A Methodology for Evaluating Advanced Legal Data Infrastructures

Shefali Virkar*
Department for E-Governance and Administration, Danube University Krems, Austria
shefali.virkar@donau-uni.ac.at

Euripidis Loukis
Department of Information and Communication Systems Engineering, University of the Aegean, Greece
eloukis@aegean.gr

Charalampos Alexopoulos*
Department of Information and Communication Systems Engineering, University of the Aegean, Greece
alexop@aegean.gr

ABSTRACT

*Legal information is of paramount importance for social, economic and political activities. Sophisticated legal information retrieval (LIR) systems are designed and built to facilitate user access to accurate and current legal information. The advent of new types of capabilities in the legal information systems domain has necessitated a closer look at the performance and usability of such advanced computer systems, and brought into sharp focus the requirement for an effective evaluation of the available functionalities and quality of the software. The purpose of this research paper is to develop a framework for the effective monitoring and evaluation of an advanced legal information system from differing salient perspectives, and to critically examine the basic set of attributes that characterize the important dimensions of platform functionality. Through the application of the developed evaluation model to a novel legal data platform, we propose both a rigorous validation of the framework itself, together with the identification of the strengths and shortcomings associated with the use of the system under consideration.

CCS CONCEPTS

• **Applied computing**; • **Computers in other domains**; • **Computing in government**; • **E-government**;

KEYWORDS

open data, legal data, evaluation methodology, usability

ACM Reference Format:
Shefali Virkar*, Euripidis Loukis, and Charalampos Alexopoulos. 2021. A Methodology for Evaluating Advanced Legal Data Infrastructures. In *14th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2021), October 06–08, 2021, Athens, Greece.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3494193.3494205

*Senior Research Associate, Department for E-Governance and Administration, Danube University Krems, Dr-Karl-Dorrek-Strasse 30, 3500 Krems an der Donau, Austria.

ICEGOV 2021, October 06–08, 2021, Athens, Greece

ACM ISBN 978-1-4503-9011-8/21/10…$15.00
https://doi.org/10.1145/3494193.3494205

1 INTRODUCTION

The critical nature of legal information for social, economic and political activities necessitates comprehensive capabilities for accessing such information to be provided to citizens, businesses and governments. Usually, this need for access to laws is addressed by traditional portals providing basic search and retrieve functionality. However, the European Union (EU) vision of a unique market, and in general the internationalization of economic activity, pose additional and more complex requirements concerning access to legal information. Only recently, these advanced capabilities are starting to be developed and offered. Some more sophisticated systems are developed that provide new functionality in the legal information and analytics domain, while most of them are proprietary and limited to one country's legislation [1].

Nevertheless, the advent of this new type of capabilities in the legal information systems domain needs careful consideration and analysis. The aim of a sophisticated information retrieval (IR) system is to facilitate the transfer of the right information in the right format at the right time to the right user population. In order to appropriately build and maintain this type of system, an effective evaluation of the available functionalities and quality of the software is required, which enables the assessment and identification of the strengths and weaknesses that have to be addressed.

The purpose of this paper is to develop a framework for the monitoring and evaluation of an advanced legal information systems from differing salient perspectives; one that provides a basic set of attributes to characterize the important dimensions of platform functionality through the explicit linking of process and product aspects with the ultimate utility of the system. It is widely recognized that the performance of an information retrieval system depends on the correct identification and fulfilment of user requirements, particularly those concerned with the need for information access; and then on the effective evaluation of the degree of the system meeting these requirements. In this paper we develop such an evaluation methodology, and apply it to an advanced legal data infrastructure developed as part of a European project [2] in order to validate it and demonstrate how it can be successfully deployed for a quantitative assessment of the performance and the response quality of a fully functioning legal information system.

Monitoring may be defined as "…the systematic process of collecting, analyzing and using information to track a programme's progress toward reaching its objectives and to guide management decisions" [3]. When taken as a process, monitoring may be considered as a continuous function that involves the systematic collection of data based on specified indicators, to provide management and the main stakeholders of an ongoing programme or intervention

with an indication of the extent of progress and the achievement of objectives [4]. Effective monitoring is, therefore, an ongoing, continuous process that requires the collection of data at multiple points throughout the program cycle, including at the beginning, to provide a baseline. In consequence, it can be used to determine whether activities need adjustment during the intervention to improve desired outcomes [5].

Evaluation, on the other hand, is the process of measuring the outcomes from a near-complete or completed project in order to assess the extent to which it is achieving its intended purpose. Evaluations may be considered under two broad categories: (a) formative and (b) summative [6]. Based on the above definitions it may be inferred that the former type of assessment is essentially concerned with the monitoring of ongoing critical project activities and procedures, while the latter focuses on assessing the output and outcome of a project against expected results at the close of all activity.

