Jurisdictional Culture and Memory Digitization of the "Government of Justice." Data Modeling and Digital Approach for the Legal History of Ibero-America

Víctor Gayol[1] and Jairo Antonio Melo Flórez[2]

[1] Professor-researcher. Centro de Estudios Históricos de El Colegio de Michoacán, A.C. Martínez de Navarrete 505,
Las Fuentes, Zamora de Hidalgo, 59699, Michoacán, México
e-mail: vgayol@colmich.edu.mx
ORCID iD: http://orcid.org/0000-0002-2442-3193

[2] Doctoral student. Centro de Estudios Históricos de El Colegio de Michoacán, A.C. Martínez de Navarrete 505,
Las Fuentes, Zamora de Hidalgo, 59699, Michoacán, México
e-mail: jairom@colmich.edu.mx
ORCID iD: http://orcid.org/0000-0002-2020-1163

ABSTRACT: Can a machine retrieve the cultural meaning from a corpus of sources? This article addresses the scope and restrictions that digitization, transcription, and data modeling represents for machine-mediated readings of legal, historical records, particularly those derived from the cultural context of Hispanic empire. It compares the dichotomy between the ambiguous language of ancient regime legal texts and the unambiguity required by machine-readable files. Besides, is problematize the corporal reading and the strategy of distant-reading and visualizations as a model for interpretation of vast bulk of textual data. We propose a strategy for segmentation and data modelling for the approach the textual logic of ancient regime legal records based on their hierarchization, interrelation with non-judiciary sources (theological, historical, philosophical, etc.), its internal segmentation, the non-linear logic of the normative, and the authoritative requirements of compilations and relevant legal works. Its conclude that the advantages of automation are attached to the ability to manipulate files without distorting the original meaning of the texts, therefore, it proposes the necessity to develop standardized vocabularies that help to avoid anachronistic approaches regarding Modern Age legal sources.

KEYWORDS: Digital History; Distant reading; Web semantic; Spanish colonial law; Critical legal history; Judiciary tradition; Digitization.

RESUMEN: *Cultura jurisdiccional y digitalización de la memoria del "gobierno de justicia." Modelado de datos y enfoque digital para la historia del derecho de Iberoamérica.-* ¿Puede una máquina recuperar el significado cultural de un corpus de fuentes? Este artículo aborda el alcance y las restricciones que representan la digitalización, la transcripción y el modelado de datos para las lecturas automatizadas de registros legales e históricos, en particular aquellos derivados del contexto cultural del imperio hispano. Compara la dicotomía entre la ambigüedad característica de los textos legales del antiguo régimen y la precisión requerida para la legibilidad automatizada. Además, problematiza la lectura corporal, la estrategia de lectura distante y las visualizaciones como un modelo para la interpretación de la gran mayoría de los datos textuales. Se propone un modelo de segmentación y modelado de datos que

aborde la lógica textual de los registros legales del antiguo régimen con base en su jerarquización, interrelación con fuentes no judiciales (teológicas, históricas, filosóficas, entre otras), su segmentación interna, la lógica de lectura no lineal de la normativa, así como los argumentos de autoridad requeridos en compilaciones y trabajos legales relevantes. Concluye que las ventajas de la automatización están asociadas a la capacidad de manipular archivos sin distorsionar el significado original de los textos, por lo tanto, propone la necesidad de desarrollar vocabularios estandarizados que ayuden a evitar enfoques anacrónicos con respecto a las fuentes legales de la Edad Moderna.

PALABRAS CLAVE: Historia digital; Lectura distante; Web semántica; Derecho indiano; Monarquía Hispana; Historia crítica del derecho; Tradición judicial; Digitalización

INTRODUCTION

Decades of digitization and web publication of manuscripts and printed documents have increased the public access to historical legal records. The incorporation of digital tools, allow historians to search, download, copy, translate, and edit legal-historical sources; besides enriched their toolboxes, and present a new horizon in their analytical perspectives. Progressively, the access to almost any ancient book or manuscript, formerly safekeeping for the libraries guardians, are now openly reached via the "Infinite Archive" (Turkel, 2005). The vast "Library of Babel," as Borges and Asimov envisioned, is now a gargantuan repository that devours information at a much higher rate than humanity as a whole can consume it. Even so, the transition from scarcity to abundance has a limit. Digitalization of all ancient books and manuscripts matches a finite number of items, although its amount is still terrifying we anxiously await its conclusion.

Concomitant with the progressive process of digitalization, historians, hand in hand with other humanists, challenge the primary methods of historical science, propose new strategies to model information and expand the temporal, spatial, and documentary limits of their research. Those strategies become in the field named as Digital History (Rosenzweig, 2003: 738–739; Ayers, 1999; Lines Andersen, 2002; Cohen and Rosenzweig, 2006). Humanist consider Digital History as a branch of the Digital Humanities, and historians regard it as "an approach to examining and representing the past that works with the new communication technologies of the computer, the internet network, and software systems" (Seefeldt and Thomas, 2009). Initially conceived as a strategy for democratization of history through gathering, preserving, and presenting the past on the web (Cohen and Rosenzweig, 2006), in the short time, Digital History moves his focus towards quantitative, culturomics (computational lexicology), and linguistic analysis of digital historical sources, what made the limits with the digital humanities increasingly diffuse[1].

Although digital history persists as a computational approach within historical science, the initial expectation has been weakened over the years, mainly due to its inability to reach new conclusions to old questions or at least generate innovative inquiries to the past (Weingart, 2016). Our hypothesis considers that which prevents digital history from generating significant queries or provide new answers is not necessarily their technical or methodological weakness. The intricacy lies in the very material that nourishes the historiography, formed by the varied types of historical documents that the historian face in his daily work. Quantitative and digital historians confront the same issue: convert the text-based data into a machine-readable or electronic format. There is no surprise then that the most successful results in digital history are those encompassed with quantitative history methodology (Anderson, 2007: 249–250). Digital historians work with data subjected to serialization: vital records, commercial information, travel routes, court trials, slave sales, migrations, and so forth. Perhaps, the distinction between quantitative and digital history resides in the non-numerical semantic approach to data, in other words, the necessity to understand how a machine can retrieve natural language from a historical perspective, i.e., the use, usefulness, and interactions between language and the subjects from the past. Therefore, one of the challenges of digital history lies in the computational treatment of unstructured or semi-structured data that can reveal those cultural trends inside segments within the measureless digital archive.

This article addresses the idea of "corporea" readingl[2], understood as the reading of a corpus "that cannot be read by a human" (Michel et al., 2010: 176), beside the conception that manipulation of massive textual corpora lies in "new approaches to understanding language, literature, and culture." (Kayman, 2016: 356). In recent years it is evident the improvement in the technics for machine-reading the digital archives, retrieve information from selected repositories, and apply computational methods for historical analysis (Graham, Milligan and Weingart, 2016; Turkel, 2015). Meanwhile, the necessary skills required to explore the vast sea of digitized information become more specialized for the average historian[3]. The corporeal reading also involved the development of abstract representations of the analysis results from the infinite archive. In that regard, Franco Moretti's (2007, 2013) idea of a distant reading has permeated, not uncritically, into the mainstream of digital humanities. His assumption of "explanation over interpretation," mediated by graphics, maps, and trees; appears to appeal to blackboxing, "a

process that makes the joint production of actors, and artefacts entirely opaque" (Latour, 2000: 183). Without trying to go too deeply into this idea, this article modestly aims to contribute to the discussion about digital hermeneutics and digital history.

This work scope some exercises and reflections made as part of an ongoing project focused on the construction of a digital corpus of Hispanic-American legal sources with the purpose to explore which possibilities have automatized analysis in the study of jurisdictional culture. Specifically, it examines some general technics of automatized textual analysis (world clouds and trees, KWIC and topic-modeling), the potentialities of data modelling, specifically the XML-TEI schema, and a preliminary approach to Linked Open Data (LOD). Those explorations are still in a preliminary stage of experimentation, although it takes advantage from former reflections related with the analysis and reading of digitalized sources wrote in Castillian language (Gayol, 2016, 2017b; Gayol and Melo Flórez, 2017; Melo Flórez, 2017). Finally, it problematizes the possibility of distant reading in long-term perspectives, as proposed by Jo Guldi and David Armitage (2014), mainly from the perspective of the Iberic-American legal culture from 15th to 18th centuries.

DIGGING INTO LEGAL HISTORICAL CORPUS

One of the central aspects of the practice of the critical or cultural legal history lies in the method of approaching sources, their analysis and interpretation[4]. The objective of this historiographical perspective pretend to reconstruct the history of political power and its dispositions or the discourses and practices of political power understood as a phenomenon located beyond the restrictive framework of the statist paradigm. The considered "classical" approach to legal history, preconceive the past from a teleological perspective where the ancient regime normative order was only an imperfect or obsolete version of the contemporary judiciary system that unavoidably evolves towards the current legal model (Garriga Acosta, 2004a: 4). Therefore, is not a surprise that the primary sources of the traditional legal history have been restricting to "law" texts (*cédulas*, *reales órdenes*, *ordenanzas*) and to a lesser extent the criminal trials, jurisprudence, and authoritative works (knew as *doctrina*).

To overcome the anachronistic interpretations of judiciary texts António Manuel Hespanha, inspired by Clifford Geertz's ethnographic method and the studies of Pietro Costa and Paolo Grossi, proposed a "thick reading" method for approaching legal sources. This strategy, that attempts a close and contextualized reading of yesteryear legal documents, including a series of principles formulated to "respect the logic of the sources," which is represented by the underlying sense of which appears to be evident in the documents. As Hespanha remarks (2002: 45), "it is must strive to recover the strangeness of what is said and not the familiarity." Hence, the exploration into the deepness of the documents can be achieved through the reading, rereading and constant interrogation of each

possible "evidence" found into texts with the purpose of revealing its logic rather than imposing reader presumptions. Such aspiration, apparently close to the method of the vituperated positivist historiography, paradoxically constitutes an argument against legal positivism, since it is not the statutory text itself that speaks about its logic but rather the cultural context within which was immersed which reveals its possible original meaning (Hespanha 2002: 43-45).

