica Palmirani</td>
<td>Reviewer</td>
<td>Legal text examples</td>
</tr>
</tbody>
</table>
Disclaimer: The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. MIREL consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Table of Contents

List of Figures .................................................................................................................................................. 5
Executive Summary.............................................................................................................................................. 6
1 Introduction.................................................................................................................................................... 6
2 Questionnaire: design and results ................................................................................................................... 8
   2.1.1 Design choices and final questions ..................................................................................................... 8
   2.1.2 Results .................................................................................................................................................. 15
3 Usage scenarios: description and examples.................................................................................................. 25
   3.1.1 Licenses and Contracts ...................................................................................................................... 26
   3.1.2 Technical Documents ......................................................................................................................... 30
   3.1.3 Multilingual corpora of norms ........................................................................................................... 32
4 Conclusions..................................................................................................................................................... 41
References ......................................................................................................................................................... 42
Appendix .......................................................................................................................................................... 43
List of Figures

Figure 1 - Demographics: male vs female ................................................................. 16
Figure 2 - Demographics: age................................................................................. 16
Figure 3 - Demographics: school grade................................................................. 17
Figure 4 - Demographics: country....................................................................... 17
Figure 5 - Demographics: interaction with legal data............................................. 18
Figure 6 - Demographics: issues with legal data................................................... 18
Figure 7 - Categories of users............................................................................... 19
Figure 8 - Number of employees ......................................................................... 19
Figure 9 - Spread of legal technology.................................................................... 20
Figure 10 - Investment in legal technology............................................................ 20
Figure 11 - Legal issues for developers................................................................. 21
Figure 12 - Difficulty in understanding legal data............................................... 22
Figure 13 - Investment in legal technology for developers...................................... 22
Figure 14 - Spread of legal technology.................................................................. 23
Figure 15 - Investment in information technology for legal experts ...................... 24
Figure 16 - Importance of information technology to manage legal data ............... 24
Figure 17 - An example of technical document from the TCP Code. ..................... 31
Executive Summary

The goal of this deliverable is twofold: first, we report about the users’ needs for applications to be developed in legal informatics, and second, we describe the three usage scenarios envisioned in the project, highlighting how these scenarios have been tackled in the project after the first year.

Concerning the first point, we have designed a questionnaire for users (being them legal experts of not) to investigate about their needs for the legal informatics applications of the future. In this deliverable, we report about the questionnaire and the results we have collected from a balanced set of users.

Concerning the second point, we describe the usage scenarios envisioned in the project by detailing the concrete instances, i.e., the concrete legal documents, considered in each scenario for the design of the very preliminary prototypes worked on during the first year of the project.

1 Introduction

The overall goal of the MIREL project is to mine and reason with legal text. In order to address this issue, which involves a range of research fields from natural language processing (NLP) to defeasible reasoning to ontologies, a vital question to answer concerns the users. More precisely, given the aim to develop methods and techniques to mine and reason with legal texts, what are the requirements the potential users of such applications may identify?

Task 4.1 in the project addresses this issue with the aim to listen, understand and collect users’ needs for the applications to be developed. What they would like to do with such applications, and what sort of services would they like to see. This task looks, in particular, for interaction and feedback with both common users (i.e., not legal experts) and users working in the legal domain: What do they need? How will they build applications? How would they visualize the legal data?

To answer these questions, we have designed, with the support of legal and marketing experts, a questionnaire to be submitted to the potential users of the future applications developed in the framework of the MIREL project. In this deliverable, we first report about our design choices and the resulting questionnaire, and then we will discuss the results we have collected from the users’ answers (Section 2).

Another key issue to support the definition of innovative methods for mining and reasoning with legal text consists in the identification of a number of case studies or usage scenarios to concretely
develop, test and evaluate such methods. At the beginning of the project, the following case studies have been identified:

1) **Licenses and contracts**: licenses and contracts are two particular instances of legal documents we aim to take into account. More precisely, licenses are legal texts stating what are the set of actions you are obliged/prohibited/permitted to perform in case of use and reuse of some material or data. Contracts are legal texts establishing an agreement among two or more parties, with the aim to define a set of legal obligations between them. In both cases, a (semi-)automated translation from such natural language texts into their machine-readable formulation would provide a valuable support to humans dealing with such texts. A specific usage of such a kind of texts is that of manually extracting the set of obliged/prohibited/permitted actions in order to check whether certain business processes are compliant with the legal text they should refer to.

2) **Technical documents**: technical documents describe how to handle and what are the functionalities of a certain product. They are intended to provide information about what can/cannot be done with the product. Given the huge dimension of this particular kind of legal texts and their diffusion in the companies, these documents have to be taken into account in our project. The advantages of extracting a machine-readable representation of such texts would allow to have in this representation a kind of summary of the main constraints expressed in the documents, with invaluable time saving for every person that is expected to read the whole document before obtaining this information.

3) **Multilingual corpora of norms**: the translation of legal texts into a shared uniform machine-readable representation is of fundamental importance in the legal multilingual scenario. In this case, like in the European Union, slightly different norms are proposed by the different countries to regulate a certain aspect of interest for the EU. Having one shared language-agnostic representation for such a kind of norms would help a lot in avoiding costly and time consuming translations, to avoid misunderstandings, and to focus directly on the content of the norms.

During the first year of the project, concrete instantiations of these usage scenarios have been considered, in order to investigate the definition of methods to mine and reason with such texts. In this deliverable, we report about these concrete instantiations of the usage scenarios providing examples extracted from these legal texts, and highlighting what are the issues the very preliminary prototypes we developed during the first year have tackled (Section 3).

The remainder of this document is as follows: we start with the description of the users’ questionnaire (Section 2.1) and its results (Section 2.2), and then we describe the concrete instances of the licenses and contracts (Section 3.1), technical documents (Section 3.2), and multilingual corpora of norms (Section 3.3) we have considered till now in the project. Concluding remarks end the document.
2 Questionnaire: design and results

This Section reports about the users’ questionnaire and the results we collected.

