


The legal macroscope: Experimenting with visual legal analytics

Information Visualization
2017, Vol. 16(4) 332–345
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1473871616681374
journals.sagepub.com/home/ivi


Nicola Lettieri¹, Antonio Altamura² and Delfina Malandrino²

Abstract

This work presents Knowlex, a web application designed for visualization, exploration, and analysis of legal documents coming from different sources. Understanding the legal framework relating to a given issue often requires the analysis of complex legal corpora. When a legal professional or a citizen tries to understand how a given phenomenon is disciplined, his attention cannot be limited to a single source of law but has to be directed on the bigger picture resulting from all the legal sources related to the theme under investigation. Knowlex exploits data visualization to support this activity by means of interactive maps making sense out of heterogeneous documents (norms, case law, legal literature, etc.). Starting from a legislative measure (what we define as Root) given as input by the user, the application implements two visual analytics functionalities aiming to offer new insights on the legal corpus under investigation. The first one is an interactive node graph depicting relations and properties of the documents. The second one is a zoomable treemap showing the topics, the evolution, and the dimension of the legal literature settled over the years around the norm of interest. The article gives an overview of the research so far conducted presenting the results of a preliminary evaluation study aiming at evaluating the effectiveness of visualization in supporting legal activities as well as the effectiveness of Knowlex, the usability of the proposed system, and the overall user satisfaction when interacting with its applications.

Keywords

Visual analytics, legal informatics, data mining, visualization, evaluation

Introduction

Social life is regulated by sets of heterogeneous and closely intertwined legal sources that span from laws and other legal regulations to administrative measures, from case law to legal literature: an intricate universe of documents forming a unitary and complex whole. Picking the way in this universe is difficult from different points of view.

Retrieving and manipulating all the relevant information is the first stumbling block. When a citizen tries to understand how a given issue is legally disciplined, when a legal professional tries to see how a specific area of a legal system evolves over time, their attention cannot be limited to a single source of law. Specifically, it has to be directed on the bigger picture resulting from all the legal sources related to the theme taken

into account, a complex set of information that are often difficult to be identified, retrieved, and gathered in the same context. Once we have the information, the next step is the analysis of the retrieved materials, an analysis whose content can vary depending on the user (scholar, professional, layman, etc.) and the pursued aim: identify and study connections between the retrieved documents, track the evolution of case law

¹ISFOL, Rome, Italy

²Department of Computer Science, University of Salerno, Fisciano (SA), Italy

Corresponding author:

Delfina Malandrino, Department of Computer Science, University of Salerno, Via Giovanni Paolo II, 84084 Fisciano (SA), Italy.
Email: dmalandrino@unisa.it

and legal doctrine over time, find and recover papers relating to specific topics, and understand which are the sectors of the legal system on which the norm of interest has had greater impact. All these analyses can be crucial in many different fields spanning from academia to legal professions. Visualization techniques can represent an interesting innovation in this regard: they do not simply make easier and more intuitive information retrieval—a feature that is particularly useful when one has to handle large quantities of data—but they also offer new insights in the legal world as it is already happening in many other research fields. The amount of web-accessible legal documents has dramatically grown in the last few years, thanks to the spread of public database that offers free access to laws and regulations as well as to case law and legal literature. Unfortunately, these repositories are independent, and users have to separately access each of them: without adequate exploring system, data often remain in the repositories without exploitation. Moreover, traditional list-based search engines do not allow contextual visualization of the documents and, more importantly, do not usually offer analytics tools. In this scenario, a tool gathering in a single context the most important legal sources related to a certain topic and allowing an easy exploration (select groups of data, aggregate them, and perform comparisons over the data via an intuitive interface) could be a precious ally for both legal professionals and laymen.

This work presents Knowlex, a web application designed for the visualization, the exploration, and the analysis of legal documents coming from different sources. Our attention so far has been focused on two applications. The first one, named “Norm Graph Navigator,” is the visual browsing of a map that we define as the “Reference Network of a Norm” (RNN from now on), a graph-based interactive visualization of all the legal sources (laws, Supreme Court judgments, constitutional judgments, preparatory works, and legal literature) connected to a given law. The second one, named “Legal Doctrine Treemap,” is a zoomable treemap¹ showing the topics, the evolution, and the dimension of the legal literature settled over the years around the norm of interest.

The main contributions of our work are as follows:

- Investigate the effectiveness of information visualization techniques during the process of analyzing a huge corpus of legal documents;
- Present a tool, named Knowlex, that both legal professional and layman can use to analyze legal documents coming from different and independent sources;
- Assess the effectiveness of our visual-based approach, in terms of time to complete domain-related tasks and their corresponding

correctness (i.e. reduced number of errors) against the standard method of performing the same tasks. We also analyzed the system usability and the overall user satisfaction.

The rest of the article is organized as follows. Section “Related work” discusses some relevant work in this field. In section “Knowlex: visual browsing and analytics for law,” we describe Knowlex, its architecture, and its main applications. In section “Evaluation,” we present the results of a preliminary evaluation study aiming at assessing the effectiveness of Knowlex, its usability, and its overall user satisfaction. Finally, in section 5, “Conclusion and future work,” we conclude with some final remarks and future directions.

Related work

Our work deals with the exploration of legal database storing thousands of heterogeneous legal documents. Knowlex is an experimental attempt to devise innovative ways to support legal search and analysis offering professional, scholars, and laymen new opportunities to explore the legal world. Instead of displaying lists of results, Knowlex exploits interactive visualization to enable the visual exploration and the analysis of the repositories’ contents, at the same time, supporting data filtering and selection tasks.

Actually, recent years have witnessed a growing interest toward the application of visualization techniques in the legal field, an interest that resulted first of all in the development of domain-specific tool. Ravel Law² is a visual search engine designed by two students from Stanford Graduate School of Design that exploits natural language processing, machine learning, and graph visualization to help lawyers in sorting through legal information. Whereas traditional legal databases present results in a column, often hiding important cases pages back in search results, Ravel Law visually represents the most important cases on a particular topic as the node of a network, with numerous edges pointing to subsequent cases that have cited it. The size of the hub reflects the relative number of cases that cite it. The frequency with which courts cite a particular case often signals the influence of the case over a given area of law or its relevance to a particular legal concept.

Another interesting (even if simpler) experiment is Lexmex,³ an online system developed in France at the Compiègne University of Technology that allows to visually display the relations between texts of law. The attention of the experiment is focused on the French Civil Code and related legislation; Lexmex “translates” the legislation in nodes and links: two texts (an article, a law, decree, or even a prescription) are connected in the graph if one mentions, modifies, or creates another. For now, Lexmex only includes a single

“view” on the data, 2928 nodes and 6574 links. The semantics of the visualization is simple: a node is bigger depending on the number of connections it has with other nodes of the graph. The colors correspond to the cluster detected by means of an algorithm allowing to detect of clusters or “communities.” The clusters detected automatically correspond roughly to the paragraphs of the civil code. Technically, the tool incorporates essential functionality for dynamic navigation: zooming, selecting a node on the fly, display contextual, and search by keywords.