By placing the texts in their production and appropriation contexts, the legal document is "released" from its strictly functional box (the law interpreted regarding its application), thereby recognizing interdependence with other areas of normative meaning, such as the theological, the moral and the ethical. The legal order, understood then as a formula to guarantee the order of society (*aequitas*), acquires a meaning that is foreign to the contemporary reader but close to the lifeworld of the inhabitant of the vast Modern Age Hispanic empire (Hespanha, 2002: 30–57, 1987). By applying the abovementioned principles, the critical legal historians have developed the expression "jurisdictional culture," understood as "a manner of organization and power handling verified with few changes in all European political spaces since Low Middle Ages through the late 18th century (Agüero Nazar, 2007: 24)." In that regard, justice works as a series of actions that transcend the merely legal to involve it within the exercise of political power.

Pietro Costa's research of the medieval concept of *iurisdictio* allowed identifying the key for understanding the judiciary and political power during Middle and Modern Age in the frame of the *ius commune*, which was the jurisdictional authorities with the *potestas* to say right (*ius-dicere*)[5]. In Costa's terms, *iurisdictio* denoted the semantic axis which arranges the experience –from its cultural and historical context– of the use of political power towards the Middle Ages[6]. Those were jurisdictional powers defined by the faculty of each authority to say –or determine– the applicable law. Thus, the memory of the practice of authority lies in the texts that recorded the exercise of the various *potestas*. Therefore, critical or cultural legal history comprise the exercise of the hermeneutical understanding of historical legal texts –treatises, normative, or documents generated by judicial practice such as trials records– embedded into semantic networks centrally linked into the concept of *iurisdictio*.

According to Bartolomé Clavero (2002), Vallejo (1992) and Hespanha (1997), jurisdictional culture prescribed by medieval writers prefigures a variable geometry of the act of power. Consequently, the memory of middle ages process of power resides in the corpus of the legal texts who transcends the epochal boundaries defined by the historiography and have sense and use into the Modern Age, not only in Europe but also in their imperial extension. Unlike the logic of contemporary law, which is mainly prospective, the right of middle and Modern Ages took from the biblical account its historical principle and the search for original order. The right embodied in the normativity was a part, substantial but not unique, within

the normative order that sought to maintain and recover the natural order (*ordo*). Besides, for readers of the ancient regime, the survival of one particular law or rule was considered as evidence of its wisdom, an indication of the reflection (*speculo*) of the divinity and its underlay reason. Therefore, a proverb, an adage, or a biblical verse, could have similar "value" to the legal text since they came from the same origin: the divine creation (Garriga Acosta, 2004b: 31).

Semantic approach from ibero-american judiciary culture

Besides the most technical approach of historical legal semantics (Costa, 1969) it is necessary to point here what it is identified as the main characteristics of the legal textuality of the Modern Age: his ambiguity, the tendency to "decontextualize", and the coexistence between oral and writing culture in judiciary tradition. Unlike contemporary legal systems, in Modern Age legal tradition technical language shared his power with traditional knowledge; therefore, a single trial record can share in its several stages different approaches to legal knowledge. Some lawyers spent hundreds of folios only to display its in-depth expertise in a particular matter, abound Latin excerpts and quotations from doctrinal authorities, and profusely comment specific laws selected from compilations, a royal charter, or a papal bull. However, typical proceedings frequently resorted to most prosaic language, although following some general rules like the unavoidable oath to god and the king, gestures and rituality's associated with the judiciary process, scribe formalities reflected in the structuration of the document, and others[7].

Despite recognition of semantic and semiotic values of legal language for understanding the ancient regime judiciary (Costa, 1972), merely technical-linguistic approaches are uncommon. Semantic networks are expressed but rarely presented, and semantic changes are still studied almost exclusively by philologist. That scenario reveals the separation between history and linguistics in Ibero-American legal history, which cautiously avoids linguistic technicality and addresses the European legal textuality of the low middle and Modern Age from the perspective of discursive regimes (Vallejo, 1992: 153) or discursive systems (Hespanha, 1987: 538–544). As is known, those are approaches proposed by critical discourse analysis which addressing to language in use within social interactions usually related with power intersections (Martin, 1992; Scollon, 2001). Is not surprising whatsoever that the critical legal history moves away from the functional and structural perspectives of linguistics that, as Jesús Vallejo remarked (1992: 38), represent the most significant contribution and at the same time the main reason for the "neglected forgetfulness" of Pietro Costa's work.

António Manuel Hespanha proposed a "discursive analysis" supported in a "distant" not-trivializing reading of sources but is not a technical model for analysis of text or discourses. Instead of linguistics, there is an ethno-graphic approach to legal historical text (Clavero, 1986, 1991; Hespanha, 2002: 42–43) and further contributions from conceptual and "critical-conceptual" history that overcomes a traditional history of ideas and definitions (Costa and Zolo, 2007; Petit Calvo, 1995). However, linguistics is not explicitly rejected by legal historians, hermeneutics is still a widely accepted strategy to understanding legal historical sources, and conceptual history is considered likewise as a semantic of historical time (Koselleck, 2006, 2015). Besides, semiotics allows expanding the legal discourse from merely textual expressions towards images, gestures, sounds, and so forth, which allows legal historians to talk to the history of art and culture.

The challenge for legal historians resides in the semantic innovation not reflected in language that affects the pragmatic use of terms and expressions, but also lies in the capability of repetition of a concept that embraces previous experiences, which besides allows and additionally restrict the discursive field of action of each expression (Koselleck, 2006: 30). The hurdle to overcome for an automation of the reading process lies in the construction of linguistic models from and for the corporal reading of the legal texts of the old regime, without evading in the process the discursive interconnection with the illiterate culture, the ethical-moral normativity, the political power (*iurisdictio*), the symbolic assessment of legal principles, among other complex manifestations of the jurisdictional culture that could not be overlooked in a macro-analysis process that seeks to go beyond mere disclosure (critical or not) of documents.

Corporal order, corporal text

The concept of iurisdictio configured a "corporeal" shape for Spanish monarchy. From ancient regime symbolic, the anthropomorphic metaphor of society was fundamental to understand the harmony of the political "body." Metaphor represented the Prince as "head" of society which "arms" were the clergy and nobility. The body of the king (as depicted in Hobbes Leviathan frontispiece[8]) symbolically was constituted by the states. Well organized represented the health of the monarchy as the organs of a healthy body; corrupted, perfidious, rebellious states represented the harm of the empire, its ruin (Martínez, Beck Varela and Agüero Nazar, 2012: 104; Kantorowicz, 1997).

Is corporeal shape reflected in legal corpora? Compilators structured books beginning with ecclesiastical followed by the temporal organization, which constitutes a reflection of the two "arms" of the monarchy. Both organizations represent the main *fueros* (jurisdictions) of the monarchy (ecclesiastical and secular), followed by the *foral* rights, the privilege of each state to be judge by its own courts (Sánchez, 1834: 45–46; Hespanha, 2002: 68–69). That jurisdictional order develops a cornucopia of laws presented in different shapes and authority. Local laws overlapped with corporative judges, royal cédulas rivalled with papal bulls, bandos and autos for good gov-

ernment defined legislation for governorates, smallest as a province or biggest as a viceroyalty, besides, in America the republic of Spanish has its own justice and government which produced ordinances relating to *politeia* and good governance of the cities, towns and villages; however, republic of Indias was subject to legislative dispositions produced by several jurisdictions like governors, viceregals, councils and the King (Barrientos Grandon, 2004).

Therefore, a digital corpus of Hispanic-American legal sources must reflect those corporal characteristics in order to accurately represent the judiciary culture of Modern Age Spaniard domain in the new continent. As most of those legislative dispositions repose in local archives and only a few were printed, it is expected that the process of digitalization and transcription of related regulations take a longer time. In any case, proposals for data modelling and hierarchization of local provisions for general, regional, and local archives can represent an advance in future digitization endeavours.

INFORMATION RETRIEVAL FOR LEGAL CORPORA

One of the issues associated with digitization projects, particularly for the Ibero-America scope, consist in their object approach: many of the documents change their format from paper to digital, but instead of became a machine-readable text there are transformed into digital images. The public efforts for the publication of documents focus on the construction of databases (documentary and bibliographic) that are then disseminated through tools that mimic the structure of the file (sections - funds - series - bundles) but neglect the logic of the document as text itself presenting it as a digital object unreadable to the machine. Besides, the development of digitization processes and the improvement of Optical Character Recognition (OCR) technology have followed different paths, in that regard, libraries subjected materials digitized in the early stages to OCR technologies that are now obsolete, which repercutes in the accuracy of transcription and recognition of Latin characters and old fonts. Therefore, it is necessary to recover records through metadata schemas built by contemporary archivists who fill the element texts within these schemas with the already made descriptions available from previous cataloguing instruments, made for paper-based documents, or by new information recovered from the records identification elements. In that regard, the inquiry for information is made not in the text itself but in the metadata associated with it[9]. Consequently, the machine is unable to read even the superficiality of the records.

Without detracting from the effort and investment made by a multitude of individuals and institutions, it is clear that the digitization process remains at a primitive stage that requires redoubling efforts to allow information retrieve and textual automatized analysis in Ibero-American sources. Foreword scenario for Hispanic-American legal sources implies that a possible macroanalysis of le-

gal document demand a previous task of segmentation, modelization, and interaction of data starting from the smallest linguistic units and progressively make machine-readable paratextual elements of the documents itself and even incorporate experimental schemas for conceptual reading.

