

Jakub Harašta / Jaromír Šavelka / František Kasl / Jakub Míšek

Automatic Segmentation of Czech Court Decisions into Multi-Paragraph Parts

The authors describe a tool for automatic segmentation of the Czech top-tier court decisions (Supreme Court, Supreme Administrative Court, and Constitutional Court) into multi-paragraph parts. The tool allows segmenting a decision into Header, Party Response, Proceeding Summary, Court Argumentation, Footer, Dissent, and Footnotes. Segmenting text into multi-paragraph parts allows to treat different parts differently even when they contain similar linguistic or other features. Eventually, this is useful in data processing pipelines, as this tool is planned for use in automatic reference recognition purposes.

Category: Articles

Region: Czech Republic

Field of law: LegalTech; E-Justice

Citation: Jakub Harašta / Jaromír Šavelka / František Kasl / Jakub Míšek, Automatic Segmentation of Czech Court Decisions into Multi-Paragraph Parts, in: Jusletter IT 23 May 2019

Table of contents

1. Introduction
2. Related Work
3. Task
4. Dataset
5. Results and Discussion
 - 5.1. Automatic Segmentation
 - 5.2. Baselines
 - 5.3. Results
6. Future Work and Conclusion
 - 6.1. Amount of Annotated Documents
 - 6.2. Reference Recognition Pipeline
7. Contribution Statement & Acknowledgment
8. References

1. Introduction

[Rz 1] In this paper we present a tool for automatic segmentation of the Czech top-tier court decisions (Supreme Court, Supreme Administrative Court, and Constitutional Court) into multi-paragraph subtopic parts. The tool allows segmenting court decision into parts presenting a specific part of the decision. The tool distinguishes between Header, Party Response, Proceeding Summary, Court Argumentation, Footer, Dissent, and Footnotes. We describe these categories in detail in Section 3. We annotated these categories into 350 top-tier court decisions, and we provide detailed account of resulting dataset in Section 4. Section 5 presents and discusses validated results of automatic segmentation using Conditional Random Fields. Section 6 concludes the paper by focusing on possible further use of segmentation tools.

[Rz 2] Our motivation to engage in this task is twofold. First, a large volume of legal information is available in unstructured form, which makes its processing a challenging task for computers. Second, both academics and practitioners alike often mention certain tasks stemming from their work with unstructured information as time-consuming and tedious. In the past, SCHWEIGHOFER called for generic tools allowing for document segmentation, mentioning arguments and ruling as desirable recognition of facts (SCHWEIGHOFER 2015).

[Rz 3] Therefore, readily available segmentation of documents could be beneficial to both humans and computers. Humans could benefit from using color-coded segments for easier work with large unstructured texts, while computer applications could benefit from providing better structure for legal texts as a prerequisite for further applications. For example, segmentation of documents could allow us to treat different parts of documents differently for further analysis (reference recognition, argument mining). Similar linguistic features or text patterns carry different meanings and segmentation allows us to unravel those more precisely.

2. Related Work

[Rz 4] Topic-based segmentation focuses on finding sentences that are borderlines of subtopical segments. There are three main methods for identifying these sentences (PRINCE 2007).

[Rz 5] Similarity-based methods measure proximity between sentences by using cosine of the angle between vectors representing sentences (CHOI 2001, HEINONEN 1998) with pivotal work

of HEARST (1997). Graphical method represent term frequencies graphically and use representations to identify subtopical segments. Most common is the dotplotting algorithm introduced by REYNAR (1994) and further refined by YE (2005). Finally, Lexical-chain-based methods link multiple occurrences of a term (BARZILAY 1997, ERCAN 2007, SILBER 2000).

[Rz 6] Understandably, automatic segmentation does not present an individual goal. It always serves as a prerequisite for further tasks that require structured data. Segmentation is required for text summarization (BARZILAY 1997, HEARST 1997), keyword extraction (ERCAN 2007), audiovisual (GALUÁKOVÁ 2012) and textual information retrieval (PRINCE 2007), and other tasks requiring structured data input. The segmentation task specifically targeted to legal domain appeared as part of the pipeline for legal document summarization in works of FARZINDAR (2004) and SARAVANAN (2006, 2010). Following the work of LAFFERTY (2001), SARAVANAN used Conditional Random Fields for text segmentation.

[Rz 7] Our work differs from previous work in terms of the predefined segments we look for in the document. The segmentation tool then does not search for any borderline sentences, but for borderline sentences between predefined and known segments. These segments and *ratio* for their existence is described in Section 3.

