Social Science

Geography of legal water disputes in Chile

Roberto Costumero, Jesús Sánchez, Ángel García-Pedrero, Diego Rivera, Mario Lillo, Consuelo Gonzalo-Martín and Ernestina Menasalvas

ABSTRACT

Water resources are the main productive source in Chile. Growing competition for water use causes conflicts which end up in courts (Courts of Appeal and The Supreme Court). Legal disputes end when the courts issue a judgement which is recorded. The volume of court decisions made makes the task of searching, analyzing, and extracting knowledge from these texts challenging. This research is aimed at developing a mapping tool to explore spatio-temporal patterns from legal records, as court decisions are an objective proxy for conflicts. Natural Language Processing techniques are applied in order to process and extract information from the court decisions to help the map visualization process. The mapping tool allows information to be visualized in different layers ranging from the whole text of the courts’ decisions to maps in which one can easily find the location of the conflict or any other places mentioned in the courts’ decisions being analyzed by the tool by filtering the courts’ decisions through different terms and concepts. To implement the mapping tools, we combined a geographic information system, along with a search engine (Elasticsearch) and an analytic dashboard (Kibana).

1. Introduction

In terms of Chilean regulation, water is regarded as a national asset for public use. The government provides water rights to individuals through an administrative action. A water right is granted by considering that (Melo and Retamal, 2012): (i) there is no preference between different uses, (ii) the Chilean Water Authority (Dirección General de Aguas, DGA) is obligated to grant the requested water rights wherever applicable from a legal standpoint. The Chilean water management system has shown an increase in conflicts related to water rights, water markets and water uses within a neoliberal system (Bauer, 2005, 2015; Hearne and Donoso, 2014; Rivera et al., 2016; Valdés-Pineda et al., 2014), as water rights can be traded and transferred between users, with no involvement of the State.

The reallocation of current water rights is accomplished by trading water rights through market mechanisms. As noted by the Estrategia Nacional de Recursos Hídricos (National Water Resources Strategy) 2012–2025 (MOP, 2012), a healthy water market requires reliable and up-to-date information regarding the demands and supplies of water rights, as well as accessible information. Currently, the Catastro Público de Aguas (CPA – Public Water Register) is the official database for water rights at the DGA. Stored information includes the essential data on each water right, such as the holder, date, intake location, discharge, water source and related claims (changes in the diversion coordinates, transfers of water rights, changes in ownership, etc.). However, the CPA has drawbacks. Even though private and public institutions such as Real State Registries (Conservadores de Bienes Raíces), which record real estate titles and transactions in different localities, and water user organizations are obligated to provide updated information, the CPA is not updated or complete, posing barriers for equal access to information (Boettiger, 2012). According to Melo and Retamal (2012), under the current assignment system, water right holders should check new requests published in newspapers and analyze whether the conditions of supply and demand for water is likely to affect their future business strategy.

Conflicts in water use are first brought to the attention of the administrative agencies – DGA and water user organizations. If the resolution from these agencies does not satisfy the parties, the courts of justice required to resolve the conflicts. However, within the current framework, there are large imbalances in accessing relevant information in the decision-making process, with the most affected individuals being small farmers and water user organizations. Thus, the system is prone to inequalities as the financial costs curbs the ability of small farmers to go to court (Rivera et al., 2016).
On the other hand, court decisions related to water disputes have information regarding the location of the conflict, the parties and other contextual information. However, there is a need to include more accurate and reliable information related to the location of the conflicts. For instance, in most of the cases, first instance decisions contain a complete description of the location of the conflict, while records from the upper courts make no explicit reference to locations. Therefore, researchers and practitioners typically have to read through lots of documents in order to find specific information, and the simple indexation of the terms in a judgement may not be enough when looking for the cause of the conflict, the date and location of the court or even the parties involved or other geographical information. Thus, a tool that allows large databases of courts’ decisions to be analyzed and mapped would become a valuable asset to professionals such as lawyers, engineers and government officials.

From our knowledge, this is the first approach to generating a cartography of water conflicts by considering an extensive and objective dataset. We propose a system that makes automatic extraction of data from court decisions written in natural language. Geographical data is then visualized interactively using maps. Prior work has focused on smaller and biased datasets following the course of specific conflicts [http://www.derechoalagua.cl/mapa-de-conflictos/] (Larrain, 2010). The Supreme Court maintains a scale for moral damage (Baremo de daño moral por muerte) that acts as a searching tool from a database, with no mapping capabilities.