Segmentation

Previous to textual analysis the text requires being tokenized, segmented into linguistic units with the intention to know the metrics of the sources (Mikheev, 2005). This process has the purpose of group alphanumerical characters into words, differentiate types (number of different words into a corpus), the frequency of each word represented as tokens, and apply stemming (reducing tokens to their stems) and lemmatization (inflected forms of a word). Therefore, at this stage, the analysis is reduced to the basic structure of the text, its construction and weight measurement of its elements. This preliminary approach to the logic of textual and data analysis consists in applies technics who allows to "scratch the surface" of a corpus of documents. The idea of superficiality in digital humanities does not necessarily presuppose a deficient application of computational techniques in historical sources; it is only the reflection that "[e]ven as digital text became more readily available, the computational methods for analyzing them remained quite primitive" (Jockers, 2013: 4). Even so, the use of computational technics designed to analyze big data can represent a possibility to infer some patterns that could escape regular reading.

With the purpose to evaluate the options of those technics applied into historical legal documents, we take advantage of a corpus of court fees derived from three series created for the Mexico *Real Audiencia* already processed for printed publication. The preparation of the text suffered the traditional process for a transcription and printed commented edition, with the addition of explicative footnotes, a glossary, a index, and an introductory study (Gayol, 2017a). Subsequent task corresponded to make those records machine-readable, migrate the data from word processors and spreadsheet software to plain text with the purpose of being destructured and represented with computational strategies. Lastly, these results can be quickly validated thanks to the previous work of presenting the documentary collection.

The set of documents is compound with three series of court fees (aranceles) formed by the royal Audiencia to arrange the collection of professional payments intended for the delivery of the services for several public officials who intervened in the procedures carried out in various courts and offices. Two tribunal judges (*oidores*) constructed the first series, between 1697 and 1699; a *junta* (a provisional governing body) created *ad hoc* by a royal cedula in 1738 redacted several court fees who formed the second series for the corpus, published by proclamation (*bando*) since 1741 till 1759. In 1759 the marquess de las Amarillas compiled and ordered the printing of the court fees with the addition of some unpublished records.

The third series was the work of the regent Herrera who introduces court fees for the collection of judicature rights of governors, *corregidores,* and *alcaldes mayores*, besides other officials from outside Mexico City. The regent ordered the print of the group of court fees in a single volume, dated March 29, 1784. For 1833, already in the independent Mexico, some anonymous editors published the series from 1738 and 1784 almost in its whole with some modifications (the suppression of some deciduous occupations), with the explicit purpose of guiding the establishment of rights and wages in the courts and offices of the Mexican government.

In pursuit of practicality, the application Voyant Tools (https://voyant-tools.org/) served as an application for the initial experimentation. The developers of this software describe it as "a web-based text reading and analysis environment,[10]" and digital humanist widely used it thanks to the "plethora of text visualizations" that is unfolding for almost any text encompased (Fankhauser, Kermes and Teich, 2014; Sinclair and Rockwell, 2012, 2016; Welsh, 2014). The interface is pretty intuitive and allows the user to visualize its analysis in a straightforward mode through word cloud, frequencies of terms, document segments, text correlations, word tree, and others. It also presents a summary of the corpus segmentation with a count of words and tokens (unique word forms), and a primary semantical analysis represented in vocabulary density, average word per sentence, and a list of frequent words in the corpus who works as a complementary to the word cloud. It also performs the setemming and lemmatizing process in the background, which allows users to perfom queries by using wildcards.

There is necessary to perform some previous task to represent data most accurately because the automatized tokenization does not fulfill the segmentation of those terms written in old Spanish. The accentuation of conjunctions (/á/, /é/, /í/, /ó/), the use of consonants as vocals and vice-versa (for instance, /vno/ as /uno/), the app interpretation of Roman numerals as letters, are some of the issues manually solved to obtain most precise visualizations. Voyan Tool's cloud representation barely indicates some trends in the logic of the corpus, most of them irrelevant. For instance, in the visualization protrude terms as *peso*, *real*, and right, but are at a glance evident, even without reading the document, it is what the reader presuppose for a corpus of court fees and collection of taxes. Nevertheless, is not necessarily an application problem but one derived from the size of the corpus. Several studies evidence that word cloud properties (font size, weight, and color) have an essential effect in the average user (Heimerl et al., 2014), therefore, enriching the text with a semantic descriptive code could enhance the results given by those visualizations.

Manipulation of the filters brings new results. Excluding the most frequent "types" for the analysis provide new insights for the corpus providing the following count: *partes* (513 times); *autos* (415); *personas* (336); *audiencia* (317); *oficiales* (312); *escrivano* (311); *san* (305); *oficial* (297); *indios* (296); *escripto* (291); *escrito*

(276); *oficio* (268); *parte* (268); *aranzel* (254); *cosa* (253). Take advantage of stemming and lemmatizing allows to group some terms that differ only from its writing or quantity (*escripto-escrito*, *oficiales-oficial*, and *parte-partes*) and bring to viasualization a new shape. However, can these preliminary results represent any nature of the corpus? The function of the tariffs for the collection of rights was related to a series of judicial or governmental litigations that were carried out in courts. That judiciary acts produced proceedings (*autos*), documents (*escritos*) and testimonies (*testimonios*), with the intervenience of scribes (*escribanos*), officials (*oficiales*) and the court (*audiencia*), and the opposition of litigants or requestors, legally differenced by its rights (persons and Indians) although in many times named as *partes*. That last expression can be associated with a person or can reference the usual construction of the documents (x lines for y parts).

Disambiguation, therefore, requires another strategy. For instance, a term like /san/ denotes a religious meaning as apocope of Saint, but in context, its connotation is revealed as a toponym associated with a spatial political organization like a town or congregation (i.e. san Juan, san José). Some tools can help to put in context some ambiguous or polysemic words resulted from the tokenization, likewise the KWIC or word trees, nevertheless use some additional strategies is required in order to retrieve discursive context and identify homonymies and polysemies. KWIC was developed as an automatized mechanism to find concordance in a book or a library collection to create a searchable, alphabetical, contextual index (Fischer, 1966) and its one of the tools to contextualized and find trends in automatized queries proposed by the *Programming Historian* team (Gibbs, 2015; Turkel and Crymble, 2012a, 2012b; Wiegand, Mahlberg and Stockwell, 2017).

Raw application of KWIC represents some advantages regarding a cloud word, like the possibility of grouping terms[11]. For instance, the stem /parte/ in a simple KWIC query result in 584 coincidences, besides sorting it by the coincident words at left and right will result in a practical way to find coincidences (Gibbs, 2015). Additionally, the identified polysemy of *parte* as subjects and as a piece of the document, it is evident the lack of coincidence between the books of court fees from 1699, 1727 respect to the books of 1738 and 1784. The same result is obtained from other queries with /*justicia*/ and /*indio*/, moreover, the term "Indian" is predominant in the book of tariffs of 1727, as a subject of rights tied to the community, but also associated with the position "*cacique*" and in others with the quality "*mazehual*". The book of 1738 presents a more direct connection between the Indian and the *cacique* ("*Caziques y demás indios*", "*indios y caziques*"), while the book of 1784 deals with the communities of Indians and their caciques ("*Caziques ó comunidades de indios*").

It is evident that segmentation of a few documents can provide clues that lead to a careful review of certain aspects that did not necessarily caught the attention of the reader at first. Our example used a text previously read and analyzed by the usual historiography methodology which facilitates its validation. Prior reading compared with the

text processing reveals the evident, but that is a preliminary confirmation that the processing technics can be helpful to locate what is foreseeable and what is atypical in a group of historical sources (Robertson and Mullen, 2017: 15). At this stage of digitization for Ibero-American legal sources, it is possible to "scratch the surface" from the big data gathered into repositories like *Google Books*, *Europeana*, *Biblioteca Digital Hispánica*, among others which present sources with searchable text transcriptions. Maybe not all the queries and exercises reveals "hidden patterns", notwithstanding, the discoveries made by using automatized technics, like topic modeling, can be submitted to the human inquiry to be validated or further explored (Ophir, 2016; Melo Flórez, 2017: 172–178).

Besides, a quick comparison between the segmentation of Spanish and English texts (mainly stemming and lemmatizing) reveals the difficulty of these prefabricated applications in interpreting the historical logic of the Castillian language. Lexical examination of legal text requires, therefore, the application of customized software. Fortunately, each time more projects share their codification through repositories like GitHub, which allows new projects to learn from previous experiences and collectively propose alternatives to textual analysis from historical texts. Simple transcription and object identification by metadata enable user access to materials sheltered in national or even regional and local archives and libraries, therefore the application of advanced treatment to data could represent not only the possibility to open access but also of interpretation and correlation between records in larger scales.

Semantic Data Modelling

Data modelling define the strategies used in informational sciences to organize and standardize relations between data. Broadly, this term refers the structure of entities-relationship and its set of concepts (entities, attributes, relations, tables). In this article, we cover some concepts defined for the Semantic Web, especially the XML markup, the XML/TEI schema, and technologies like RDF and OWL. Our intention is merely problematizing the possibilities of semantic data modelling in the construction of a digital corpus of historical legal sources[12].

Roughly speaking, digital textual modelling represents the strategies and technics applied to single documents or corpus to make its structure machine-readable and to allow the retrieval of structured information via query methodologies. The possibilities of modelling go beyond the technology used and are sustained by the complexity of the model. Undoubtedly, the choice of programming language or project approach (f.i. server-side, static web) will affect the possibilities for retrieve and handle the information previously modelled. Therefore, segmentation in conjunction with previous knowledge of the textual nature of the corpus represents necessary prior steps for semantic data modelling.

