EuroVoc-based Summarization of European Case Law

Florian Schmedding, Peter Klügl, David Baehrens, Christian Simon, Kai Simon, and Katrin Tomanek

Workshop on Language and Semantic Technology for Legal Domain
September, 10 2015
Hissar, Bulgaria
• Harmonization of the legal systems across the European Union
  – National Case Law decisions become effective in all member states, for example

• Context research must consider
  – Several legal systems (different terms)
  – Multiple legal repositories (different formats)
  – Different languages

→ Need for comprehensive access to pan-European legislation and jurisdiction
1. Summaries provide a huge benefit for legal professionals

2. Automated summarization simplifies the operation of legal information systems
   - Controlled keyword summaries
     - Descriptive
     - Relate documents across languages
   - Extractive sentence summaries
     - Exact wording is important
EUROVOC

- Multilingual & multidisciplinary thesaurus from the Publications Office of the European Union
- European Parliament and national governments use it for indexing
- ~7000 concepts
- 21 area fields with 127 micro thesauri
- 20 languages
- We use Bulgarian, English, French, German, and Italian
- Blacklisted some general concepts like „law“
MULTILINGUAL NLP-PIPELINE SETUP

Input document

Language detection

bg

Sentence detection
Tokenizer
Stopword Tagger
Lemma-
izer
Part-of-
Speech Tagger

en

Sentence detection
Tokenizer
Stopword Tagger
Stemmer
Morpho-
Semantic Segmenter
Part-of-
Speech Tagger

de

Sentence detection
Tokenizer
Stopword Tagger
Stemmer
Morpho-
Semantic Segmenter
Part-of-
Speech Tagger

fr

Sentence detection
Tokenizer
Stopword Tagger
Stemmer
Morpho-
Semantic Segmenter
Part-of-
Speech Tagger

it

Sentence detection
Tokenizer
Stopword Tagger
Lemma-
izer
Part-of-
Speech Tagger

EuroVoc Concept Mapper

TextRank Keyword Extraction

Summarizing keywords

TextRank Sentence Extraction

Summarizing sentences
CONCEPT MAPPING

- Extension of the ConceptMapper (Tanenblatt, Coden, and Sominsky) developed by Averbis
- Different lookup techniques
  - Exact matches
  - Normalized matches (stems, lemmas, segments)
  - Order-dependent/independent matches
  - Contiguous/non-contiguous matches
  - Restrict to matches containing nouns
- Disambiguation
  - Canonical form is preferred
  - Remove all concepts if still ambiguous
  - Statistics and dictionary based components also available
For nearly half a century, legal aid provided out of public funds was the main source of funding for those of modest means who sought to make or (less frequently) defend claims in the civil courts and who needed professional help to do so. By this means access to the courts was made available to many who would otherwise, for want of means, have been denied it. But as time passed the defects of the legal aid regime established under the Legal Aid and Assistance Act 1949 and later statutes became more and more apparent. While the scheme served the poorest well, it left many with means above a low ceiling in an unsatisfactory position, too well off to qualify for legal aid but too badly off to contemplate incurring the costs of contested litigation. There was no access to the courts for them. Moreover, the effective immunity against adverse costs orders enjoyed by legally-aided claimants was always recognised as a burden on a privately-funded defendant resisting a legally-aided claim, since he would be liable for both sides' costs if he lost and his own even if he won. Most seriously of all, the cost to the public purse of providing civil legal aid had risen sharply, without however showing an increase in the number of cases funded or evidence that legal aid was directed to cases which most clearly justified the expenditure of public money.
TEXTRANK (MIHALCEA AND TARAU)