This research is aimed at developing a mapping tool to explore spatio-temporal patterns from legal records, as courts’ decisions are an objective proxy for water-related conflicts. For a thorough analysis of the use of legal disputes as a proxy for regional conflicts related to water rights in Chile, refer to Rivera et al. (2016).

2. Methods

This paper presents a tool for data extraction from court decisions by providing different visualization options. The system shows different views that allow users to process and extract the data by querying or interacting with different maps in order to filter out records or find patterns in the courts’ decisions. This particular tool analyzes courts’ decisions written in Spanish, but the method is also applicable to other languages. The main issue we approach is finding which location the courts’ decisions are linked to. In order to resolve this challenge, we used text mining techniques.

To assure that locations shown on maps really exist, the nouns representing locations resulting from the parsing process of the sentences were analyzed using a geographic information system (GIS) to obtain the coordinates of the possible places mentioned and to verify that the given noun phrases corresponded to actual cities, provinces or regions in Chile. The extraction of coordinates and mapping was implemented using OpenStreetMap (Haklay and Weber, 2008), which is embedded in the functionalities offered by Kibana (2015). This annotation has been added to the JSON (JavaScript Object Notation) (Crockford, 2006) files generated in order to visualize the decisions’ text in the maps properly.

2.1. Technologies and methodologies

In this section a set of technologies and the methodologies used in order to produce the final maps and the interactive tool are explained. Figure 1 shows the sequence of the following steps. Apache OpenNLP (Baldridge, 2005) has been used to train and apply models which extract key information from the judgement texts, imposed by the structure of the courts’ decisions: the plaintiff, the defendant, the location of the court, date of the decision, other places of interest, the type of court and the type of dispute.

1 Court decisions were gathered by purchasing them from a regional provider of legal information (Microjuris, 2015).
2 The documents, which are in PDF format, were transformed by custom scripts into plain text in order to facilitate the processing of the raw information and enable the removal of metadata that the PDF files may have.
3 Natural Language Processing models trained and applied to the database using Apache OpenNLP (Baldridge, 2005).
4 Extracted information was transformed into JSON files (Crockford, 2006) in order to obtain a common representation. JSON is a well-known format of object representation for information interchange in modern web-based applications. This notation is useful because it is descriptive and adaptable to the developer’s needs. In our particular case, the JSON files generated contain all the aforementioned information.
5 The JSON files were indexed by Elasticsearch (Banon, 2011) to manage the information that they contain. Elasticsearch is a full-text, schema free, search engine which is built on top of Apache Lucene (2005), a well-known Information Retrieval

![Figure 1. Methodology followed in order to create the tool.](image-url)
library written in Java. Elasticsearch is developed in order to make powerful queries to the system without losing any performance or compromising the data, and can be deployed easily across different nodes, thus it can be scaled if necessary.

Kibana (2015) accesses the Elasticsearch index and presents the information on an informative front-page. Elasticsearch is typically deployed with the developer’s two other products: Logstash (2015) and Kibana (2015). Kibana is an open-source (Apache License v2.0), browser-based analytics, and search dashboard for Elasticsearch. Kibana is easy to set up and start using. Written entirely in HTML and JavaScript it only requires a plain web server, whilst Kibana requires no bespoke server side components. It strives to be easy to get started with, while also being flexible and powerful.

The integration capabilities of Elasticsearch with Kibana allow full visualization, focusing on the presentation of the data and the map using a web page to provide all services (see Figure 2). The web page has a box to introduce queries, a table where all the judgements are detailed (the main left-hand panel), and one map (the main right-hand panel) showing the location of the courts’ decisions. There are two more maps with a different level of detail where one can see the locations to which the courts’ decisions refer (regions – bottom left-hand panel, and provinces – bottom right-hand panel). The main left-hand panel shows a brief summary of the legal records indexed on the platform. The user can then click on a document to expand the information and see the complete text of the court’s decision and, if a search has been carried out, the relevant keywords are highlighted.

Figures 3 and 4 show the map at the region and province levels, respectively. It is worth noting that if the first map refers to the location of the courts where the courts’ decisions were made, the latter shows all the locations that are mentioned throughout the entire text written in the judgement, typically mentioning regions or provinces related to that particular case. For Chile, province boundaries are closely related to the boundaries of main river basins. Thus, the end-user can gauge the conflict intensity for specific watersheds – the main unit for water management (Rivera et al., 2016).