A digitalization project for legal sources must take into account how to make texts readable by the machine, not only for the OCR technology used in the process. Maybe most important that make characters recognizable is the decision of how to structure text which was intended to be printed in a specifical way and unbind them to being represented in plain text and web readable code. It is relevant to understand how public presentation in this regard is relevant as the capability for the machine to recognize a standard vocabulary for the structure of text: titles, authors, creators, dates, publisher, and language, among others that are identified unmistakably in the document itself. In that regard, Dublin Core or another metadata set could be useful to bring interoperability to each bibliographical object. Nevertheless, even taken the objects in its full dimension – as printed or manuscript books – the logic of the layout differs significantly from the logic of the computers.

A representative example is the Old Bailey Project, which is composed of 125 million words in an archive of 197,000 court trials covering roughly 250-years (1674 to 1913). There is possible to perform data mining technics in the corpus because its documents were structured in XML base which allows identifying not only tokens but also specifical meanings of words and phrases, tagging categories of crimes, punishment, verdicts, names, between others textual elements which will enable to perform visualizations and then reveal some "hidden patterns" that could inspire new interpretations (Cohen et al., 2011; Kelly, 2013). Codification process concluded in March 2012 (version 7.0) and since then the team released two minor upgrades, version 7.1 in April 2013 to correct tagging errors and broken links, and version 7.2 in March 2015 which equally corrects some punctual tagging errors product of mistranscribed data or blunders during automatized tagging tasks (Hitchcock et al., 2012). Researchers can retrieve data through the Old Bailey API (OBAPI) developed by Jamie McLaughlin, and users can perform queries with its demonstrator interface, which allows integrating Old Bailey data with other repositories, projects, and studies, besides to perform advanced task with machine learning methods (Cohen et al., 2011; Hitchcock and Turkel, 2016; Turkel, 2008; Watkins, 2015).

The advantage of XML is represented in its capacity to describe the original format of books and manuscripts (size, volume, medium), the original organization of records (books, volumes, papers, folios, codices, scrolls), the hierarchical structure of documents (chapters, paragraphs, pages, lines), particular caracteristics that define the nature of source (f.i. books, titles, articles, numerals), and expand the description to identification or conceptual elements in text (names, relationships, locations, qualities, etc.). Although XML is for definition an extensible language the description structure must be represented as a hierarchical tree, therefore, no element can be represented as independent from the root. Hence, documents must fulfill the conditions of being well-formed and valid to be understood by the machine as a hierarchical sequence of elements (Birnbaum, 2017). Structure complexity also depends on the objectives of the project. A detailed description of the elements of the original docu-

ment (pages, images, stamps, rubrics, spots, capitals) will require a higher depth and consistency in the labeling of these characteristics.

However, structuring a well-formed document does not imply that accurately reflects the intent of each section of the text. The labels are constructed to address a significant generality of textual elements, but the modeling would have to question the label exactitude to accurately define the nature of each part of the document. For instance, Gregorio López intention for its glosses at Siete Partidas is not as footnotes but as marginal authority commentaries to text that extend, interprets and constitutes a new document into the legal codex[13]. Anthony Grafton recalled that the glosses are not identical to footnote as a form of annotation, "historical footnotes resemble traditional glosses in form. But they seek to show that the work they support claims authority and solidity from the historical conditions of its creation" (Grafton, 1997: 32). Describe accurately the book layout is problematic as encode glosses as footnotes because would imply modifying the original intention of the glossator whose interest was to probe the divinely inspired, wisdom, and antiquity of the judiciary text (Hespanha, 2002: 110). The question is, how to respect the logic of the original text making readable for the machine?

As previously mentioned, the choice of a scheme for semantic modeling is fundamental for the development of a project, considering the scalability and possibilities of migration. XML is undoubtedly the most widespread technology currently in the field of digital history, but this does not mean that it is the only one, or the best, technology. Other options such as JSON and YAML are more straightforward and interact better with technologies such as Python or PHP-frameworks such as Symfony. Also, within XML the TEI consortium developed a set of Guidelines to represent the texts on the web adequately. Furthermore, RDF coupled with the OWL vocabularies (RDF+OWL) can be structured independently or as an XML structure with the OWL vocabularies functioning as schemas. As Ciotti and Tomasi remark (2016) there is anything but a lack of consensus about what is the most suitable technology for semantic modelling.

Currently leading solution proposed for digital edition consists of the application of XML-TEI schema (TEI Consortium, 2017). The use of description frameworks or schemas enjoys a known tradition in digital humanities, especially in the fields associated with digital publication and semantic web explorations. Critical digital editions became an alternative to transform digital historical text that in its digital version contains variations, additions, amendments, sources, altogether with another paratext included by the editor or copyist. As Antonio Rojas (2017: 6) remarks "the purpose of critical edition is that readers can reconstruct the history of the text and appreciate the editorial interventions." In that regard, digital critical edition reveals to e-readers the process of creation of the text and the relevance of the editorial process in their construction.

Nevertheless, digital critical edition represents a further edition of the document which allows readers to visualize textual shapes and structures from several (all or a representative sample) editions of the same work. One example of digital critical edition applied to a legislative corpus is represented by the *7 Partidas Digital* project, led by an interdisciplinary researchers team from Spanish, American and British universities[14]. Due to the complexity that implies use OCR technologies over manuscripts and early printed books, the team chose transcript directly from the digital images, which allows coding and transcription to be carried out concomitantly. The project embraces TEI as the encoding schema and constructs the critical digital edition for the *Siete Partidas*.

The structure for each *Partida* is as follows:

```
<text>
  <body>
    <gb n="B"/>
    <pb n="258r"/>
    <fw type="encabezado">Texto encabezado</fw>
    <cb n="a"/>
    <div n="1,2...7" type="partida">
      <div n="x.x" type="titulo">
        <div n="x.x.x" type="ley">
        <p>
          <lb></lb>
          <hi></hi>
          <del></del>
          <corr></corr>
          <supplied></supplied>
          <ex></ex>
          <foreign xml:lang="la, ar, gr"></foreign>
          <persName></persName>
          <placeName></placeName>
          <title></title>
          <sic></sic>
          <gap/>
        </p>
      </div>
    </div>
  </div>
  </body>
</text>
```

The structure is quite simple, but it reflects the hierarchy of the document: *Partida*-Title- Law unequivocally. Likewise, within the paragraph element (text of the law) a series of attributes that allow describing with greater depth, such as abbreviations, are listed. <ex>, languages <foreign xml: lang = "">, personal names <persName>, places <placeName>, errata <sic>, among others. A similar structure is adequate to be applied in several compilatory legal books (*Siete Partidas*, *Nueva Recopilación de Castilla*, *Recopilación de Leyes de Indias*, *Novísima Recopilación*, even early Constitutions), besides legal dispositives like the abovementioned books of court fees, ordinances, or instructions to officials appointed to Indies. Complex models are represented by doctrinal or commented legal books widely used by judges and lawyers towards Spanish rule in America. In order to

addapt the structure above proposed, we encode the glosses of Gregorio López following the next structure:

```
<text>
  <front>
   <p/>
  </front>
  <body>
   <div n="1" type="partida">
    <div n="1.1" type="titulo">
     <fw type="encabezado">Título...</fw>
     <div type="prologo" xml:id="prGLp1t1">
     <p>
      <name type="deity"/>
      <span type="norma"/>
      <span type="iurisdictio"/>
     </p>
    </div>
    <div n="1.1.1" type="ley" xml:id="GLp1t1l1">
     <head>Ley primera...</head>
     <p>
      <term xml:id="glossGLp1t1l1c" rend="sc">
      a la buena vida </term>
     </p>
     <div type="glosa" xml:id="glossGLp1t1l1">
      <p xml:lang="latin">
       <gloss target="#glossGLp1t1l1">
       <emph/>
       <name role="filósofo" type="Aristóteles"/>
       <sic/>
       <corr/>
       <title/>
      </gloss>
      <cit>
       <quote/>
       <note>
        <bibl/>
       </note>
      </cit>
     </p>
    </div>
   </div>
  </div>
 </div>
</body>
</text>
```

In our example, the structure differs concerning the assumed by the "7 Partidas digital" project. Its preserved the nested structure (P-T-L) with slight variations as in the case of the element <front>, dedicated to containing the *Partida*'s foreword, the @type "prologo" (prologue), and @xml:id for each <div> in order to reduce ambiguity. The main variation consists in the addition of @type "glosa" and elements <gloss> and <cite>, both nested inside @type "ley," and linked with its note by a @target to @xml:id. The decision to nest the glosses into the laws is due to the necessity to remark that glosses, in this case, are at the same hierarchical level as the legal text. When a

reader consulted Gregorio Lopez' glosses to the Partidas, they were viewing both documents at the same time, the statutory reference and the authorized interpretation given by the commentator.

Considering that the interest of the legal corpus project does not correspond to a critical digital edition, the attention focused by pointing out some semantic indicators identifiable in the text. For instance, its used the element <name> with the @type "deity" to represent the names of God and Jesus Christ. To betoken a law or norm, we use the element <span> with the @type "norma." The Same procedure was carried out with the term "iurisdictio" relative with expressions of political power. Those labels correspond to experimental identifications that will be subjected to more detailed analyzes but are currently functional to demonstrate the possibilities and limitations of the TEI labeling to describe the language of the Hispanic jurisdictional culture. According to Øyvind Eide (2014) there are at least six methods to include ontologies into a TEI document using the <relation> element which allows enhancing descriptions by using RDF-OWL ontologies. Nevertheless is a sophisticated method that still needs to be tested, but the perspective of success will allow integrating into TEI documents vocabularies like FOAF, LKIF Core, Bio Vocabulary, and even experimental ontologies for historical documents (Adorni et al., 2015).