2.1.1 Design choices and final questions

The goal of this questionnaire was to listen, understand and collect users’ needs for the legal informatics applications of the future. What users would like to do with these applications, and what sort of services would they would like to see. This survey asks for feedback both standard users (i.e., not legal experts) dealing with legal-related issues, and users working in the legal domain: what do they need? how will they build applications? how would they visualize the legal data?

Behind the definition of the questions to be posed to the users, the first issue was to identify a reliable and balanced set of users to submit the questionnaire. Three high-level categories of users have been identified:

- Professional service firms (e.g., lawyers, business consultants)
- Data publishers and system developers
- Legal scholars and legal experts

These three categories cover the range of potential users of future legal informatics applications exploiting the results of the MIREL project. It is worth noticing that these three classes of users have (as confirmed by the results of the questionnaire) express different requirements, and allow us to cover multiple scenarios of usage. This is reflected in the set of questions asked to each class.

The second issue to be addressed in the identification of the set of users is to ensure the balance among the categories of users in order to provide reliable results. For this reason, the questionnaire has been proposed to a set of 20 users covering in equal portions the three classes mentioned above. The questionnaire will be however open for some more months in order to collect further answers, and an update of this deliverable (v 2.0) is planned for the end of the second year of the MIREL project.

The third problem we faced was the definition of the questions themselves. We designed the questionnaire following the standard guidelines, i.e., by starting with a section concerning demographic information, and then proposing a set of specific questions to the three kinds of users. The questionnaire ends with a question with a free text answer to collect additional information the user may want to provide.
The questionnaire is as follows:

[PART A] - Demographics

1. You are a
   - male
   - female

2. What is your approximate age?
   - 18 - 24
   - 25 - 34
   - 35 - 44
   - 45 - 54
   - 55 - 64
   - 65 and over

3. What was the last grade you completed in school?
   - Some high school
   - High school graduate
   - Some college
   - College graduate
   - Postgraduate/professional
   - Doctorate
   - Other (Please Specify)

4. In which country do you operate?
   - Europe [if so, specify the country from the list]
   - US
   - Asia
• Other (Please Specify)

5. You deal with legal data
   • Daily
   • Weekly
   • Monthly
   • Rarely
   • Almost never

6. What is the issue you most often realize when dealing with legal data?
   • Visualization issues
   • Editing issues
   • Update issues
   • Legal language issues
   • Other (Please Specify)

7. Please give us more details about the issues you mentioned above.
   -- free text answer --

*****

[PART B]

0. To which of the following categories do you belong:
   • Professional service firms (e.g., lawyers, business consultants)
   • Data publishers and system developers
   • Legal scholars and legal experts

N.B.: from this point on, questions are related to the above categories.
[PART B - Professional service firms]

1. Including yourself, how many employees work at this professional service firm?
   - 1
   - 2 to 7
   - 8 to 19
   - 20 to 49
   - 50 to 99
   - 100 or more
   - Other (Please Specify)

2. What is the degree of spread of legal technology in the context of your professional service firm?
   - Extremely popular
   - Very popular
   - Somewhat popular
   - Not so popular
   - Not at all popular

3. What is the interest to invest in legal technology of your professional study in the next two years?
   - Extremely interesting
   - Very interesting
   - Somewhat interesting
   - Not so interesting
   - Not at all interesting

4. Do professionals know their professional service firms?
• Strongly agree
• Agree
• Neutral
• Disagree
• Strongly disagree
• Not sure

5. What are the organizational models and current business and which, instead, are the emerging ones?
-- free text answer --

6. What are the causes that hinder, according to professionals, the spread of legal technology in the professional service firms?
-- free text answer --

7. In what type/s of legal technology would you invest (e.g., office management, cloud computing, legal billing, contract software)?
-- free text answer --

8. Do you think technology could help you in performing your tasks?
-- free text answer --

9. What type/s of legal tech could/should be developed?
-- free text answer --

[PART B - Data publishers and system developers]

1. What is the kind of legal information you need to manage with respect to your data/system?
• Licenses and rights
• Privacy statements
• Other (Please Specify)

2. Do you find it easy to manage legal information up to your needs?
• Strongly agree
• Agree
• Neutral
• Disagree
• Strongly disagree
• Not sure

3. What is the interest to invest in legal technology in your company?
• Extremely interesting
• Very interesting
• Somewhat interesting
• Not so interesting
• Not at all interesting

4. What are the causes that hinder, according to you, the spread of legal technology in your company?
-- free text answer --

5. In what type/s of legal technology would you invest (e.g., office management, cloud computing, legal billing, contract software)?
-- free text answer --

6. Do you think technology could help you in performing your tasks?
-- free text answer --
7. What type/s of legal tech could/should be developed?

-- free text answer --

[PART B - Legal scholars and legal experts]

1. What is the degree of spread of information technology in the context of your working environment?
   - Extremely popular
   - Very popular
   - Somewhat popular
   - Not so popular
   - Not at all popular

2. What is the interest to invest in information technology in your working environment?
   - Extremely interesting
   - Very interesting
   - Somewhat interesting
   - Not so interesting
   - Not at all interesting

3. Do you think that information technology would improve the management of legal documents and data?
   - Strongly agree
   - Agree
   - Neutral
   - Disagree
   - Strongly disagree
   - Not sure
4. If so, could you please specify how?
   -- free text answer –

5. In what type/s of legal technology would you invest (e.g., office management, cloud computing, legal billing, contract software)?
   -- free text answer --

6. Do you think technology could help you in performing your tasks?
   -- free text answer –

7. What type/s of legal tech could/should be developed?
   -- free text answer --

*****

[PART C]

Do you have any thoughts you want to share with us about legal technology?
   -- free text answer –

2.1.2 Results

In this section, we report about the answers we collected after the publication of the questionnaire. In total, up to now, we have invited to contribute and collected the answers of 24 persons belonging to one of the three categories described above. We report now about the collected answers.
The first part of the questionnaire was about demographic questions. It may be noted that our participants are mostly males, but the participation of females is however relevant.

**You are a** (24 responses)

![Pie chart showing gender distribution with 58.3% males and 41.7% females.]

*Figure 1 - Demographics: male vs female*

The age of the participants ranges from 25 to 64, with the majority of the participants being around 35-44 years old.