On these two maps, a color gradient indicates the number of judgements in which each region or province is mentioned. Lower intensity indicates a lesser number of judgements in the corresponding areas, while a higher intensity indicates a larger number of judgements. These maps change intensity depending on the actual search carried out.

2.2. Map generation

Since the result of each query is a series of decisions that contain information relative to it, the tool automatically locates each court decision in three different ways, generating a different map for each of them:

Figure 2. Capture of the tool developed contains a search bar that filters the content of the documents (main left-hand panel) and the different maps.
Court location: Where each decision refers to the location of the court where the judgement was made.

Regional map: This map represents the number of cases per region mentioned throughout the judgement texts.

Province map: This map represents the number of cases per province mentioned in the judgement texts.

The tool automatically modifies the maps according to the search criteria. However, we have generated a static version of the maps including all the court decisions in the tool. These maps are shown in Figure 6. Region and province level maps are key components of the system as one can see the number of disputes that have arisen per region or province. Thus this visualisation makes it possible to analyze those regions in which the water rights of use are more problematic.

2.3. Use example

As the mapping tool was designed to map different queries and specific parts of the water code, we present an exercise on how this tool is used for analyzing water conflicts using several query terms based on 1000 records. The quality of the extracted information was checked manually considering the number of correctly geotagged records from the total number of records. The extraction process was able to geotag 90% of the records. The remaining records were not geotagged as the record itself does not provide any geographical information. However, in order to improve the analyses, missed records were tagged to the corresponding court’s locations.

To exemplify the use of the API, Figure 5 shows the results from a series of search terms (filter criteria). The first two terms are explotación de aguas subterráneas and exploración de aguas subterráneas (groundwater exploration and groundwater use) as the Chilean Water Code requires these steps before granting water rights; the third term is hidroelectricidad (hydroelectric power), as the main non-consumption use is for hydroelectric power. Legal disputes are used as a proxy for water conflicts, as disputes reaching the higher courts are an objective measure of conflicts that were not settled on other occasions, such as administrative appeals before the DGA or water user organizations. As provinces in the Chilean administrative system are closely related to a basin, the number of disputes gives a measure of intensity of formal disputes. The fourth term is Dirección General de Aguas, as this
agency is the first instance to administratively solve conflicts, such as conflicts between water users or between water rights holders and the State. Thus, the number of legal disputes reaching the Courts of Appeal and the Supreme Court gauges the intensity of conflicts as the number of legal disputes in each region.

The Northern region has much mining activity which contributes 15% to gross domestic product (GDP) and uses 6% of total available water. This region has arid and semi-arid conditions. The Central valley, with a Mediterranean climate, concentrates on agricultural activities contributing to 8% of GDP, but consuming 75% of total available water. The Southern regions, with wet conditions and high streamflows, have low population density. Thus, Chile presents an uneven distribution of precipitation, resources and population, under a single environmental regulation and water code.

**Figure 5.** Conflict intensity, using the number of legal disputes as a proxy for four different queries.

**Figure 6.** From left to right: the location of the Courts, the decisions of the Courts by regions, the decisions of the Courts by provinces.
3. Conclusions

The work presented shows the potential that Natural Language Processing techniques have for textual analysis, especially in legal domains, where the high volume of information stored in free-text is not usually exploited.

It is usual in law research to extract relevant information manually from legal records such as the date of the sentence, location of the court, type of court and the parties involved. However, this procedure is time consuming and is followed to prove a predefined hypothesis. This approach allows contextual information to be extracted, but is not free from bias from those analyzing the record. The proposed technique does not make any prior assumption regarding the underlying pattern, but it requires a set of keyword terms to be defined. Researchers involved in research have shown interest in the application of the technique as it allows hypotheses to be constructed with a rapid assessment tool, but further developments are needed to include contextual information, as the current version of the API does not fully satisfy the required level of granularity. Therefore, more research is needed to improve the current knowledge base (the storage of structured and unstructured data) for better disambiguation and/or to identify the legal nature (individual, community, private, government) of the parties.

Interactive maps are a powerful tool for visualizing the courts’ decisions and to filter the different cases shown in the table just by selecting the area in which the court is located, or by selecting all the cases related to a particular region or province. However, it is worth noting that water conflicts have multiple drivers and consequences, so our approach has to be seen in the context of a plethora of approaches needed to analyze a given water situation.