Markiewicz (2018) notes that a rigorous theory-based and evaluation-led monitoring and evaluation framework will enable the mapping out of the programme theory and programme logic, and allow for project managers to use these conceptual models to guide the development of a set of key evaluation questions (KEQs) and performance indicators [7]. In this context, Key Evaluation Questions refer to the high-level questions that an evaluation is designed to answer, and are developed by considering the type of evaluation being done, its intended users, its intended uses (purposes), and the evaluative criteria being used [8]. For the purposes of monitoring and evaluating the development of an advanced computer system, the following research questions are asked: (i) What is the problem that needs to be solved?; (ii) How does the software system address the identified problem?; (iii) What criteria will be used to determine the system's success or failure?;(iv) Does the design of the system allow for meaningful evaluation?; and (v) Are there enough units of study to ensure statistical significance?.

This paper is structured as follows: section 2 presents the advanced functionality of the legal data infrastructures used to build the basic metrics for our model; section 3 elaborates on the current evaluation theories based on which our evaluation methodology has been developed; section 4 describes the research approach and the derived evaluation methodology; section 5 presents the results from the abovementioned application of the developed methodology, while section 6 concludes the paper.

2 ADVANCED LEGAL DATA INFRASTRUCTURES

The main vision of the European Union is the establishment of a well-functioning Digital Single Market (DSM), wherein European citizens can move freely and trade with their counterparts in other EU member states [9]. The adoption of the DSM strategy in 2015 can be considered as a significant step towards the development of a European Single Market based upon seamless access to digital information, tools and services across different environments within the economy and society [10]. Unhindered access to different types of information, and especially legal data pertaining to different European member states, can be considered vital for achieving this vision of seamless cross-border mobility and trade. However, the European legal landscape is multi-layered and complex, with a large quantity of legal documentation produced since its inception currently remaining fragmented across multiple national databases, hidden within inaccessible systems, and scattered across public data silos [1]. Moreover, although present in abundance, unfettered access and comprehension of legal information remains the preserve of a handful of legal experts who possess the specialist theoretical and practical knowledge required to retrieve, follow, and comprehend the latest legislation and policy evolutions produced by ministries, parliaments, and courts at different levels of government [11].

Advanced legal information retrieval systems (LIRs) developed to facilitate the accurate search and retrieval of different sorts of legal texts - including legislation, case law, policy documents, expert commentary, and scholarly works – offer the opportunity to make legal information more easily accessible and comprehensible to actors in society. In response to the challenges posed to universal access to legal information inherent in the European legal system, the ManyLaws project upon which this paper is based initially investigated the requirements that an advanced legal data infrastructure has to fulfill, and then has developed a suite of user-centric services that will meet these requirements, and ensure the real-time provision and visualization of cross-country and multilingual legal information to citizens, lawyers, businesses and administrations. The system under consideration is based on a platform supported by the proper environment for semantically annotated Big Open Linked Legal Data (BOLLD). The ultimate objective is to provide the technical foundation and the tools for the development of a second-generation legal data e-infrastructure, making cross-country and multilingual legal information available to everybody, in a customizable, structured and easy-to-handle way, as well as all the required processing and analysis of it in order to become practically manageable and highly useful.

From the abovementioned requirements' analysis, it has been concluded that the services/capabilities that have to be offered by an advanced legal data infrastructure are:

- Parallel search in multiple EU member-state legal frameworks (including European legislation or EU directives). This process will be effectuated through the parallel translation of queried search terms, using a suitable legal vocabulary.
- Different language environment. The portal will be available in three different languages (English, Greek, German). By language environment, we are referring to general functions of the UI, as the menu, the filter names, the legends of the charts etc., but not the legal content.
- Translations of legal texts. For each law, an automated English translation, developed from the e-Translation DSI, will be provided to the user.
- Comparative analysis of related/connected laws from the same national legal framework. This will be a text visualization and will give the user the ability to identify correlations, dependencies and conflicts between different laws.

- Timeline analysis for all legal elements. This functionality will provide a visualization of the progress and current status of a specific piece of national legislation (after amendment/extensions) over time, including preparatory acts and agreements.
- Graph-based visualization based on different correlation types. The user will have access to a representation of the relations between the selected law and other laws. This visualization will display a main graph representation where all the available relations for the selected law are included.
- Editing functionality on the graph. An additional editing functionality will allow the user to delete nodes from the graph, if they are not needed for their use.
- Visualization of the connection between an EU directive and a national legal framework. This visualization will be presented through the system as a graph, in which the connection would be clearly identified. This functionality will allow the user to assess the degree of transposition of an EU Directive into national law.
- Export of legal artefacts based on search and editing. All the relations for the selected law can be exported in a csv file.
- Authentic and consolidated versions of law (Austria only). Since, for the Austrian legal framework, the portal uses the section as base unit, an additional functionality of the aggregated version will be provided for the Austrian case.
- Parliamentary data evolution. For each law, the evolution of the parliamentary process will be provided, which is displayed through a table describing all the parliamentary steps.
- Manual annotation tool. This service aims to gather feedback for the automated text mining processes from expert users. It is a form in which users with an editor role can report inaccuracies in the automatically generated legal data.
- Public opinion service (Greece only). For each law a dedicated page regarding the public opinion will be provided. It is a table with user comments corresponding to specific parts in the legal document.

