Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority

Julia A. Ekstrom
University of California, Santa Barbara
4526 Bren Hall
Santa Barbara, CA 93106
jekstrom@bren.ucsb.edu

Gloria T. Lau
Stanford University
380 Panama Mall
Stanford, CA 94305-4020
glau@stanford.edu

ABSTRACT
In this paper, we describe an innovative application of mining laws to identify and measure overlapping jurisdictions of government agencies. Laws (statutes and regulations) were used to represent ocean and coastal management for four geopolitical jurisdictions of federal and state levels (Washington, Oregon, and California). We developed preliminary metrics of overlap based on the number of statutes, regulations, and agencies associated with any given topic. The utility of these metrics was tested on 46 topics representing a range of activities and resources across ocean-related sectors within the geographic scope of laws investigated. We found the preliminary results of the overlaps metrics to reveal results similar to a recent review of federal ocean management [1]. In addition, a network diagram graphical display of the data revealed multiple dimensions to facilitate interpretation of results.

General Terms
Algorithms, Management, Experimentation, Legal Aspects.

Keywords
Text mining, ocean management, overlap, legal jurisdiction.

1. INTRODUCTION
In the United States, the large and growing number of laws related to any given issue or domain, such as building construction, water quality, emergency response, and oceans, is leading to inefficient and inconsistent management. The huge number of statutes and regulations can hinder decision-making involving both current activities and emerging uses. For example, in the early 1980s, while conducting mineral exploratory assessments, seismic survey vessels unintentionally cut lines of fishing traps set along the Southern California coast. This cutting resulted in derelict lost traps scattered along the ocean floor, which directly impacted fishermen economically and threatened future fish populations since traps would continue to catch fish with no escape route (pers. comm. John Richards). Commercial fishermen had set these traps under the permission of the California Department of Fish and Game, and the survey vessels operated under the authorization of the California State Lands Commission. The shared spatial jurisdiction between the Department of Fish and Game and the State Lands Commissions became problematic when the two agencies permitted activities that functionally interfered with one another. Although the situation was eventually remedied through a collaborative process, the State Lands Commission could have avoided permitting an incompatible use if decision-makers had information about management regulations of other agencies.

Regulatory measures permitting a new activity should be developed in the context of existing legislation. However, decision-makers have depended on qualitative legal evaluations to provide information about existing legislation. While this may be sufficient in small-scale issues that occur in relatively small locations, such as within one or two counties, a tool to provide comprehensive and quantitative data is needed for larger scale issues. This is particularly a concern for the near future as the number of laws continues to grow[1].

Locating all applicable laws and their authoritative agencies is no easy task. Still, to avoid inconsistent and conflicting law-making, government agencies and other stakeholders need objective baseline information about existing legislation. Additionally, these data must be transparently produced so that decision-makers not familiar with advanced information retrieval techniques can easily interpret necessary information. In this paper, we propose to employ information retrieval techniques along with social network graphical representations to reveal quantitative information about selected topics in the domain of ocean-relevant law.

1.1 Problem in context of oceans
As the health of Earth's oceans is pushed to its limits by increasing anthropogenic stressors, it is vital that we more effectively manage uses and abuses of the marine environment. Uncoordinated ocean management is a major source of deteriorating ocean health and will continue to be a problem under the current sector-based management system [5]. In the United States, decision-making for the marine environment is divided into sectors such as fishing, mining, and transportation, among others. Growing coastal populations, combined with technological advances, have greatly increased ocean use, which in turn has led to a massive body of government ocean regulation. Coupled with the morass of law, the fragmented approach has resulted in overlapping jurisdictions, gaps in management, and inconsistent regulation [17]. As a consequence of the sectoral divisions, the agencies with authority to manage often do not consult or cooperate with one another to ensure permitted activities are compatible. Lacking sufficient coordination, the jurisdictional overlaps have become major culprits in damaging ocean health [5, 15]. Policy-makers cannot begin to strategically fill problematic gaps in coordination without a comprehensive

evaluation of the problem. Focusing on the challenge of overlap, this paper presents a simple but powerful use of text mining and social network analysis to systematically identify and characterize who manages what in the oceans. As the tool is further developed, the intended users are advisory boards of ecosystem-based management (EBM) programs along the west coast of the United States.

In recent decades, the problems of uncoordinated overlapping laws and agency jurisdictions have been highlighted by a number of actors, including proponents of ecosystem-based management (EBM) [4, 5] and marine protected areas [10]. For instance, EBM is a management approach developed to address problems of sectoral management [15]. The implementation of this integrated approach requires tactical coordination of marine management decisions between agencies. Thus far, advocates of EBM articulate fragmented management problems through a mix of cases. These cases thoroughly and qualitatively describe instances of uncoordinated overlaps, inconsistent regulations, incompatible activities, and cumulative impacts [5, 11, 15]. However, for both marine protected area and EBM efforts, no comprehensive analysis has existed to compare the degree of overlap across sectors. Such an analysis of overlaps would equip decision-makers with baseline information so that they can identify gaps in coordination and incompatible regulations. Identification of key agency control and regulatory overlaps relating to any given management topic can assist effective stakeholder communication, participation, and decision-making.

This paper demonstrates a simple, but formal, analysis of ocean and coastal law that aims to answer the following questions:

- What ocean issues are the most fragmented in terms of overlap?
- A. What laws functionally overlap? B. What agencies are involved in implementing these laws?

Answering these questions provides data to: 1. determine the severity of fragmentation by place; 2. prioritize problems based on location and severity of fragmentation; 3. serve as a baseline for monitoring institutional performance; and 4. measure the impact of management changes on ecosystem health. More generally, the information generated from a comprehensive and quantitative analysis of ocean laws can assist decision-makers to define high priority areas more precisely to improve government cooperation based on empirical information.

Exploration of text mining applications to answer questions about overlap required a collection of documents to represent ocean and coastal management. We used a compilation of laws that were manually collected by Ekstrom in 2006 from publicly accessible websites from four geopolitical jurisdictions (federal and three states). The following Dataset section presents the data and the metadata used. Section 3 presents the analysis methods used to explore the data. The results of the preliminary analysis are presented in Section 4. In Section 5, we present interpretation of preliminary findings and suggested future work to fine tune the algorithm.