- Unsupervised approach for ranking elements of text documents
- Independent from content and language
- Represent the text as graph and weight its nodes
- PageRank (Brin and Page):
  \[ S(V_i) = (1 - d) + d \times \sum_{j \in \text{In}(V_i)} \frac{1}{|\text{Out}(V_j)|} S(V_j) \]
  - Choose \( d := 0.85 \) as usual
- Natural language preprocessing improves results
For nearly half a century, legal aid provided out of public funds has been the main source of funding for those of modest means who sought or (less frequently) defend claims in the civil courts and who could not otherwise obtain professional help to do so. By this means access to the courts was available to many who would otherwise, for want of means, have been denied it. But as time passed the defects of the legal aid scheme established under the Legal Aid and Advice Act 1949 and later, became more and more apparent. While the scheme served the purpose well, it left many with means above a low ceiling in an unsatisfactory position, too well off to qualify for legal aid but too badly off to contemplate incurring the costs of contested litigation. The access to the courts for them. Moreover, the effective immunity from adverse costs orders enjoyed by legally-aided claimants was a source of concern that it would place an unfair burden on a privately-funded defendant resisting a legally-aided claim, since he would be liable for costs if he lost and his own even if he won. Most serious of all, the cost to the public purse of providing civil legal aid soared sharply, without however showing an increase in the number of funded or evidence that legal aid was directed to cases which did not clearly justify the expenditure of public money.
KEYWORD EXTRACTION

- Map nodes to concepts:
  aid, legal → eurovoc:3969

- Score the concepts:
  \[ \text{Score(eurovoc:3969)} = \frac{2.46 + 2.80}{2} = 2.63 \]

- Cluster all mentions of a concept

- Score the clusters: Highest score among the contained concepts

- Main synonym becomes the keyword:
  eurovoc:3969 → legal aid
• Take the sorted list of keywords
• Starting with the best, select the enclosing sentence
• Continue with the next best keyword
• Skip keywords that are already contained within a selected sentence

For nearly half a century, legal aid provided out of public funds was the main source of funding for those of modest means who sought to make or (less frequently) defend claims in the civil courts and who needed professional help to do so. By this means access to the courts was made available to many who would otherwise, for want of means,
CASE STUDY

• 100 questionnaires about case law decisions from different sources

• Each questionnaire contains
  – 12 automatically selected keywords
  – 8 automatically selected sentences

• Participants divide them into four categories
  – Very useful
  – Useful
  – Not useful
  – Misleading

• Participants choose five keywords and sentences
# DOCUMENT SOURCES

<table>
<thead>
<tr>
<th>Language</th>
<th>Source</th>
<th>Documents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>EUR-Lex</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Judiciary.gov.uk</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supreme Court (UK)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>House of Lords Judgments</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>EUR-Lex</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Legifrance</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>Guistizia Amministrativa</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Corte Costituzionale</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Bulgarian</td>
<td>EUR-Lex</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Varhoven Kasatsionen Sad</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Varhoven Administrativen Sad</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Portal &quot;Sadebni aktove&quot; - Vish Sadeben Savet</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>EUR-Lex</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Rechtsinformationssystem (at)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundesverfassungsgericht</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundesgerichtshof</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundesverwaltungsgericht</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundesfinanzhof</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundesarbeitsgericht</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundessozialgericht</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bundespatentgericht</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Summary Sentences

1. Please rate the utility of each preselected summary sentence (yellow) by checking the radio buttons in the blue columns.
2. Please select five sentences that summarize the document in your opinion by ticking five checkboxes in the green column. You may choose from all available sentences (yellow or white).

<table>
<thead>
<tr>
<th>Very useful</th>
<th>Useful</th>
<th>Not useful</th>
<th>Misleading</th>
<th>Human (5) Summary Sentence</th>
</tr>
</thead>
</table>
| ![Rating](image1)
| ![Rating](image2)
| ![Rating](image3)
| ![Rating](image4)
| ![Rating](image5) |

For nearly half a century, legal aid provided out of public funds was the main source of funding for those of modest means who sought to make (or less frequently) defend claims in the civil courts and who needed professional help to do so.

By this means access to the courts was made available to many who would otherwise, for want of means, have been denied it.

But as time passed the defects of the legal aid regime established under the Legal Aid and Advice Act 1949 and later statutes became more and more apparent.

While the scheme served the poorest well, it left many with means above a low ceiling in an unsatisfactory position, too well off to qualify for legal aid but too badly off to contemplate incurring the costs of contested litigation.

There was no access to the courts for them.