Therefore, the metrics that a methodology for the evaluation of an advanced legal data infrastructure should include concern the assessment of system performance, accuracy of results, system ease of use, quality of the system and data, and also the assessment of both the traditional and advanced (see above) capabilities it provides.

Furthermore, in the abovementioned requirements' analysis we also looked at the users of such as advanced legal data infrastructures. There are many different types of users who are potentially interested in having this improved access to legal information [11]. As a first step to investigating user requirements, we opted to construct initial use case scenarios using data collected from semi-structured interviews with five archetypal users [12]. Following from a series of interviews and focus groups, it was recognized that the top-level user categories could be nuanced as follows: Legal Professionals could encompass both practicing legal professionals and legal scholars, Government Users could be refined and expanded, and two new groups of potential users were identified - the scientific community and third sector actors. The final top-level categories of users carried over into the Requirements Engineering phase of the study were: Citizens, the Scientific Community, Legal Professionals, Businesspersons, Public Servants, and Third Sector Actors [11].

3 THEORETICAL BACKGROUND

The central purpose of this chapter is to present an organized summary of major contemporary approaches to advanced computer system evaluation.

3.1 Evaluating Complex Computer Systems

McNamara & Kirakowski (2006) argue that Functionality, Usability, and User Experience represent the three most important measurable aspects of computer system performance that need to be considered when designing and evaluating technology [13]. Their characteristics are summarized below, in Table 1.

3.2 Usability and Performance Evaluation Basics

Broadly speaking, the term 'usability' refers to the degree to which a product or service is easy to use [13]. Saket et.al. (2016) refer to the term usability as encompassing "... the effectiveness, efficiency, safety, utility, and learnability of a design" [14]. For a long time, the notion of usability has been dominated by the 'quality of use' perspective; one that considers the interaction between the user and the product, circumscribed by particular aspects collectively known as the 'context of use' [15]. Through the lens of quality of use, the concept of usability is usually taken to consist of a combination of factors, including [16]:

1. Intuitive design: a measure of the user's understanding of the architecture and navigation of the system;
2. Ease of learning: the speed at which a user who has never seen the user interface before can accomplish basic tasks;
3. Efficiency of use: the rate at which an experienced user can accomplish tasks;
4. Memorability: the degree to which a user, after using the system, can remember enough to use it effectively in future visits;
5. Error frequency and severity: the frequency at which users make errors while using the system, the degree of severity of the errors made, and the manner in which users recover from the errors;
6. Subjective satisfaction: the degree to which the user likes using the system.

More recently, the concept of usability expressed in terms of 'quality of experience' has entered the discussion, with proponents of this perspective proposing to address those aspects of technology usage - including aesthetics, image, impression, and ambience - that are hitherto omitted by the quality of use perspective [13].

Usability Evaluation, otherwise referred to as Usability Testing, assesses the extent to which an interactive computer system is easy and pleasant to use [17]. In other words, this sort of assessment focuses on the manner and degree to which users can learn and use a product to achieve their goals. The efficacy of usability evaluation methods relies on the appropriate delineation of a so-called 'usability problem' to both specify the scope of the evaluation and

Table 1: Dimensions of Advanced Computer System Performance

Dimension	Nature of Issue	Key Questions	Issues to be Investigated
Functionality	Technical	(i) What does the product do? (ii) What can the product do?	Number and type of features; usefulness of device features; product maintainability; product reliability.
Usability	User	(i) How easy is it to use the product? (ii) How easy is it to customize the product?	User support; user goal facilitation.
User Experience	Experiential	(i) How does the user feel about the product? (ii) What aspects of the product are considered important? (iii) What does the user dislike about the product?	Overall user impressions of the product; willingness to recommend product; willingness to pay a subscription fee.