2. DATASET

Two sets of information were used in the overlaps analysis: 1. term and phrase frequencies extracted from a set of ocean and coastal laws; and 2. record of authoritative agencies for each law. These data and metadata were integrated for 46 topics representing various issues related to the marine environment along the Pacific coast of the United States (see Figure 4 key for list of topics). We recognize that analyzing laws to represent management constrains the analysis only to formal rules, rights, and decision-making procedures. However, until a dataset including non-governmental and informal institutions is compiled to represent all sectors across multiple jurisdictions, the laws provide a free and publicly available dataset to begin quantitative examination of fragmented management.

2.1 Data filtering

In order to generate term and phrase frequencies, we used a set of ocean and coastal laws representing the state and federal laws relevant to the west coast of the United States[6]. Choosing a set of laws for analysis required identifying and applying a set of criteria. To be included in the analysis of this project, a law had to fulfill three criteria: geographic scope, scale of social organization, and type of document. Collecting within the defined criteria produced a consistent collection of laws for quantitative examination of overlap relevant to federal and state levels, as well as among multiple topics.

2.1.1 Geographic scope

The scope of this project was the Northern California Current Large Marine Ecosystem. Therefore, documents with power or influence over managing the activities that affect resources in this region were selected. International organizations and national and state governments have adopted the Large Marine Ecosystem (LME) concept to improve management of the marine environment. On the magnitude of 200,000 km^2, LMEs "are regions of ocean space encompassing coastal areas from river basins and estuaries to the seaward boundaries of continental shelves, enclosed and semi-enclosed seas, and the outer margins of the major current systems," [14]. LMEs spatially cover the most economically, politically, and ecologically important portions of the oceans worldwide [16].

The California Current Large Marine Ecosystem consists of one of the most well-documented marine ecosystems in the world [9]. Located from the Washington State-Canada border to just south of Baja California Sur, Mexico, the California Current LME extends seaward to approximately 300-600 nautical miles from the continent. The northern portion of this LME includes the coast and offshore regions of northern California, Oregon, and Washington [13].

2.1.2 Scale of social organization

The second criterion was that the laws were limited to national and state levels. The inclusion of additional levels of management, such as county, regional, and city, would have provided a finer scale of analysis, but there are thousands of localities within the geographic scope. Therefore, due to time constraints, it was not feasible to identify and gather laws from the smaller-scale jurisdictions.

2.1.3 Type of law

The third criterion was that laws were in the format of codified statutes or administrative code (regulations) for state and federal levels. Codified versions of laws were used because these were the most accessible. Additionally, the publicly accessible digital format throughout all relevant jurisdictions is updated regularly for codified versions of law. For example, the updated code does not include repetitive text from a reauthorized act that existed in the original version. The aim of this collection was to gather relevant laws for one point in time, for which codified laws were the most appropriate. For each jurisdiction, we included any law that mentioned at least one of the terms "ocean," "coast," or "marine." Laws referring only to "marine" were manually filtered out if they only applied to issues relating to the United States Marines (i.e., insurance or retirement regulations, or other issues unrelated directly to uses of the ocean). The remaining list of laws was compiled in their hierarchical units to be as parallel as possible among each jurisdiction within the constraints of digital availability (Table 1). The number of laws meeting the criteria varied with jurisdiction (Figure 1).

Ideally the legal units would have been compiled consistently, such as in chapters. However, the hierarchies varied slightly across geopolitical jurisdictions and the California code was more readily available at the Article level than the Chapter level. There were two types of legal units used in this analysis. Documents containing regulations are referred to as *Regulatory Units* (U.S. Code of Federal Regulations, Washington Administrative Code, Oregon Administrative Rules, and California Code of Regulations), and the codified statute documents are referred to as *Statutory Units* (U.S. Code, Revised Code of Washington, Oregon Revised Statutes, and California Code).

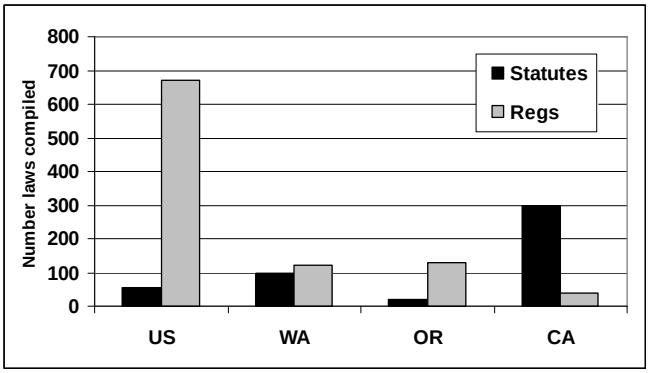


Figure 1. Number of ocean and coastal laws compiled for overlaps analysis (see Table 1 for hierarchical unit of law compiled for each geopolitical jurisdiction).

Table 1. Jurisdictions, format of law, and units collected for marine-related law dataset.

Geopolitical jurisdiction	Law type	Codification hierarchy	Compiled document (Statutory/ Regulatory Unit)
Federal United States law	U.S. Code (statutes)	Title/Chapter/Section	Chapter
	U.S. Code of Federal Regulations	Title/Volume/Chapter/Part/Section	Part
State of Washington	Revised Code of Washington (RCW)	Title/Chapter/Section	Chapter
	WA Administrative Code (WAC)	Title/Chapter/Section	Chapter
State of Oregon	Oregon Revised Statutes (ORS)	Title/Chapter/Section	Chapter
	Oregon Administration Rules (OAR)	Chapter/Division/Section	Division
State of California	California Code	Code/Division/Chapter/Article/Section	Article
	California Code of Regulations	Title/Division/Chapter/Section	Division

2.2 Metadata - Agency authority tables

The agency authority metadata for each law were in part supplied by the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center and in part compiled by Ekstrom. The NOAA Coastal Services Center Digital Legislative Atlas Program (http://www.csc.noaa.gov/legislativeatlas/) had the agency authority list for each federal ocean-related statute publicly available on its website. Their website listed authority to the most specific level of program or agency that was apparent from reading the law. For the state statutes, Ekstrom obtained agency authority by skimming laws. These metadata were stored in the format of an agency by document matrix (Table 2).