Interesting results can be obtained also with general purpose tools. Due also to the rise of Big Data, many solutions have been developed allowing to analyze, display, manipulate, and explore huge amount of documents. Radiance,⁴ just to give an example, is a powerful visual analytics platform offering various kinds of visualization and allows to visually identify patterns and relations in large data sets residing also in cloud services like Office 365, Google, Dropbox, Box, and Slack. Other examples include both standalone applications like Vosviewer⁵ and web-based solutions like Hypergraph⁶ allowing to generate more or less interactive visualizations starting from raw data. A major stumbling block to the use of these kinds of tools in the legal field is represented anyway by the need of customizations that are more difficult with general purpose software solutions. When dealing with legal documents, customizations are needed for at least three goals: (1) develop complex workflows for data gathering, since legal documents often have different structures and formats and are stored in different and independent repositories; (2) design personalized visualizations tailored for lawyers, legal scholars, and general audience (with low legal and technical skills); and (3) implement domain-specific features (e.g. in-text citation browsing).

In more general terms, all the mentioned tools and projects can be brought back to the efforts made by information visualization⁷ and visual analytics research^{8,9} to allow people to turn data into knowledge facilitating analytical reasoning by means of interactive visual interfaces. The issue is connected with “exploratory search,” a topic emerged in the information retrieval area^{10,11} that has devoted a great deal of research to design systems and interaction techniques overcoming the limits of classical lookup search where the results are presented in list-based widgets. The idea underlying the exploratory paradigm is to more actively involve users providing them with a more direct control over the search process combining the traditional search activities with higher-level goals (e.g. comparison, analysis, synthesis, and evaluation). Exploratory systems allow users to formulate and reformulate queries, give information on the search space and clues for further possible search directions¹² allowing the constant exploration and filtering of retrieved results.¹¹

The way in which the query results are presented plays a key role in this vein: exploration systems exploit visualization to display groups of items and provide the users with an initial overview. Interacting with the visual representation of the query results, users can formulate new refined queries and update the visualization. This interaction (“overview”, “zoom and filter”, “details-on-demand”) is known as “visual information-seeking mantra”.¹³ A tough problem, in this scenario, is overcoming the limits of classical list-based widgets that are often unable to effectively show all the items of large data set and the relationships among them. Graphs are a first, powerful solution for this kind of information retrieval problem:¹⁴ they not only make better use of screen space facilitating access to large amounts of data but also effectively depict both the structure and the properties of the relations occurring between the entities displayed (value/dimension of entities/node connections, subgroups, density of the network, etc.). Two-dimensional (2D) space-filling visualization like Treemap, on the other hand, represents another popular solution to this problem: they not only provide at a glance intuitive overview of entire data sets, but they also offer an effective way to visualize intrinsically hierarchical data.^{15,16}

Knowlex: visual browsing and analytics for law

In this section, we introduce Knowlex, an online “visual analytics toolkit,” composed of different modules implemented to experiment new ways of interaction with legal materials for both professional and scientific purposes. Starting from a legislative measure chosen by the user, what we call “Root” norm, Knowlex gathers data from several sources (discussed later on in this section). Our work so far has been focused on two modules implementing the basic operations of exploratory search: overview, navigation, manipulation, and interaction needed to analyze query results. The first one is the Norm Graph Navigator, a module allowing to visually explore the RNN and the documents belonging to it. The second one is the Legal Doctrine Treemap, a module aiming to achieve two main results: visually support the semantic navigation in the legal literature connected to a given norm and offer to the user further features for the analysis of the data. In the following, after a brief overview of the system sketching both the architecture and the workflow, we present the functionalities implemented in both modules.

Architecture

Knowlex is based on a client/server architecture, as shown in Figure 1, with a zero-configuration on the

client-side, since users have only to type a specific URL in their browser and start with the interactions.

From the technological point of view, Knowlex exploits mainstream technologies. On the client-side, it has been developed using JavaScript open-source libraries (i.e. Sigma.js, Linkurious.js, and D3.js). On the server-side, data are gathered through HTTP requests using cURL,¹⁷ a PHP extension that provides access to libcurl, a multiprotocol file transfer library. The server-side scripts are wrappers that parse different external sources and produce structured data in JSON format (a wrapper exists for each source). The application has been designed so that it can be easily expanded with new document sources by developing pluggable wrappers. We support live remote requests by storing results in a MySQL database (no other network connection is needed). The goal is to build a time-based cache in order to accelerate the response time for successive requests. The database also stores one of the resources used by the application, that is, the Constitutional Court data set. It is officially available in XML format and has been converted in MySQL in order to improve the server response time and significantly increase performance.

Document base

Knowlex exploits a document base spread over several online legal databases, whereas each of them is composed of different categories of document and data. In the following, we will describe each of them, with a

particular emphasis on the extracted information. We have to remark that in the following, we will refer to a legislative measure, chosen by an end user, with the term *Root*.

Legislation. The integral texts of the laws are retrieved using “Normattiva,” a public database managed by the Italian Prime Minister’s Office. Given a law (the Root), Normattiva provides us with several information, such as reference details and full texts of the norms amending it and reference details and full texts of the norms citing it. To recover this data, Knowlex exploits Uniform Resource Name (URN) schema (named NIR by the research project Norme In Rete¹⁸ during which it has been defined) used in Italy as a technical standard for the identification and the automatic retrieval of normative documents in electronic format. The standard makes it possible to automatically build, starting from the reference details of a normative act, the link to the page containing its full text. Figure 2 shows data gathered from Normattiva.

Preparatory works. The site of the Italian Chamber of Deputies (<http://www.italgiure.giustizia.it/sncass/>) makes accessible all the documents produced by the Parliament during the legislative process (the so-called “preparatory works”) including references to any of the various materials generated, such as bills, committee reports, minutes of committee hearings, and parliamentary debates.

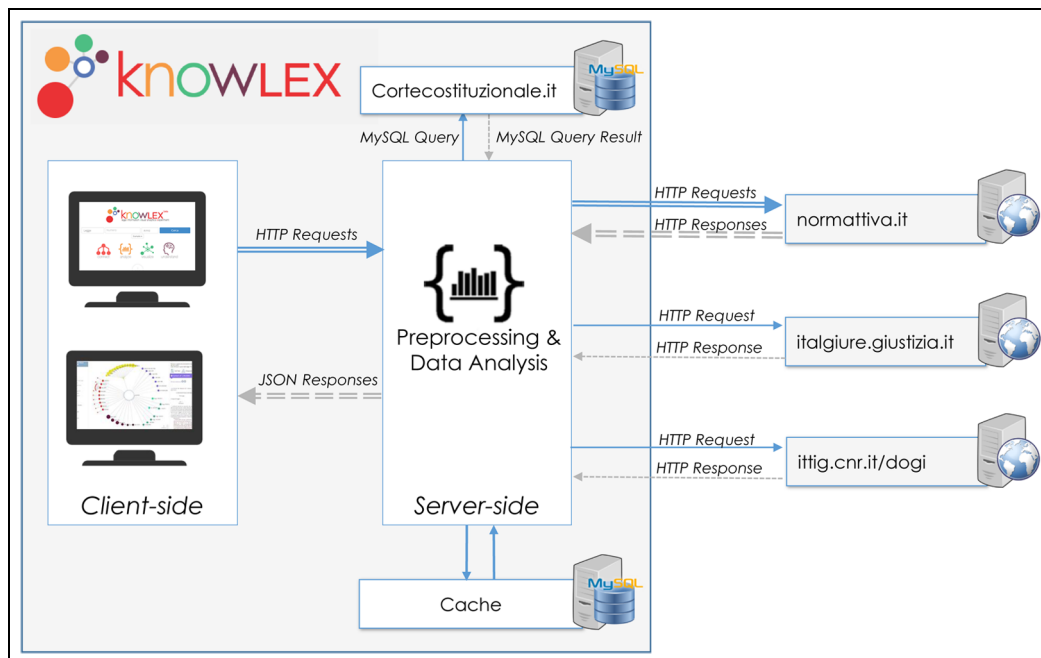


Figure 1. Knowlex architecture and the interactions' flow between components.