3. Task

[Rz 8] We focus on segmentation of court decisions by the three top-tier courts in the Czech Republic – Supreme Court, Supreme Administrative Court, and Constitutional Court. We focus on these courts because their decision are publicly available. Moreover, importance of their decisions extends beyond individual cases.

[Rz 9] Trying to understand the structure and textual units in court decisions has been an important part of GROVER (2004) and SARAVANAN (2010). Both authors segment court decisions into coherent chunks of text based on the rhetorical roles played by the individual sentences. Labels are shown in Table 1 for GROVER (2004) and in Table 2 for SARAVANAN (2010).

[Rz 10] However, both works focused on court decision from common-law systems, and aimed for automatic summarization. Therefore, labels are fine-grained and tuned to common law, where it is crucial to tie in facts together with used legal acts and precedents.

[Rz 11] Our own work focuses on court decision from the continental legal system, where careful distinguishing of facts plays virtually no role whatsoever. Moreover, our aim is not to achieve summarization of the court decision, but use this tool in a pipeline for information extraction. We therefore differ from previous work of GROVER (2004) and SARAVANAN (2010) in two important regards. First, our labels do not strictly follow rhetorical roles, but mix rhetorical roles with structural features of the text (eg. footnotes). Second, our label system does not categorize individual sentences, but focuses on classifying paragraph-sized chunks of text, because these are present in the Czech court decisions in largely uninterrupted manner.

Label	Description
Fact	The sentence recounts the events or circumstances, which gave rise to legal proceedings.
Proceedings	The sentence describes legal proceedings taken in the lower courts.
Background	The sentence is a direct quotation or citation of source of law material.
Framing	The sentence is part of the law lord's argumentation.
Disposal	A sentence which either credits or discredits a claim or previous ruling.
Textual	A sentence which has to do with the structure of the document or with things unrelated to a case.
Other	A sentence which does not fit any of the above categories

Table 1: Rhetorical Annotation Scheme for Court Decisions according to GROVER (2004).

Label	Description
Identifying the case	Identification of issues to be decided for a case.
Establishing facts of the case	Facts relevant to the present proceedings that are required for proper applications of correct legal principles and laws.
Arguing the case	Application of legal principles and laws advocated by contending parties.
History of the case	Chronology of events with factual details that led to the present case.
Argument (analysis)	The court discussion on the law that is applicable to the set of proved facts.
<i>Ratio decidendi</i> (ratio of the decision)	Central generic reference of text where court applies the correct law to a set of facts.
Final decision (disposal)	Ultimate decision or conclusion of the court.

Table 2: Rhetorical Annotation Scheme for Court Decisions according to SARAVANAN (2010).

Label	Description
Header	Identification of participants, brief case description, and ruling of court in the case.
History	Procedural history containing summary of previous court decisions.
Submission / Rejoinder	Submission of petitioner, and rejoinder of opposing party.
Argumentation	Court's argumentation about the case.
Footer	Legally required information on further appeal or calculation of legal costs.
Dissent	Dissenting opinion of single or multiple judges.
Footnotes	Footnotes containing additional information or references.

Table 3: Current version of annotation scheme for automatic segmentation of Czech top-tier court decisions.

[Rz 12] We have established a pattern that predominantly appears throughout Czech top-tier court decisions (for brief description see Table 3). Detailed description of individual labels follows:

- **Header** is an introduction of the given case. It contains identification of the parties, their legal representatives, deciding judge (or judges), and the ruling. Headers are generally very structured and appear at the beginning of the decision. Certain deviations appear, because decisions were made available in 1990s in heavily edited form.
- **History** contains the procedural history. Since we are concerned with the top-tier court decisions, there is always a part that contains procedural history. Decisions of lower court prior the appeal (in case of Supreme Court), cassation (Supreme Administrative Court) or constitutional complaint (Constitutional Court) are summarized in different degree of details. It provides us with facts established by lower courts, with their legal assessment of events, and with verdict.
- **Submission/Rejoinder** summarizes the petition that led to the top-tier court hearing the case. It contains petitioner's arguments (being it plaintiff or defendant), response of lower court to the petition, arguments of other parties (*amicus curiae*), and other relevant bodies (administrative bodies, Parliament in case of constitutional review of legal acts etc.). Certain deviations appear, as in the 1990s and early 2000s, some judges used multiple single- or multi-paragraph segments dispersed throughout the decision, instead of more frequent single chunk of text.
- **Argumentation** contains the argumentation of the court hearing the case. This part is usually the most important for legal professionals or students reading the decision. It is also often the longest. It usually appears as a single multiple-paragraph segment. Deviations appear when single- or multi-paragraph segments of other types might interrupt it. This often happens when the court wants to tie together the argumentation of the parties or the lower instance courts together with its own reasoning.