Table 2. Excerpt of document-agency matrix metadata compiled for each law in dataset. Ones indicate where an agency has authority to implement the law. A full list of agencies and acronym definitions can be found in Figure 2.

Agency / Document	EPA	DOC	DHS	ACE
Clean Water Act (33 USC 1251 et seq.)	1	0	1	1
Magnuson Stevens Fisheries Management & Conservation Act (16 USC 1801-1883)	0	1	0	0
Invasive Species Act (16 USC 4701 et seq.)	0	0	1	0

To ensure consistency, the higher department level of the agency was recorded for the metadata. With this generalization, an agency was recorded as its parent department, in which it is embedded. For instance, the National Oceanic and Atmospheric Administration (NOAA) was recorded as the Department of Commerce, and the National Park Service was recorded as the Department of the Interior. However, the more specific agency authority information will be used in forthcoming analysis to investigate needs for intra-agency coordination.

Agency authority for the national and state regulations was available on the U.S. Code of Federal Regulations website (http://www.gpoaccess.gov/uscode/index.html) and on the relevant State government administrative code (regulations) websites. Authorities were scaled up consistently in parallel of the statute authorities, as described above.

3. PRELIMINARY OVERLAPS ANALYSIS

Preliminary text analysis was performed to map overlapping functions among laws of relevant agencies. To demonstrate the technique's utility and test its accuracy, we selected 46 issues related to ocean and coastal management (see key to Figure 4) to represent key ocean topics. A topic for overlap analysis can be anything related to the marine environment, such as an activity, resource, species, or ecosystem stressor. Several of these topics were associated with well documented management arrangements [15], thus enabling verification of results. In addition, the 46 topics were selected as a representative sample of activities and resources that span all major marine-related sectors within the geographic scope of the laws investigated.

3.1 Data – Topic by document matrix

To establish the baseline analysis, the 46 topics are each represented by a term or a phrase (see Figure 4 for list of topics investigated). In the future, we plan to utilize multiple terms, synonyms, and related phrases to improve results. A script[1] was developed to identify and count any term (word or phrase) occurrence in the law collection. Querying the law collection with the selected term or phrase produced a topic by document matrix of raw frequencies for each legal unit (Table 3).

Table 3. Sample of topic-document matrix.

Topic freq / Document	Trans-portation	Pollut*	Fishing	Ballast
Clean Water Act (33 USC 1251 et seq.)	102	986	2	1
Magnuson Stevens Fisheries Management & Conservation Act (16 USC 1801-1883)	32	5	726	0
Invasive Species Act (16 USC 4701 et seq.)	8	3	1	79

The frequencies were used to represent the degree to which a law is involved in managing issues related to each topic. Although the frequency, as used here, cannot precisely indicate a law's jurisdiction, it can reflect a law or agency's relative involvement. For example, if one law references 'fishing' two times and a different law references the term 700 times, it is evident the latter is more concerned with fishing activities. Alternatively, the fact that two laws contain a term 15 times does not necessarily reveal that they are equally involved in management relating to the topic.

To determine what agencies were involved in a given topic, the topic-document matrix (Table 3) was integrated with the agency-document matrix (Table 2) resulting in a topic by agency matrix (Table 4). The number of agencies associated with laws containing a topic represented a second dimension of overlap. As such, a relatively high number of agencies involved in a topic indicated a likely complicated case for coordination.

Table 4. Excerpt of topic-agency matrix compiled from combination of document-agency and document-topic matrices. See Figure 2 for agency acronyms defined.

Agency / Topic	EPA	DOC	DHS	ACE
Transportation	1	1	1	1
Pollut*	1	1	1	1
Fishing	1	1	1	1
Ballast	1	0	1	1

[1] The script was implemented by Daniel Spiteri.

Using the topic-document and topic-agency matrices, the following two subsections present preliminary variables developed to calculate the degree of overlap of laws and agencies.

3.2 What topics are most fragmented from overlapping jurisdictions?

We developed preliminary metrics to indicate the degree of overlap as a function of topic and geopolitical jurisdiction. The degree of overlap was calculated using the number of laws involved and the number of associated agencies that were linked to laws involved in each topic. The topics were then ranked for each geopolitical jurisdiction based on these variables.

We used three variables to indicate the degree of overlap that occurs for each given topic. The first variable was derived from the number of statutes that contain a given topic, referred to as Statute Overlap (SO). The topic with the highest number of laws ranked as having the highest overlap from this statute variable. To compare the variable across multiple geopolitical jurisdictions, we normalized the statute overlap variable by the total number of possible statutes in the ocean law compilation for the given geopolitical jurisdiction.

$$SO(T,GP) = \frac{SU(T,GP)}{\sum SU(GP)}$$

SO= Statute Overlap; T= Topic; GP = Geopolitical jurisdiction; SU= Statutory units

The second variable was derived from the number of regulations that contain a given topic, referred to as Regulation Overlap (RO). The topic with the highest number of laws ranked as having the highest overlap from this regulation variable. To compare the variable across multiple geopolitical jurisdictions, we normalized the RO variable by the total number of possible regulations in the ocean law compilation for the given geopolitical jurisdiction.

$$RO(T,GP) = \frac{RU(T,GP)}{\sum RU(GP)}$$

RO= Regulation Overlap; T= Topic; GP = Geopolitical jurisdiction; RU= Regulatory units

The third variable was derived from the agency authority metadata for each law. To calculate this agency overlap variable, referred to as Agency Overlap (AO), the agencies associated with the overlapping laws (statutes and regulations) for a given topic were summed. To compare the variable across multiple geopolitical jurisdictions, we normalized the AO variable by the total number of agencies represented in the ocean law compilation for the given geopolitical jurisdiction.

$$AO(T,GP) = \frac{A(T,GP)}{\sum A(GP)}$$

AO= Agency Overlap; T= Topic; GP = Geopolitical jurisdiction; A = Agencies

In our preliminary development of an overarching index of overlap, the three variables were averaged as the Overlap Index (OI). Within any geopolitical jurisdiction for any given topic, this Overlap Index demonstrates the legal and agency complexity involved in managing the topic. For each jurisdiction, the number of laws and the number of agencies were normalized by their corresponding total possible laws and agencies. Then the average sum of the normalized variables was calculated as follows:

$$OI(T,GP) = \frac{SO + RO + AO}{3}$$

OI= Overlap Index; T= Topic; GP = Geopolitical jurisdiction

This overlap measurement provides an index that allows the systematic comparison of overlap between topics within and among jurisdictions. The index can range from zero to 100%. A topic involving a high number of laws and a high number of associated agencies would result in a number closer to 100%. Alternatively, with zero number of laws and with consequently no agencies associated, the index result would be zero.