As is noticed, glosses represent a challenge to describe in a standardized language. First, López wrote them in Latin, which represents an additional hurdle for an accurate description. Caroline Barriére (2016: 18–19) remarks the intrinsic difficulties that imply approach data from multilingual sources yet with RDF. We attempt to solve that issue with the help of the global attribute @xml:lang in the element <p>. Besides textual elements like emphasis, sic, correction and title, its included the identifier <name> with the attribute @role and @type in order to describe who in the gloss number three is called as philosophus. Next table shows original gloss text and its encoding:

The gloss is structured in several elements:

1. Gloss identifier: "3c"
2. The call phrase: "A la buena vida."
3. The identification of authority: "Dicit philosophus"
4. The doctrinal reference: "9. Ethicorum"
5. And, the quote: "plegalia iusta dicimus factiva, & conservativa fœlicitatis, & particularium ipsius, politica communicatione."

Like most of the early Modern Age treatises, the use of punctuation signals was a stylistic decision of the author or copyist; therefore the reference to an authority trigger the recognition to a possible textual or paraphrased quote. Nevertheless, the quote is not an Aristotle's Nicomachean Ethics textual note but an excerpt from Thomas Aquinas' Summa Theologica. Therefore, after describing the language opened the element <gloss> which have as attribute the @target. Include the name of the author, Aristotle, as an attribute @type for the element <name>

TABLE 1. Gloss text and econding gloss in TEI schema.

Gloss original text	Gloss encoded
3 c. A la buena vida. Dicit philosophus. 9. Ethicorum, plegalia iusta dicimus factiva, & conservativa fœlicitatis, & particularium ipsius, politica communicatione.	\<p xml:lang=”latin”>\<gloss target=”#glossGLp1t1l1c”>3 c \<emph xml:lang=”es”>A la buena vida.\</emph> Dicit \<name role=”filósofo” type=”persona” key=”Aristóteles”>philosophus\</name>. \<sic>9.\</sic>\<corr>5.\</corr>\<title key=”Ethica Nicomachea”>Ethicorum\</title>,\</gloss> \<cit>\<quote>plegalia iusta dicimus factiva, \<abbr xml:lang=”latin”>&\</abbr> conservativa fœlicitatis, \<abbr xml:lang=”latin”>&\</abbr> particularium ipsius, politica communicatione.\</quote> \<note>\<bibl>Tomás de Aquino, Summa Theologica, I-II, q. 90, a. 2\</bibl>\</note>\</cit>\</p>

and also include the attribute @role to describe its "occupation." The doctrinal reference was incorrect; hence, we tagged the indicator "9" with the element \<sic> and added the correct reference with the element \<corr>. Its included "Ethicorum" as \<title> with the full title in the attribute @key. Finally, we include the quote in the element \<cite> which allows nesting the elements \<quote>, \<note> and \<bibl> were we included the textual citation besides the Aquinas reference.

Some questions emerge like, how standardize the glosses structure? There is possible to construct an algorithm which can automatize the identification of elements in glosses text? Is possible bring semantic attributes accurately with existing vocabularies or is necessary to develop one in particular for Ibero-American Modern Age legal sources? Address this kind of questions will help further authoritative and critical editions in the process of migrating from traditional analog supports to digital. For instance, Max Planck Institute for European Legal History is developing a project for the study of Solorzano de Pereyra's work that includes close reading and digital humanities approaches[15]. The dialogue between projects could represent the possibility to build new strategies to digitize, modelling and interact with juridical and canonical sources from Modern Age Ibero-America to connect Spanish-Latin-Greek legal speech with other European traditions.

Linking data

Data is meaningful for digital history when the reader has the capability to integrate them into an argumentative narrative. The weighted count, abstract visualization or semantic tagging are inaccessible for most readers, in the same way as natural language transcriptions are available for "end-users" but are not machine-readable (Koho et al., 2017: 369). Regardless, the law is not significant because it was written, its significance dwells in the use or usefulness of it for subjects integrated into a specific jurisdictional context. Likewise, the digitized corpus of Ibero-American laws can reveal, or at least suggest, the sense within the texts when is linked with its use, scilicet with other "digital objects" represented as different types of text or acts of speech.

Strategies aforementioned are necessary to link the data into a discourse. Modeling information by itself represents an advance into the preservation of digitized historical information, besides text-mining, and other statistical tools can help to suggest possible inconsistencies or confirm previous assumptions about beforehand studied documents, visualizations contribute to producing new trends for analysis and presentation of historical arguments. The question which remains is, how to link all this separated results into a historical argumentation subject to validation for the academic community?

That question is stemmed from several researchers and developers concern. For instance, the last release of Omeka web publishing platform (Omeka S) includes into its features the capability to publish items with Linked Open Data (LOD), which means to use the RDF model to publish data on the Web and to interlink with different sources (Yu, 2011: 410). Omeka developers included that feature in order to confront the issue that source description of digitized sources made by libraries and archives have the purpose of arranging historical sources instead of "advance an argument" through interaction with different data sources (Leon, 2016; Robertson and Mullen, 2017: 5–6). As this is an ongoing unresolved issue, we only want to present some considerations about the implications to address this kind of strategy.

Likewise Semantic Web, LOD designate a common ground concept for web development. It represents a complex but at the same time straightforward method to construct relationships between data. Three technologies sustain this strategy: RDF, SPARQL, and URIs. Beyond merely technic language, LOD allows the connection between a subject and an object through a predicate, which roughly represents the underlying semantics of natural language. Those relationships are named "triplets" and could be represented, following the example of last Gregorio López's gloss, as next: "Aristotle"-"his role was"-"Philosopher" / "Aristotle"-"was born in"-"Stagira" / "Aristotle"-"wrote"-"Nicomachean Ethics". Nevertheless, those triplets cannot be textually translated into code and must be adjusted to standarized vocabularies or ontologies. Allemang and Hendler (2011: 323) remark that although RDF structure reminds the grammatical construction of subject-predicate-object, the model must accomplish what the modeler intended. Lousy modeling could inaccurately retrieve information, for instance, interpret Aquinas as the author of the Nicomachean Ethics.

Despite TEI schema, which purpose is to describe the structure and elements of each text, RDF model their in-

terconection. If it is taken into account that the legal and authoritative texts were prepared for segmented reading with the purpose to be used as a specific reference to resolve doubts or act according to the right, an RDF modelling is adequate to link *articuli* without losing the identification of their original production location. In this regard, TEI allows describing the internal elements of each text of the law with reasonable precision, integrating their comments, clarifications, and additions; while RDF connects the norms beyond its original binding foreseeing their dispersion and posterior unintelligibility.

The purpose of exploring LOD for a Legal Historical Source corpus also comprehends the feasibility to interconnect with repositories as *Europeana* and retrieve information via the SPARQL end-point, besides several archives and libraries of historical sources which establish Linked Data recipes (HTTP URIs). Furthermore, it could facilitate the understanding of the interrelations between different types of documents, authoritative works, and legal sources, besides helping to construct visualizations that contribute to the distant reading of legal historical sources corpora.

Nevertheless, the scope of a project of this magnitude demands the evaluation of its possibilities previously to attempt to achieve more than is possible with current technology and methodology. Although interrelated, segmentation, data modeling, and linking open data, represents three not necessarily interoperative strategies. Moreover, the unambiguity as the prerequisite for the machine-reading lead to questioning if computers are helping us to understand the sources or are we who help machines to understand the documents.

DISTANT READING FOR CULTURAL-LEGAL HISTORY

The distancing of the historian from its natural language transform its process for questioning the past, but its transformation differs from quantitative and cliometric paradigms from the 1960s. The aim of digital history must vary from a mechanic pretension of neutral interpretation of the past or the approach of "explanation over interpretation" (Moretti, 2007: 91)[16], but neither can resign to its pretensions of objectivity in historical analysis. From Jorn Rüsen perspective, historical objectivity is not tantamount to an absence of subjectivity. Instead, recent "reconciliation" with objectivity and historical truth[17] attach interpretation with "molding the source information in a historical narrative, significative and with sense." (Rüsen, 2014: 227) "Thick reading" proposed by Hespanha aspire to retrieve "fidelity" from the historical text, "identify the spiritual dispositions there encrusted, the origin of the authentic senses of the practices." (Hespanha, 2002: 45) If "thick reading" can discover the strangeness, represented in the "authentic senses" from the historical legal text, can the singularity of programming language help to reveal of the same sense of past practices by "moulding" the semi-structured textual data?

The presumption of a distant reading, especially after the immersion into the academic zeitgeist of Franco Moretti's proposal, cautiously assumed, represents the possibility to combine analysis and macroanalysis of big text-data. In this kind of corporeal approach, the observation scales (micro and global) are connected, to enhance a theoretical-distant approach combined with a semiological-thick reading (Pons, 2013: 124). The machine in that scenario represents a tool to assist abstraction, something that it inevitably does inasmuch in computational language ambiguity means illegibility[18]. As Franco Moretti remarked, distance is the condition of knowledge for distant reading, "it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems." (Moretti, 2000) However, also we must point, as John Sceski remarks in its interpretation of Popper, that "'Objectivity' connotes the idea of distance," (Sceski, 2007: 4) therefore it is relevant to remark and make evident the "modelling" associated with the distant reading of historical sources.

Black-boxing is a problematic and unresolved issue in Digital Humanities[19]. Paradoxically, the more public and accessible the sources are the code used for their interpretation become more obscure. Although Open Access and Freeware culture are spreading over Digital Historians and Humanist, there is still a significant constriction for institutions to protect their intellectual property, therefore, propietary software and limited access to programming code imply an essential impediment to questioning methodologies used in the textual analysis (Röhle, 2012: 75–77). There is also a fiction of neutrality in software, a sensation of objectivity produced by lines of unambigous code, nevertheless, paths and networks are previously defined by schemas and editions can use functional but anachronic vocabularies to facilitates machine-reading.

