Article

Evaluation of Diversification Techniques for Legal Information Retrieval †

Marios Koniaris [1,*], Ioannis Anagnostopoulos [2] and Yannis Vassiliou [1]

[1] Knowledge and Database Systems Laboratory, Divison of Computer Science, School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou 9, Politechnioupoli Zographou, 15780 Athens, Greece; yv@cs.ntua.gr

[2] Department of Computer Science and Biomedical Informatics, School of Sciences, University of Thessaly, Papassiopoulou 2-4, 35131 Lamia, Greece; janag@dib.uth.gr

[*] Correspondence: mkoniari@dblab.ece.ntua.gr, Tel.: +30-210-7721602

† This paper is an extended version of our paper published in Koniaris,M; Anagnostopoulos,I; Vassiliou,Y. Diversifying the Legal Order. Proceedings of Artificial Intelligence Applications and Innovations: 12th IFIP WG 12.5 International Conference and Workshops, AIAI 2016 (Springer International Publishing), 2016, pp 499-509

Abstract: "Public legal information from all countries and international institutions is part of the common heritage of humanity. Maximizing access to this information promotes justice and the rule of law". In accordance with the aforementioned declaration on Free Access to Law by Legal information institutes of the world, a plethora of legal information is available through the Internet, while the provision of legal information has never before been easier. Given that law is accessed by a much wider group of people, the majority of whom are not legally trained or qualified, diversification techniques, should be employed in the context of legal information retrieval, as to increase user satisfaction. We address diversification of results in legal search by adopting several state of the art methods from the web search, network analysis and text summarization domains. We provide an exhaustive evaluation of the methods, using a standard data set from the Common Law domain that we subjectively annotated with relevance judgments for this purpose. Our results i) reveal that users receive broader insights across the results they get from a legal information retrieval system, ii) demonstrate that web search diversification techniques outperform other approaches (e.g., summarization-based, graph-based methods) in the context of legal diversification and iii) offer balance boundaries between reinforcing relevant documents or sampling the information space around the legal query.

Data Set: https://github.com/mkoniari/LegalDivEval

Keywords: diversity; algorithms; legal information retrieval

1. Introduction

Nowadays, as a consequence of many open data initiatives, more and more publicly available portals and datasets provide legal resources to citizens, researchers and legislation stakeholders. Thus, legal data that was previously available only on a specialized audience and in "closed" format is now freely available on the internet.

Portals as the EUR-Lex[1], the European Union's database of regulations, the on-line version of the United States Code [2], United Kingdom [3], Brazil [4] and the Australian [5], just to mention a few, serve as

[1] http://eur-lex.europa.eu/
[2] http://uscode.house.gov/
[3] http://www.legislation.gov.uk/
[4] http://www.lexml.gov.br/
[5] https://www.comlaw.gov.au/

an endpoint to access millions of regulations, legislation, judicial cases, or administrative decisions. Such portals allow for multiple search facilities, as to assist users to find the information they need. For instance the user can perform simple search operations or utilize predefined classificatory criteria e.g., year, legal basis, subject matter to find relevant to her information needs legal documents.

At the same time, however, the amount of Open Legal Data makes it difficult, both for legal professionals or the citizens to find relevant and useful legal resources. For example, it is extremely difficult to search for a relevant case law, by using boolean queries or the references contained in the judgment. Consider, for example, a patent lawyer who want to find patents as reference case and submits a user query to retrieve information. A diverse result, i.e. a result containing several claims, heterogeneous statutory requirements and conventions -varying in the numbers of inventors and other characteristics- is intuitively more informative than a set of homogeneous results that contain only patents with similar features. In this paper, we propose a novel way to efficiently and effectively handle similar challenges when seeking information in the legal domain.

Diversification is a method of improving user satisfaction by increasing the variety of information shown to user. As a consequence, the number of redundant items in a search result list should decrease, while the likelihood that a user will be satisfied with any of the displayed results should increase. There has been extensive work on query results diversification (see Section 2), where the key idea is to select a small set of results that are sufficiently dissimilar, according to an appropriate similarity metric.