This basic calculation of OI adjusts appropriately for cases where one variable is high and the other is relatively low; however, the separate variables of SO, RO and AO provide a more detailed depiction of the overlapping information. For example, in cases where a topic has many laws that are implemented through one agency, the OI may be a high number only based on the high results of the SO and RO. Only by comparing the individual variables will the researcher see that the AO is low or null and therefore, the topic is not at risk of interagency overlap (though intra-agency overlap may be revealed through further investigation). The aggregated OI and more granular components are likely to be of interest to different users, and we plan to perform usability evaluations in the future to determine their usages.

3.3 What laws and agencies overlap?

To visualize ocean management overlaps, we demonstrate here a graphical representation of the previously defined data and metadata matrices. For this task, we used the social networking software UCINET version 6.170 [3] and NetDraw version 2.064 [2]. The document-agency authority metadata matrix served as the primary data input (Table 2). Agencies and documents were displayed as individual nodes with agencies labeled and each document (legal unit) represented by a circular node. A line was drawn from each document to its associated agency (or multiple agencies) (Figure 2). For example, the National Environmental Policy Act (42 U.S.C. 4321 et seq.) was connected to its authority agency of the Council on Environmental Quality (CEQ) because this agency has jurisdiction to implement the statute. Some statutes are under the authority of multiple agencies, such as the Clean Water Act. This Act is under the authority of the Environmental Protection Agency (EPA), the Army Corps of Engineers (ACE), and the Department of Homeland Security (DHS). Regulation nodes were linked to the agency that wrote them. Lastly, the topic (represented by a term or phrase) frequencies were added as attributes. Document nodes were resized to reflect relative frequency of each topic. These diagrams visually demonstrate what laws overlap, and consequently what agencies overlap given their authority over the topic-associated laws.

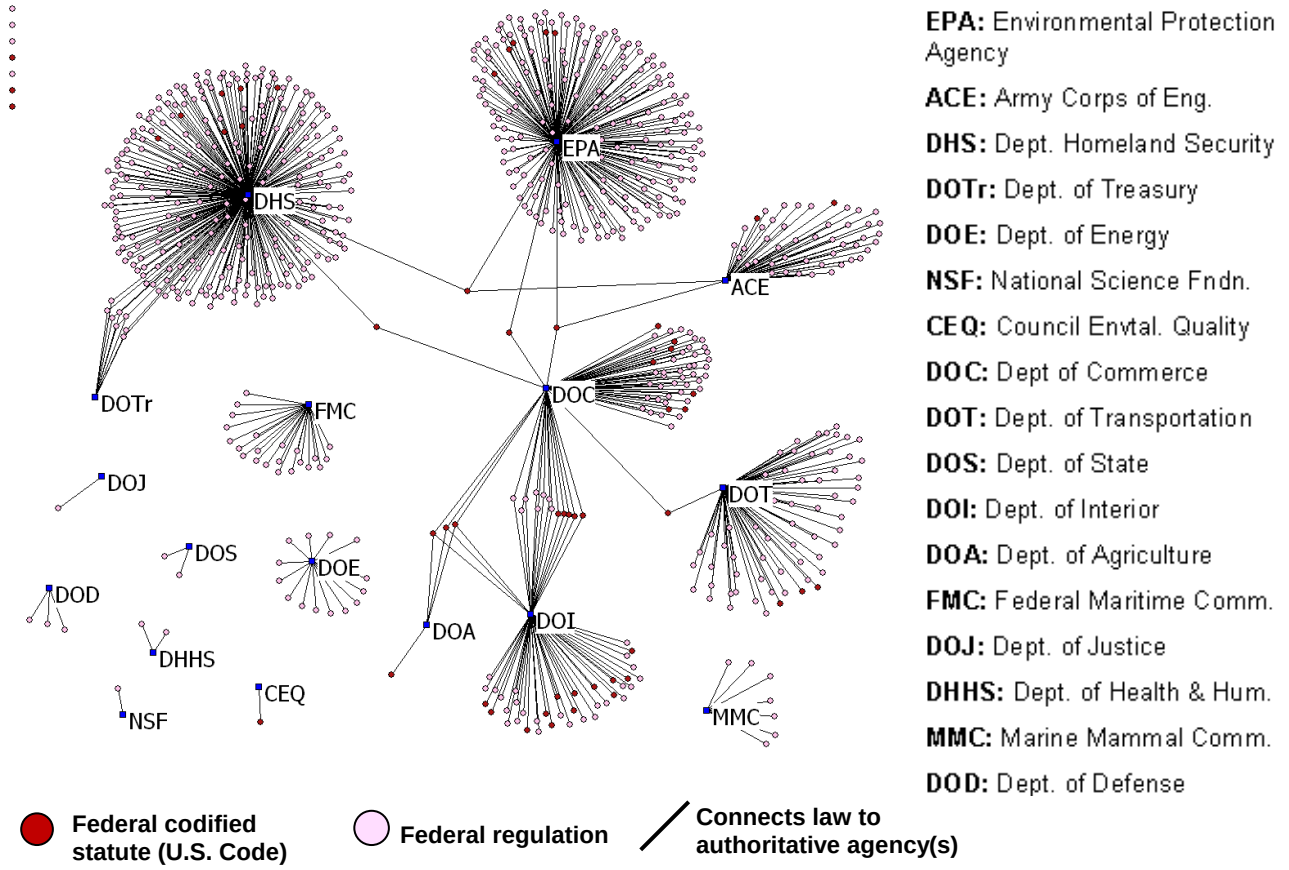


Figure 2. Metadata of agency authority for federal statutes and regulations. Laws (circular nodes) linked to their authoritative and/or implementing agencies (square nodes labeled with agency acronyms). The placement of agencies and length of lines are randomly generated. This is the foundational map from which the diagrams in Figure 5 were generated. In Figure 5, the law nodes are re-sized by the frequency in which a selected term occurs in the law.