Although macroanalysis encompasses quantities of text measureless for a single person or even a team, the units of its approach are the smallest structures of discourse: morphemes, words, phrases, syntagms. Even though, the approach to large-corpus allows to "reconnect" those small pieces of language into significant discourse structure and processes (Beaugrande, 2011: 36–43). Automatization of reading, even in its most primitive form, revives the wish of retaking the long-term perspective in history without having to abandon the micro-local view of historical phenomena. Jo Guldi and David Armitage clearly expressed above idea in "The History Manifesto" when they proposed the "new trends" in historical writing: "first, a need for new narratives capable of being read, understood, and engaged by non-experts; second, an emphasis on visualisation and digital tools; and third, a fusion between the big and the small, the 'micro' and the 'macro', that harnesses the best of archival work on the one hand and bigpicture work about issues of common concern on the other." (Guldi and Armitage, 2014: 117–118)

The use of tools for digital text analysis allow separating, depurate, and sort these minimum units of the language (generally words or small phrases). The visualization, likewise, helps the interpretation and explanation of the sense of the tendencies formally shown as graphs,

maps, trees, clouds or networks. However, visualizations can only beckon interpretation because they cannot replace the historical narrative. For instance, a review by Adam Crymble of Paper Machines, a topic-modeling plug-in for Zotero (currently deprecated), makes the following annotation about its visualization outputs:

> […] users will find appealing Paper Machines' ability to generate different types of visualizations very quickly. As the tool does not currently make it easy to export raw data or validate the analyses conducted, users would be wise not to risk their academic reputations on the tool's outputs unless verified through more verbose and transparent tools. This project's great contribution is probably not to research, but to pedagogy and skills training. (Crymble, 2012)

Despite Guldi and Armitage presumption that visualizations can act as a rhetorical trope (Guldi and Armitage, 2014: 119; Edelstein, 2016: 246), Big Data analysis and visualization are not necessarily attached to grand narratives. Crymble critic to Paper Machines visualizations is a call for attention in the sense that abstraction can be public outstanding, but most transparent displays are preferable to present narratives instead of nebulous interpretations of an illegible image. For instance, narrative cartographies (Caquard and Cartwright, 2014) represents the possibility of mapping histories in long-term perspectives after applying distant reading analysis into textual data.

Black-boxing can be overcome by the honest presentation of code limitations, remarking the necessity of in-depth reading for historical interpretation, evaluating how digital methodologies not replace but enhance source criticism, heuristic, and synthesis. Besides, being aware of allocating elements that accurately and critically correspond with authors intentions (Smith, 2004), and validate not only the functionality of code but the relevance of its results. The scale analysis must correspond with the deep of description, hence, distant reading is constrained by the complexity of data modelling. Decisions made in that regard are fundamental to possible results.

The corporeal reading approach of historical legal sources implies structuring a data modelling that, although it does not represent the linguistic and symbolic complexities of the text, does not hide them either. The strategy proposed in this paper consists of moving from the obvious to the ambiguous, that is, from the thematically better-defined legislation (i.e., court fees) to more complex constructions (i.e., Gregorio López glosses). The possibilities of automation lie in the ability to identify and manipulate documents without distorting the original meaning of the texts. The advantage lies therefore not so much in the application of preexistent models, but the possibilities of creating that computer languages represent for digital historians.

NOTES

1 Unlike Digital Humanities, Digital History is still an emergent field whose practitioners are grouped into centers and, so-called, laboratories. Undoubtedly, one of the most prestigious is the Roy Rosenzweig Center for History and New Media (RRCHNM) of George Mason University, and the European Luxembourg Center for Contemporary and Digital History (C2DH). Still, there is no journal devoted to digital history; even the RRCHNM published for a short time a "Journal of Digital Humanities," although with an evident focus on historical research. Likewise, there is no annual conference similar to the "Digital Humanities conference" carried out by the Alliance of Digital Humanities Organizations since 1989. Nevertheless, small conferences are held continuously in several spaces like the annual one day conference "Current Research in Digital History" hosted by the RRCHNM, or the DHNORD2017 conference dedicated this year to discussing the (de)construction of Digital History, co-organized by the *Maison européenne des sciences de l'homme et de la société* (MESHS) and the C2DH.

2 Kayman (2016) clarifies that this term refers to the term *corpus* rather than the "body and bodily life" formulated by Maxine Sheets-Johnstone (2015).

3 A British and Canadian group of digital historians propends for a *programming historian* and the development of a series of "new technical skills" for the humanist research process. Essentially a historian is also a programmer when is capable of going beyond software limitations, in Turkel and MacEachern words "if you don't program, your research process will always be at the mercy of those who do" (Graham, Milligan and Weingart, 2016: 59).

4 The critical legal history in the Ibero-America tradition differs from the perspective constructed in the United States by scholars like Robert W. Gordon (1984, 2012). For a description of the critical legal history of Hispanic and Portuguese tradition, see Hespanha (2002: 21–26). For the cultural legal history approach, see Garriga (2016).

5 Following to Costa and Vallejo, the direct translation of the term *iurisdictio* (as *imperium*, *potestas*, and other similar concepts stemming from the medieval legal language) would be useless for its comprehension outside its Latin medieval political-juridical language context. (Costa, 1969: 99–101; Vallejo, 1992: 40)

6 In Costa's perspective, "la lingua organizza l'esperienza e non semplicemente la cataloga" (Costa, 1969: 14) It is evident there the coincidence with Koselleck's *Begriffsgeschichte* where the concepts had as a function to integrate the experiences (*Erfahrungen*) of their contemporaries and fix in the language such experiences that were diluted in time to face the past and prepare the future. (Koselleck, 2006: 58–59)

7 The formalities of the courts were defined, reformed and permanently adjusted by the Crown through different cédulas and orders that responded to particular projects or were responses to complaints raised to the several councils of the king. As was the case with legislation, some judges and erudite lawyers undertook the task of elaborating "guidebooks" for the judges' use of the "style and practice" recommended for carrying out the different processes of each court (Gayol, 2007). Each *audiencia* defined its ceremonial, which comprehends the trial schedules, the mode to begin and finish the sessions, the ceremonies in the courtrooms, the customs to celebrate agreements, hearings, trials, visits to prisons, besides the behavior of the judges at the assistance to public and religious ceremonies (Ximénez de Embún, 2009: 332).

8 Although several interpretations of Leviathan's frontispiece point at symbolic elements within the engraving (Kristiansson and Tralau, 2014), Hobbes gave it an explanation at introductory first page when he depicts the Common-Wealth or State (*civitas*) as an "Artificiall Man" created by man as imitation of natural "Body" (Hobbes, 1909: 8).

9 Perhaps the most notable example related to legal sources is Miguel Artola's Spanish historical legislation database available at http://www.mcu.es/archivos/lhe/. The repository gathers digitized images from normativity promulgated for peninsular and american territories since 10th century till the first half 19th century. The database consists of 35,355 referenced norms

which 26,831 digital images are available without any transformation. In order to allow users interact with the database items, developers constructed a search form which retrieves metadata for each document collected in a "thesaurus".

10 Although Voyant Tools is primarily a web-based software, there is available a version to run locally who is very helpful to handle large corpus or just for keep it confidential avoiding to be cached in Voyant Tools servers. http://docs.voyant-tools.org

11 Early exercises were done with Python scripts. The results presented here were obtained with the help of ANTCONC software available via http://www.laurenceanthony.net/software/antconc/.

12 Web semantics and data modeling are two specialized fields in the computational science in what we do not pretend to be experts. Several resources are available online, for instance, the World Wide Web consortium https://www.w3.org.

13 The *Siete Partidas* were wrote in the 13th century by the King Alfonso X "The Wise" with the purpose to unify the scattered legislation of the kingdoms and cities from Castille. An interesting study for English public was made by Robert I. Burns (2001). A digital copy of the edition from 1576 with the glosses original layout can be retrieved from the *Bayerische Staatsbibliothek digital* via the permalink http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb11304365-2

14 Project website: https://7partidas.hypotheses.org/. Code is available at https://github.com/7PartidasDigital.

15 Overall project presentation is available via http://www.rg.mpg.de/1388428/event17-11-09-solorzano.

16 About the possible neo-positivist pretentions of Digital Humanities (Drucker, 2012; Pons, 2013: 125–126)

17 Some examples of this "reconciliation" are (Førland, 2017; Rüsen, 2005; Ricœur, 2001)

18 "If natural language were structurally unambiguous with respect to some comprehensive, effectively parsable grammar, our parsing technology would presumably have attained human-like accuracy some time ago, instead of levelling off at about 90% constituent recognition accuracy." (Schubert, 2015) In that regard, disambiguation is one of the most challenging strategies from the Semantic Web and Natural Language Understanding (Barrièrre, 2016) and are also the main reason for the development of vocabularies or ontologies (W3C Consortium, n.d.).

19 A recent approach about this issue in Milligan and Warren (2018).

REFERENCES

Adorni, Giovanni, Maratea, Marco, Pandolfo, Laura and Pulina, Luca (2015) "An ontology for historical research documents", In *International Conference on Web Reasoning and Rule Systems*. Springer, Berlin. 11–18.

Agüero Nazar, Alejandro (2007) "Las categorías básicas de la cultura jurisdiccional", In *De justicia de jueces a justicia de leyes: hacia la España de 1870. Cuadernos de derecho judicial*. edited by Lorente Sariñena, M. Consejo General del Poder Judicial, Madrid.