The screenshot shows the NORMATTIVA website interface. At the top, there are navigation tabs: 'Full text (article by article)', 'Norm title', 'Norm amendments', 'Preparatory jobs', and 'Cited norms'. Below these, the search results for 'LEGGE 19 febbraio 2004, n. 40' are displayed. The main content area includes the law's title, its entry into force date (10-3-2004), and a table of contents with 12 articles. The text of Article 1 is visible, along with a constitutional amendment note from the Corte Costituzionale dated 14/5/2015.

Figure 2. Example of a search on <http://www.normattiva.it>. Data gathered about the Root 19 Febbraio 2004, n. 40.

Supreme Court case law. The Italian Court of Cassation has recently makes accessible online the database of its most recent provisions (<http://www.cortedicassazione.it>). Through a full-text search in the database, the texts of the judgments that cite the Root are obtained in PDF format.

Legal Doctrine. Legal Doctrine, the scientific literature produced by legal scholars commenting, analyzing, and interpreting rules and judgments are found on DoGi (Dottrina Giuridica), a reference database created by the Institute of Legal Information Theory and Techniques of the National Research Council contains references and abstracts of articles published in Italian legal journals since 1970 (www.ittig.cnr.it/dogi). DoGi classifies its resources according to a consolidated classification scheme covering all areas of national (Italian) law and international law, divided into 24 micro thesauri covering general areas of law, each structured in a three-level hierarchy of systematic descriptors.¹⁹

Constitutional case law. All the decisions (judgments and orders) adopted by the Italian Constitutional Court

are made available through the institutional website that publishes a data set structured in XML format containing the texts of all the Courts' decisions integrated by a rich series of metadata (<http://www.cortecostituzionale.it>). With the full text of each decision, there is an abstract of the principle affirmed by the law.

In summary, currently our retrieval process involves around 720,000 documents: laws (88,336), Constitutional Court case law (56,518), Court of Cassation case law (161,269), legal doctrine (404,827), and preparatory works (16,141).

Workflow

The series of activities in our workflow can be summarized as follows (see also Figure 3):

- *User query.* The procedure starts with a query that uses as keywords the reference details of the normative act under investigation (law or legislative decree, year, and number). The query activates a search inside different databases and data sets.
- *Data gathering.* Once the user defines the norm of interest, that is, the Root, the application, using



Figure 3. Application workflow.

server-side wrappers that are called automatically by the tool, retrieves the data that will be used for the analysis process.

- *Data analysis.* First, given the web scraping, data are cleaned from redundancy (e.g. information that are not needed for the analysis) and potential errors (e.g. network connections). Depending on the active module, the related server-side wrappers are queried, returning a JSON file as result and different kinds of analysis are made (we will describe them in the following sections).
- *Data visualization.* Using an asynchronous approach, the application retrieves the data produced from the server-side wrappers and visualizes them (through the Norm Graph Navigator or the Legal Doctrine Treemap) (see Figures 5 and 10).

Norm Graph Navigator

The Norm Graph Navigator module has been designed with three main goals: (1) visualize and explore the RNN and all related documents, (2) use the RNN to experiment visual information retrieval^{20,21} features, and (3) evaluate the advantages of quantitative analysis and visualization to support domain experts in the analysis of the impact of a norm on the legal system over time. The RNN is the network of connected legal sources related to a same norm, a large group of documents belonging to different categories (legislation, case law, legal doctrine, and preparatory works) produced by different subjects (parliaments, courtrooms, and legal scholars) often belonging to different levels of government (from local to state level) or even to different legal orders and spread in different publications. In the work so far done, for practical reasons, we focused on an “operational” version of the RNN characterized by some limitations compared to the “complete” RNN. The first limitation is due to the fact that we build the RNN starting from a law considered as a unique entity. We know that in many cases, the elementary unit to be considered for the reconstruction of the RNN is not represented by the law considered as a whole, but rather by internal partitions like the article. Anyway, we were forced to make this choice because of the level of coarseness of the available data structure and metadata. The second limitation is related to the categories of documents taken into account: we limited ourselves to considering only the most important documents

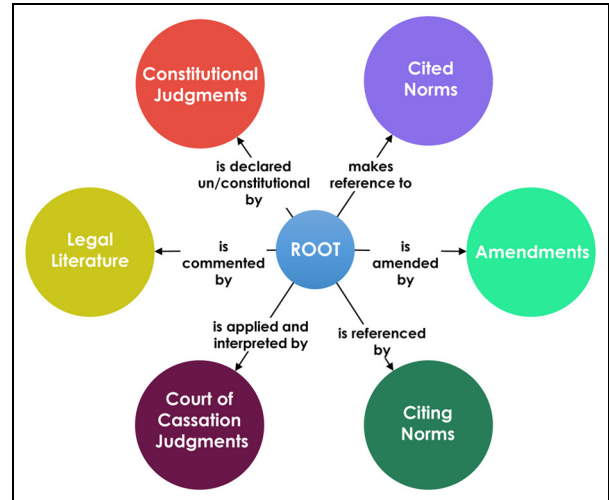


Figure 4. Categories of document taken into account and their relation to the Root.

(i.e. more relevant in the hierarchy of sources of law) and the ones more easily attainable: Legislation, Constitutional Court case law, Court of Cassation case law, preparatory works, and legal literature. Such documents give rise to seven different typologies of relations with the Root that we show in Figure 4.

The Norm Graph Navigator visualizes the RNN aforementioned connecting in a unique graph different categories of legal sources and represent them in a network of connections. In this context, relations between different legal systems at different levels of government can be effectively represented. The network is represented by a node graph with radial layout, with a central node standing for the Root, and several sectors around it, each representing a specific category of documents (see Figure 5).

The number of documents retrieved in each category is often high, and therefore, in order to make easy its readability, we decided to set a threshold limit (default is 20) for the number of nodes that can be displayed for each sector. This value can also be customized according to user preferences as option to provide (*Options* form, in the Menu positioned at the left side of Figure 5).

Beyond this threshold, all documents that have the same publication year are merged. Thereafter, if the number of nodes to show is still greater than the threshold, an iterative hierarchical agglomerative clustering is applied. In each iteration, the two nodes with minimum distance are merged. The distance is defined as the euclidean distance of the centroids of the nodes (using the “publication year” as one-dimensional datapoint). Hence, the linkage criteria used to determine the distance between nodes is the average linkage.²² The algorithm (also shown in Figure 6) iterates until the threshold limit is satisfied.