4. RESULTS

This section first presents initial results justifying use of term and phrase counts to reveal a law's involvement in a given topic. Then we present results for the measurements of overlap, which are followed by graphic display of overlaps using the laws, topic frequencies, and associated agencies in network diagrams.

4.1 Topic frequencies

Initial results showed that federal United States laws that ranked as most involved for each of the topics accurately corresponded to the descriptions of the recent U.S. Commission on Ocean Policy report [15]. For example, the U.S. law containing the most references to the term 'fishing' (frequency = 726) was the Magnuson-Stevens Fishery Conservation and Management Act (16 U.S.C. 1866 et seq.). The laws that ranked second and third by their raw count of the same term were the regulations written by NOAA to implement the Magnuson-Stevens Act (50 CFR 600 et seq., 50 CFR 660 et seq.). Similarly, the authoritative agencies that ranked highest for each topic accurately corresponded to the USCOP report descriptions. For instance, the Department of Commerce (DOC), which in many cases then delegates authority to the NOAA, had authority over most of the laws for the topic of fishing (see Table 3, Figure 5c).

The general observation regarding a law's relative involvement provided sufficient justification that simple text analysis can be used to represent law and agency jurisdictions. As such, the following presents a summary of results for what topics ranked as the highest degree of overlap and a sample of what graphic display of these data can illustrate.

4.2 What topics are most fragmented from overlapping jurisdictions?

This subsection presents a summary of federal level results of the three individual variables (SO, RO, and AO) and then the results of the Overlap Index (OI). Results of the state levels of jurisdiction are briefly summarized for the Overlap Index. Table 4 provides excerpts of data used to calculate these three variables for the federal laws. For example, for the topic of 'fishing,' there were 31 statutory units. To obtain the Statute Overlap variable, we divided 31 by the total number of statutory units (55) for the geopolitical jurisdiction of the federal United States level.

Table 4. Sample of data used to calculate overlap variables for federal geopolitical jurisdiction

Units (federal only)	Units in collection	# units that refer to topic			
		Trans-portation	Pollut*	Fishing	Ballast
Statutes (USC)	55	43	35	31	5
Regulations (CFR)	670	265	260	114	86
Agencies	18	17	15	12	9

For the federal laws, the topics of 'transportation' (78%), 'fisher*' (69%), and 'pollut*' (64%) ranked as having the highest Statute Overlap. The top three topics ranked by Regulation Overlap were 'discharge' (48%), 'shipping' (43%), and 'navigation' (43%) for the federal laws. In terms of Agency Overlap, the topics of 'transportation' (94%), 'public health' (88%), 'pollut*' (83%), and 'discharge' (83%) measured the highest. To follow the examples of four topics of 'transportation', 'pollut*', 'fishing', and 'ballast', Figure 3 presents the variables measured for each for federal laws.

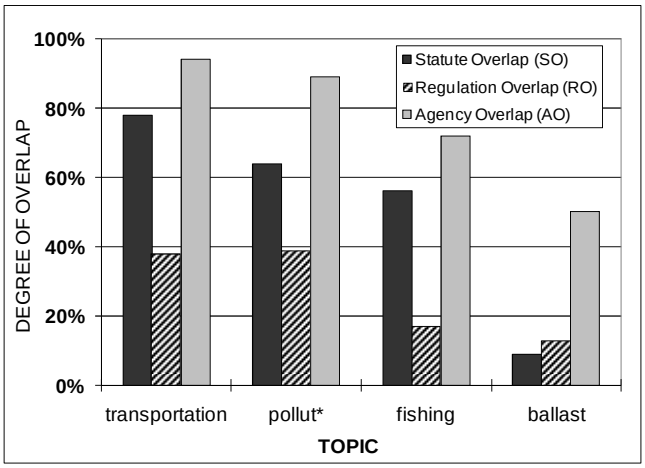


Figure 3. Three variables of overlap for sample of four topics in the U.S. federal level

Based on the Overlap Index from the three combined variables of the number of statutes, regulations, and agencies per topic, the issue of 'transportation' measured as the highest overlap for the U.S. federal level and all three states examined (Figure 4). For 'transportation,' the U.S. had 43 statutes, 256 regulations and 17 agencies involved, which resulted in an OI of 72%. Following the same computation, the OIs of the states of Washington, Oregon and California are 50%, 55% and 66% respectively. The topic 'agricultur*' ranked second in the Overlap Index for the states of California and Oregon, while OIs that ranked second for federal level and Washington were 'pollut*' (64%) and 'discharge' (49%) respectively. Figure 4 presents Overlap Index for the 46 topics for each of the four geopolitical jurisdictions investigated.