Diversification techniques in legal information systems can be helpful not only for citizens but also for law issuers and other legal stakeholders in companies and large organizations. Having a big picture of diversified results, issuers can choose or properly adapt the legal regime that better fits their firms and capital needs, thus helping them operate more efficiently. In addition, such techniques can also help lawmakers, since deep understanding of legal diversification promotes evolution to better and fairer legal regulations for the society [1].

In this work, we address result diversification in the legal IR. To this end, we adopt various methods from the literature that are introduced for text summarization [LexRank [2] and Biased LexRank [3]], graph-based ranking [DivRank [4] and Grasshopper [5]] and web search result diversification [MMR [6], Max-Sum [7], Max-Min [7] and MonoObjective [7]]. We evaluate the performance of the above methods on a legal corpus subjectively annotated with relevance judgments using metrics employed in TREC Diversity Tasks. To the best of our knowledge none of these methods were employed in the context of diversification in legal IR and evaluated using diversity-aware evaluation metrics.

Our findings reveal that i) diversification methods, employed in the context of legal IR, demonstrate notable improvements in terms of enriching search results with otherwise hidden aspects of the legal query space and ii) web search diversification techniques outperform other approaches e.g., summarization-based, graph-based methods, in the context of legal diversification. Furthermore, our accuracy analysis can provide helpful insights for legal IR systems, wishing to balance between reinforcing relevant documents, result set similarity, or sampling the information space around the query, result set diversity.

The remainder of this paper is organized as follows: Section 2 reviews previous work in query result diversification, diversified ranking on graphs and in the field of legal text retrieval, while it stresses out the differentiation and contribution of this work. Section 3 introduces the concepts of search diversification and presents diversification algorithms, while section 4 describes our experimental results and discuss their significance. Finally, we draw our conclusions and future work aspects in Section 5.

2. Related Work

In this section, we first present related work on query result diversification, afterwards on diversified ranking on graphs and then on legal text retrieval techniques.

2.1. Query Result Diversification

Result diversification approaches have been proposed, as a means to tackle ambiguity and redundancy, in various problems and settings e.g., diversifying historical archives [8], diversifying user comments on news articles [9], diversifying microblog posts [10,11], diversifying image retrieval results [12], diversifying recommendations [13], utilizing a plethora of algorithms and approaches e.g., learning algorithms [14], approximation algorithms [15], page rank variants [16], conditional probabilities [17].

Users of (Web) search engines typically employ keyword-based queries to express their information needs. These queries are often underspecified or ambiguous to some extent [18]. Different users who pose exactly the same query may have very different query intents. Simultaneously the documents retrieved by an IR system may reflect superfluous information. Search result diversification aims to solve this problem, by returning diverse results that can fulfill as many different information needs as possible. Published literature on search result diversification is reviewed in [19,20].

The maximal marginal relevance criterion (MMR), presented in [6], is one of the earliest works on diversification and aims at maximizing relevance while minimizing similarity to higher ranked documents. Search results are re-ranked as the combination of two metrics, one measuring the similarity among documents and the other the similarity between documents and the query. In [7] a set of diversification axioms is introduced and it is proven that it is not possible for a diversification algorithm to satisfy all of them. Additionally, since there is no single objective function suitable for every application domain, the authors propose three diversification objectives, which we adopt in our work. These objectives differ in the level where the diversity is calculated, e.g., whether it is calculated per separate document or on the average of the currently selected documents.

In another approach, researchers utilized explicit knowledge as to diversify search results. In [21] the authors proposed a diversification framework, where the different aspects of a given query are represented in terms of sub-queries and documents are ranked based on their relevance to each sub-query, while in [22] the authors proposed a diversification objective that tries to maximize the likelihood of finding a relevant document in the top-k positions given the categorical information of the queries and documents. Finally, the work described in [23] organizes user intents in a hierarchical structure and proposes a diversification framework to explicitly leverage the hierarchical intent.

The key difference between these works and the ones utilized in this paper is that we do not rely on external knowledge e.g. taxonomy, query logs to generate diverse results. Queries are rarely known in advance, thus probabilistic methods to compute external information are not only expensive to compute, but also have a specialized domain of applicability. Instead, we evaluate methods that rely only on implicit knowledge of the legal corpus utilized and on computed values, using similarity (relevance) and diversity functions (e.g., tf-idf cosine similarity) in the data domain.

