LexrideLaw: An Argument Based Legal Search Engine

Matthew Gifford
Lexride
22580 Murietta Rd., Salinas, Ca 93908
matthew.gifford@lexride.com

ABSTRACT

Legal research search engines are overwhelmingly defined by adherence to the appellate case-law organizational model, whereby cases are discovered by relational keyword searches and case files are returned as results. We are proposing a new legal research search engine model where arguments are extracted from appellate cases and are accessible either through selecting nodes in a litigation issue ontology or through relational keyword searches.

CCS CONCEPTS

• **Computing methodologies → Information extraction**; **Discourse, dialogue and pragmatics**; **Lexical semantics**; **Ontology engineering**; • **Applied computing → Law**;

KEYWORDS

Natural Language Processing, Text Mining

ACM Reference format:
Matthew Gifford. 2017. LexrideLaw: An Argument Based Legal Search Engine. In *Proceedings of ICAIL '17, London, United Kingdom, June 12-16, 2017,* 2 pages.
https://doi.org/10.1145/3086512.3086548

1 INTRODUCTION

We will present and demonstrate an argument based legal search engine called LexrideLaw. It is designed for discovering and presenting legal arguments rather than entire appellate cases. An argument is an independent legal issue consisting of claims, facts and analysis, statutory citations, case law citations and holdings. A case may contain multiple arguments as well as background information and even dissents. A relational keyword search on a case-file search engine returns matches on all of a case's words, possibly resulting in non- optimal matches when searching for distinct legal arguments. We solve the non-optimality problem by mining cases for arguments and storing them as separate argument entities in a hybrid relational/hierarchal database. Arguments can be found or discovered by an end user either by selecting them from a set litigation ontology or from a relational keyword search. The keyword search can be performed over all arguments in the issue tree, or over any specific branch of the tree. This organizational idea most closely approximates the old Yahoo hierarchal model, enabled by modern Google-like relational capabilities. The old Yahoo model failed because the Internet was both too large and distinctly non-hierarchal. Appellate case-law suffers from neither of these conditions. It has a relative small corpus, written using distinct, if at times poorly observed, rules and is naturally hierarchal. An optimal legal search engine should exploit these intrinsic properties of case law.

We employ a process known as componentization of legal knowledge whereby knowledge, here appellate law cases, are atomized into the smallest possible knowledge quanta, the linguistic phrase, and then progressively reassembled back up to arguments. Arguments being the largest coherent legal knowledge structure with internal meaning and logic.

ICAIL '17, June 12-16, 2017, London, United Kingdom

ACM ISBN 978-1-4503-4891-1/17/06…$15.00
https://doi.org/10.1145/3086512.3086548

2 ARGUMENT ISOLATION

The first task in creating the database is to separate cases into distinct proto-arguments. Argument isolation algorithms have been developed to assist in this task. The algorithms are heuristic, meaning practical and possibly non-optimal but sufficient for the task at hand. They rely on Natural Language Processing (NLP), discovering and exploiting intrinsic case law structures and the correct classification of argument elements. In other words, proto-arguments are later optimized to final arguments, either cropped or expanded, based on meaning discovered through knowledge componentization described herein.

3 ARGUMENT ELEMENT CLASSIFICATION

A critical task is the correct classification of argument elements: claims, facts and analysis, statutory citations, case law citations and holdings. To accomplish this task, algorithms were developed that employ NLP parsing, Artificial Intelligence (AI) classification and normalization to symbolic representations. Many different technologies are employed, sometimes in series and sometimes parallel, and are focused to specific conditions in the argument element classification task. To facilitate this task, we have developed a unique visual training engine that facilitates the development, use and realtime optimization of our classification algorithm cascades. Any appellate case hosted on Google Scholar can be loaded into the engine and with a single click, parsed and classified. The classifications are immediately presented to the trainer who has an interface through which he or she can submit an incorrect classification sample to optimize a specific algorithm in the cascade. The trainer then immediately presses the classification button again to see the change on the entire case.

4 LITIGATION ISSUE ONTOLOGY CREATION

Each law has a natural litigation issue ontology. This is not simply a hierarchal representation of the law's elements or issues as would be found in black-letter law-school outlines. It is an actual representation of the issues as they have been litigated. The creation of the litigation ontology for a law or a legal field is a difficult process and cannot easily be machine automated. It is an expert representation of the actual litigation issues an attorney or judge working in the specific legal field would recognize.

In the sample provided below in Figure 1, the node tree CEQA-Particular Impacts-Air Quality-Adequate Discussion looks like it is not a real legal issue, but for a practitioner in CEQA law (California Environmental Quality Act), it is an exact match for a frequently litigated issue. The legal issue is: whether the required Environmental Impact Report contained a legally adequate discussion of air quality impacts. The CEQA practitioner knows and understands this ontology as representing his intrinsic workspace. Imagine the ontology nodes as the result of an expert designed keyword query where all of the returned arguments have been vetted to assure compliance with the specific issue. Another analogy is that the LexrideLaw ontology takes the place of the paralegal, who collects relevant cases for an attorney based on both statutory legal issues [CEQA][EIR][legally adequate discussion containing substantial evidence] and facts [air quality]. The litigation strategy for this issue is very different from one that has, for instance, a fact base of [aesthetic impacts] even though the underlying statutory foundations are identical.