Moreover, the effective immunity against adverse costs orders enjoyed by legally-aided claimants was always recognised to place an unfair burden on a privately-funded defendant resisting a legally-aided claim, since he would be liable for both sides' costs if he lost and his own even if he won.
RESULTS

• Ratings received for
  – 1070 keywords and
  – 660 sentences

• Participants have chosen
  – 366 keywords and
  – 468 sentences

• Distinguish between
  – Helpful (Very useful, Useful) and
  – Not helpful (Not useful, Misleading)

• Keywords are mostly not helpful 😞
• Sentences are helpful 😊
<table>
<thead>
<tr>
<th>Keyword</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>claim</td>
<td>Not useful</td>
</tr>
<tr>
<td>costing</td>
<td>Useful</td>
</tr>
<tr>
<td>fees</td>
<td>Very useful</td>
</tr>
<tr>
<td>insurance</td>
<td>Very useful</td>
</tr>
<tr>
<td>succession</td>
<td>Not useful</td>
</tr>
<tr>
<td>insurer</td>
<td>Very useful</td>
</tr>
<tr>
<td>insurance premium</td>
<td>Very useful</td>
</tr>
<tr>
<td>operating cost</td>
<td>Very useful</td>
</tr>
<tr>
<td>insurance claim</td>
<td>Very useful</td>
</tr>
<tr>
<td>insurance policy</td>
<td>Very useful</td>
</tr>
<tr>
<td>insured risk</td>
<td>Useful</td>
</tr>
<tr>
<td>administrative cost</td>
<td>Very useful</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>ДДС</td>
<td>Useful</td>
</tr>
<tr>
<td>доставчик</td>
<td>Not useful</td>
</tr>
<tr>
<td>данък</td>
<td>Useful</td>
</tr>
<tr>
<td>договор</td>
<td>Not useful</td>
</tr>
<tr>
<td>предоставяне на услуги</td>
<td>Useful</td>
</tr>
<tr>
<td>производство</td>
<td>Not useful</td>
</tr>
<tr>
<td>услуги</td>
<td>Not useful</td>
</tr>
<tr>
<td>данъчна основа</td>
<td>Not useful</td>
</tr>
<tr>
<td>данъчна администрация</td>
<td>Not useful</td>
</tr>
<tr>
<td>ревизия</td>
<td>Useful</td>
</tr>
<tr>
<td>документ</td>
<td>Not useful</td>
</tr>
<tr>
<td>трудов договор</td>
<td>Not useful</td>
</tr>
</tbody>
</table>
For nearly half a century, legal aid provided out of public funds was the main source of funding for those of modest means who sought to make or (less frequently) defend claims in the civil courts and who needed professional help to do so.

While the scheme served the poorest well, it left many with means above a low ceiling in an unsatisfactory position, too well off to qualify for legal aid but too badly off to contemplate incurring the costs of contested litigation.

The main instruments upon which it was intended that claimants should rely to achieve the second and third of the aims are described by my noble and learned friend: they are conditional fee agreements and insurance cover obtained after the event giving rise to the claim.

It was however evident that the success of the new funding regime was threatened by two contingencies which, had they occurred, could have proved fatal.

Another possible abuse was that lawyers would be willing to act for claimants on a conditional fee basis but would contract for a success uplift grossly disproportionate to any fair assessment of the risks of failure in the litigation, again knowing that the burden of paying this uplifted fee would never fall on their client but would be borne by the defendant or his insurers.

Under the new regime, a claimant who makes appropriate arrangements can litigate without any risk of ever having personally to pay costs either to those acting for him or to the other side and without any risk of ever having to pay an after the event insurance premium whatever the outcome: the practical result is to transfer the entire cost of funding this kind of litigation to the liability insurers of unsuccessful defendants (and defendants who settle the claims made against them) and thus, indirectly, to the wider public who pay premiums to insure themselves against liability to pay compensation for causing personal injury.

Likewise with the premium payable for after the event insurance: if the claim is successful, the premium will be payable by the other side's liability insurers.