**What is your approximate age?** (24 responses)

![Pie chart showing age distribution with 29.2% for 18-24 years, 37.5% for 25-34 years, 16.7% for 35-44 years, and 16.7% for 45-54 years.]

*Figure 2 - Demographics: age*
The school grade of participants is postgraduate or doctorate.

Despite some participation from the US and Asia, the vast majority of our participants operate in Europe.

Figure 3 - Demographics: school grade

Figure 4 - Demographics: country
Due to our selection of participants, the majority of participants deals with legal data with a daily frequency. No participant selected the “Almost never” answer. This information shows that all our participants show some frequent interaction with legal data.

**Figure 5 - Demographics: interaction with legal data**

![Pie chart showing interaction frequency]

The most important issue they realize in dealing with legal data are visualization and legal language complexity, followed by updates and editing issues. The participants also highlighted as issues: the current visualization of legal documents is not hypertext, and there are ambiguous deals such that one is unable to find the source IP asset.

**Figure 6 - Demographics: issues with legal data**

![Bar chart showing issues frequency]
In the Figure below, you may note that the set of participants is balanced over the three classes of users we are interested in.

![Figure 7 - Categories of users](image)

**Professional service firms**

The first set of questions aim at understanding the working environment of the professional answering the questionnaire. It may be noted that we mainly dealt with small (2 to 7) professional service firms.

![Figure 8 - Number of employees](image)
It is worth noticing that the majority of the professionals who filled in our questionnaire affirm that legal technology is not that popular in their professional service firm. This is an interesting information underling the fact that a big gap exists between professional service firms and legal technology scholars and developers. This means that the needs of these users have probably not been captured in an appropriate way by current legal technology, and it is actually here where the MIREL project may play a significant role.

**What is the degree of spread of legal technology in the context of your professional service firm?**

(8 responses)

![Figure 9 - Spread of legal technology](image)

The answers to the question about the investment in legal technology of the firm reflects the same issue, highlighting that legal technology is a somehow interesting investment for such firms. However, we may note that none of them reported that legal technology is not interesting.

**What is the interest to invest in legal technology of your professional study in the next two years?**

(8 responses)

![Figure 10 - Investment in legal technology](image)
Concerning the free-text answers, the users reported that:

- **Standard/emerging needs** are bar associations vs online legal marketplaces, emerging: online legal marketplace, bank account management, privacy regulation, bank account regulations, privacy contract vs digital contracts

- **Reasons that hinder the diffusion of legal technology** are costs and technicalities, the fact that the results are unsure, and the mean age of persons working in professional service firms.

- **Legal technology** of most interest for professional service firms is contract software and intelligent office management.

- **Legal technology could help in performing their tasks (they all agree on this point).**

- The kind of legal technology to be developed in their opinion includes: drafting of contracts, search tools to extract the most used legal argument from precedent, systems to alert about updates able to describe the impact of the update on the current regulation.

**Data publishers and system developers**

Data publishers and system developers mainly deal with licenses and rights information about the data/software they deal with. However, also privacy issues arise.

![What is the kind of legal information you need to manage with respect to your data/system?](image)

*Figure 11 - Legal issues for developers*
Mainly, data publishers and developers find it (very) difficult to manage and understand legal data, and it is shown that there is more interest in investing in legal technology from their companies than for the case of professional service firms.

Figure 12 - Difficulty in understanding legal data

Figure 13 - Investment in legal technology for developers
The highlighted causes that hinder the diffusion of legal technology are: the fact that the legal language requires legal experts to be understood, the business model, the lack of money, the low uptake of tech across the sector, the complexity of legal language.

The kind of legal technology to invest on are: cloud computing, contract software, software to manage/distribute license information, legal billing, and the blockchain.

The legal technology of the future should develop: system to ease or summarize legal documents, a better version of EURLex, licensing software, back by open distributed ledger (e.g., Blockchain), and digital contracts.

All participants agreed about the fact that legal technology would help them in addressing their tasks.

**Legal scholars and legal experts**

Concerning this third class of users, the spread of legal technology is rather popular, with no negative answers.

![Graph showing the degree of spread of information technology in the context of your working environment](image)

*Figure 14 - Spread of legal technology*

In this case, differently from the previous ones, the interest in investing in information technology is very high, but there are also some negative opinions reporting a “not so interesting” answer.
All participants (strongly) agreed about the importance of information technology to manage legal documents and data.

Figure 16 - Importance of information technology to manage legal data
The kind of legal technology to invest upon is, in this case, office management, cloud computing, webinars, legal advice, legal billing, and support for intelligent editing of legal documents.

The legal technology to be developed concerns student examination tools, case data management, expert systems, and privacy regulation enforcement systems.

All participants agreed about the fact that legal technology would help them in addressing their tasks.

3 Usage scenarios: description and examples

In this Section, we provide concrete examples that have been considered during the first year of the project to address the three usage scenarios envisioned in the MIREL project, namely licenses and contracts, technical documents, and multilingual corpora of norms. For each of these scenarios, we also highlight how the very preliminary prototypes developed during the first year of the project have dealt with such legal texts.

The following table provides some of the terms that will be used in the reminder of the report, and the intended meaning we assume.

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright</td>
<td>A copyright is a set of exclusive rights (to copy, distribute and adapt) granted by a state to the creator of an original work or their assignee for a limited period of time in exchange for public disclosure of the work.</td>
</tr>
<tr>
<td>Digital signature</td>
<td>A digital signature is a mathematical scheme for demonstrating the authenticity of a digital message or document, and it gives a recipient reason to believe that the message was created by a known sender, and that it was not altered in transit.</td>
</tr>
<tr>
<td>License</td>
<td>A license is granted by a licensor to a licensee as an element of an agreement between those parties. It is an authorization by the licensor to use the licensed material by the licensee. It allows an activity that would otherwise be forbidden.</td>
</tr>
<tr>
<td>Policy</td>
<td>A policy is a principle or a rule to guide decisions. Policies differ from law. While law can compel or prohibit behaviours, policies guide actions toward those that are most likely to achieve a desired outcome.</td>
</tr>
<tr>
<td>Waiver</td>
<td>A waiver is the voluntary relinquishment or surrender of some known right or privilege. In the legal field, to waive means to give up voluntarily a right or privilege to something.</td>
</tr>
</tbody>
</table>
3.1.1 Licenses and Contracts