5 LITIGATION ISSUE ONTOLOGY NODE ASSIGNMENT

Once the litigation issue ontology is created, the isolated arguments are assigned as nodes in the litigation issue ontology tree. The node assignment classification engines are still under development. Arguments are assigned based on argument element (claim, holding, case, statute) matches to similar elements in trainer assigned arguments under each node. The argument nodes are identified by the full case citation linked to the Google Scholar case and also by a prevailing party notation. Prevailing parties are determined through a separate claim-hold classification engine which is also still under development. This is probably the most difficult machine classification task as it requires the engine to understand element meaning, not just vocabulary matching. (see Figure 1)

6 END USER: VIEWING DISCOVERED ARGUMENTS

When discovered arguments are reviewed by the user, she can choose what argument elements to display. She can turn on only the argument claims and holdings if she desires. Thus a 10,000-word argument can be reduced to only the essential elements so she can make a quick determination of legal or factual suitability to her case at hand. She can access the original case file and any statutory or case law citations as embedded hyper-links in the argument body. She can also keyword search within individual arguments to quickly isolate and view only those argument elements of interest to her. (see Figure 2)

7 CONCLUSION

The argument-based legal research search engine is optimized to facilitate the discovery and display of legal arguments. What would normally be a multi-day legal research project with indeterminate results can be reduced to a few minutes with a practical guarantee that all applicable arguments have been discovered and reviewed.

Alpha Demonstration Site: www.lexridelaw.com.

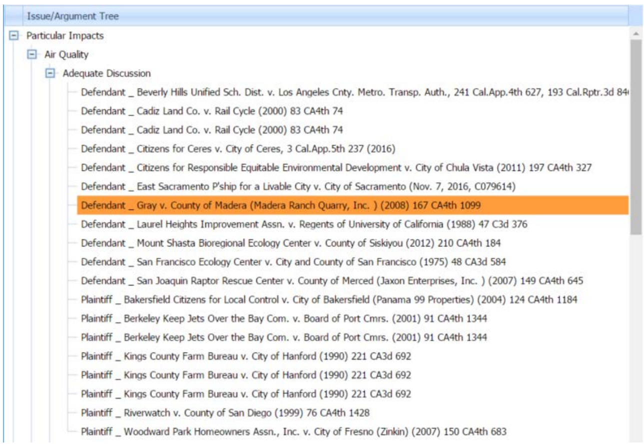
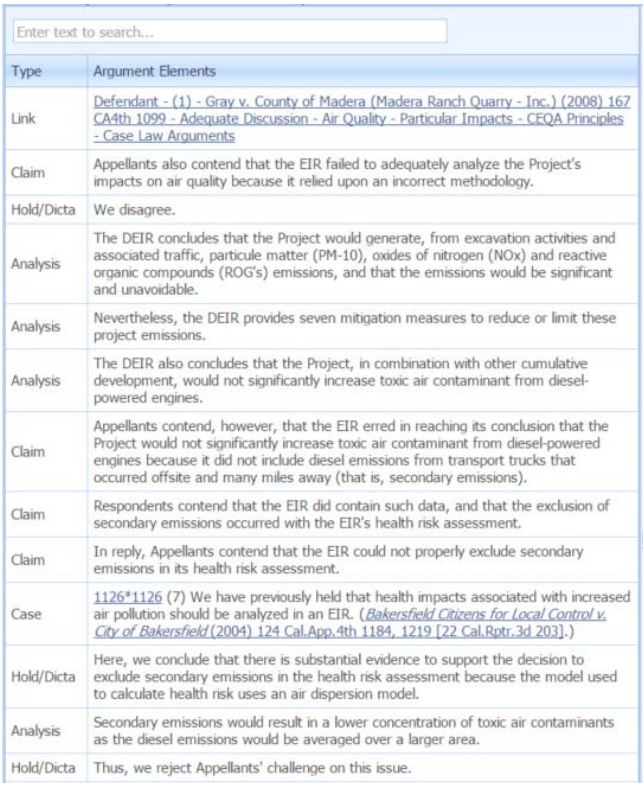


Figure 1: Litigation Issue Ontology



Enter text to search...

Type	Argument Elements
Link	Defendant - (1) - Gray v. County of Madera (Madera Ranch Quarry - Inc.) (2008) 167 CA4th 1099 - Adequate Discussion - Air Quality - Particular Impacts - CEQA Principles - Case Law Arguments
Claim	Appellants also contend that the EIR failed to adequately analyze the Project's impacts on air quality because it relied upon an incorrect methodology.
Hold/Dicta	We disagree.
Analysis	The DEIR concludes that the Project would generate, from excavation activities and associated traffic, particule matter (PM-10), oxides of nitrogen (NOx) and reactive organic compounds (ROG's) emissions, and that the emissions would be significant and unavoidable.
Analysis	Nevertheless, the DEIR provides seven mitigation measures to reduce or limit these project emissions.
Analysis	The DEIR also concludes that the Project, in combination with other cumulative development, would not significantly increase toxic air contaminant from diesel-powered engines.
Claim	Appellants contend, however, that the EIR erred in reaching its conclusion that the Project would not significantly increase toxic air contaminant from diesel-powered engines because it did not include diesel emissions from transport trucks that occurred offsite and many miles away (that is, secondary emissions).
Claim	Respondents contend that the EIR did contain such data, and that the exclusion of secondary emissions occurred with the EIR's health risk assessment.
Claim	In reply, Appellants contend that the EIR could not properly exclude secondary emissions in its health risk assessment.
Case	1126*1126 (7) We have previously held that health impacts associated with increased air pollution should be analyzed in an EIR. (*Bakersfield Citizens for Local Control v. City of Bakersfield* (2004) 124 Cal.App.4th 1184, 1219 [22 Cal.Rptr.3d 203].)
Hold/Dicta	Here, we conclude that there is substantial evidence to support the decision to exclude secondary emissions in the health risk assessment because the model used to calculate health risk uses an air dispersion model.
Analysis	Secondary emissions would result in a lower concentration of toxic air contaminants as the diesel emissions would be averaged over a larger area.
Hold/Dicta	Thus, we reject Appellants' challenge on this issue.

Figure 2: Argument Element Representation