2.2. Diversified Ranking on Graphs

Many network-based ranking approaches have been proposed to rank objects according to different criteria [24] and recently diversification of the results has attracted attention. Research is currently focused on two directions: a greedy vertex selection procedure and a vertex reinforced random walk. The greedy vertex selection procedure, at each iteration, selects and removes from the graph the vertex with maximum random walk based ranking score. One of the earlier algorithms that address diversified ranking on graphs by vertex selection with absorbing random walks is Grasshopper [5]. A diversity-focused ranking methodology, based on reinforced random walks, was introduced in [4]. Their proposed model, DivRank, incorporates the rich-gets-richer mechanism to PageRank [25] with reinforcements on transition probabilities between vertices. We utilize these approaches in our diversification framework considering the connectivity matrix of the citation network between documents that are relevant for a given user query.

2.3. Legal Text Retrieval

In respect to legal text retrieval that traditionally relies on external knowledge sources, such as thesauri and classification schemes, various techniques are presented in [26]. Several supervised learning methods have been proposed to classify sources of law according to legal concepts [27–29]. Ontologies and thesaurus have been employed to facilitate information retrieval [30–33] or to enable the interchange of knowledge between existing legal knowledge systems [34]. Legal document summarization [35–37] has been used as a way to make the content of the legal documents, notably cases, more easily accessible. We also utilize state of the art summarizations algorithms but under a different objective: we aim to maximize diversity of the result set for a given query.

Finally, a similar approach with our work is described in [38], where the authors utilize information retrieval approaches to determine which sections within a bill tend to be outliers. However, our work differs in a sense that we maximize the diversify of the result set, rather than detect section outliers within a specific bill.

In another line of work citation analysis has been used in the field of law to construct case law citation networks [39][6]. Case law citation networks contain valuable information, capable of measuring legal authority [40], identifying authoritative precedent[7] [41], evaluating the relevance of court decisions [42] or even assisting summarizing legal cases [43], thus showing the effectiveness of citation analysis in the Case law domain. While the American legal system has been the one that has undergone the widest series of studies in this direction, recently various researchers applied network analysis in the Civil law domain as well. The authors of [44] propose a network-based approach to model the law. Network analysis techniques where also employed in [45] demonstrating an online toolkit allowing legal scholars to apply Network analysis and visual techniques to the entire corpus of EU case law. In this work, we also utilize citation analysis techniques and construct the Legislation Network, as to cover a wide range of possible aspects of a query.

3. Legal Document ranking using diversification

At first, we define the problem addressed in this paper and provide an overview of the diversification process. Afterwards, legal document's features relevant for our work are introduced and distance functions are defined. Finally, we describe the diversification algorithms used in this work.

3.1. Diversification Overview

Result diversification is a trade-off between finding relevant to the user query documents and diverse documents in the result set. Given a set of legal documents and a query, our aim is to find a set of relevant and representative documents and to select these documents in such a way that the diversity of the set is maximized. More specifically, the problem is formalized as follows:

Definition 1 (Legal document diversification)**.** *Let q be a user query and N a set of documents relevant to the user query. Find a subset $S \subseteq N$ of documents that maximize an objective function f that quantifies the diversity of documents in S.*

$$S = \underset{\substack{|S|=k \\ S\subseteq N}}{\operatorname{argmax}}\, f(N) \tag{1}$$

Figure 1, illustrates the overall workflow of the diversification process. At the highest level, the user submits his/ her query as a way to express an information need and receives relevant documents.

[6] case documents usually cite previous cases, which in turn may have cited other cases and thus a network is formed over time with these citations between cases.

[7] legal norm inherited from English common law that encourages judges to follow precedent by letting the past decision stand.