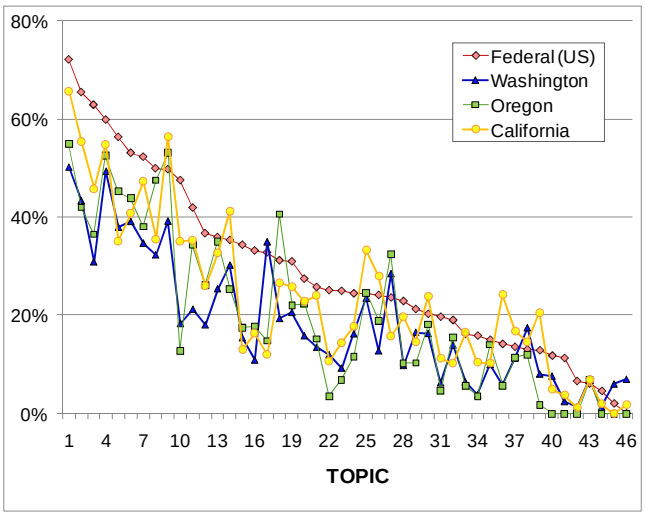


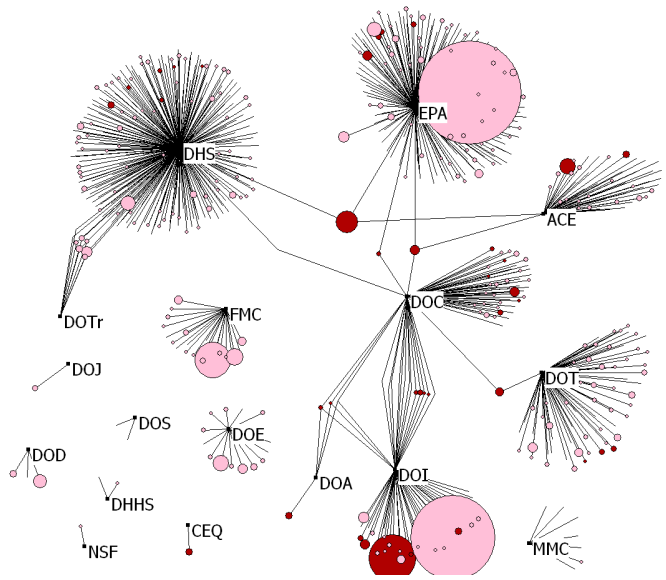
Figure 4. Overlap Index (OI) for topics investigated for each geopolitical jurisdiction. Key to topics: 1. transportation, 2. pollut*, 3. navigat*, 4. discharge, 5. fisher*, 6. port(s), 7. public health, 8. fishing, 9. agricultur*, 10. shipping, 11. mineral, 12. dredg*, 13. water quality, 14. contamina*, 15. ecosystem, 16. mammal, 17. shellfish, 18. estuar*, 19. bird, 20. sediment, 21. pesticide, 22. bulkhead, 23. ballast, 24. wastewater, 25. sewage, 26. climat*, 27. salmon, 28. oil spill, 29. aquaculture, 30. boating, 31. armor 32. spawn, 33. herbicid*, 34. sea level, 35. crab, 36. mercury, 37. nutrient, 38. oyster, 39. cattle, 40. invasive spec*, 41. sea otter, 42. algal bloom, 43. kelp, 44. nonindigenous spec*, 45. spartina, 46. geoduck.

Although the degree of overlap varied slightly for some topics, the results among jurisdictions were highly correlated. The topic of 'discharge' ranked within the five highest overlapping issues for each jurisdiction. Similarly, for all four jurisdictions investigated, the topics of 'fishing' and 'fisher*' ranked within the top ten. For California, Washington, and the U.S., the Overlap Index of 'pollut*' measured within the top five of each jurisdiction.

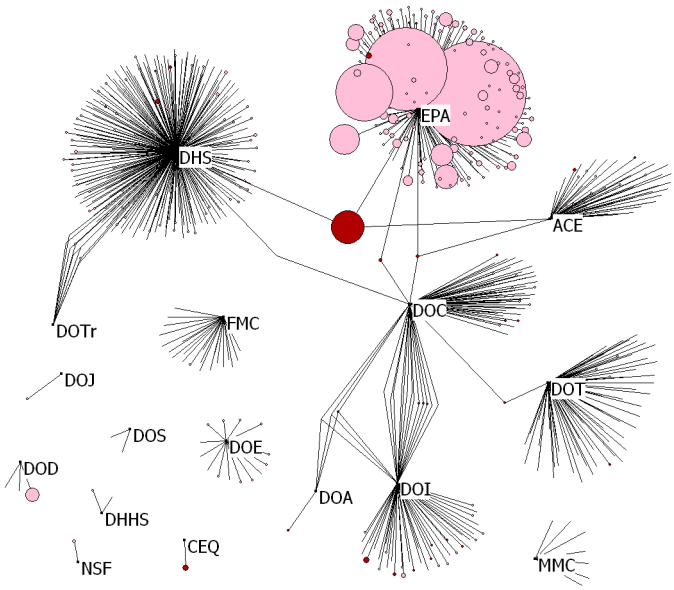
4.3 What laws functionally overlap, involving what agencies?

Numerical values of term frequencies revealed the laws overlapping for each topic. However, these long laundry lists of laws in tabular form are difficult and unpleasant to synthesize. As such, visual display of these data in network diagrams exposed multiple dimensions of the data, allowing for a more thorough and attractive interpretation. Diagrams were produced using the metadata table of agency authority to laws (see Table 2). Labeled nodes represent federal government agencies and lines were drawn from agencies to laws, which are represented by circular nodes (pink = regulations, red = statutes). These law nodes were then sized by the frequency of topic contained in the law (see Table 3). A sample of four topics for the federal level is presented in Figure 5 to demonstrate the utility of the graphical display.