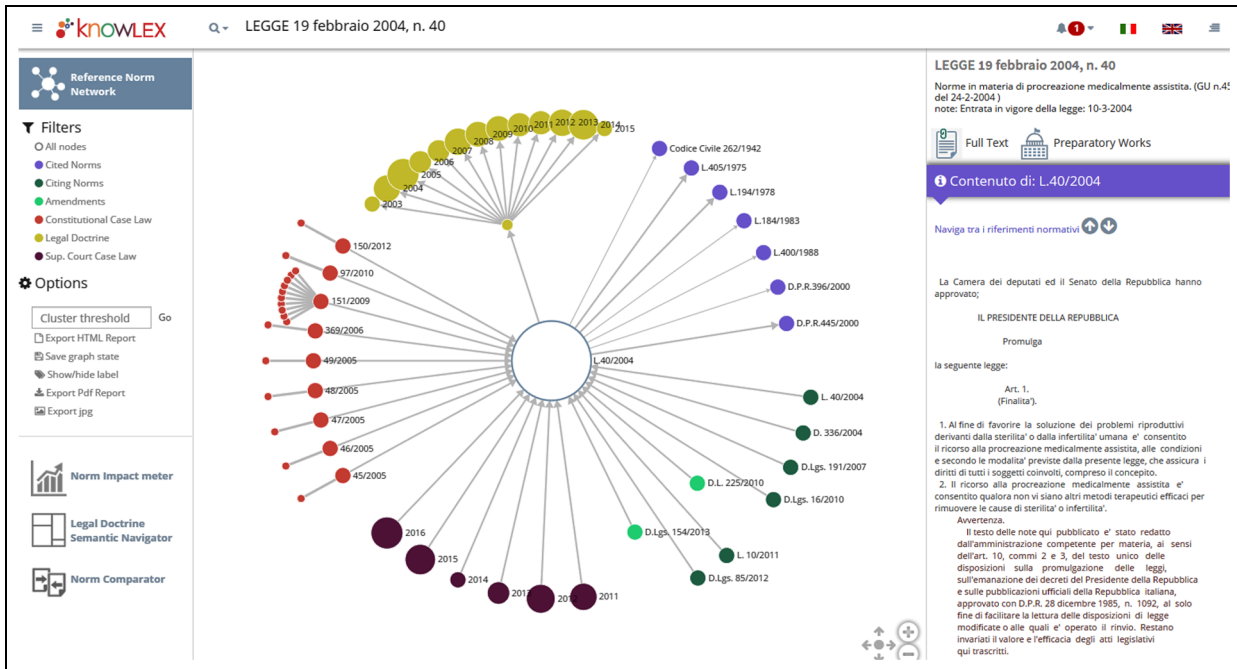


Figure 5. Norm Graph Navigation GUI. Root: Law 19 Febbraio 2004, n. 40.

The algorithm is applied for each sector of the RNN. An application is shown in Figure 7. Finally, we have to emphasize that the year is of relevance in our case study, since lawyers conventionally browse/search regulations according to a chronological criterion. The application of legal rules is indeed governed by the “tempus regit actum” principle according to which an act is governed by the law that is in effect when the act occurs.

Starting from an initial view, the Norm Graph Navigator allows users to manipulate the graph (zoom, drill-down and roll-up, drag and drop) exploring the RNN and its components. By clicking on a node, the focus switches from the general overview to the active node. We show both contextual information (document data) and, if needed, the full text of the selected documents. When the node represents a cluster of documents, the user can explode it and browse its content. The graph can be filtered to show a single sector in order to focus on a specific category of documents. Furthermore, the user can explore the sector of norms cited by the Root and he or she can also conduct iteratively an in-depth navigation in order to identify all norms cited by each of them. The user, in this way, is able to reconstruct backward in time, the path that has led him to the norm searched. As an example, as we can see from Figure 8, the Root cites the norms in Level 1. Starting from the Law A in Level 1, the user can explore all the norms cited by it in Level 2, and so on.

The data collected can be exported in an HTML or PDF report. Users can also take a snapshot of the graph saving it in JSON format.

The Norm Graph Navigator also contains an interactive mixed (i.e. line chart and bar chart) chart, as depicted in Figure 9, showing the temporal distribution of the documents connected to the Root. Largely inspired by a computational and quantitative approach to the investigation of legal problems,^{23–25} the module has two main goals: on one hand, allow users to observe a quantitative representation of the trends in the production of documents connected to the Root; on the other hand, support domain experts in evaluating the impact of the Root norm on the legal systems on the basis of objective indexes. We indeed believe that the evolution and the nature of the impact of a norm can be partially inferred, thanks to an even simple representation depicting, year by year, the quantity and the categories of documents (judgments, doctrine articles, etc.) connected to the Root.

Legal Doctrine Semantic Navigator

The Legal Doctrine Semantic Navigator, as shown in Figure 10, is an attempt to visually support a semantic navigation of the legal literature connected to the Root. A tough problem, for users trying to pick their way in an often overwhelming legal literature, is to find relevant materials (for scholarly or professional use) when they are not thinking to a specific and known paper, but they simply have an idea of the topic of interest. Classification schemes (vocabularies of hierarchically structured descriptors) help users dealing with this kind of issue allowing them to make queries using simply a term that semantically defines the scope

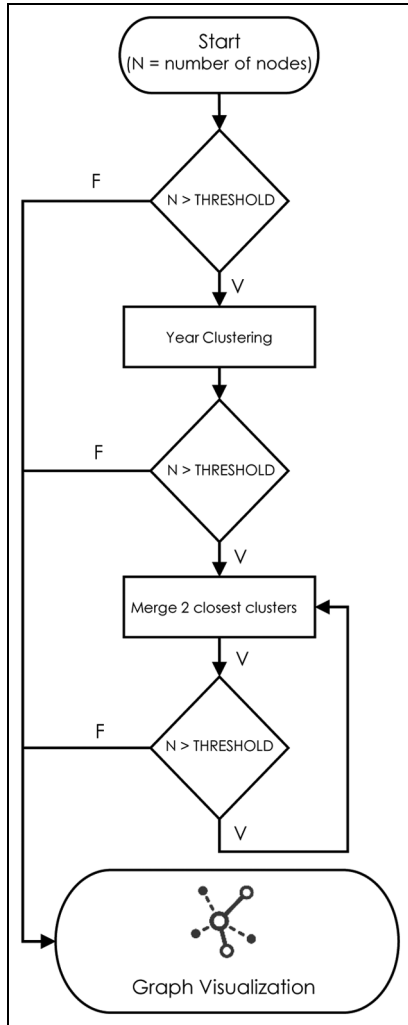


Figure 6. Clustering algorithm flowchart.

of the research. Thanks to classification schemes, often used in legal reference database like DoGi, the search becomes a step-by-step semantic process based on the use of more and more specific search keys. Against this background, we decided to exploit the potentialities of Treemaps¹ that, as highlighted above, are a perfect fit to ease the navigation of hierarchical structured data.

In the Legal Doctrine Semantic Navigator, users can interactively explore the legal literature data set through an in-depth navigation¹⁴ of the topics performing drill-down and roll-up operations. The documents are categorized by the DoGi classification scheme as explained in the “Document base” section. The treemap is generated by processing the list of doctrine articles according to the following algorithm (see also Figure 11):

- Analysis of the classification of each document.
- *Tree population.* A node is created for each classification topic hierarchically (the papers with tag

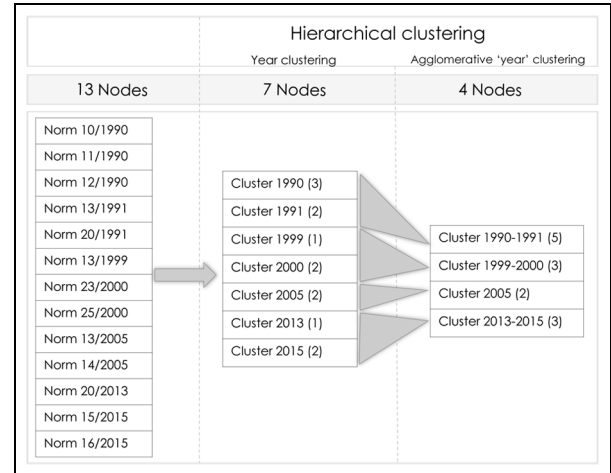


Figure 7. Application of the clustering algorithm. The initial set was composed of 13 nodes. After the application of the year clustering, the number of nodes becomes 7, while the iterative clustering will reduce the initial set to 4 nodes.