- **Footer** contains predefined and legally required information on the possibility of further proceedings. It contains information for the parties on possible further appeal and often contains the calculation of the costs of legal representation.
- **Dissent** appears only when allowed by the procedural rules. It contains the argumentation of judges not agreeing with the majority opinion. This is very frequent for Constitutional Court, but could appear in decisions of all three courts.
- **Footnotes** occur rarely, as references to case law and literature and other further information are mostly contained in the decision itself. However, some judges in 2010s started to use footnotes to make the decisions more «clear» and more akin to doctrinal works, such as academic papers.

[Rz 13] We identified these parts based on our study of the Czech case law. With the exception of specific procedures (e.g. the authorization of wiretapping or the order of enforcement of a decision), court decisions generally follow the structure above. Annotated parts often appear in specific order and often appear as uninterrupted multi-paragraph text spans (with deviations outlined above). The annotators manually annotated 350 top-tier court decisions as is explained in detail in Section 4.

4. Dataset

[Rz 14] Our dataset consists of 350 decision of the Czech top-tier courts: 160 decision of Supreme Court, 115 decision of Supreme Administrative Court and 75 decision of Constitutional Court. The shortest decision has 4,746 characters, while the longest has 537,470 characters with average length of 36,148 characters.

[Rz 15] Out of 350 annotated decisions, 280 were annotated by a single annotator and then curated by a single editor. 70 additional documents were annotated by two annotators, and then curated by a single editor. The editorial layer was added to the annotation task to ensure high quality of data. All participating annotators and editors are listed as co-authors of this paper, and an editor could not be assigned with a document that he himself has annotated. Gold standard dataset containing all the correct annotations resulting from curation is available for download via LINDAT/CLARIN repository at <http://hdl.handle.net/11372/LRT-2901>.

[Rz 16] Table 4 reports the inter-annotator agreement on the 70 double-annotated decisions. We report the agreement as the percentage of annotations where annotators agreed completely on assigned labels, including the first and last characters of any annotation. Reported IAA for all labels suggested the annotation task is reasonably easy with the exception of the label Footer. Error analysis of the low IAA for this label suggest the task was not particularly difficult, but improperly defined.

Label	IAA
Header	.96
History	.82
Submission/Rejoinder	.81
Argumentation	.95
Footer	.49
Dissent	.94
Footnotes	1.00

Table 4: Inter-Annotator Agreement for different labels over 70 double-annotated decisions.

[Rz 17] We annotated documents with the types explained in Section 3. We did not annotate sentences, but focused on paragraphs. Amount of labels of individual types used over the whole dataset is available in the second column of Table 5, while average percentage of document covered by annotation type (ratio of annotation size to document size) is available in the third column of Table 5.

[Rz 18] We used described dataset for machine learning and evaluation for automatic segmentation task. Results are discussed in Section 5.

Label	Amount of annotations in dataset	Average ratio of annotation size to document size
Header	349	.024
History	392	.145
Submission/Rejoinder	384	.188
Argumentation	399	.584
Footer	332	.023
Dissent	19	.034
Footnotes	6	.001

Table 5: Overall statistics of our dataset including number of annotations per label and average part of document covered by annotations per label.

5. Results and Discussion

5.1. Automatic Segmentation

[Rz 19] We attempted to automatically segment the text into pre-defined parts explained in Section 3. As a prediction model, we use first/order conditional random fields (CRF).¹ A CRF is a random field model that is globally conditioned on an observation sequence O . The states of the model correspond to event labels E . We use a first-order CRF in our experiments (observation O_i is associated with E_i) (LAFERTY 2001, OKAZAKI 2007). We train a CRF model for each of the 7 labels.

¹ We use the CRFSuite implementation of first-order CRF which is available at www.chokkan.org/software/crfsuite/.