Table 2: Key Relevance Dimensions of Advanced Information Retrieval

Dimension	Brief Description
Algorithmic or System Relevance	The computational relationship between a user query and the information objects contained within a system database, based on matching or a similarity between them
Topical Relevance	The relationship between the 'topic' (defined as concept and/or subject) of a user request and the information objects retrieved by the system about that topic.
Bibliographic Relevance	The relationship between a specific user query and the bibliographic closeness of the information objects retrieved by the system.
Cognitive Relevance or Pertinence	The degree of correlation between the information needs of a user as specified in a particular query and the information objects in the system, together with how these are interrelated.
Situational Relevance or Utility	The relationship - and by extension degree of correlation - between the need, problem or task of the user and the information objects in the system.
Domain Relevance	The relevance of constituent information objects within a broader, given domain context, and not merely as they relate to a work task or current problem.

determine the format of the report. Manakhov & Ivanov (2016) define the concept of a usability problem as being "...a set of negative phenomena, such as user's inability to reach his/her goal, inefficient interaction and/or user's dissatisfaction, caused by a combination of user interface design factors and factors of usage context." [18]. The correct articulation of a usability problem serves three key functions: filtering, or the identification of core issues and problems by the evaluator; comparison, or the active juxtaposition by the evaluator of different evaluation methods against each other; and communication, or the precise articulation of a given problem and associated solution. In specifying the parameters of a usability problem, usability evaluators aim to capture all relevant phenomena attributed to a particular human-computer interaction. However, as [18] point out, the majority of problem definitions are vague and/or fail to encompass all pertinent information, leading to poor usability evaluations. The techniques commonly used to evaluate usability include [19]: (i) open-ended comments or reactions; (ii) objective questions or surveys; (iii) task performance measures such as simulation and role plays; (iv) multiple choice or similar tests; (v) participant self-assessment.

Usability is, however, no longer the dominant paradigm in computer systems evaluation. [17] argues that what is currently understood by usability differs significantly from the early studies of Human Computer Interaction in the 1980s. In particular, he states that usability is no longer a binary paradigm encompassing the extent or degree of ease-of-use. The rise of networked digital media has added an emotional dimension to HCI evaluation, resulting in the emergence of the term 'user experience' as a new, measurable indicator, with usability now being considered as one aspect of user experience. The advantage of considering user experience is that the concept places usability in context.

3.3 Information Retrieval Evaluation

In relation to legal information retrieval, and the advanced computer systems that facilitate this process, Van Opijnen & Santos (2017) outline several so-called 'relevance dimensions' that need to be taken into account when developing an evaluation framework [20]. These are summarized in Table 2, below.

4 PROPOSED EVALUATION METHODOLOGY

Requirements engineering (RE) is the process by which individual stakeholder requirements and needs are first elicited, identified, and modelled; and subsequently developed into detailed, agreed functional and non-functional requirements that are documented and communicated in a manner that facilitates the progression of further system development activities. The RE process is considered a pivotal phase of the systems development workflow, and

consists of five main sequential activities: Elicitation, Analysis and Negotiation, Documentation, Validation, and Management [21].

Each stage is associated with a number of tools and techniques, drawn from across various subject domains and disciplines, that are employed by requirements engineers to successfully realize each task. A series of Requirements Elicitation and Requirements Analysis exercises – questionnaire, expert interviews, target user workshops – were undertaken by researchers on the project team between October 2018 and May 2019, often in immediate collaboration with colleagues and senior members of the Greek and Austrian national parliamentary administrations representing the project's two pilot environments and our primary target group. These have been comprehensively documented in [2].

4.1 Basic Requirement Categories

From the data gathered through Requirements Engineering exercises, the ten functional system requirements, and ten non-functional systems requirements were identified. Based on these, the following key performance indicators or basic requirements for system performance evaluation were outlined:

I. Recall: whether a particular item is retrieved or the extent to which the retrieval of wanted items occurs.

- BR_1_Performance: The system must always respond quickly to user queries, even during peak traffic times.
- BR_2_Usability: The new solution is easy to learn, operate, and to use in the preparation of inputs and the interpretation of results.

II. Precision: the proportion of documents retrieved that is relevant.

- BR_3_Accuracy: The platform needs to be accurate and deliver precise results.
- BR_4_Reliability: The system consistently performs its specified functions and delivers appropriate results without failure.

III. Fallout: the proportion of non-relevant items that has been retrieved in a given search.

- BR_5_Security: The system can resist unauthorized, accidental or unintended usage and provide access only to legitimate users.
- BR_6_Supportability: The solution must be cost-effective to support and use.
- BR_7_Verifiability: The system and supporting infrastructure must be validated to the highest reasonable commercial reliability standards.

IV. Generality: is the proportion of relevant items that have been retrieved in a given search.

- BR_8_Availability: The platform needs to be available all the time.
- BR_9_Customisation: The system allows the user to make changes and personalize their usage of the platform.
- BR_10_Internationalization: The user can work in different languages, and switch languages during the same session.