Licensing of data needs to be explicit to avoid any ambiguity in terms of use and reuse. There are many differences worldwide relating to the copyright of data. For example in the European Union database rights exist automatically, whereas, in the USA, data is not covered by any existing copyright law. A guidance is available from a number of sources including the following:

1. Creative Commons licenses: http://creativecommons.org/
2. GNU Free Documentation License: http://www.gnu.org/copyleft/fdl.html
3. Open Data Commons licenses: http://www.opendatacommons.org/

Bizer et al. [7] observe about licensing: “the applications that consume data from the Web must be able to access explicit specifications of the terms under which data can be reused and republished. Availability of appropriate frameworks for publishing such specifications is an essential requirement in encouraging data owners to participate in the Web of Data, and in providing assurances to data consumers that they are not infringing the rights of others by using data in a certain way. Initiatives such as the Creative Commons have provided a framework for open licensing of creative works, underpinned by the notion of copyright”.

The main legal frameworks to set out the conditions for reuse are:

- Terms of APIE\(^1\) (October 2010), a pedagogic reformulation of the framework of law of 1978.
- The types of licenses-APIE\(^2\), which are frames to assert that the data reuse is subject to special conditions (e.g., update) and / or the payment of a fee.
- The license ODbL\(^3\) (also known as ODC-ODbL, 2010), an international license issued by the Open Data Commons organization.
- The license ODC-By\(^4\) (2010) that is also produced by the Open Data Commons organization, but it does not require the "share alike" restriction as in ODbL.
- The license PDDL 1.0\(^5\) (2010) which is also produced by the Open Data Commons organization as ODbL.

---

\(^1\) https://www.apiefrance.fr/sections/actualites/des-conditions-generales-pour-la-reutilisation-des-informations-publiees/view
\(^2\) https://www.apiefrance.fr/sections/acces_thematique/reutilisation-des-informations-publiees/licence-type/view
\(^3\) http://www.opendatacommons.org/licenses/odbl/
\(^4\) http://www.opendatacommons.org/licenses/by/
\(^5\) http://www.opendatacommons.org/licenses/pddl/
• Creative Commons licences\textsuperscript{6} (since 2004), a group of licenses and definitions for an author to facilitate the reuse of his work, in the field of the legislation on copyright. They can be applied to the field of literary and artistic works, but these licenses seem less suited to the field of databases.

• The IP license from the Department of Justice\textsuperscript{7} (May 2010). One of the first licenses in France inspired by Creative Commons licenses, but suited for databases, and, in particular, public data.

These different licenses have common features, but also differ from each other. The requirement to mention the author (attribution - BY) seems to be one of the best shared features, since it is absent only in the license PDDL 1.0 - which in essence, exempt any obligation. Most legal frameworks allow commercial use: that is, they make it possible for re-users to sell public data without transforming or enriching them. The license IP is an exception and prohibits reuse as is of its data for commercial purposes: to make a profitable commercial reuse, the data must be enriched in some way. Other features also differentiate certain licenses. The ban on commercial use of data is a feature of some Creative Commons licenses (such as CC-NC), the only ones that adopt it. The data covered by this obligation cannot be reused in a commercial setting.

In the context of MIREL, we have started by considering mainly Creative Commons licenses and Open Data Commons licenses, plus a number of proprietary licenses like the GNU and Apache ones.

Creative Commons was publicly launched in December 2002, but its genesis traces to summer 2000 and discussions about how to promote a reasonable and flexible copyright regime for the Web. There was no standard legal means for creators to grant limited rights to the public for online material, and obtaining rights often required difficult searches to identify rights-holders and transaction costs to negotiate permissions.

Concerning Open Data Commons licenses\textsuperscript{8}, the Public Domain Dedication and Licence (PDDL)\textsuperscript{9} allows the information provider to freely share, modify, and use this work for any purpose and without any restriction. This license is intended for use on databases or their contents ("data"), either together or individually. Many databases are covered by copyright. These sets of rights, as well as other legal rights used to protect databases and data, can create uncertainty or practical difficulty for those wishing to share databases and their underlying data but retain a limited amount of rights under a "some rights reserved" approach to licensing as outlined in the Science Commons Protocol for Implementing Open Access Data. The Attribution License (ODC-By)\textsuperscript{10} has

\textsuperscript{6} http://creativecommons.org/licenses/
\textsuperscript{7} http://www.rip.justice.fr/information_publique_librement_reutilisable
\textsuperscript{8} http://opendatacommons.org/licenses/
\textsuperscript{9} http://opendatacommons.org/licenses/pddl/
\textsuperscript{10} http://opendatacommons.org/licenses/by/
been developed for the databases, and it allows to copy, distribute, and use the database, to produce works from the database, and to modify, transform and build upon the database. In the Open Database Licence (ODbL)\(^{11}\), any public use of the database or works produced from the database must be attribute in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, the license of the database has to be clear, and any notices on the original database must be kept intact. If any adapted version of this database is publicly used, the adapted database has to be offered under the ODbL.

Let us consider now the overall structure of a licenses, e.g., the Creative Commons BY 4.0\(^{12}\)

---

**This is a human-readable summary of (and not a substitute for) the license.**

**Disclaimer.**

This license is acceptable for Free Cultural Works.

You are free to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

- **No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

**Notices:**

---

\(^{11}\) http://opendatacommons.org/licenses/odbl/

\(^{12}\) Full legal code available at https://creativecommons.org/licenses/by/4.0/legalcode
Following the line of research started in 2013 [4][5], the INRIA and Cordoba partners are working together to mine such a kind of legal text to extract their main features, i.e., the deontic elements like obligations, permissions and prohibitions, sub-licensing information if any, temporal constraints. The overall idea behind this approach is the following: we applied different Active Learning strategies to Information Extraction from licenses in English, with highly repetitive text, few annotated or non-annotated examples available, and very fine precision needed. We showed that the most popular approach to active learning, i.e., uncertainty sampling for instance selection, does not provide a good performance in this setting. They showed that they can obtain a similar effect to that of density-based methods using uncertainty sampling, by just reversing the ranking criterion, and choosing the most certain instead of the most uncertain instances. This approach is currently under implementation in the Licentia platform with the aim to add a new service able to receive as input a license in natural language (English) and returning a machine-readable version of such licenses to be attached to the data it licenses. More details are available in [4][5].