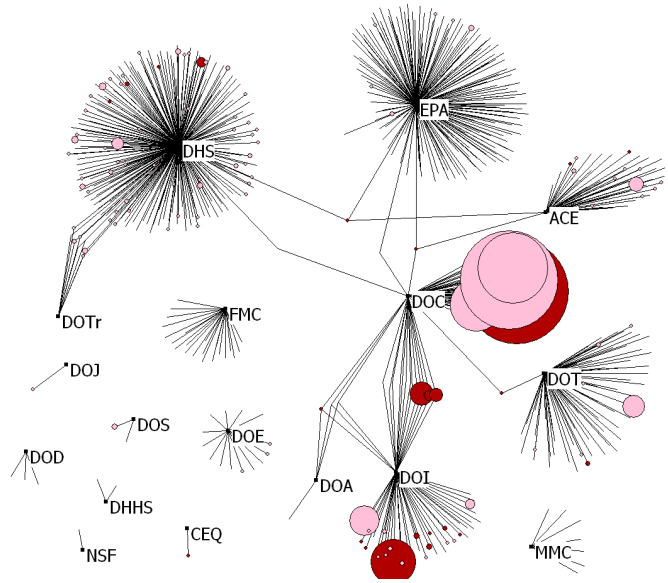
a. OI('transportation', Federal) = 72%



b. OI('pollut*', Federal) = 64%



c. OI('fishing', Federal) =49%



d. OI('ballast', Federal) = 24%

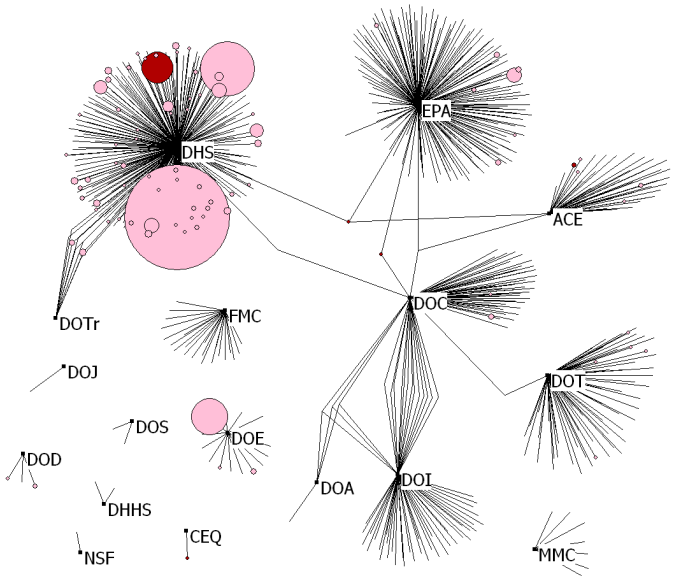


Figure 5. Overlapping United States federal laws and agencies for sample of four topics. Relative frequency of term or phrase in each law (document node size varies with frequency). Refer to Figure 2 for legend.

In the network diagrams, relational patterns and multiple dimensions were revealed that cannot be easily captured from tables or bar charts. For instance, the diagrams reflected that the topic of 'transportation' appeared to be more complex in its management relative to the topics of 'pollut*', 'fishing', and 'ballast' (Figure 5). Large nodes point to laws that contain a high frequency of references to the topic (represented by a word or phrase). Similarly the laws with no reference to the topic are eliminated, but the lines remain. There were several laws containing high frequency of the term 'transportation'. On the other hand, there were relatively few laws that refer to the term 'ballast' with high frequency. The complexity of each topic was revealed through the associated agencies that are linked to the laws. For example, the largest nodes in the 'transportation' diagram (Figure 5a) were connected to the Environmental Protection Agency (EPA) and the Department of Interior (DOI). In addition, medium size nodes were connected to several more agencies. In contrast, the majority of laws containing high frequency of the term 'pollut*' are under the authority of the EPA, conveying a relatively low complexity in terms of agency overlap

for this topic (Figure 5b). The agency primarily involved in 'fishing' appeared to be the Department of Commerce (Figure 5c) because the statute and regulations containing the highest frequency of the topic were linked to the DOC. Although there were relatively few document nodes for the topic 'ballast', the largest of these nodes were primarily connected to the DHS, which is the parent department of the U.S. Coast Guard (Figure 5d). Although the relative complexity displayed in the diagrams accurately matched the Overlap Index measurement, the visual depiction of the raw data provided the results in a more transparent manner, which can be used by policy-makers and other ocean-related stakeholders.

5. DISCUSSION

The discussion section includes a brief interpretation of key results, our plan of evaluation of usability and accuracy for the overlaps metric, and related work.

5.1 Interpretation of results

Results demonstrate the utility of text mining, even in its simplest form, for untangling overlapping jurisdictions in ocean management. Although government agencies report on their policies, functions, and duties, etc., generation of a baseline understanding of ocean management requires an objective overview. Of the 46 topics investigated, the one that ranked as having the highest Overlap Index was 'transportation' for each of the four geopolitical jurisdictions. This result was consistent with the findings of the recent U.S. Commission on Ocean Policy. After a multi-year examination of ocean management by government-appointed experts, the Commission found that management of the Shipping and Transportation sector was so fragmented that it needs to be restructured: "Statutory, regulatory, and policy differences among federal agencies with roles in marine transportation lead to fragmentation, competition, and in some cases, an inability to work collaboratively due to conflicting mandates" [15].

The quantitative aspects of a baseline assessment enable objective comparison across sectors. Combining the Overlap Index measurement with the graphical display of the overlap provided a comprehensive picture of the data. In comparing these results, we were able to see discrepancies between the generically calculated Overlap Index, which does not take into consideration term frequency or the relative involvement of multiple agencies illustrated in the network diagrams.

The simple but comprehensive tool has enormous potential, for example, to assist ecosystem-based management initiatives in defining priorities from data collection to stakeholder communication. Present applications identified where jurisdictional relationships and functions dictate the need for management coordination. Even from the prototype text analysis with transparent methods, the lucid identification of the multiple agencies involved in management of various topics provides policy-makers with a roadmap for locating where (between whom) coordination should exist.

5.2 Evaluation

Initial testing of the accuracy of results has begun through a series of interviews with approximately 25 experts in ocean and coastal management. These experts included government agency representatives, academic scientists (both social and ecological disciplines), and non-governmental organizations. Conducted in 2007, these meetings were used to steer the line of inquiry to produce useful and accurate information about ocean management overlap. Based on the last set of interviews, suggestions for improvement will be woven into the analysis in future work, including a more thorough survey to evaluate accuracy of results. This future study could survey the degree to which each agency finds itself involved in the given topics. These survey results would be compared to the text analysis results to determine degree and patterns of error that text analysis reveals.

From the input of experts, it is also apparent that future work needs to include input of synonyms for topics investigated. Inclusion of multiple terms or phrases to represent a single topic could improve the accuracy of results. This improvement could also be intertwined with the verification of result survey to test how much inclusion of synonyms (and what rules are needed for synonyms) can increase the accuracy of results.

Once the algorithm is fine tuned to meet the needs of coastal and ocean management stakeholders, automation of the overlaps tool will require additional surveys to establish usability for the potential users.