4.2 Integrated Evaluation Framework

Building on one hand on the theoretical background concerning computer systems evaluation outlined in section 3, and on the other hand on the articulation of proposed system's basic requirements and key performance indicators outlined above in section 4.1, the following integrated evaluation framework has been developed and described below in Table 3. The full list of evaluation dimensions and their indicators is provided in the Appendix. The main evaluation dimensions are: usability, performance, availability, security, customization, information quality, capabilities, as well as overall evaluation. We can see that we have used/exploited all the major evaluation dimensions proposed by the theoretical background outlined in section 3 (functionality, usability, user experience and also aspects of information retrieval quality), and adapted them to the specificities and the functional requirements of advanced legal data infrastructures. The key output variables correspond directly to the top-level parameters used to describe the basic requirements. Each basic requirement was then associated with one or more key output variable: correctness or soundness of the algorithm (BR_1, BR_3, BR_7); correctness of the integrated system (BR_6, BR_8, BR_10); completeness of query results (BR_3, BR_4, BR_10); and performance (BR_1, BR_2, BR_5, BR_9). From this clustering, the three most significant key output indicators are listed, and a set of desirable outcomes derived. The system is then evaluated against the attainment of certain measurable key outcome indicators.

5 APPLICATION AND RESULTS

A first application of the proposed methodology was conducted for the legal data infrastructure developed as part of the above-mentioned project. For the purposes of monitoring and evaluation within the purview of the project under consideration, a self-administered questionnaire [22] was developed to collect primary data in the form of attitudes, opinions, and perspectives from those individuals participating in the User Acceptance Tests, pilot tests and associated workshops. The decision to use a questionnaire as the research data collection instrument of choice was taken based on the method's appropriateness to the task at hand, prior experience from previously organized user-led exercises, and the resources at our disposal. Key strengths and weaknesses of the questionnaire method were identified through an analysis of the relevant research methods literature [22-24]. The nature of the investigation and the characteristics of the target population to be surveyed were central to the selection of the questionnaire method as the research instrument of choice. In balance, it was concluded that, if designed and administered properly, the advantages of using a questionnaire would far outweigh any disadvantage. Further, following [24], the selection of the questionnaire method was justified on the grounds that an awareness of the disadvantages to using this tool is the first step towards improving the quality of data obtained from it.

The questionnaire was based on the evaluation methodology presented in the previous section, and divided into two parts. Part 1 captured demographic data and user expectations of what the system might entail. This was administered prior to user testing of the prototype. Questions in Part 2 captured user experiences following their interaction with the system, and was structured so as to mirror the four key output categories and ten basic requirements discussed

Table 3: Monitoring and Evaluation Framework

Key Outputs	Key Output Indicators	Key Outcomes	Key Outcome Indicators
Correctness (soundness) of the implementation of the algorithms	1. Query definition time. 2. Query answering time. 3. System response time	1. Improved query definition time. 2. Improved query answering time. 3. Improved system response time.	1. Search results list available in less than two seconds. 2. Search results no less than 98% accurate. 3. Low proportion of non-relevant responses to queries. 4. System response rate is similar across different browsers.
Correctness of the integrated system	1. System response time for interactive tasks. 2. Number of continuous queries the system can handle for stream-based data. 3. Number of users online simultaneously.	1. Improved system response time for interactive tasks. 2. Increased number of continuous queries handled. 3. Increased number of parallel user sessions.	1.The system is consistently accessible. 2. System functionalities are consistently functional. 3.The system is continuously operational. 4. Multiple users can access the system simultaneously.
Completeness of the query results	1. Query language adequacy. 2. Query formulation adequacy. 3. Relevance of returned results.	1. Improved query language adequacy. 2. Improved query formulation 3. Increased relevance of returned results.	1. The system is able to recognize different keywords from the same legal domain. 2. Quality of search results compares favorably with other available systems. 3. Quality of translations provided compares favorably with other available systems. 4. Quality of visualizations provided compares favorably with other available systems.
Performance	1. System administration adequacy. 2. Extensibility of system architecture. 3. User satisfaction.	1. Improved range of system administration tools. 2. Increased vertical or horizontal extension of the system. 3. Increased user satisfaction.	1. The system provides access only to legitimate users. 2. The system can resist unauthorized, accidental or unintended usage. 3. The system allows the user to customize their experience. 4. The user is satisfied that the system provides all required functionalities.

above, Consequently, this half of the questionnaire included five groups of questions concerning the usability, the performance, the availability, the information quality and the capabilities provided by this legal data infrastructure, and also an additional group of overall evaluation questions, which assessed the overall perception of the respondent concerning the value of this infrastructure in comparison with other pre-existing competitive legal information infrastructures. These overall evaluation questions enable also the assessment of the importance of each of the individual capabilities and characteristics of the legal data infrastructure for the users, through the calculation of the correlations of the variables that correspond to them with the overall assessment variables. Our two-part questionnaire, designed to capture both expectations and actual experience, was filled out by 42 users of the system (legal professionals, legal researchers and public servants), who implemented a series of usage scenarios of it in a number of workshops organized as part of the project.