Partners INRIA and UNITO are currently working over a dataset of natural language licenses to develop a question-answering system able to answer natural language questions of the users about how to licenses their data. This approach is under definition, and no result is available yet.

Business contracts are mutual agreements between two or more parties engaging in various types of economic exchanges and transactions. They are used to specify the obligations, permissions and prohibitions that the signatories should be held responsible for and to state the actions or penalties that may be taken when any of the stated agreements are not being met. Given the increasing efforts by organizations to carry out their business via the Internet, it is crucial to model contracts in terms of workflows, such that all relevant tasks of contracts can be described as elements of business processes, where business processes are constrained by business rules, statements or policies listed in business contracts or other legal documents that are used by organizations to run the activities, to provide an understanding of how a business operates, and to direct the behavior of the organization.

During the first year, no concrete business contract legal text has been taken into account for mining tasks. This is planned for the second year of the MIREL project.
3.1.2 Technical Documents

Technical documents describe how to handle and what are the functionalities of a certain product. They are intended to provide information about what can/cannot be done with the product. Given the huge dimension of this particular kind of legal texts and their diffusion in the companies, these documents have to be taken into account in our project. The advantages of extracting a machine-readable representation of such texts would allow to have in this representation a kind of summary of the main constraints expressed in the documents, with invaluable time saving for every person that is expected to read the whole document before obtaining this information.

In MIREL, an example of technical document used to address the mining and reasoning tasks is the Australian Telecommunications Consumer Protections Code\textsuperscript{13}. The Telecommunication Consumer Protections Code is the Australian industry code for the telecommunication industry and mandates that every operator has to provide annual compliance statements with the Code. The Code was enacted in September 2012, and it entered in force in April 2013. In 2015, the Code was revised and a new version enacted with some amendments entered in force in 2016.

An example from the Code is provided in Figure 17. This document expresses deontic elements (i.e., obligations, prohibitions and permissions), but the main difficulty of such a kind of documents with respect, for instance, to licenses is that in technical documents the deontic component is often coupled with more complex conditions expressing temporal and location constraints and thus complexifying both the extraction (or mining) of the legal knowledge herein and the definition of reasoning methods over such knowledge.

A first attempt to deal with the problem of mining legal information from the TCP Code has been addressed by the NICTA-Data61 and INRIA partners of MIREL\textsuperscript{[8]}. The idea is of combining different Natural Language Processing techniques towards the extraction of rules from legal documents. More precisely, our framework for automated rules generation exploits the Stanford Parser to obtain the grammatical representation of the sentences, and WordNet\textsuperscript{5} to deal with the variability of the language in expressing the deontic components in natural language legal texts. We combine this syntactic-based rules extraction approach, relying on the well-known Stanford Parser, together with a logic-based approach, exploiting the Boxer framework for the extraction of logical dependencies between chunks of text. The results of the evaluation of the combined framework on a section of the Australian TCP Code (Section 8) show the feasibility of the proposed approach, and foster further research in this direction.

\textsuperscript{13} http://www.commsalliance.com.au/Documents/all/codes/c628
8.1.1 A Supplier must take the following actions to enable this outcome:

(a) **Implement a process**: implement, operate and comply with a Complaint handling process that:

(i) is Consumer focused and easy to use;

(ii) is free of charge, other than for:

A. the call costs referred to in clause 8.1.1(a)(vii)A;

B. the provision of information where:

   - a Consumer or former Customer requests access to information held by the Supplier about the Consumer or former Customer which was collected by the Supplier more than 2 years prior to the date of the request, unless the Complaint relates to an interference with the privacy of the Consumer under the Privacy Act by the Supplier; or

   - the free provision of the information in the form or quantities requested is inconsistent with the Supplier’s Standard Form Customer Contract or the relevant summary of the Offer referred to in clause 4.1.2, unless the Complaint relates to an interference with the privacy of the Consumer under the Privacy Act by the Supplier;

in which circumstances, the Supplier may levy a charge to recover its costs. The Supplier must inform the Consumer or former Customer of the proposed charge and notify the Consumer or former Customer of the option to pursue the Complaint and pay the charge or to discontinue the Complaint. The Supplier must tell the Consumer or former Customer about the options for external dispute resolution before levying any Charge under this paragraph.

(iii) is approved by, and provides that, the Chief Executive Officer (or equivalent) of the Supplier is responsible for ensuring its implementation, operation and compliance in accordance with the requirements of this chapter;

(iv) is under the direction of a senior manager who is responsible for maintaining its effective and efficient operation in accordance with the requirements of this chapter;

*Figure 17 - An example of technical document from the TCP Code.*
Secondly, an approach to reason over this legal information has been proposed by partners NICTA, UNIBO and INRIA. They presented an application of the semantic business process regulatory compliance checking where we relay on the semantics of LegalRuleML\textsuperscript{14} for the representation of the norms and their dynamics, as expressed in such a kind of technical documents. They discussed and analyzed different but comparable ways to model the semantics of norms as well as their dynamics (e.g., new versions of certain regulations are proposed). Moreover, they showed how this semantic modeling phase, with tasks coupled with semantic annotations, can be exploited to address and improve the regulatory compliance checking process, and answer companies’ needs about compliance checking. For the experimental evaluation, they consider the 2012 and 2016 versions of the Code, and model them using LegalRuleML. More details about this approach are available in [3].

3.1.3 Multilingual corpora of norms

The translation of legal texts into a shared uniform machine-readable representation is of fundamental importance in the legal multilingual scenario. In this case, like in the European Union, slightly different norms are proposed by the different countries to regulate a certain aspect of interest for the EU. Having one shared language-agnostic representation for such a kind of norms would help a lot in avoiding costly and time consuming translations, to avoid misunderstandings, and to focus directly on the content of the norms.