[Rz 20] In tokenization we consider an individual token to be any consecutive sequence of either letters, numbers or whitespace. Each character that does not belong to any of these constitutes a single token. Each of the tokens is then a data point in a sequence a CRF model operates on. Each token is represented by a small set of relatively simple features. Specifically, the set includes:

- *position* – position of a token within a document.
- *lower* – a token in lower case.
- *stem* and *aggressive stem* – two types of token stems²
- *sig* – a feature representing a signature of a token.
- *length* – token’s length.
- *islower* – true if all the token characters are in lower case.
- *isupper* – true if all the token characters are in upper case.
- *istitle* – true if only the first of the token characters is in upper case.
- *isdigit* – true if all the token characters are digits.
- *isspace* – true if all the token characters are whitespace.

[Rz 21] For each token we also include *lower*, *stem* and *aggressive stem*, *sig*, *islower*, *isupper*, *istitle*, *isdigit*, and *isspace* features from the five preceding and five following tokens. If one of these tokens falls beyond the document boundaries we signal this by including *BOS* (beginning of sequence) and *EOS* (end of sequence) features.

5.2. Baselines

[Rz 22] It is required to set up a baseline performance for automatic segmentation. We set baselines by use of the two following methods:

- Zero Rule Algorithm
- Average Distribution Algorithm

[Rz 23] The Zero Rule Algorithm labels every part of the document with label for majority class. Given the percentage of dataset covered by individual labels, it is apparent from the last column of Table 5 that majority class is the label Argumentation covering .584 of our dataset. Hence, by labelling every part of document with the label for Argumentation, we achieve a precision of .584, recall of 1.0 and F1 measure of .737. This is our first baseline to evaluate performance of our model.

[Rz 24] The Average Distribution Algorithm labels individual documents by average distribution of labels throughout the whole dataset. Labels are distributed in accordance with the third column of Table 5. By using this algorithm, we achieved precision, recall and F1 measure for individual labels as reported in the fifth, sixth and seventh column of Table 6.

² A stemmer for Czech implemented in Python by Luís Gomes was used for stemming. The stemmer is available at http://research.variancia.com/czech_stemmer/.

5.3. Results

[Rz 25] To evaluate the performance, we use a 10-fold crossvalidation. Table 6 summarizes the results of the experiment. The second to seventh column provide precision, recall and F1 measure for both baselines elaborated in Section 5.2. The eighth to tenth column provide precision, recall and F1 measure for automatic segmentation via our model.

[Rz 26] As is evident from Table 6, our model outperforms the two baselines consistently across all labels. Our model achieves F1 measure of:

- .959 for Header (outperforming .703 provided by Average Prediction Baseline)
- .737 for History (.518 by Average Prediction Baseline)
- .734 for Submission (.433 by Average Prediction Baseline)
- .915 for Argumentation (.737 by Zero Rule Baseline and .840 by Average Prediction Baseline)
- .803 for Footer (.129 by Average Prediction Baseline)
- .523 for Dissent (.170 by Average Prediction Baseline)
- .250 for Footnotes (values not available for either baseline).

Label	Zero Rule Baseline			Avg. Prediction Baseline			Automatic Segmentation		
	P	R	F1	P	R	F1	P	R	F1
Header				.703	.703	.703	.969	.950	.959
History				.518	.518	.518	.764	.712	.737
Submission/Rejoinder				.433	.432	.433	.768	.703	.734
Argumentation	.584	1.0	.737	.841	.840	.840	.885	.950	.915
Footer				.129	.129	.129	.845	.765	.803
Dissent				.170	.170	.170	.693	.420	.523
Footnotes							.988	.140	.250

Table 6: Results containing two baselines and performance of our model.

6. Future Work and Conclusion

[Rz 27] Future Work contains two distinct tasks. The first is to annotate more documents to account for low representation of some labels in the dataset. The second is to include this tool into our reference recognition pipeline, once we are content with results.

6.1. Amount of Annotated Documents

[Rz 28] As is reported in the second column of Table 5 some labels are underrepresented in the annotated dataset. Header, History, Submission/Rejoinder, Argumentation, and Footer appear regularly throughout the whole dataset. On the other hand, rare labels, such as Dissent and Footnotes, appear only in 19, respectively 6, instances. Inclusion of more documents to fix this disparity and underrepresentation of certain labels in dataset is therefore advisable.

6.2. Reference Recognition Pipeline

[Rz 29] We intent to add this tool to our automatic reference recognition pipeline. The references to other court decisions or literature appear frequently in various part of the court decisions. Importantly, references contained in the summary of previous proceedings or brought up by the parties in their petitions have different meaning than references in the court argumentation. We believe that only references made by the court while arguing the case are to be considered as relevant from the perspective of the case law. Overall, it is very difficult to decide automatically which references to capture. As such, automatic segmentation into multi-paragraph passages allows us to treat different parts of the text differently, despite their syntactic similarity.