In Table 4, below, we can see the average ratings of the users for all questions/variables and also the average rating of each group. Furthermore, we can see for each question/variable its correlation (Pearson's Correlation Coefficient) with the average of the overall evaluation questions, which is a measure of the importance of the corresponding capability/characteristic of this legal data infrastructure for the user (as they quantify its association with the overall level of value (in comparison with other similar infrastructures) perceived by the user.

From the second and fifth columns of the above Table we can see that technical quality, with respect to availability and performance, is assessed as high (average ratings 4.14 and 4.05 respectively), while capabilities, usability and information quality are perceived as moderate to high (average ratings 3.74, 3.73 and 3.52 respectively). Availability has received the highest average rating (4.14), so it can be viewed as a strength of the infrastructure, while the customization capabilities have received the lowest average rating (3.40), so it constitutes a weakness of it, and the improvement of it should be considered. Furthermore, we remark that among the capabilities the CAP7 (concerning the assessment of the conflicts,

Table 4: Average ratings and correlations of value dimensions and measures

Measure/Dimension	Averageratings	CorrelationOverall Evaluation	Measure/Dimension	Averageratings	CorrelationOverall Evaluation
USAB	**3.73**	**0.750****	**INFQ**	**3.52**	**0.613****
USAB1	3.64	0.508**	INFQ1	3.57	0.611**
USAB2	3.88	0.608**	INFQ2	3.26	0.285*
USAB3	3.81	0.599**	INFQ3	3.50	0.563**
USAB4	3.59	0.699**	INFQ4	3.74	0.537**
PERF	**4.05**	**0.090**	**CAP**	**3.74**	**0.723****
PERF1	3.86	0.002	CAP1	3.86	0.675**
PERF2	3.74	0.074	CAP2	4.02	0.375**
PERF3	4.55	0.185	CAP3	3.90	0.565**
AVAIL	**4.14**	**0.209**	CAP4	3.81	0.707**
AVAIL1	4.12	0.213	CAP5	3.71	0.622**
AVAIL2	4.07	0.169	CAP6	3.79	0.625**
AVAIL3	4.24	0.209	CAP7	3.45	0.564**
SEC	**3.47**	**0.364****	CAP8	3.86	0.580**
SEC1	3.38	0.280*	CAP9	3.93	0.589**
SEC2	3.55	0.258*	CAP10	3.74	0.574**
SEC3	3.48	0.400**	CAP11	3.36	0.575**
CUST	**3.40**	**0.486****	CAP12	3.48	0.588**
CUST1	3.38	0.335*	**OVE**	**3.42**	
CUST2	3.24	0.441**	OVE1	3.52	
CUST3	3.58	0.472**	OVE2	3.33	

comparisons or dependencies between different laws), CAP11 (capability to report inaccuracies and manually annotate text) and CAP12 (concerning access to relevant public opinion data) have received the lowest ratings (3.45, 3.36 and 3.48 respectively), so they constitute functional weaknesses and their improvement should be considered. The capabilities CAP2 (concerning the retrieval of a particular law or legal document), CAP3 (concerning access to accurate translations of a law or legal document in user's own language) and CAP 9 (concerning highly informative visualizations depicting the comparisons and contrasts among laws) have received the highest ratings – 4.02, 3.90, 3.93- so they constitute definite strengths. Finally, there is a moderate to high level of overall assessment of this advanced legal data infrastructure in comparison with pre-existing competitive ones (average rating 3.42).

The third and sixth column (showing the Pearson's Correlation Coefficients with the average of the overall evaluation questions) we can see that usability is the most important aspect for users' overall value perception (correlation 0.750), followed by functional capabilities (0.723), and then information quality (0.613). Lower still is the importance of the security (0.364) and customization capabilities (0.486). We remark that the technical quality, concerning availability and performance (non-significant correlations) do not considerably influence users' overall value perception. Furthermore, CAP4 (for comparing laws on the same subject within the same country) and CAP1 (concerning the search for legal information on a particular topic in different EU Member States' legislations) are regarded as the most important (correlations 0.707 and 0.675 respectively), followed by CAP5 (for comparing laws on the same subject from different countries) and CAP6 (assessment of the degree of transposition of EU directives into national legislation) (correlations 0.622 and 0.625 respectively).