Allemang, Dean and Hendler, James A. (2011) *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*, 2nd ed. Morgan Kaufmann/Elsevier, Waltham, MA.

Anderson, Margo (2007) "Quantitative History", In *The SAGE Handbook of Social Science Methodology*. edited by Outhwaite, W. and Turner, S. SAGE Publications Ltd, London. 248–264. http://methods.sagepub.com/book/the-sage-handbook-of-social-science-methodology/n14.xml [accessed 16/November/2017].

Ayers, Edward L. (1999) "The Pasts and Futures of Digital History", *University of Virginia*. http://www.vcdh.virginia.edu/PastsFutures.html [accessed 16/November/2017].

Barrientos Grandon, Javier (2004) *El gobierno de las Indias*, Colección historia. Fundación Rafael del Pino, Marcial Pons, Madrid.

Barrièrre, Caroline (2016) *Natural Language Understanding in a Semantic Web Context*, Springer, Berlin.

Beaugrande, Robert de (2011) "The Story of Discourse Analysis", In *Discourse as structure and process. Discourse studies*. Reprinted. edited by Dijk, T. A. van. SAGE, London. 35–62.

Birnbaum, David J. (2017) "What is XML and why should humanists care? An even gentler introduction to XML", *Obdurodon Digital humanities*.

Burns, Robert Ignatius (2001) "The Partidas: Introduction", In *Las siete partidas. The Middle Ages series*. edited by Burns, R. I. University of Pennsylvania Press, Philadelphia. xi–xxix.

Caquard, Sébastien and Cartwright, William (2014) "Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping", *The Cartographic Journal*. 51(2). 101–106.

Ciotti, Fabio and Tomasi, Francesca (2016) "Formal Ontologies, Linked Data, and TEI Semantics", *Journal of the Text Encoding Initiative*. (Issue 9). http://jtei.revues.org/1480?lang=en [accessed 27/November/2017].

Clavero, Bartolomé (1986) *Tantas personas como estados: por una antropología política de la historia europea*, Colección Derecho, cultura y sociedad. Tecnos, Fundación Cultural Enrique Luño Peña, Madrid.

Clavero, Bartolomé (1991) *Antidora: antropología católica de la economía moderna*, Per la storia del pensiero giuridico moderno. Giuffrè, Milano.

Clavero, Bartolomé (2002) "Iurisdictio nello specchio o el silencio de Pietro Costa", In *Iurisdictio. Semantica del potere politico nella repubblica medievale, 1100-1433*. Giuffrè, Milán. xix–lxxv.

Cohen, Dan, Gibbs, Frederick, Hitchcock, Tim, Rockwell, Geoffrey, Sander, Jorg, Shoemaker, Robert, Sinclair, Stefan, Takats, Sean, Turkel, William J. and Briquet, Cyril (2011) "Data mining with criminal intent", *White paper*.

Cohen, Daniel J. and Rosenzweig, Roy (2006) *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*, University of Pennsylvania Press, Philadelphia.

Costa, Pietro (1969) *Iurisdictio. Semantica del potere politico nella pubblicistica medievale (1100-1433)*, Giuffrè Editore, Milano.

Costa, Pietro (1972) "Semantica e storia del pensiero giuridico", *Quaderni fiorentini per la storia del pensiero giuridico moderno*. 1(1). 45–87.

Costa, Pietro and Zolo, Danilo (eds.) (2007) *The Rule of Law. History, Theory and Criticism*, Law and Philosophy Library. Springer, Dordrecht. //www.springer.com/gp/book/9781402057441 [accessed 24/November/2017].

Crymble, Adam (2012) "Review of Paper Machines, produced by Chris Johnson-Roberson and Jo Guldi", *Journal of Digital Humanities*. 2(1). http://journalofdigitalhumanities.org/2-1/review-papermachines-by-adam-crymble/.

Drucker, Johanna (2012) "Humanistic Theory and Digital Scholarship", In *Debates in the Digital Humanities*. edited by Gold, M. Universidad de Minnesota, Minneapolis. 85–95. http://dhdebates.gc.cuny.edu/debates/text/34 [accessed 20/June/2016].

Edelstein, Dan (2016) "Intellectual history and digital humanities", *Modern Intellectual History*. 13(1). 237–246.

Eide, Øyvind (2014) "Ontologies, Data Modeling, and TEI", *Journal of the Text Encoding Initiative*. (Issue 8). https://jtei.revues.org/1191 [accessed 27/November/2017].

Fankhauser, Peter, Kermes, Hannah and Teich, Elke (2014) "Combining macro-and microanalysis for exploring the construal of scientific disciplinarity", *Proceedings of Digital Humanities*.

Fischer, Marguerite (1966) "The KWIC index concept: A retrospective view", *American Documentation*. 17(2). 57–70.

Førland, Tor Egil (2017) *Values, Objectivity, and Explanation in Historiography*, Taylor & Francis.

Garriga Acosta, Carlos (2004a) "Historia y derecho, historia del derecho", *Istor. Revista de historia internacional*. IV(16). 3–8.

Garriga Acosta, Carlos (2004b) "Orden jurídico y poder político en el Antiguo Régimen", *Istor. Revista de historia internacional*. IV(16). 13–44.

Garriga Acosta, Carlos (2016) "Mientras tanto. El Manual de Tomás y Valiente: una obra de y para la transición", In *Francisco Tomás y Valiente. Memoria y legado de un maestro*. edited by Alonso Romero, M. P. Universidad de Salamanca, Salamanca. 49–73.

Gayol, Víctor (2007) *Laberintos de justicia: procuradores, escribanos y oficiales de la Real Audiencia de México (1750-1812)*, *Colección Investigaciones*. El Colegio de Michoacán, Zamora.

Gayol, Víctor (2016) "Exploring big historical data. The historian's macroscope", *Virtualis*. 7(13). 102–105.

Gayol, Víctor (2017a) *El Costo del gobierno y la Justicia: Aranceles para tribunales, juzgados, oficinas de justicia, gobierno y real hacienda de la Corte de México y lugares foráneos (1699-1784)*, El Colegio de Michoacán, Zamora.

Gayol, Víctor (2017b) "The Programming Historian en español", *Humanidades Digitales*. Blog. http://humanidadesdigitales.net/blog/2017/03/17/the-programming-historian-en-espanol/ [accessed 18/November/2017].

Gayol, Víctor and Melo Flórez, Jairo Antonio (2017) "Presente y perspectivas de las humanidades digitales en América Latina", *Mélanges de la Casa de Velázquez*. 47(2). 281–284.

Gibbs, Fred (2015) "Corpus Analysis with Antconc", *Programming Historian*. https://programminghistorian.org/lessons/corpus-analysis-with-antconc [accessed 16/November/2017].

Gordon, Robert W. (1984) "Critical Legal Histories", *Standford Law Review*. 36(57). 57–125.

Gordon, Robert W. (2012) ""Critical Legal Histories Revisited": A Response", *Law and Social Inquiry*. 37(1). 200–215.

Grafton, Anthony (1997) *The footnote: a curious history*, Rev. ed. Harvard University Press, Cambridge, Mass.

Graham, Shawn, Milligan, Ian and Weingart, Scott B. (2016) *Exploring big historical data: the historian's macroscope*, Imperial College Press, London.

Guldi, Jo and Armitage, David (2014) *The history manifesto*, Cambridge University Press, Cambridge.

Heimerl, Florian, Lohmann, Steffen, Lange, Simon and Ertl, Thomas (2014) "Word Cloud Explorer: Text Analytics Based on Word Clouds", In *IEEE*. 1833–1842. http://ieeexplore.ieee.org/document/6758829/ [accessed 16/November/2017].

Hespanha, António Manuel (1987) "Da iustitia à disciplina, textos, poder e política penal no antigo regime", *Anuario de historia del derecho español*. (57). 493–578.

Hespanha, António Manuel (2002) *Cultura jurídica europea: síntesis de un milenio*, Tecnos, Madrid.

Hitchcock, Tim, Shoemaker, Robert, Emsley, Clive, Howard, Sharon and McLaughlin, Jamie (2012) "The Old Bailey Proceedings Online, 1674-1913", https://www.oldbaileyonline.org/ [accessed 7/August/2017].

Hitchcock, Tim and Turkel, William J. (2016) "The Old Bailey Proceedings, 1674-1913: Text Mining for Evidence of Court Behavior", *Law and History Review*. 34. 929.

Hobbes, Thomas (1909) *Leviathan*, Reprinted from the edition of 1651. edited by Pogson Smith, W. G. Clarendon Press, Oxford. http://archive.org/details/hobbessleviathan00hobbuoft [accessed 25/November/2017].

Jockers, Matthew Lee (2013) *Macroanalysis: digital methods and literary history*, *Topics in the digital humanities*. University of Illinois Press, Urbana.

Kantorowicz, Ernst Hartwig (1997) *The king's two bodies: a study in mediaeval political theology*, *Princeton paperbacks*. Princeton University Press, Princeton, N.J.

Kayman, Martin A. (2016) "Corpus Juris, Habeas Corpus, and the "Corporeal Turn" in the Humanities", *Law & Literature*. 28(3). 355–378.

Kelly, Jason M. (2013) "Mining in the Old Bailey Project", *Digital Public History*. https://digitalpublichistory.wordpress.com/2013/02/25/mining-in-the-old-bailey-project/ [accessed 7/August/2017].

Koho, Mikko, Hyvönen, Eero, Heino, Erkki, Tuominen, Jouni, Leskinen, Petri and Mäkelä, Eetu (2017) "Linked Death—Representing, Publishing, and Using Second World War Death Records as Linked Open Data", In *The Semantic Web: ESWC 2017 Satellite Events*. *Lecture Notes in Computer Science*. edited by Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., and Hartig, O. Springer, Berlin. 369–383.