This usage scenario adds to the previous ones a new degree of complexity which lies in multilingualism. The ideal solution to this issue would be to provide methods to mine and reason over corpora of norms written in different languages. At the present stage, we have only limitedly tackled this issue in MIREL during the first year, but this will be a major focus in our research for the next years.

An example of multilingual corpus of norms is provided by the JRC-Acquis EU corpus\textsuperscript{15}. The Acquis Communautaire (AC) is the total body of European Union law applicable in the EU Member States. This collection of legislative text changes continuously and currently comprises selected texts written between the 1950s and now. As of the beginning of the year 2007, the EU had 27 Member States and 23 official languages. The Acquis Communautaire texts exist in these languages, although Irish translations are not currently available. The Acquis Communautaire thus is a collection of parallel texts in the following 22 languages: Bulgarian, Czech, Danish, German, Greek, English, Spanish, Estonian, Finnish, French, Hungarian, Italian, Lithuanian, Latvian, Maltese, Dutch,

\textsuperscript{14} LegalRuleML\textsuperscript{6} is an effort to create a standard for the representation of norms. More details at https://tools.oasis-open.org/version-control/browse/wsvn/legalruleml/

\textsuperscript{15} https://ec.europa.eu/jrc/en/language-technologies/jrc-acquis
Polish, Portuguese, Romanian, Slovak, Slovenian and Swedish. The data release by the JRC is in line with the general effort of the European Commission to support multilingualism, language diversity and the re-use of Commission information. The JRC did not receive an authoritative list of documents that belong to the Acquis Communautaire. In order to compile the document collection JRC-Acquis, all those CELEX documents that were available in at least ten of the twenty EU-25 languages were selected. The corpus is thus an approximation of the Acquis Communautaire called JRC-Acquis.

An example from this corpus is available below (XML format):

```xml
<p n="2">COUNCIL DECISION of 3 March 1981 on the granting of daily allowances and the reimbursement of travelling expenses of members of the Economic and Social Committee, alternates and experts (81/121/EEC)</p>

<p n="3">THE COUNCIL OF THE EUROPEAN COMMUNITIES,</p>

<p n="4">Having regard to the Treaty establishing a Single Council and a Single Commission of the European Communities, and in particular Article 6 thereof,</p>

<p n="5">Whereas procedures should be laid down for granting daily allowances and reimbursing the travelling expenses of members of the Economic and Social Committee, and of alternates and experts within the meaning of the Rules of Procedure of the Economic and Social Committee of 13 June 1974 (1), as amended on 22 July 1980 (2),</p>

<p n="6">HAS DECIDED AS FOLLOWS:</p>

<p n="7">Article 1</p>

<p n="8">Members of the Economic and Social Committee, alternates and experts shall be entitled to a daily allowance for days spent at meetings or travelling and to the reimbursement of travelling expenses in accordance with the provisions below.</p>

<p n="9">Article 2</p>

<p n="10">The daily allowance per day of meeting shall amount to: - Bfrs 3 300 for members,</p>

<p n="11">- Bfrs 2 475 for alternates and experts.</p>

<p n="12">Article 3</p>

<p n="13">1. The time taken into consideration for the reimbursement of days travelling shall be established on a flat-rate basis for return journeys in accordance with the following calculation:</p>
```
Users’ Needs and Usage Scenarios for Future Legal Applications

- three-quarters of a day for a distance of over 100 kilometres up to a maximum of 200 kilometres,

  - one day for a distance of over 200 kilometres up to a maximum of 500 kilometres,

  - one day and a half for a distance of over 500 kilometres up to a maximum of 1000 kilometres,

  - two days for a distance of over 1000 kilometres.

2. However, travelling time shall continue to be reimbursed at the rates paid under the financial rules of the Economic and Social Committee in force at 1 March 1981 until such time as the method of calculation described in paragraph 1 enables the same level of reimbursement to be attained.

   Article 4

   1. Travelling expenses shall be reimbursed in accordance with the financial rules of the Economic and Social Committee in force at 1 March 1981, it being understood that the alternates shall receive the same treatment as members and experts.

   2. However, air travel of members, alternates and experts shall be reimbursed on the basis of the tourist class fare, except in the case of the chairman on the one hand and the vice-chairman on official missions on the other, who are entitled to reimbursement of the first class fare.

   Article 5

   This Decision shall take effect on 1 March 1981.

   Done at Brussels, 3 March 1981.

   For the Council

   The President

The corpus has been used by the UNITO and UL partners to develop the Eunomos System. Eunomos is a legal document and knowledge management system. Differently than other systems, it firstly recognizes the need for a stricter coupling between legal knowledge and its legislative sources, associating the concepts of its legal ontology with the part of regulations defining them, structured using legislative XML. This solution allows to ground concepts of legal ontologies to their sources, making ontologies more acceptable to practitioners and synchronizing their meaning with the evolution of the text of the law across its modifications. The main goal of the Eunomos software is to help users to cut through the information overload to get the legal information they need in an organized and structured way and keep track of the state of the relevant law on any given topic. For more details about Eunomos, we refer the reader to [2].

Another example of (multilingual) corpus of norms that has been exploited in the project consists of the corpus of legal questions from Japanese Legal Bar exams, and the relevant Japanese Civil Law articles. The corpus is translated in English. An example of the Japanese Civil Law articles from such corpus is provided below:

(Seller’s Warranty in cases of Superficies or Other Rights) Article 566 (1) In cases where the subject matter of the sale is encumbered with for the purpose of a superficies, an emphyteusis, an easement, a right of retention or a pledge, if the buyer does not know the same and cannot achieve the purpose of the contract on account thereof, the buyer may cancel the contract. In such cases, if the contract cannot be cancelled, the buyer may only demand compensation for damages. (2) The provisions of the preceding paragraph shall apply mutatis mutandis in cases where an easement that was referred to as being in existence for the benefit of immovable property that is the subject matter of a sale, does not exist, and in cases where a leasehold is registered with respect to the immovable property. (3) In the cases set forth in the preceding two paragraphs, the cancellation of the contract or claim for damages must be made within one year from the time when the buyer comes to know the facts.