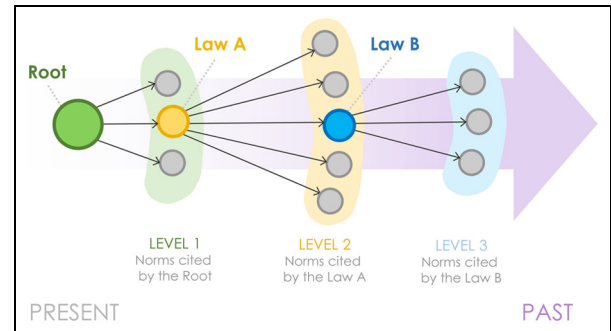


Figure 8. In-depth search. Starting from the Root, users can explore the path of the cited norms, level by level, until the searched one.

“Civil law” are associated with the node “Civil law”).

- *Tree processing.* The resulting tree is processed in order to obtain statistical data, and the treemap is composed according to them.

Every aspect of the treemap is defined according to a specific metric and conveys information. The size of each sector is proportional to the percentage of articles associated with it compared to the total amount of papers for each level. The sectors are distributed according to a descending order from the biggest (top-left) to the smallest (bottom-right). Each level of the treemap has a different color gradient palette. The color of each sector is chosen using the statistical mode calculated according to the feature “publication year”: each sector takes the color associated with the

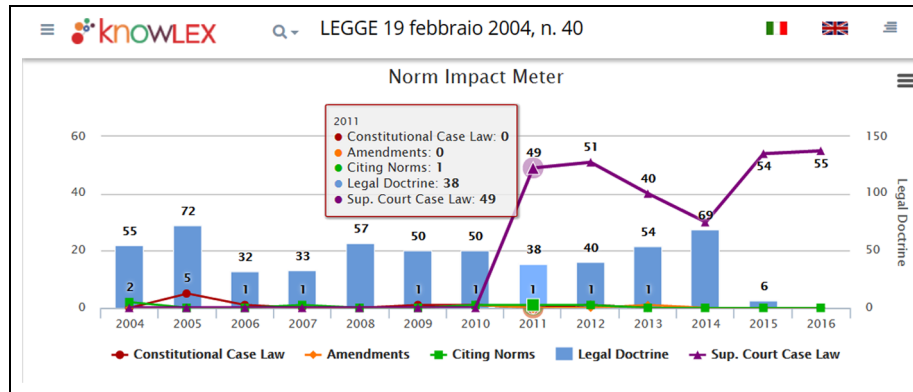


Figure 9. Analysis of the impact of the norm. A specific data set can be interactively included/excluded by the visualization by simply clicking on the corresponding series label shown in the Legend.

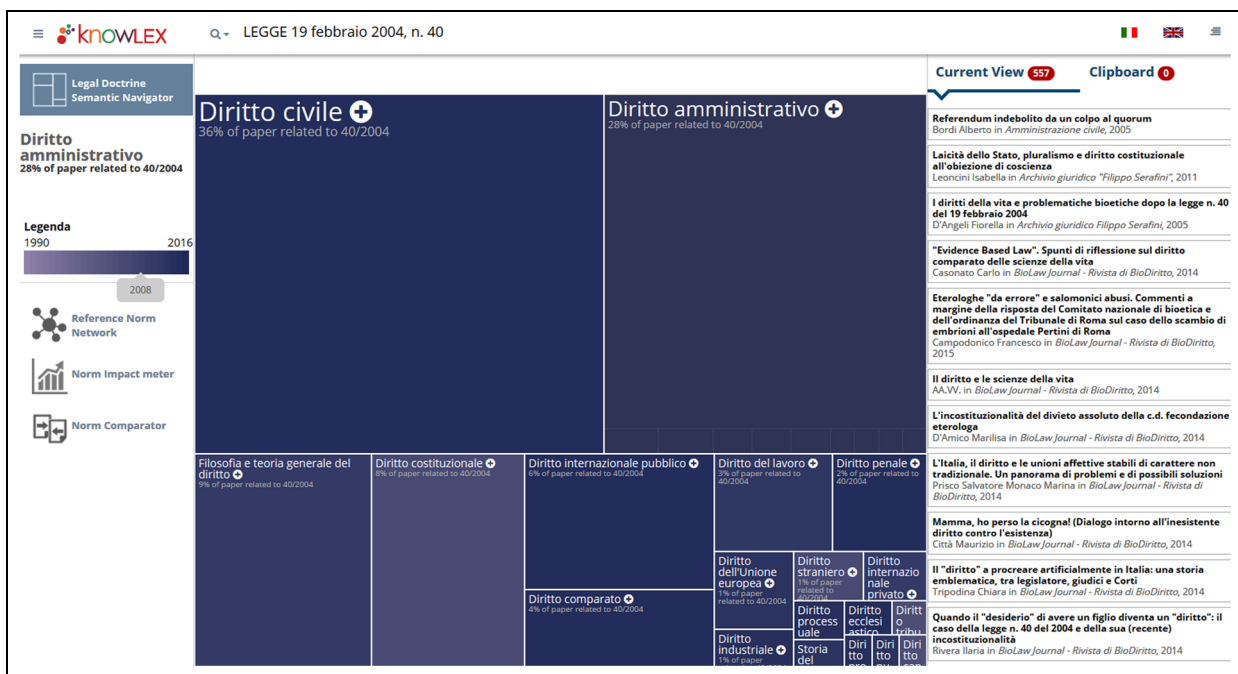


Figure 10. Legal Doctrine Semantic Navigator.

year of publication with greater frequency (the newer year the darker in the related palette). Watching the size and the color of the sectors, the user could deduce in which sector of the legal literature (and therefore, somehow, of the legal system) the Root norm had major impact (see Figure 12).

The Legal Doctrine Treemap user interface, as shown in Figure 10, shows the treemap and the options allowed by the module. While exploring the map, the user can see the list of related documents of the active sector in section “Current view.” There is also a “Clipboard” section in which the user can create customized lists of documents according to the areas

of jurisprudence of interest and export them in custom report in PDF format.

Evaluation

In this section, we first describe the methodology that we employed for our evaluation study; afterward, we discuss the results obtained when a group of 13 students was involved in testing Knowlex. In our evaluation study, we followed the standard human-computer interaction (HCI) methodology,²⁶ commonly applied in different contexts,^{27,28} as well as visualization.²⁹⁻³¹

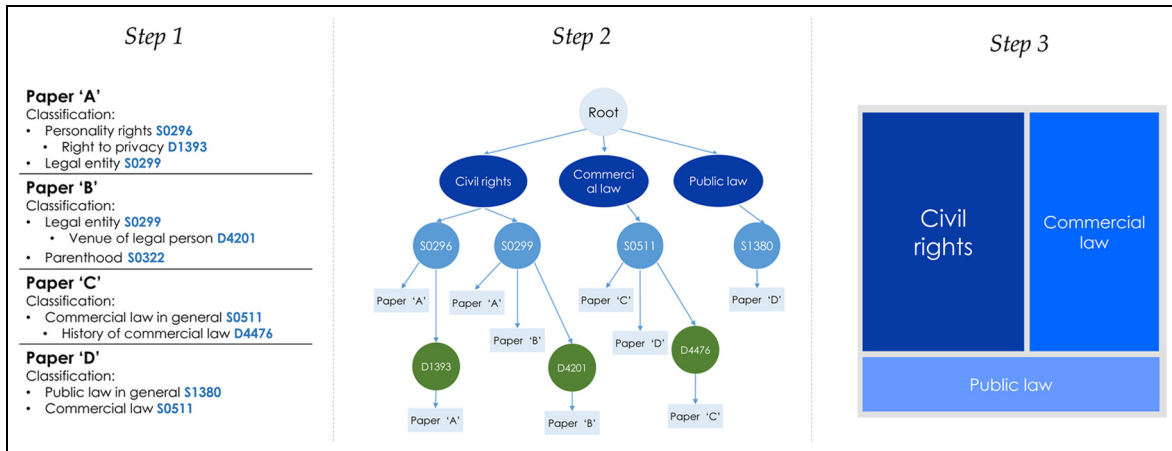


Figure 11. The topic tree generation process. How data can be represented and then visualized to users.