7. Contribution Statement & Acknowledgment

J.H. and J.Š. developed the annotation scheme. J.H., J.Š., and F.K. annotated decisions. J.H., F.K., and J.M. curated/edited the decisions. J.Š. programmed the annotation environment, and prepared dataset statistics. J.H. and J.Š. wrote the paper with input from all authors.

We presented earlier versions of this work at *Legal Data Analysis 2017* workshop held in conjunction with JURIX2017 in Luxembourg, and at 2nd MET-ARG workshop within WAW2018 in Warsaw.

J.H., F.K., and J.M. gratefully acknowledge the support from the Czech Science Foundation under grant no. GA-17-20645S.

8. References

REGINA BARZILAY, MICHAEL ELHADAD, Using Lexical Chains for Text Summarization. Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization, pp. 10–17 (1997).

FREDDY CHOI, PETER WIEMER-HASTINGS, JOHANNA MOORE, Latent Semantic Analysis for Text Segmentation. Proceedings of the 6th Conference on EMNLP, pp. 109–117 (2001).

GONENC ERCAN, ILYAS CICEKLI, Using Lexical Chains for Keyword Extraction. *Information Processing & Management*, 2007, vol. 43, no. 6, pp. 1705–1714.

ATEFEH FARZINDAR, GUY LAPALME, LetSum, an automatic legal text summarizing system. Proceedings of Jurix, pp. 11–18 (2004).

PETRA GALUÁKOVÁ, Application of Topic Segmentation in Audiovisual Information Retrieval. WDS'12 Proceedings of Contributed Papers: Part I, pp. 118–122 (2012).

CLAIRE GROVER, BEN HACHEY, IAN HUGHSON, The HOLJ Corpus: Supporting Summarisation of Legal Texts. Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora, pp. 47–54 (2004).

MARTI A. HEARST, TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages. *Computational Linguistics*, 1997, vol. 23, no. 1, pp. 33–64.

OSKARI HEINONEN, Optimal Multi-Paragraph Text Segmentation by Dynamic Programming. *Proceedings of the 17th COLING*, pp. 1484–1486 (1998).

JOHN LAFFERTY, ANDREW MCCALLUM, FERNANDO C.N. PEREIRA, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Proceedings of International Conference on Machine Learning*, pp. 282–289 (2001).

NAOAKI OKAZAKI, CRFSuite: a fast implementation of Conditional Random Fields (2007), <http://www.chokkan.org/software/crfsuite/>.

VIOLAINE PRINCE, ALEXANDRE LABADIÉ, Text Segmentation Based on Document Understanding for Information Retrieval. *Proceedings of 12th International Conference on Applications of Natural Language to Information Systems*, pp. 295–304 (2007).

JEFFREY C. REYNAR, An Automatic Method of Finding Topic Boundaries. *Proceedings of ACL'94 (student session)*, pp. 331–333 (1994).

MURALI SARAVANAN, BALARAMAN RAVINDRAN, Identification of Rhetorical Roles for Segmentation and Summarization of a Legal Judgment. *Artificial Intelligence and Law*, 2010, vol. 18, no. 1, pp. 45–76.

MURALI SARAVANAN, BALARAMAN RAVINDRAN, SHIVANI RAMAN, Improving Legal Document Summarization Using Graphical Models. *Proceedings of Jurix*, pp. 51–60 (2006).

ERICH SCHWEIGHOFER, The Role of AI & Law in Legal Data Science. *Proceedings of Jurix*, pp. 191–192 (2015).

H. GREGORY SILBER, KATHLEEN F. MCCOY, Efficient Text Summarization Using Lexical Chains. *Proceedings of the 5th International Conference on Intelligent User Interfaces*, pp. 252–255 (2000).

NA YE, JINGBO ZHU, HAITAO LUO, HUIZHEN WANG, BIN ZHANG, Improvement of the Dotplotting Method for Linear Text Segmentation. *Proceedings of Natural Language Processing and Knowledge Engineering*, pp. 636–641 (2005).

JAKUB HARAŠTA is the corresponding author (at jakub.harasta@law.muni.cz) and works with FRANTIŠEK KASL, and JAKUB MÍŠEK at the Institute of Law and Technology, Faculty of Law, Masaryk University, Czech Republic. JAROMÍR ŠAVELKA is a graduate student at the Intelligent Systems Program, University of Pittsburgh, USA.