5.3 Future and related work

The term-document matrix data yielded by this technique affords excellent opportunities to use information retrieval statistics and other advanced text analysis methods, such as the vector space model and other content analyses [7, 12]. However, even raw frequencies provide information that pre-empts the need to read hundreds of documents to ascertain an extremely detailed, relative assessment of statutory and regulatory overlap. In addition, text analysis can be employed with any set of laws or policy documents on any subject. As already seen with work on construction and water quality law [8], the application of text analysis can help untangle management in different domains. Recognizing the growing problem of increasing legislation requiring review, a small group of computer scientists and engineers has been developing algorithms using information retrieval statistics and methodologies for navigating through legal documents [13].

Future research also will further develop the technique to prioritize what agencies need to coordinate around any given topic. With more topics, graphic display through network diagrams of these data could provide a valuable teaching tool for marine policy courses. In addition, text analysis is being applied to the collection of ocean related legal documents to investigate gaps in management in the context of a given conceptually modeled ecosystem. Combining the overlaps analysis with gaps analysis may prove to be the most useful for marine management initiatives because it could be used to locate what agencies and through what laws gaps in management could be filled.

6. CONCLUSION

Text analysis of the laws has the potential to provide a thorough synopsis of which agencies and laws manage various topic issues in the ocean. The approach to measuring overlaps demonstrates how an interdisciplinary integration of methods and perspective

can be used to illuminate the black box of ocean management. It is our expectation that by providing a systematic and repeatable technique, policy-makers and other stakeholders will be better equipped to make new laws consistent with existing ones. Rather than passing new legislation or writing new regulations that unintentionally conflict with existing ones, if necessary, policy-makers will be able to address the inconsistency in new law. With improved knowledge of management, policy-makers can implement and adapt future regulation of the marine environment, in particular for emerging uses, in a more integrated and consistent manner. Furthermore, this tool can be used to define high priority areas for alleviating uncoordinated ocean management overlaps.

Though contributing through the lens of ocean management, this prototype text analysis technique can be applied to any set of problems of legal and government agency overlap. With more and more regulations created and increased competition for agency authority, overlapping jurisdictions and the need for improved cooperation will continue to increase. By supplying policy-makers with cross jurisdictional information about overlaps, this information can assist them to begin untangling and alleviating not only overlapping jurisdictions, but also the subsequent inefficiencies and ineffectiveness in existing management.

7. ACKNOWLEDGEMENTS

This project was partially supported by funding from the University of California Coastal Environmental Quality Initiative Graduate Fellowship**,** California Sea Grant and Ocean Protection Council, and the National Center for Ecological Analysis and Synthesis (NCEAS) graduate student support. Ekstrom was also supported an assistantship by the working group on ocean zoning/Ecosystem-Based Management through NCEAS.

8. REFERENCES

[1] Baron, J. R. and Thompson, P. The Search Problem: Posed by Large Heterogeneous Data Sets in Litigation: Possible Future Approaches to Research. in *Proceedings of the 11th international conference on Artificial Intelligence and Law* (Stanford, CA, 2007), 141-147.

[2] Borgatti, S. P. *NetDraw: Graph Visualization Software*. Analytic Technologies, Harvard, MA, 2002.

[3] Borgatti, S. P., Everett, M. G. and Freeman, L. C. *Ucinet for Windows: Software for Social Network Analysis*. Analytic Technologies, Inc., Harvard, MA, 2002.

[4] Cortner, H. J., Wallace, M. G., Burke, S. and Moote, M. M. Institutions matter: the need to address the institutional challenges of ecosystem management. *Landscape and Urban Planning, 40* (1-3). 159-166.

[5] Crowder, L., Osherenko, G., Young, O., Airame, S., Norse, E. A., Baron, N., Day, J. C., Douvere, F., Ehler, C. N., Halpern, B. S., Langdon, S. J., McLeod, K. L., Ogden, J. C., Peach, R. E., Rosenberg, A. A. and Wilson, J. A. Resolving Mismatches in U.S. Ocean Governance. *Science, 313* (5787). 617-618.

[6] Ekstrom, J. A. Evaluating Current Ocean Management Systems to Facilitate the Development of Ecosystem-Based Management. in *California and the World Ocean Book of Abstracts*, Long Beach, CA, 2006.

[7] Krippendorff, K. *Content Analysis*. Sage Publications, Thousand Oaks, 2004.

[8] Lau, G. T., Law, K. and Wiederhold, G. A Relatedness Analysis of Government Regulations using Domain Knowledge and Structural Organization. *Information Retrieval, 9* (6). 657-680.

[9] Lluch-Belda, D., Luch-Cota, D. B. and Luch-Cota, S. E. Interannual Variability Impacts on the California Current Large Marine Ecosystem: Trends in Exploration, Protection, and Research. in Hempel and Sherman (eds) *Large Marine Ecosystems of the World*. Elsevier, Boston, MA, 2003.

[10] McArdle, D. A. *California marine protected areas*. California Sea Grant, University of California, La Jolla, CA, 1997.

[11] McLeod, K. L., Jane Lubchenco, Palumbi, S. R. and Rosenberg, A. A. Scientific Consensus Statement on Marine Ecosystem-Based Management. *Communication Partnership for Science and the Sea*, 2005.

[12] Salmon, G., Wong, A. and Yang, C. S. A vector space model for automatic indexing. *Comm. ACM, 18* (11). 613-620.

[13] Sherman, K. The Large Marine Ecosystem Concept: Research and Management Strategy for Living Marine Resources. *Ecological Applications, 1* (4). 349-360.

[14] Sherman, K. The Large Marine Ecosystem Approach for Assessment and Management of Ocean Coastal Waters. in Hennessey and Sutinen (eds) *Sustaining Large Marine Ecosystems: The Human Dimension*. Elsevier Boston, MA.

[15] USCOP (U.S. Commission on Ocean Policy). *An Ocean Blueprint for the 21st Century Final Report of the U.S. Commission on Ocean Policy*. 2004.

[16] Wang, H. Ecosystem Management and Its Application to Large Marine Ecosystems: Science, Law, and Politics. *Ocean Development and International Law, 35*. 41-74.

[17] Young, O. R. *The Institutional Dimensions of Environmental Change: Fit, Interplay, and Scale*. MIT Press, Cambridge, MA, 2002.