In the example presented, the extraction of the different locations mentioned together with the aid of a good visualization platform ends up by giving researchers and lawyers much more information on the importance of water disputes all over Chile. It must be stressed that for a more in-depth policy and institutional analysis, the use of maps must be complemented with elements of policy assessment, political geography and regulatory framework.

After using the platform in researching the causes of water conflict intensity, we have identified some necessary improvements. First, for approximately 90% of the records, the name of the province or county was not recorded. This means that improvements in the extraction of geographical entities are necessary. Also, the tool would benefit from the possibility of allowing manual input when the current algorithms do not state the location or when supervised disambiguation is required. Finally, we will implement a report tool for temporary data (i.e. the number of occurrences of different terms for selected years or periods).

Software

The map shown in Figure 6 was constructed entirely using QGIS 2.6.0 (QGIS Development Team, 2014) from the information generated by the aforementioned NLP process. The rest of the maps mentioned throughout the article were generated using OpenStreetMaps (Haklay and Weber, 2008) included in the Kibana (2015) framework.
Acknowledgments

We are grateful for all comments from the Water & Society cluster. Special thanks to Roto Quezada for supporting our research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the Water Resource Center for Agriculture and Mining CONICYT/FONDAP-15130015.

ORCID

Roberto Costumero http://orcid.org/0000-0002-0069-5960
Consuelo Gonzalo-Martín http://orcid.org/0000-0002-0804-9293
Ernestina Menasalvas http://orcid.org/0000-0002-5615-6798

References

Apache Lucene (2005). A high-performance, full-featured text search engine library. Retrieved from January 25, 2015, http://lucene.apache.org
Baldrige, J. (2005). The OpenNLP project. Retrieved February 2, 2012, http://opennlp.apache.org/index.html
Banon, S. (2011). ElasticSearch: An open source, distributed, RESTful search engine. Retrieved January 25, 2015, http://www.elasticsearch.org
Bauer, C. J. (2005). In the image of the market: The Chilean model of water resources management. International Journal of Water, 3, 146–165.
Bauer, C. J. (2015). Water conflicts and entrenched governance problems in Chile’s market model. Water Alternatives, 8, 147–172.
Boettiger, C. (2012). Del catastro público de aguas: A propósito de una sentencia del tribunal constitucional. Actualidad Jurídica, 25, 557–560.
Crockford, D. (2006, December). JSON: The fat-free alternative to XML. Paper presented at the proceedings of XML (Vol. 2006), Boston, USA.
Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. IEEE Pervasive Computing, 7, 12–18.
Hearne, R., & Donoso, G. (2014). Water markets in Chile: Are they meeting needs? In W. Easter & Q. Qiuqiong (Eds.), Water markets for the 21st century: What have we learned? Global issues in water policy (Vol. 11, pp. 103–126). Dordrecht: Springer.
Kibana. (2015). Retrieved from January 25, 2015, http://www.elasticsearch.org/overview/kibana
Larrain, S. (2010). Conflictos por el Agua en Chile: entre los derechos humanos y las reglas del mercado, Gráfica Andes.
Logstash. (2015). Retrieved from January 25, 2015, http://logstash.net
Melo, O., & Retamal, M. (2012). The water users organizations in Chile. In D. Rivera (Ed.) Environmental, political and social issues (pp. 1–32). New York, NY: Nova Science.
Microjuris. (2015). Retrieved from January 25, 2015, http://cl.microjuris.com
MOP. (2012). Estrategia Nacional de Recursos Hídricos 2012-2025. Ministerio de Obras Públicas. Retrieved from January 25, 2015, http://www.mop.cl/documents/enrh_2013_ok.pdf
QGIS Development Team. (2014). QGIS geographic information system. Open Source Geospatial Foundation. Retrieved from January 25, 2015, http://qgis.osgeo.org
Rivera, D., Godoy-Faúndez, A., Lillo, M., Álvez, A., Delgado, V., Gonzalo-Martin, C., ...García-Pedrero, Á. (2016). Legal disputes as a proxy for regional conflicts over water rights in Chile. Journal of Hydrology, 535, 36–45.
Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J. B., Olivares, C., Vera, M., ...Helwig, B. (2014). Water governance in Chile: Availability, management and climate change. Journal of Hydrology, 519, 2538–2567.