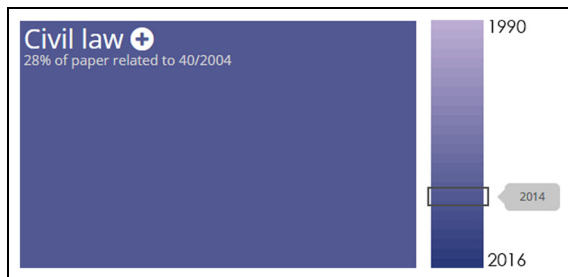


Figure 12. A sector of the treemap. The color matches the color gradient in corresponding value of 2014.

Method

The objective of this study was to explore the intuitiveness and the effectiveness of Knowlex when a huge amount of legal documents, from different sources, have to be accessed and analyzed to understand how given issues are legally disciplined. Specifically, we compared the “Enhanced” approach that exploits the functionalities provided by Knowlex against the “Standard” approach, generally employed to perform this process. Specifically, the Standard approach consists in making a set of independent queries on different databases without any kind of visual support for the end users.

We analyzed the usability of the tool and the overall user satisfaction. We also studied the relationships of users’ intentions to use the tool with selected constructs from technology acceptance model (TAM) model,³² such as their attitudes, perceived usefulness, ease of use, and finally, playfulness.

The study was conducted in a computer science laboratory of the Law department at the University of Benevento. The study envisioned three different phases in which we carried out: (1) a Preliminary Survey, (2) a Testing Phase, and (3) a Summary

Survey, as defined in other contexts.^{33,34} At the beginning of the evaluation, participants were briefly introduced to the purpose of the study and the upcoming tasks. In the first phase, we collected information about demographics, technical information communication technology (ICT) experience, and familiarity with visual navigation tools. In the Testing Phase, we asked users to perform 7 tasks for each approach, for a total amount of 14 tasks performed by each participant.

To mitigate the impact of individual differences and to increase the output of the test results, a within-group design²⁶ was selected, and therefore, each participant tested both conditions (Enhanced and Standard). However, to mitigate learning and fatigue effects, the tested approaches were counterbalanced and tasks randomized. Specifically, seven participants first tested the Enhanced approach and then the Standard approach, while for the second group (six participants), we reversed this procedure. We fixed a time limit (90 min) for testing each approach. The order of the tasks was not fixed, each user could decide which task to execute first, and to go forth and back among the administered tasks. Tasks were chosen by interviewing five field experts. At the end of each task, we asked participants to rate its easiness. A description of the tasks is shown in Table 1. At the end of the Testing Phase, we asked users to spend another 10 min to answer to Technology Acceptance Model (TAM) questionnaire.

Further 15 min was required to fill out the standard Computer System Usability Questionnaire (CSUQ),³⁵ the standard Questionnaire for User Interface Satisfaction (QUIS),³⁶ and finally, a Summary questionnaire. The goal of these standard questionnaires is to evaluate system usability and user satisfaction. The Summary questionnaire asked users to express their general opinion about the intuitiveness of the

Table 1. Tasks submitted in the Testing Phase.

Task #	Question
Task 1	<i>Define and analyze the legal context of the Root by identifying all the norms in it cited and the number of citation for each norm</i> Given the Law n. 40 of 19 February 2004, list all the norms cited by the Root and the number of citations for each norm
Task 2	<i>Make an in-depth exploration of legal context of the Root identifying the norms cited by the norms cited in the Root</i> Given the Law n. 40 of 19 February 2004, identify, among the norms cited by Root, those that contain within them the highest number of normative references
Task 3	<i>Retrace the historical evolution of the legal context taken into account identifying all the amendments introducing changes in the Root norm</i> List all the amendments to Law n. 40 of 19 February 2004
Task 4	<i>Find the way through the case law generated by the Root norm</i> Retrieve all the relevant data (case number, Judge-rapporteur, name of the parties, etc.) of the last judgment by the Italian Supreme Courte of Cassation about the Law n. 40 of 19th February 2004
Task 5	<i>Reconstruct the background of the Root norm retrieving its preparatory works</i> Retrieve the preparatory works of the Law n. 40 of 19th February 2004
Task 6	<i>Explore "by topic" and analyze the legal doctrine connected to the Root norm</i> List the number of legal doctrine papers written about the Law n. 40 of 19th February 2004 that discuss the issue of maternity protection
Task 7	<i>Identify the profiles of the Root norm that have been more discussed by the legal doctrine</i> Identify the legal profiles of the Law n. 40 of 19th February 2004 (e.g. those related to Civil, Administrative, or Labor law) that have been more discusses by the doctrine

functionalities provided by Knowlex. The administered questionnaires were made available online through Google Form (<https://goo.gl/vNEODX>).

Participants were recruited from students of a course at the Law Department of the University of Benevento, and they were not compensated for taking part at the evaluation study. Participants were also informed that all the information they provided would remain confidential.

Non-parametric tests were applied to study differences between the visualizations. The Shapiro–Wilk goodness-of-fit test was used to assess the normality of data.³⁷ Using regression analysis, we analyzed the influence of the independent variables' usefulness, ease of use, attitude toward the use, and playfulness (PU, EOU, ATT, PP), on the dependent variable behavioral intention (BI). The internal consistency reliability among the multi-item scales was examined with Cronbach's alpha.³⁸ Finally, questionnaire responses were analyzed using SPSS version 20.

Results

Results of the *Preliminary Phase* showed that the sample was mostly female (i.e. 23% male, 77% female) with a mean age of 22 (standard deviation (SD) = 3) years. The education level included 85% with a bachelor degree, 15% with a master degree in Law. Finally, 38% of the participants were acquainted with computers, while only 30% had used software tools to perform visual navigation of the norms.

In the second part of the study, we collected performance data, in terms of time to complete the tasks and the overall accuracy (error rate). We also consider the error rate as indicator of performance, since the completion time could also refer to situations in which questions were wrongly answered.³⁹ Finally, the error rate was calculated for each task by considering the number of uncorrected responses over the total number of responses. The completion time and the error counts are shown in Figures 13 and 14, respectively.

As we can see from Figure 13, the completion time for tasks performed with Knowlex is significantly lower than time obtained with the Standard approach ($p < .05$).

We did not show the results for the remaining tasks, since the participants who tested the Standard approach were not able to complete them in a reasonable time (to remain within 90 min to complete the assigned tasks²⁶). We have to emphasize that when participants used Knowlex, they were able to complete them in an average time of 109, 271, and 110 s, respectively.