(Seller’s Warranty in cases of Mortgage or Other Rights) Article 567 (1) If the buyer loses his/her ownership of immovable property that is the object of a sale because of the exercise of an existing statutory lien or mortgage, the buyer may cancel the contract. (2) If the buyer preserves his/her ownership by incurring expenditure for costs, he/she may claim reimbursement of those costs from the seller. (3) In the cases set forth in the preceding two paragraphs, the buyer may claim compensation if he/she suffered loss. There is a limitation period on pursuance of warranty if there is restriction due to superficies on the subject matter, but there is no restriction on pursuance of warranty if the seller’s rights were revoked due to execution of the mortgage.

This is instead an example of yes/no questions from the Japanese Legal Bar exams, where the first text snippet represents the question, and the following four snippets represent the proposed
An entrepreneur from State A decided to sell hot sauce to the public, labeling it “Best Hot Sauce.” A company incorporated in State B and headquartered in State C sued the entrepreneur in federal court in State C. The complaint sought $50,000 in damages and alleged that the entrepreneur’s use of the name “Best Hot Sauce” infringed the company’s federal trademark. The entrepreneur filed an answer denying the allegations, and the parties began discovery. Six months later, the entrepreneur moved to dismiss for lack of subject-matter jurisdiction. Should the court grant the entrepreneur’s motion?

- No, because the company’s claim arises under federal law. The claim asserts federal trademark infringement, and therefore it arises under federal law. Subject-matter jurisdiction is proper under 28 U.S.C. § 1331 as a general federal-question action. That statute requires no minimum amount in controversy, so the amount the company seeks is irrelevant. (YES)

- No, because the entrepreneur waived the right to challenge subject-matter jurisdiction by not raising the issue initially by motion or in the answer. Under Federal Rule 12(h)(3), subject-matter jurisdiction cannot be waived and the court can determine at any time that it lacks subject-matter jurisdiction. Therefore, the fact that the entrepreneur delayed six months before raising the lack of subject-matter jurisdiction is immaterial and the court will not deny his motion on that basis. (NO)

- Yes, because although the claim arises under federal law, the amount in controversy is not satisfied. There is no amount-in-controversy requirement for actions that arise under federal law. Although diversity jurisdiction requires an amount in controversy of $75,000 or more, when diverse parties are litigating a federal claim, the action is treated for jurisdictional purposes as a federal-question action, not a diversity action. The claim here asserts federal trademark infringement and therefore it arises under federal law. The fact that the action does not meet all the requirements for diversity jurisdiction is irrelevant. (NO)

- Yes, because although there is diversity, the amount in controversy is not satisfied. (NO)

This corpus has been used by the UNITO partner that participated at COLIEE 2016 Information Retrieval and Legal Question Answering challenge. The proposed Question Answering solution answers YES or NO to a question, i.e., YES if the question is entailed by a text and NO otherwise. With recent exploit of Multi-layered Neural Network systems at language modeling tasks, they presented a Deep Learning approach, which uses an adaptive variant of the Long-Short Term Memory. Moreover, even though they used some features that have performed well from similar
works, they also introduced some semantic features for performance improvement. The overall result confirms the competitiveness of the proposed approach. For more details, we refer the reader to [6].

Another corpus of norms from the EU has been used in MIREL, namely the corpus of judgments from the European Court of Human Rights\textsuperscript{16}. The HUDOC (Human Rights Documentation) database provides access to the case-law of the Court (Grand Chamber, Chamber and Committee judgments and decisions, communicated cases, advisory opinions and legal summaries from the Case-Law Information Note), the European Commission of Human Rights (decisions and reports) and the Committee of Ministers (resolutions). This collection contains all the press releases issued by the Registry since 1 January 1999, and it is available in English and French. The corpus includes the summaries of judgments and decisions delivered by the Court and information about cases pending and about the Court’s activities in general.

An example from the corpus is as follows (CASE OF B.S. v. SPAIN):

\begin{center}
\begin{tabular}{|p{20cm}|}
\hline
\textbf{THE LAW} \\
1. \\
\textbf{ALLEGED VIOLATION OF ARTICLE 3 OF THE CONVENTION} \\
29. \\
The applicant complained, firstly, that the national police had both verbally and physically abused her when they had stopped and questioned her. \\
She alleged that she had been discriminated against on account of her skin colour and her gender, whereas other women with a “European phenotype” carrying on the same activity in the same area had not been approached by police. \\
The applicant also complained about the language used by Palma de Mallorca investigating judge no. 9, who, in his decision of 10 June 2007, had referred to the “shameful spectacle of prostitution on the public highway”. Relying on the provisions of Article 3, the applicant alleged that the domestic courts’ investigation of the events had been inadequate. \\
30. \\
The provisions relied on are worded as follows: \\
Article 3 \\
“No one shall be subjected to torture or to inhuman or degrading treatment or punishment.”
\hline
\end{tabular}
\end{center}

\textsuperscript{16} http://www.echr.coe.int/
Article 14

“The enjoyment of the rights and freedoms set forth in [ the ] Convention shall be secured without discrimination on any ground such as sex , race , colour , language , religion , political or other opinion , national or social origin , association with a national minority , property , birth or other status .”

...

B .

The merits

1 .

Effectiveness of the investigations carried out by the national authorities

a ) The parties’ submissions

i .

The Government

31 .

The Government disputed , at the outset , the seriousness of the injuries sustained by the applicant and pointed out that their cause had not been proved .

32 .

The Government also submitted that the police interventions in the area in question had not in any way targeted the applicant personally or discriminated against her , but had been preventive security measures designed to respond to public alarm caused by prostitution and to combat networks operating in the Balearic Islands which exploited immigrant women , in particular in the El Arenal district in which the applicant carried on her activity.

The Ministry of the Interior had already implemented measures to combat such networks under Institutional Law no .1/1992 on the protection of urban security.

The Government observed in that connection that whilst prostitution was not in itself a criminal offence in Spain , forced prostitution was an offence under the Criminal Code .

33 .