As a result, it was said, the new arrangements, as they are currently working, are unbalanced and unfairly prejudicial to liability insurers and the general body of motorists whose insurance policy premiums provide the money with which liability insurers meet these personal injuries claims and costs.
<table>
<thead>
<tr>
<th>Sentence</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Неизпълнението на задължението за своевременно начисляване на данъка, води единствено до възникване на задължение за заплата на лихви, както и за реализиране на административно-наказателна отговорност на данъчния субект, но не може да бъде предоставка за непризнаване на данъчен кредит.</td>
<td>Useful</td>
</tr>
<tr>
<td>Договорът за предоставяне на услуги е такъв за доставки с периодично и непрекъснато изпълнение.</td>
<td>Useful</td>
</tr>
<tr>
<td>Производството е по реда на чл. 156 и сл. от Данъчно-осигурителния процесуален кодекс /ДОПК/.</td>
<td>Useful</td>
</tr>
<tr>
<td>- главен инспектор по приходите при ТД на НАП - гр. Хасково, потвърден с Решение № 747/15.10.2008г. на Директор на дирекция &quot;Обжалване и управление на изпълнението&quot; - гр. Пловдив при Централно управление на Национална агенция за приходите в обжалваната част за непризнато право на данъчен кредит в размер на 87237 лв. и прилежащи лихви за забава в размер на 38930, 45 лв.</td>
<td>Not useful</td>
</tr>
<tr>
<td>В тази връзка се твърди, че неправилно е отказано право на данъчен кредит с мотивите, че дружеството е получавало ежемесечно услуги и поради това е следвало да издава фактури за всяка една конкретна услуга на края на всеки месец, когато е прието, че е възникнал данъчното събитие по смисъла на чл. 25 от ЗДДС /отм./, във вр. с чл. 31 от ППЗДДС /отм./.</td>
<td>Useful</td>
</tr>
<tr>
<td>В допълнение към тях се позовава на нарушение на разпоредби на Дириктива 77/388 на Съвета относно хармонизиране на законодателствата на държавите-членки относно данъците върху оборота - обща система на данъка върху добавената стойност: единна данъчна основа /Шеста директива/, като се твърди, че с оглед директивата щом не е изтекъл данъчния период за събиране на данъчното задължение от страна на данъчната администрация факта на закъснение при начисляване на ДДС не следва да се третира във вреда на данъчния субект.</td>
<td>Useful</td>
</tr>
<tr>
<td>Разплащането по фактурите е извършено по банков път на 28.03.2005г. Към писмото са приложени заверени копия на следните документи - договор за предоставяне на услуги, сключен на 19.01.2004г. между &quot;Д.</td>
<td>Useful</td>
</tr>
<tr>
<td>В даденото обяснение е посочено, че служителите на дружеството са извършили услугите, упоменати в сключения договор между двете дружества, лицата притежават необходимата квалификация, същите работят по трудов договор в &quot;Д.</td>
<td>Useful</td>
</tr>
</tbody>
</table>
CONCLUSIONS

• Keywords mostly fail to describe the documents

• Nevertheless they are good indicators for useful sentences

• Utility measure is more important than precision/recall
  – Most positive feedback for English sentences (“manually selected”)
  – But lowest precision and recall
  – Low interrater agreement

• The approach works for different languages

• Clear separation of language-specific and language-independent processing

• Can also be used for document retrieval (boosting)
FUTURE WORK

• Improve keyword extraction
  – Revise natural language preprocessing
  – Add a specifically legal terminology

• Modify the scoring
  – Concluding sentences/keywords are more important
  – Include the scores of skipped concepts

• Take the structure into account
  – Akoma Ntoso
  – Try to detect important paragraphs using TextRank
  – Identifying important parts with a specifically legal terminology

• Increase recall of missed keywords and sentences

• Automatic evaluation with standardized metrics and datasets
ACKNOWLEDGMENTS

• Parts of this work have been supported by the European Commission under the 7th Framework Programme through the project *EUCases* - EUropean and National CASE Law and Legislation Linked in Open Data Stack (grant agreement no.~611760).

• Thanks to all legal experts for their effort in finishing the questionnaires!
Thanks for your attention!