Koselleck, Reinhart (2006) *Begriffsgeschichten: Studien zur Semantik und Pragmatik der politischen und sozialen Sprache*, Suhrkamp, Frankfurt am Main.

Koselleck, Reinhart (2015) *Vergangene Zukunft: zur Semantik geschichtlicher Zeiten*, *Suhrkamp-Taschenbuch Wissenschaft*. 9. Aufl. Suhrkamp, Frankfurt am Main.

Kristiansson, Magnus and Tralau, Johan (2014) "Hobbes's hidden monster: A new interpretation of the frontispiece of Leviathan", *European Journal of Political Theory*. 13(3). 299–320.

Latour, Bruno (2000) *Pandora's hope: essays on the reality of science studies*, Harvard University Press, Cambridge, Mass.

Leon, Sharon (2016) "What to expect in Omeka S", *Omeka S Github repository*. wiki. https://github.com/omeka/omeka-s/wiki/What-to-expect-in-Omeka-S.

Lines Andersen, Deborah (2002) "Defining Digital History", *Journal of the Association for History and Computing*. 5(1). http://hdl.handle.net/2027/spo.3310410.0005.103.

Martin, James R. (1992) *English text: system and structure*, Benjamins, Philadelphia.

Martínez, Fernando, Beck Varela, Laura and Agüero Nazar, Alejandro (2012) "La disciplina social en la cultura del ius commune. Elementos básicos", In *Manual de Historia del Derecho*. edited by Lorente Sariñena, M. and Vallejo, J. Tirant lo Blanch, Valencia. 101–140. [accessed 25/November/2017].

Melo Flórez, Jairo Antonio (2017) "Lectura distante, fragmentada y colaborativa en el archivo infinito", *Relaciones. Estudios de Historia y Sociedad*. 38(149). 169–189.

Michel, Jean-Baptiste, Shen, Yuan Kui, Aiden, Aviva P., Veres, Adrian, Gray, Matthew K., Team, The Google Books, Pickett, Joseph P., Hoiberg, Dale, Clancy, Dan, Norvig, Peter, Orwant, Jon, Pinker, Steven, Nowak, Martin A. and Aiden, Erez Lieberman (2010) "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science*. 1199644.

Mikheev, Andrei (2005) "Text Segmentation", http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199276349.001.0001/oxfordhb-9780199276349-e-10 [accessed 25/November/2017].

Milligan, Ian and Warren, Robert (2018) " Big Data and the Coming Historical Revolution: From Black Boxes to Models", In *Big data in the arts and humanities: theory and practice*. edited by Schiuma, G. and Carlucci, D. Taylor & Francis, Boca Raton.

Moretti, Franco (2000) "Conjectures on World Literature", *New Left Review*. II. (1). 54–68.

Moretti, Franco (2013) *Distant Reading*, Verso, London.

Moretti, Franco (2007) *Graphs, maps, trees: abstract models for literary history*, Verso, London, New York.

Ophir, Shai (2016) "Big data for the humanities using Google Ngrams: Discovering hidden patterns of conceptual trends", *First Monday*. 21(7). http://journals.uic.edu/ojs/index.php/fm/article/view/5567 [accessed 14/November/2017].

Petit Calvo, Carlos (1995) "El Código inexistente: Por una historia conceptual de la cultura jurídica en la España del siglo XIX", *Historia contemporánea*. (12). 49–90.

Pons, Anaclet (2013) *El desorden digital: guía para historiadores y humanistas*, Siglo XXI, Madrid.

Ricœur, Paul (2001) *Histoire et vérité*, *Points Essais*. Nouv. éd. Éd. du Seuil, Paris.

Robertson, Stephen and Mullen, Lincoln A. (eds.) (2017) "Digital History and Argument", Roy Rosenzweig Center for History and New Media. https://rrchnm.org/argument-white-paper/ [accessed 18/November/2017].

Röhle, Bernhard Rieder Theo (2012) "Digital Methods: Five Challenges", In *Understanding Digital Humanities*. Palgrave Macmillan, London. 67–84. https://link.springer.com/chapter/10.1057/9780230371934_4 [accessed 28/November/2017].

Rojas Castro, Antonio (2017) "La edición crítica digital y la codificación TEI. Preliminares para una nueva edición de las Soledades de Luis de Góngora", *Revista de Humanidades Digitales*. 1. 4–19.

Rosenzweig, Roy (2003) "Scarcity or Abundance? Preserving the Past in a Digital Era", *The American Historical Review*. 108(3). 735–762.

Rüsen, Jörn (2005) *History: Narration, Interpretation, Orientation*, Berghahn Books, New York.

Rüsen, Jörn (2014) *Tiempo en ruptura, Colección humanidades*. Universidad Autónoma Metropolitana, México.

Sánchez, Francisco de Paula Miguel (1834) *El foro español, ó sea nuevo tratado teórico-práctico del orden, modo y proceder en los tribunales de España*, Imprenta de Don Tomás Jordan, Madrid.

Sceski, John H. (2007) *Popper, Objectivity and the Growth of Knowledge, Continuum studies in British philosophy*. Continuum, London.

Schubert, Lenhart (2015) "Computational Linguistics", In *The Stanford Encyclopedia of Philosophy*. Spring 2015. edited by Zalta, E. N. Metaphysics Research Lab, Stanford University, Stanford. https://plato.stanford.edu/archives/spr2015/entries/computational-linguistics/ [accessed 14/November/2017].

Scollon, Ron (2001) "Action and Text: Towards an Integrated Understanding of the Place of Text in Social (Inter)Action, Mediated Discourse Analysis and the Problem of Social Action", In *Methods of Critical Discourse Analysis*. SAGE Publications Ltd, London. 139–184. http://sk.sagepub.com/books/methods-of-critical-discourse-analysis/n7.xml [accessed 23/November/2017].

Seefeldt, Douglas and Thomas, William G. (2009) "What Is Digital History?", *Perspectives on History*. American Historical Association. https://www.historians.org/publications-and-directories/perspectives-on-history/may-2009/intersections-history-and-new-media/what-is-digital-history [accessed 15/November/2017].

Sheets-Johnstone, Maxine (2015) *The Corporeal Turn: An Interdisciplinary Reader*, Andrews UK Limited.

Sinclair, Stéfan and Rockwell, Geoffrey (2012) "Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies", *Digital Humanities Pedagogy: Practices, Principles, and Politics*. 241–63.

Sinclair, Stéfan and Rockwell, Geoffrey (2016) "Text Analysis and Visualization", *A New Companion to Digital Humanities*. 274–290.

Smith, Martha Nell (2004) "Electronic Scholarly Editing", In *A Companion to Digital Humanities*. edited by Schreibman, S., Siemens, R., and Unsworth, J. Blackwell Publishing Ltd. 306–322. http://onlinelibrary.wiley.com/doi/10.1002/9780470999875.ch22/summary [accessed 28/November/2017].

TEI Consortium (2017) "A Gentle Introduction to XML - The TEI Guidelines", *P5: Guidelines for Electronic Text Encoding and Interchange*. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html [accessed 14/November/2017].

Turkel, William J. (2005) "Digital History Hacks", *Digital History Hacks. Methodology for the infinite archive*. Blog. http://digital historyhacks.blogspot.com/2005/12/digital-history-hacks.html [accessed 16/November/2017].

Turkel, William J. (2008) "A Naïve Bayesian in the Old Bailey", *Digital History Hacks (2005-08)*. Blog. http://digitalhistoryhacks.blogspot.com/search?q=%22naive+bayesian%22 [accessed 7/August/2017].

Turkel, William J. and Crymble, Adam (2012a) "Keywords in Context (Using n-grams) with Python", *Programming Historian*. https://programminghistorian.org/lessons/keywords-in-context-using-n-grams [accessed 17/November/2017].

Turkel, William J. and Crymble, Adam (2012b) "Output Keywords in Context in an HTML File with Python", *Programming Historian*. https://programminghistorian.org/lessons/output-keywords-in-context-in-html-file [accessed 17/November/2017].

Turkel, William J. (2015) *Digital Research Methods with Mathematica®*, William J. Turkel, University of Western Ontario, London, Ontario. http://williamjturkel.net/digital-research-methods-with-mathematica/.

Vallejo, Jesús (1992) *Ruda equidad, ley consumada: concepción de la potestad normativa, 1250-1350, Historia de la sociedad política*. Centro de Estudios Constitucionales, Madrid.

W3C Consortium (n.d.) "Vocabularies", *Semantic Web*. https://www.w3.org/standards/semanticweb/ontology.

Watkins, Emma (2015) "PhD Work in Progress- Emma Watkins &The Case of George Fenby", *The Digital Panopticon. The Global Impact of London Punishments, 1780-1925*. Blog. https://blog.digitalpanopticon.org/?p=814 [accessed 14/November/2017].

Weingart, Scott (2016) ""Digital History" Can Never Be New", *the scottbot irregular*. http://scottbot.net/digital-history-can-never-be-new/ [accessed 20/June/2016].

Welsh, Megan E. (2014) "Review of Voyant Tools", *Collaborative Librarianship*. 6(2). 96–98.

Wiegand, Viola, Mahlberg, Michaela and Stockwell, Peter (2017) "Corpus Linguistics in Action: The Fireplace Pose in 19th Century Fiction", *Programming Historian*. https://programminghistorian.org/posts/corpus-linguistics-in-action [accessed 18/November/2017].

Ximénez de Embún, Ana (2009) "El ceremonial de la Real Audiencia de Aragón en 1749", *Emblemata*. 15. 329–393.

Yu, Liyang (2011) *A Developer's Guide to the Semantic Web*, Springer, Berlin, Heidelberg. https://link.springer.com/chapter/10.1007/978-3-642-15970-1_11 [accessed 19/November/2017].