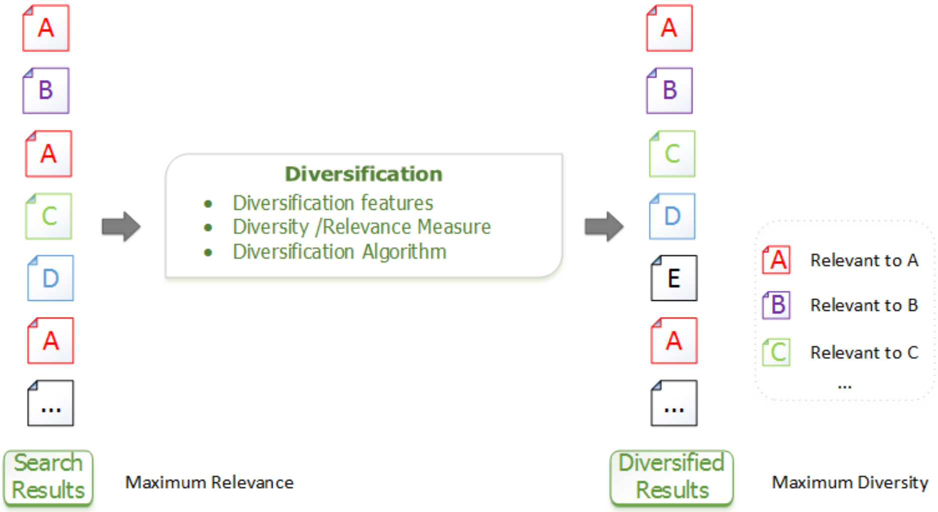


Figure 1. Diversification Overview

From the relevance-oriented ranking of documents we derive a diversity-oriented ranking, produced by seeking to achieve both coverage and novelty at the same time. Significant components of the process include:

- *Ranking Features*, features of legal documents that will be used in the ranking process.
- *Distance Measures*, functions to measure the similarity between two legal documents and the relevance of a query to a given document.
- *Diversification Heuristics*, heuristics to produce a subset of diverse results.

3.2. Ranking Features/ Distance Measures

Typically, diversification techniques measure diversity in terms of content, where textual similarity between items is used in order to quantify information similarity. In the Vector Space model [46], each document u can be represented as a term vector $U = (is_{w1u}, is_{w2u}, ..., is_{wmu})^T$, where $w_1, w_2, ..., w_m$ are all the available terms, and is can be any popular indexing schema e.g. $tf, tf-idf, logtf-idf$. Queries are represented in the same manner as documents.

Following we define:

- **Document Similarity**. Various well-known functions from the literature (e.g. Jaccard, cosine similarity etc.) can be employed at computing the similarity of legal documents. In this work, we choose cosine similarity as a similarity measure, thus the similarity between documents u and v, with term vectors U and V is:

$$sim(u,v) = \cos(u,v) = \frac{U \cdot V}{\parallel U \parallel \parallel V \parallel} \tag{2}$$

- **Document Distance**. The distance of two documents is

$$d(u,v) = 1 - sim(u,v) \tag{3}$$

- **Query Document Similarity**. The relevance of a query q to a given document u can be assigned as the initial ranking score obtained from the IR system, or calculated using the similarity measure e.g. cosine similarity on the corresponding term vectors

$$r(q,u) = \cos(q,u) \tag{4}$$

3.3. Diversification Heuristics

Diversification methods usually retrieve a set of documents based on their relevance scores, and then re-rank the documents so that the top-ranked documents are diversified to cover more query subtopics. Since the problem of finding an optimum set of diversified documents is NP-hard, a greedy algorithm is often used to iteratively select the diversified set S.

Let N the document set, $u, v \in N$, $r(q,u)$ the relevance of u to the query q, $d(u,v)$ the distance of u and v, $S \subseteq N$ with $|S| = k$ the number of documents to be collected and $\lambda \in [0..1]$ a parameter used for setting trade-off between relevance and similarity. In this paper, we focus on the following representative diversification methods:

- **MMR:** Maximal Marginal Relevance [6], a greedy method to combine query relevance and information novelty, iteratively constructs the result set S by selecting documents that maximizes the following objective function

$$f_{MMR}(u,q) = (1-\lambda)\ r(u,q) + \lambda \sum_{v \in S} d(u,v) \tag{5}$$

Algorithm 1 Produce diverse set of results with MMR

```
Input: Set of candidate results N, size of diverse set k
Output: Set of diverse results $S \subseteq N, |S| = k$
  $S = \emptyset$
  $N_i = argmax_{N v \in N}(r(v,q))$        ▷ initialize with the highest relevant to the query document
  Set $S = S \cup \{i\}$
  Set $N = N \setminus \{i\}$
  while $|S| < k$ do
      Find $u = argmax_{N v \in N}(f_{MMR}(v,q))$        ▷ iteratively select document that maximize Eq. 5
      Set $S = S \cup \{u\}$
      Set $T = T \setminus \{u\}$
  end while
```