6 CONCLUSION

This research paper defines a novel monitoring and evaluation framework for advanced legal data infrastructures. The developed methodology could be applied in the evaluation of any new legal data infrastructure since it was derived from the combination of the basic requirements of such infrastructures and the basic KPIs from the relevant monitoring and evaluation theories of complex information systems. The developed evaluation model assesses the overall satisfaction of the potential user in using a specific legal data platform. Overall satisfaction includes two dependent value measures: (a) that the system compares favorably with other available, similar legal informatics solutions and (b) that this system would be preferred over other similar legal informatics products. As a validation step, we applied the developed evaluation model in a novel legal data platform identifying the strengths and shortcomings associated with the use of the system under consideration. To do so, we calculated and combined the average ratings of the indicators and their correlation to the overall satisfaction. The weakest points of the evaluated platform are: the assessment of the conflicts, comparisons or dependencies between different laws (CAP7), the capability to report inaccuracies and manually annotate text (CAP11) and the access to relevant public opinion data (CAP12) which received the lowest ratings (3.45, 3.36 and 3.48 respectively), so they constitute functional weaknesses and their improvement should be considered. Therefore, it is necessary to improve the algorithms used for assessing the conflicts, comparisons or dependencies between

different laws, provide more capabilities for reporting inaccuracies and manually annotating text, and also collect and process data from more electronic spaces (e.g. for a, social media, etc.), in which relevant public consultations take place. Finally, the most important characteristics of a legal data platform are the capability for comparing laws on the same subject within the same country and capability to search for legal information on a particular topic in different EU Member States' legislations (correlations 0.707 and 0.675 respectively). The main limitation of our research is that the evaluation of this novel legal data platform has been conducted by a rather small number of individuals (42 users of the system - legal professionals, legal researchers and public servants), so further evaluation by a larger number of individuals, from all targeted user groups, is required. Further research is also required directed towards the evaluation of other legal data platforms using the described model, as well as the extension of the scope from general satisfaction to include users' objectives.

ACKNOWLEDGMENTS

The ManyLaws project is co -financed by the Connecting Europe Facility of the European Union, CEF-TC-2017-3 Public Open Data.

REFERENCES

[1] Alexopoulos, C., Virkar, S., Loutsaris, M.A., Novak, A.-S., and Loukis, E. 2020. Analysing Legal Information Requirements for Public Policy Making. In: Hoffmann, S. et. al. (eds.) Electronic Participation 12th IFIP WG 8.5 International Conference, ePart 2020 Linköping, Sweden (online), August 31 - September 2, 2020 Proceedings. LNCS Springer, Cham, 95-108.

[2] ManyLaws Consortium. 2019. Deliverable 1.2: Report on User Requirements Elicitation and Definition of Services.

[3] UN Women. 2010. What is monitoring and evaluation?. Retrieved April 26, 2021 from https://www.endvawnow.org/en/articles/330-what-is-monitoring-and-evaluation-.html#:~:text=Monitoring%20is%20the%20systematic%20process, and%20to%20guide%20management%20decisions.

[4] Görgens, M., and Zall Kusek, J. 2010. Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit. No. 53303. The World Bank.

[5] Frankel N., and Gage A. 2016. M&E Fundamentals: A Self-Guided Mini-Course. U.S. Agency for International Development, MEASURE Evaluation, Interagency Gender Working Group, Washington DC.

[6] Rossi, P. H., Lipsey, M. W., and Henry, G. T. 2018. Evaluation: A systematic approach. Eighth Edition. SAGE Publications, Thousand Oaks, C.A.

[7] Markiewicz, A. 2018. Why do programs benefit from developing monitoring and evaluation frameworks?. Better Evaluation, 10 January 2018. Retrieved April 26, 2021 from https://www.betterevaluation.org/en/blog/how-do-programs-benefit-from-ME-frameworks.

[8] Gullickson, A.M. 2020. The whole elephant: Defining evaluation. Evaluation and program planning, 79, 101787.

[9] Commission of the European Communities. 2011. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Open data: An engine for innovation, growth and transparent governance, COM (2011) 882 Final, Brussels.

[10] European Commission. 2015. A Digital Single Market Strategy for Europe. Eur-lex Retrieved April 27, 2021 from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:52015DC0192

[11] Virkar, S., Alexopoulos, C., Stavropoulou, S., Tsekeridou, S., Novak, A., and Charalabidis, Y. 2020. User-Centric Decision Support System Design in Legal Informatics: A Typology of Users. In Y. Charalabidis, M.A. Cunha, & D. Sarantis. ICEGOV 2020: Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance. ACM Press, 711-722.