With regard to the incidents of 15 and 21 July 2005 , the Government noted that the applicant’s allegations had been the subject of a judicial investigation by Palma de Mallorca investigating judge no .9 , during which the only investigative measure requested by the applicant had been an identity parade of the police officers behind a two-way mirror .
Besides the fact that the applicant had not lodged a complaint against the officers, the rejection of her request was justified, in the Government’s submission, on the grounds that the officers had already been identified by the police authorities.

Those proceedings had been concluded by the judgment of 11 March 2008, delivered after a public hearing, acquitting the officers in question.

34.

With regard to the second episode – of 23 July 2005 – the Government observed that this had been examined by Palma de Mallorca investigating judge no 11.

35.

The Government pointed out that the procedural obligation imposed on the States with regard to Article 3 of the Convention was an obligation of means and not of result.

In their submission, the investigative procedures brought before the two investigating judges were sufficient to consider that the Spanish State had fulfilled its obligations, irrespective of the fact that the police officers were ultimately not convicted.

ii. The applicant

36.

The applicant considered that the manner in which the investigation had been carried out before the domestic courts amounted to a breach of the State’s procedural obligations under Article 3.

In her submission, the courts had not adequately dealt with her request for certain investigative measures regarding the incidents she had alleged, such as an identity parade of the officers behind a two-way mirror which would have enabled her to recognise the police officers involved.

The applicant complained that the State shifted the obligation to investigate on to her and imposed the burden of proving the alleged offence on her, whereas according to the Strasbourg Court’s case-law, it was incumbent on the State to prove that particular treatment was not discriminatory.

This corpus has been used by the INRIA and Cordoba partners to mine legal text to populate legal ontologies. Ontologies are the main mechanism for domain-specific knowledge representation as they allow for an exhaustive characterization of the domain of interest. A special class of
ontologies are the legal ones which specify legal concepts in a formal way, such that reasoning mechanisms can then be exploited over such information. Many legal ontologies have been proposed in the literature with different purposes and applied to different sub-domains. However, their manual creation and maintenance is a very time-consuming and challenging task: domain-specific information needs to be created by legal experts to ensure compliance with the regulations we aim at modeling and capture their full semantics. This lack of coverage of legal ontologies makes it difficult to train legal Named Entity Recognition (NER) systems, and to support Entity Linking to mine legal documents. For this reason, the INRIA and Cordoba partners presented an approach to legal ontology population with the aim to find new entities to ease the next Entity Linking step. They applied a learning strategy, called curriculum learning that can deal with the lack of examples that characterizes such kind of ontologies, but that still produces few errors only. Roughly, curriculum learning is a method that trains a model by presenting increasingly more complex concepts. They implemented curriculum learning to learn progressively finer grained classes using a neural network classifier. Two are the adopted ontologies: YAGO – Yet Another Great Ontology, that is not specifically targeted to represent the legal domain, and the LKIF (Legal Knowledge Interchange Format) ontology. The results are promising. We cannot report here references to scientific contributions detailing these approaches because they are currently under submission. Also the HUD partner of MIREL is now looking at the HUDOC corpus for defeasible reasoning task evaluation.

Finally, the MIREL project, and more precisely the UL and UNIBO partners, are considering another corpus of norms, namely the General Data Protection Regulation (GDPR)\(^\text{17}\). This corpus contains the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Documents are reported in 24 languages. Mining and reasoning over such kind of data is of particular important, and this is the goal of UL and UNIBO for the coming years.

An extract from the Regulation is as follows:

(65) A data subject should have the right to have personal data concerning him or her rectified and a ‘right to be forgotten’ where the retention of such data infringes this Regulation or Union or Member State law to which the controller is subject. In particular, a data subject should have the right to have his or her personal data erased and no longer processed where the personal data are no longer necessary in relation to the purposes for which they are collected or otherwise processed, where a data subject has withdrawn his or her consent or objects to the processing of personal data concerning him or her, or where the processing of his or her personal data does not otherwise comply with this Regulation. That right is relevant in particular where the data subject has given his or her consent as a child and is not fully aware of the risks involved by the processing, and later wants to remove such personal data, especially on the internet. The data subject should be able to

\(^{17}\) http://eur-lex.europa.eu/eli/reg/2016/679/oj
exercise that right notwithstanding the fact that he or she is no longer a child. However, the further retention of the personal data should be lawful where it is necessary, for exercising the right of freedom of expression and information, for compliance with a legal obligation, for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, on the grounds of public interest in the area of public health, for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, or for the establishment, exercise or defense of legal claims.

4 Conclusions

In this deliverable, we have reported about the activity of the MIREL partners about the study of the users’ needs for the development of legal technology, and the usage scenarios envisaged in the project as well as their instantiations.

The users’ needs have been detected with the definition of a questionnaire that has been proposed to three different categories of potential users of legal technology (i.e., professional service firms, data publishers and system developers, legal scholars and legal experts). Results highlighted that all participants agreed about the fact that legal technology can be of help to them to address their tasks. This is an important feedback that confirms the relevance of the research carried out in the MIREL project. Moreover, a lack of interaction between legal technology experts and professional service firms has been highlighted, underlining the need to interact with such professionals to first explain them what legal technology can do for them (by means of concrete examples), and then ask for their more urgent needs. Another gap that has been highlighted thanks to this questionnaire is between data publishers and developers (thus the computer science side of the legal technology issue) and legal experts (the law side), meaning that for developers it is difficult to understand the legal language and also to interact with legal experts. This is one of the goals of MIREL, i.e., to ease the interaction between computer scientists and law experts in order to cooperate to the development of the next generation of legal technology. Already from the very beginning, MIREL shows to be in right direction, as witnessed by the collaborations mentioned above where computer science partners fruitfully collaborate with law partners. Concerning the legal information to invest on or to develop, some of these issues are currently considered by the MIREL partners, e.g., the licenses information management tool. We will explore how to answer the emerged needs in the coming year.

In addition, we have highlighted for each of the three usage scenarios, i.e., licenses and contracts, technical documents and multilingual corpora of norms, what are the concrete legal texts the partners of MIREL have actually exploited in this first year of the project. Furthermore, we have described the first approaches proposed in the context of the project to mine and reason from legal text. Such approaches tackle a set of heterogeneous issues in mining and reasoning from norms ranging from ontology population to legal question-answering to legal textual entailment.
References


Appendix