As we can also see from Figure 14, the analysis of the accuracy revealed that the Standard approach implies a large number of errors with respect to the Enhanced approach. Moreover, as anticipated before, the percentage of abandonment, as shown in Figure 15, was very high for participants testing the Standard approach (as an example, anyone was able to complete the Task 7). The main reason about that was the complexity of the task (with a consequent increased time to complete it)

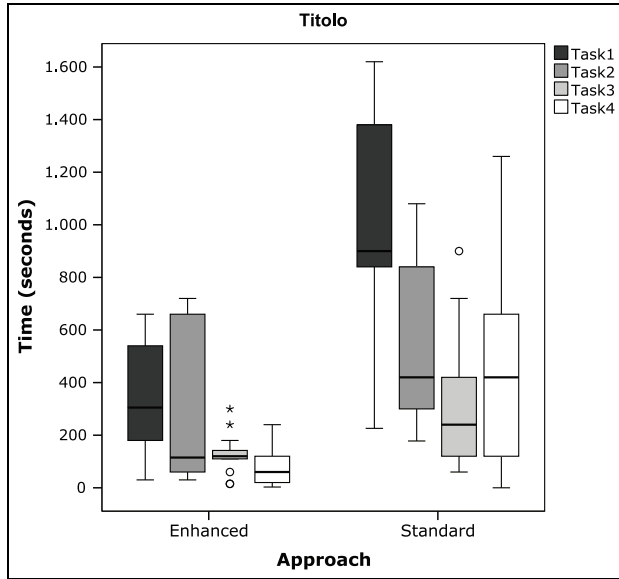


Figure 13. Performance data: completion time.

when it has to be performed without the help of any supporting system.

The analysis of the software usability showed that, on average, all posed questions were rated very positively, by highlighting how participants were highly satisfied with the software proposed (see Figure 16). Additionally, from the CSUQ questionnaire, we derived that, on average, participants have highly rated the easiness, clearness, and usefulness metrics ($M = 6.7, 6.8, 6.5$; 7-point Likert scale, Cronbach's $\alpha = .81$).

In order to identify which variables influenced the use of Knowlex, a regression analysis was carried out. The dependent variable was the BI to use metric. The independent predictor variables were the TAM subscales (i.e. usefulness, ease of use, attitude, and playfulness). The regression analysis in Table 2 shows that good predictors for the BI to use were the usefulness and the attitude toward the use.

Attitude toward a behavior is defined as an individual's positive or negative evaluation of performing the behavior. Our result shows that a positive attitude toward the proposed software as well as its usefulness could influence its final adoption.

Finally, as result of the Summary Phase, we found out that participants agreed with the intuitiveness of the Norm Reference Network ($M = 4.5, SD = .5$) and with the facilitation provided with the treemap in searching interpretation articles related to norms ($M = 4.7, SD = .5$).

Conclusion and future work

In this article, we presented Knowlex, a web application integrating two modules designed for the

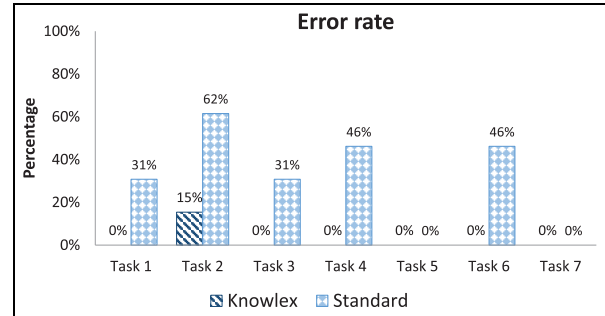


Figure 14. Error rates for both approaches. Errors were made by participants testing the Enhanced approach only for Task 2.

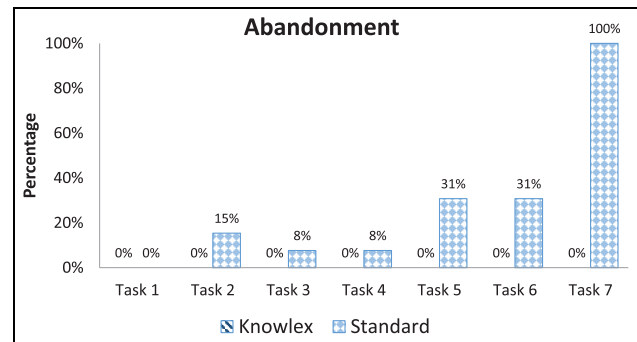


Figure 15. Percentages of unanswred responses. Participants testing the Enhanced approach completed all given tasks.

visualization, the exploration, and the analysis of legal documents coming from different sources. Understanding the legal framework relating to a given issue often requires the analysis of complex legal corpora made of heterogeneous categories of documents. When a legal professional, a citizen, or a student tries to understand how a given phenomenon is disciplined, their attention cannot be limited to a single source of law but has to be directed on the bigger picture resulting from all the legal sources related to the theme under investigation.

Knowlex exploits data visualization to simplify this activity by means of interactive maps making sense out of heterogeneous legal documents (norms, case law, legal literature, etc.). As shown by the conducted preliminary evaluation, the attempt to exploit visual analytics to support the retrieval and the analysis of complex legal corpora appears to be promising. Results of our study showed that by providing support to users through a visualization approach, they were able to accomplish their tasks quicker and in a more effective way. Moreover, attitude and usefulness were significant variables in increasing the software's acceptance.

Table 2. Results of linear regression analysis.

Predictor variables	<i>B</i>	SE (<i>B</i>)	β	<i>t</i> value	<i>p</i> value
(Constant)	1.036	2.772		.374	.718
Usefulness	−.554	.195	−5.27	−2.839	.022
Ease of use	1.082	.578	−.398	−1.870	.098
Playfulness	.444	.218	.363	2.038	.076
Attitude	1.914	.439	1.092	4.361	.002

B: unstandardized coefficient; β : standardized coefficient; SE: standard error; adjusted $R^2 = 90\%$.

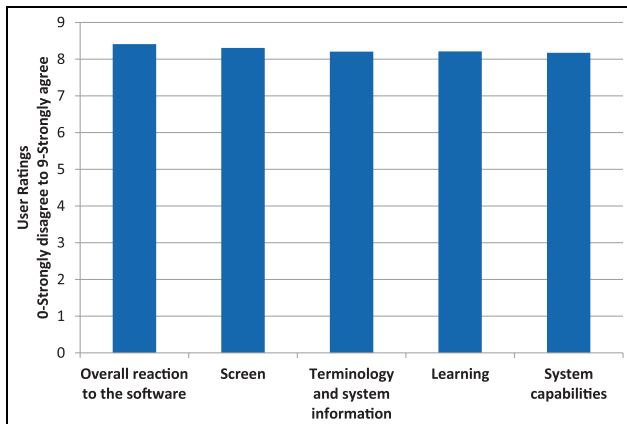


Figure 16. QUIS results. Rating on a 10-point Likert scale, Cronbach's $\alpha = .88$.

Future developments we are thinking about unfold in different directions. The first one is represented by a enhanced and more complete version of the RNN realized through the following: (1) the enlargement of the document base, so to include in the search other categories of documents (for instance, judgments adopted by lower courts or European Union (EU) legal measures) and (2) the evaluation of semantic elements making it possible to better identify content links between the different sources going beyond explicit normative references. For this purpose, we will focus first of all on the semantic metadata already associated with the documents taken into account by Knowlex (like the descriptors of the DoGi classification scheme¹⁹ or the descriptors of the thesaurus used by the Chamber Senate), a choice that could better disclose implicit relations between RNN documents. The second one is the implementation of new functionalities aiming to support the analysis of specific characteristics of the retrieved documents (e.g. relevance, evolution) and of their mutual relationships (e.g. pertinence) by means of more advanced techniques and network analysis metrics (clustering and community detection algorithms).

