D1 – End-users´ survey

Project funded by the European Union's Next Generation EU through the Slovak Republic's Recovery and Resilience Plan under project no. 09I05-03-V02-00049.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU
VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA
VÝSKUMNÁ AGENTÚRA

Introduction

The project **"Automated Legal Text Analysis based on Machine Learning"** (hereinafter "ALTAML") is a key initiative aimed at integrating innovative approaches in the field of data processing and subsequent analysis, with a focus on the field of law, which also includes information and communication technology law. The aim of this project is to develop and validate effective methods for the automated analysis of legal texts using machine learning techniques. This involves, in particular, the development of tools that facilitate the processing and analysis of large volumes of data in the form of various legal documents, as well as the extraction of relevant information (attributes) from such documents, including the identification of key terms, references to legislation and other attributes.

The aim of the ALTAML project is therefore to contribute to more efficient access to legal information and to the acceleration of legal processes, thereby ensuring a higher degree of legal certainty and improving the availability of legal texts, the results of their analysis and relevant legal information for both professionals and the general public.

As part of work package KPB1, deliverable **D1 - End-users´ survey** - summarising the requirements resulting from the survey of entities working with legal texts and identifying relevant needs and requirements

Contents

Introduction **1**

1. Introduction to legal research **3**

2. The method of systematic contextual analysis of law and its individual stages **4**

3. Methodology for the preparation of the evaluation report **5**

4. Legal teams and the formulation of appropriate research questions/hypotheses .. Chyba! Záložka nie je definovaná.

5. Identification and collection of a suitable data sample..... Chyba! Záložka nie je definovaná.

6. Coding of legal text..... Chyba! Záložka nie je definovaná.

7. Formulation of conclusions or observations based on data analysis..... Chyba! Záložka nie je definovaná.

7.1 Descriptive statistical analysis ..... ***Chyba! Záložka nie je definovaná.***

7.2 Analytical method – Comparative analysis of groups ..... ***Chyba! Záložka nie je definovaná.***

7.3 Analytical method – Correlation analysis between attributes..... ***Chyba! Záložka nie je definovaná.***

7.4 Analytical method – Cluster analysis ..... ***Chyba! Záložka nie je definovaná.***

7.5 Analytical method – Anomaly detection ..... ***Chyba! Záložka nie je definovaná.***

7.6 Analytical method – Association rules ..... ***Chyba! Záložka nie je definovaná.***

7.7 Analytical method – Predictive analysis..... ***Chyba! Záložka nie je definovaná.***

8. Reporting on the results achieved in a manner comprehensible to the legal community .... Chyba! Záložka nie je definovaná.

Conclusion **7**

Appendix 1 ..... Chyba! Záložka nie je definovaná.

1. Introduction to legal research

XXXX Defining the methodological approach and specific methodological procedures used in the research is an essential part of any scientific investigation. Despite this, one may

2. The method of systematic content analysis of law and its individual stages

The method of systematic content analysis of law comprises several interrelated stages. Hall and Wright distinguish three specific components of this research method, namely

1) the selection of court decisions,
2) the coding of court decisions and
3) analysis of coded court decisions, often using statistical methods.[1]

processing depends on the choice of method. Unlike the original proposal by Salehijam, we have added a phase of data analysis and machine learning. The results from this phase are subsequently interpreted, formulated into conclusions and presented to both specialists and the general public.

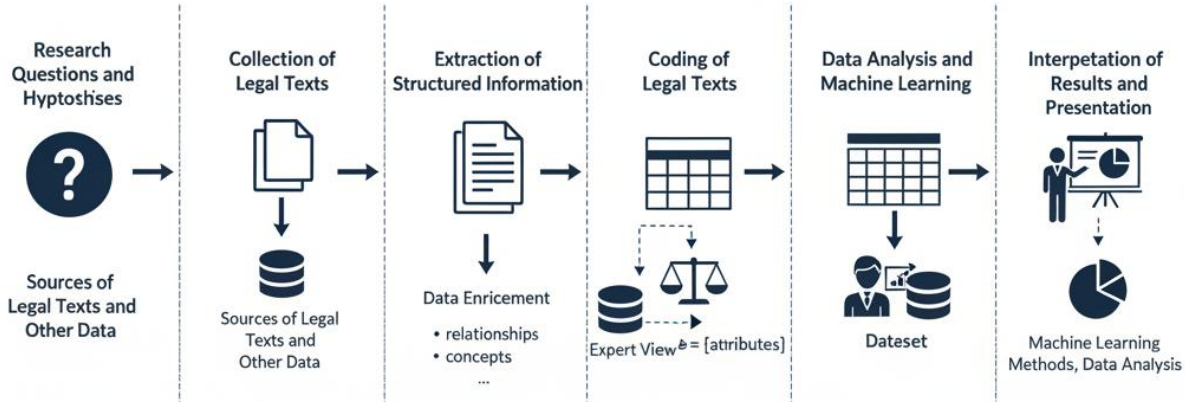


Figure1 Visualisation of systematic content analysis of law

[1] Ibid., p. 79.

3. Methodology for the preparation of the evaluation report

In this chapter, we outline the methodological procedure for conducting legal research and for the actual preparation of this evaluation report. This process consisted of the following steps, which we will examine in more detail in the following chapters:

1) formation of research teams and definition of research questions,
2) rough and fine-grained definition of the set of decisions,
3) determining attributes for coding decisions,
4) the creation of partial analyses of court decisions, and
5) feedback from the legal research teams.

1. Formation of research teams and definition of research questions

XXXx

2. Rough and fine-grained identification of the set of decisions

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXxx

analysis are usable in legal research, where their limitations lie, and what would need to be changed for future projects. The questionnaire was divided into seven thematic areas:

- **Context and background of the research** – questions exploring what research questions legal research teams were pursuing before the introduction of automated methods, what their previous methodology looked like, and what limitations arose from the content of available legal texts.
- **Attribute coding and dataset creation** – this section focuses on the extent to which the selected attributes corresponded to legal reality and established terminology, whether certain categories were incorrect, unclear or misleading, and what changes research team members would propose for the future.
- **Evaluation of individual analytical outputs** – for each of the analytical methods used, respondents assessed the substantive contribution and limitations of the output (in the form of an open-ended response), as well as its comprehensibility and usefulness on a five-point scale. The aim of this section is to identify which

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU | VAIA | VÝSKUMNÁ AGENTÚRA

methods are intuitively graspable for members of legal research teams and which require additional interpretative support.

- **Interpretation of results and connection to legal argumentation –** the questions examine whether and how the outputs can be directly integrated into legal text or argumentation, where the boundary lies between what is statistically interesting and what is legally relevant, and what role quantitative outputs play in traditional normative legal analysis.
- **Limitations of the method and manual coding** – this section focuses on identifying aspects of legal texts that are difficult or impossible to automate, on specific instances of misleading or incorrect conclusions drawn by automated analysis, and on assessing the consistency of coding across team members (inter-rater reliability).
- **Prospects and recommendations** – respondents proposed specific changes to the research process, identified legal areas with the greatest potential for this methodology, and commented on the use of large language models (LLMs). They also assessed whether they would be able to use the method independently without the support of a data analyst.
- **Concluding remarks** – an open space for any additional insights not covered by the previous questions, including ethical aspects, institutional barriers or a vision for the future development of the method.

The full text of the questionnaire is included in Appendix 1 of this evaluation report.

Conclusion

The method of systematic content analysis of law enables a different approach to legal research involving the analysis of legal texts. The distinctive feature of this method lies

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU
VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA
VÝSKUMNÁ AGENTÚRA