MMR incrementally computes the standard relevance-ranked list when the parameter $\lambda = 0$, and computes a maximal diversity ranking among the documents in N when $\lambda = 1$. For intermediate values of $\lambda \in [0..1]$, a linear combination of both criteria is optimized. In MMR Algorithm 1, the set S is initialized with the document that has the highest relevance to the query. Since the selection of the first element has a high impact on the quality of the result, MMR often fails to achieve optimum results.

- **MaxSum:** The Max-sum diversification objective function [7] aims at maximizing the sum of the relevance and diversity in the final result set. This is achieved by a greedy approximation, Algorithm 2, that selects a pair of documents that maximizes Eq. 6 in each iteration.

$$f_{MAXSUM}(u,v,q) = (1-\lambda)\ (r(u,q) + r(v,q)) + 2\lambda\ d(u,v) \tag{6}$$

where (u,v) is a pair of documents, since this objective considers document pairs for insertion. When $|S|$ is odd, in the final phase of the algorithm an arbitrary element in N is chosen to be inserted in the result set S.

MaxSum Algorithm 2, at each step, examines the pairwise distances of the candidate items N and selects the pair with the maximum pairwise distance, to insert into the set of diverse items S.

- **MaxMin:** The Max-Min diversification objective function [7] aims at maximizing the minimum relevance and dissimilarity of the selected set. This is achieved by a greedy approximation,

Algorithm 2 Produce diverse set of results with MaxSum

```
Input: Set of candidate results N, size of diverse set k
Output: Set of diverse results $S \subseteq N, |S| = k$
  $S = \emptyset$
  for $i = 1 \to \lfloor \frac{k}{2} \rfloor$ do
      Find $(u,v) = argmax_{x,y \in N}(f_{MAXSUM}(x,y,q))$        ▷ Select pair of docs that maximize Eq 6
      Set $S = S \cup \{u,v\}$
      Set $N = N \setminus \{u,v\}$
  end for
  if $k$ is odd then
      $S = S \cup \{i\}, N_i \in N$        ▷ If $k$ is odd add an arbitrary document to $S$
  end if
```

Algorithm 3, that initially selects a pair of documents that maximize Eq. 7 and then in each iteration selects the document that maximizes Eq. 8

$$f_{MAXMIN}(u,v,q) = (1-\lambda)\ (r(u,q) + r(v,q)) + \lambda\ d(u,v) \tag{7}$$

$$f_{MAXMIN}(u,q) = \min_{v \in S} d(u,v) \tag{8}$$

MaxMin Algorithm 3, at each step, it finds, for each candidate document its closest document belonging to S and calculates their pairwise distance d_{MIN}. The candidate document that has the maximum distance d_{MIN} is inserted into S.

Algorithm 3 Produce diverse set of results with MaxMin

```
Input: Set of candidate results N, size of diverse set k
Output: Set of diverse results $S \subseteq N, |S| = k$
  $S = \emptyset$
  Find $(u,v) = argmax_{x,y \in N}(f_{MAXMIN}(x,y,q))$        ▷ initially selects documents that maximize Eq. 7
  Set $S = S \cup \{u,v\}$
  while $|S| < k$ do
      Find $u = argmax_{x \in N \setminus S}(f_{MAXMIN}(x,q))$        ▷ select document that maximize Eq. 8
      Set $S = S \cup \{u\}$
  end while
```

- **MonoObjective:** MonoObjective[7] combines the relevance and the similarity values into a single value for each document. It is defined as:

$$f_{MONO}(u,q) = r(u,q) + \frac{\lambda}{|N|-1} \sum_{v \epsilon N} d(u,v) \tag{9}$$

Algorithm 4 approximates the Mono-Objective. The algorithm, at initialization step, calculates a distance score for each candidate document. The objective function weights each document's similarity to the query with the average distance of the document with the rest documents. After the initialization step, where scores are calculated, they are not updated after each iteration of the algorithm. So, each step consists in selecting the document from the remaining candidates set with the maximum score and inserting it into S.