Finally, results of the performed preliminary evaluation are really encouraging, and for that reason, we are

currently planning a larger evaluation study to compare different visualization techniques and allow participants (domain experts and general audience), with diversified technical skills and background knowledge, to rate the best in terms of both performance and aesthetic choices. We were interested in analyzing the effectiveness of colors to convey information as well as the aesthetics choices in terms of color, shapes, and type of lines (dotted, continuous, dashed), given their role in improving efficiency, effectiveness, and attractiveness.^{40,41}

Acknowledgement

The authorship of the work can be attributed as follows: Nicola Lettieri: concept and functional design of Knowlex; legal, Computational social science and Legal informatics profiles. Delfina Malandrino and Antonio Altamura: system design, user evaluation, computer science profiles. The case study is the result of a joint effort of the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Blanch R and Lecolinet E. Browsing zoomable tree-maps: structure-aware multi-scale navigation techniques. *IEEE Trans Vis Comput Graph* 2007; 13(6): 1248–1253.
2. Ravel Law—legal research and analytics, <https://www.ravellaw.com/search> (accessed 20 July 2016).
3. lexmex.fr, <http://www.lexmex.fr/> (accessed 20 July 2016).
4. Interactive visual analytics, <http://www.ftitechnology.com/radiance-visual-analytics-software> (accessed 10 September 2016).
5. Visualizing scientific landscapes, <http://www.vosviewer.com/> (accessed 10 September 2016).
6. A tool to visualise hyperbolic trees and graphs, <https://sourceforge.net/p/hypergraph/news/> (accessed 10 September 2016).

7. Borner K and Polley DE. *Visual insights: a practical guide to making sense of data*. Cambridge, MA: MIT Press, 2014.
8. Wong PC and Thomas J. Visual analytics. *IEEE Comput Graph* 2004; 24(5): 20–21.
9. Thomas JJ and Cook KA (eds). *Illuminating the path: the research and development agenda for visual analytics*. Cambridge, MA: MIT Press, 2005.
10. Marchionini G. Exploratory search: from finding to understanding. *Commun ACM* 2006; 49(4): 41–46.
11. Duarte EF, Oliveira E, Cogo FR, et al. Dico: a conceptual model to support the design and evaluation of advanced search features for exploratory search. In: Abascal J, Barbosa S, Fetter M, et al. (eds) *Human-computer interaction*. Cham: Springer, 2015, pp. 87–104.
12. Klouche K, Ruotsalo T, Cabral D, et al. Designing for exploratory search on touch devices. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems (CHI '15)*, Seoul, Republic of Korea, 18–23 April 2015, pp. 4189–4198. New York: ACM.
13. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of the IEEE symposium on visual languages (VL '96)*, Washington, DC, 3–6 September 1996. New York: IEEE.
14. Herman I, Melancon G and Marshall MS. Graph visualization and navigation in information visualization: a survey. *IEEE T Vis Comput Gr* 2000; 6(1): 24–43.
15. Demian P and Fruchter R. Finding and understanding reusable designs from large hierarchical repositories. *Inform Visual* 2006; 5(1): 28–46.
16. Bauder J and Lange E. Exploratory subject searching in library catalogs: reclaiming the vision. *Inform Technol Libr* 2015; 34(2): 92–102.
17. libcurl—the multiprotocol file transfer library, <https://curl.haxx.se/libcurl/> (15 July 2016).
18. Francesconi E, Marchetti C and Pietramala R. A URN standard for legal document ontology: a best practice in the Italian senate. In: *Proceedings of the 4th workshop on legal ontologies and artificial intelligence techniques (LOAIT 2010)*, Fiesole, 7 July 2010.
19. Agnoloni T, Marinai E, Peruginelli G, et al. Annotation schema for legal doctrine: a case study on DoGi database. *JLISit* 2013; 4(1): 227, <http://leo.cineca.it/index.php/jlis/article/view/5480>
20. Del Bimbo A. *Visual information retrieval*. San Francisco, CA: Morgan Kaufmann, 1999.
21. Gupta A and Jain R. Visual information retrieval. *Commun ACM* 1997; 40(5): 70–79.
22. Bishop CM. *Pattern recognition and machine learning*, 2006, p. 128.
23. Loevinger L. Jurimetrics the next step forward. *Jurimetr J* 1971; 12(1): 3–41.
24. Lettieri N and Faro S. Computational social science and its potential impact upon law. *Eur J Law Technol* 2012; 3(3): 1–17.
25. Faro S and Lettieri N. *Law and computational social science*. Naples: Edizioni Scientifiche Italiane, 2013.
26. Lazar J, Feng JH and Hochheiser H. *Research methods in human-computer interaction*. New York: John Wiley & Sons, 2010.
27. Leon P, Ur B, Shay R, et al. Why Johnny can't opt out: a usability evaluation of tools to limit online behavioral advertising. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Austin, TX, 5–10 May 2012, pp. 589–598. New York: ACM.
28. Malandrino D, Manno I, Palmieri G, et al. Tailorable infrastructure to enhance mobile seamless learning. *IEEE Trans Learn Technol* 2015; 8(1): 18–30.
29. Riche NH and Dwyer T. Untangling Euler diagrams. *IEEE T Vis Comput Gr* 2010; 16(6): 1090–1099.
30. Blake A, Stapleton G, Rodgers P, et al. How should we use colour in Euler diagrams? In: *Proceedings of the 7th international symposium on visual information communication and interaction*, Sydney, NSW, Australia, 5–8 August 2014, vol. 158, p. 149. New York: ACM.
31. Al-Musawi M, Ledesma A, Nieminen H, et al. Implementation and user testing of a system for visualizing continuous health data and events. In: *Proceedings of the 2016 IEEEEMBS international conference on biomedical and health informatics (BHI)*, Las Vegas, NV, 24–27 February 2016, pp. 156–159. New York: IEEE.
32. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 1989; 13(3): 319–340.
33. Malandrino D, Scarano V and Spinelli R. How increased awareness can impact attitudes and behaviors toward online privacy protection. In: *Proceedings of the 2013 international conference on social computing*, Alexandria, VA, 8–14 September 2013, pp. 57–62. New York: IEEE.
34. Fish A, Gargiulo C, Malandrino D, et al. Visual exploration system in an industrial context. *IEEE T Ind Inform* 2016; 12(2): 567–575.
35. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int J Hum-Comput Int* 1995; 7: 57–78.
36. Chin JP, Diehl VA and Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. In: *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '88)*, Washington, DC, 15–19 May 1988, pp. 213–218. New York: ACM.
37. Shapiro SS and Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965; 52(3/4): 591–611.
38. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16(3): 297–334.
39. Alper B, Riche N, Ramos G, et al. Design study of LineSets, a novel set visualization technique. *IEEE Trans Vis Comput Graph* 2011; 17(12): 2259–2267.
40. Stasko J, Catrambone R, Guzdial M, et al. An evaluation of space-filling information visualizations for depicting hierarchical structures. *Int J Hum-Comput St* 2000; 53(5): 663–694.
41. Cawthon N and Moere AV. The effect of aesthetic on the usability of data visualization. In: *Proceedings of the 11th international conference on information visualization*, Zurich, 4–6 July 2007, pp. 637–648. New York: IEEE.