[12] Charalabidis, Y., Loutsaris, M.A., Virkar, S., Alexopoulos, C., Novak, A.-S., and Lachana, Z. 2019. Use case scenarios on legal machine learning. In S. Ben Dhaou, L. Carter, & M. Gregory. Proceedings - ICEGOV 2019 12th International Conference on Theory and Practice of Electronic Governance, 3-5 April 2019, Melbourne, VIC, Australia. ACM Press, 364 - 373.

[13] McNamara, N., and Kirakowski, J. 2006. Functionality, Usability, and User Experience: Three Areas of Concern. interactions, 13(6), 26-28.

[14] Saket, B., Endert, A., and Stasko, J. 2016. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization. ACM Press, 133-142.

[15] Bevan, N. 1995. Usability is quality of use. In Advances in Human Factors/Ergonomics, vol. 20. Elsevier, 349-354.

[16] Usability.gov. 2020. Usability Evaluation Basics. Retrieved April 26, 2021 from https://www.usability.gov/what-and-why/usability-evaluation.html

[17] Cockton, G. 2012. Usability Evaluation [2nd ed.]. In: Encyclopedia of Human-Computer Interaction. Retrieved April 27, 2021 from http://www.interaction-design.org.

[18] Manakhov, P., and Ivanov, V.D. 2016. Defining usability problems. In Proceedings of the 2016 CHI Conference ExtendedAbstracts on Human Factors in Computing Systems. ACM Press, 3144-3151.

[19] Lu, H.-K., Lin, P.-C., and Wu, M.-Y. 2012. A Review of Information System Evaluation Methods. In 2012 International Conference on Software and Computer Applications (ICSCA 2012). ACM Press, 243-248.

[20] Van Opijnen, M., and Santos, C. 2017. On the concept of relevance in legal information retrieval. Artificial Intelligence and Law 25(1), 65-87.

[21] Paetsch, F., Eberlein, A., and Maurer, F. 2003. Requirements engineering and agile software development. In Proceedings - Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, WET ICE 2003. IEEE Press, 308-313.

[22] Wilson M., and Sapsford, R. 2006. Asking Questions. In Data Collection and Analysis. Second Edition, SAGE Publications, London UK, 93–123.

[23] Gillham, B. 2011. Developing a Questionnaire 2nd Edition. Continuum International Publishing Group, London, UK.

[24] Kumar, R. 2014. Research Methodology. A Step-by-Step Guide for Beginners, Fourth EditionSAGE Publications, London, UK.

A EVALUATION INDICATORS

Usability

USAB1. It was easy to find the information I needed

USAB2. The interface of the system was pleasant and easy to look at

USAB3. The output/results it provides are understandable

USAB4. The capabilities provided by the system are compliant with the work-practices and the mentality of legal professionals

Performance

PERF1. The system returned rapid results to my queries

PERF2. The speed at which the system returned results remained consistent for each login session

PERF3. The pages work in my favorite browser(s)

Availability

AVAIL1. I was able to access and browse the platform at my convenience - at any time of the day, from anywhere

AVAIL2. I was able to access and navigate through the different services at my convenience - at any time of the day, from anywhere

AVAIL3. The platform was never offline at the moment that I wanted to use it

Security

SEC1. The system provides access only to legitimate users, and can resist unauthorized, accidental or unintended usage

SEC2. I am confident that my own personal data as well as my transactions and activity in the system are protected on the site

SEC3. I can retrieve all my transactions and activity in the system in subsequent sessions

Customization

CUST1. The system allows me the flexibility to customize my user experience

CUST2. The system provides me with all functionalities I require

CUST3. My personal space on the system is user-friendly and attractive to use

Information Quality

INFQ1. The results returned by the system correspond closely to my queries

INFQ2. The proportion of 1n-relevant results to my queries provided by the system is low

INFQ3. The system is able to recognize different keywords from the same legal domain

INFQ4. The translations made by the system are reasonably accurate.

Capabilities

The system provides comprehensive capabilities for conducting the following:

CAP1. Search for legal information on a particular topic in different EU Member States' legislations.

CAP2. Retrieve a particular law or legal document.

CAP3. Access accurate translations of a law or legal document in my language.

CAP4. Compare laws on the same subject within the same country

CAP5. Compare laws on the same subject between different countries

CAP6. Assess the degree of transposition of EU directives into national legislation

CAP7. Assess the conflicts, comparisons or dependencies between different laws

CAP8. Trace the evolution of a piece of legislation over time

CAP9. Access highly informative visualizations depicting the above comparisons and contrasts

CAP10. Access different types of parliamentary data

CAP11. Report inaccuracies and manually annotate text

CAP12. Access relevant public opinion data

Overall Comparative Evaluation

OVE1. I am confident that the system compares favorably with other available, similar legal informatics solutions

OVE2. I would choose this system over other similar legal informatics products