- **LexRank:** LexRank [2], is a stochastic graph-based method for computing the relative importance of textual units. A document is represented as a network of inter-related sentences, and a connectivity matrix based on intra-sentence similarity is used as the adjacency matrix of the graph representation of sentences.

Algorithm 4 Produce diverse set of results with MonoObjective

```
Input: Set of candidate results N, size of diverse set k
Output: Set of diverse results $S \subseteq N, |S| = k$
  $S = \emptyset$
  for $x_i \in N$ do
      $d(x_i) = f_{MONO}(x, q)$                ▷ Calculate scores based on Eq. 9
  end for
  while $|S| < k$ do
      Find $u = argmax_{x_{i \in N}} d(x_i)$    ▷ Sort and select $top - k$ documents
      Set $S = S \cup \{u\}$
      Set $N = N \setminus \{u\}$
  end while
```

In our setting, instead of sentences, we use documents that are in the initial retrieval set N for a given query. In this way, instead of building a graph using the similarity relationships among the sentences based on an input document, we utilize document similarity on the result set. If we consider documents as nodes, the result set document collection can be modeled as a graph by generating links between documents based on their similarity score as in Eq. 2. Typically, low values in this matrix can be eliminated by defining a threshold so that only significantly similar documents are connected to each other. But as in all discretization operations, this means an information loss. Instead we choose to utilize the strength of the similarity links. This way we use the cosine values directly to construct the similarity graph, obtaining a much denser but weighted graph. Furthermore we normalize our adjacency matrix B, as to make the sum of each row equal to 1.

Thus, in LexRank scoring formula Eq. 10, Matrix B captures pairwise similarities of the documents and square matrix A, which represents the probability of jumping to a random node in the graph, has all elements set to $1/M$, where M is the number of documents.

$$p = [(1 - \lambda)\ A + \lambda\ B]^T p \tag{10}$$

The LexRank Algorithm 5 applies a variation of PageRank [25] over a document graph. A random walker on this Markov chain chooses one of the adjacent states of the current state with probability $1 - \lambda$, or jumps to any state in the graph, including the current state, with probability λ. Note that we interchanged $1 - \lambda$ and λ interpolation parameters in the original LexRank formula [2], as to acquire comparable results across all tested algorithms.

Algorithm 5 Produce diverse set of results with LexRank

```
Input: Set of candidate results N, size of diverse set k
Output: Set of diverse results $S \subseteq N, |S| = k$
  $S = \emptyset$
  $A_{|N||N|} = 1/M$
  for $u, v \in N$ do
      $B_{u,v} = d(u, v)$     ▷ Calculate connectivity matrix based on document similarity Eq. 2
  end for
  $p = f_{powermethod}(B, A)$     ▷ Calculate stationary distribution of Eq. 10. (Omitted for clarity)
  while $|S| < k$ do
      Find $u = argmax_{x_{i \in N}} p(x_i)$    ▷ Sort and select $top - k$ documents
      Set $S = S \cup \{u\}$
      Set $N = N \setminus \{u\}$
  end while
```

- **Biased LexRank:** Biased LexRank [3] provides for a LexRank extension that takes into account a prior document probability distribution e.g., the relevance of documents to a given query.

Biased LexRank scoring formula Eq. 11, is analogous to LexRank scoring formula Eq. 10, with matrix A, which represents the probability of jumping to a random node in the graph, proportional to the query document relevance.

$$p = [(1-\lambda)\ A + \lambda\ B]^T p \tag{11}$$

Algorithm 5 is also used to produce a diversity oriented ranking of results with the Biased LexRank method. In Biased LexRank scoring formula Eq. 11, we set Matrix B as the connectivity matrix based on document similarity for all documents that are in the initial retrieval set N for a given query and Matrix A elements proportional to the query document relevance.

- **DivRank:** DivRank [4] balances popularity and diversity in ranking, based on a time-variant random walk. In contrast to PageRank [25] which is based on stationary probabilities, DivRank assumes that transition probabilities change over time, they are reinforced by the number of previous visits to the target vertex. If $p_T(u,v)$ is the transition probability from any vertex u to vertex v at time T, $p^*(d_j)$ is the prior distribution that determines the preference of visiting vertex d_j, and $p_0(u,v)$ is the transition probability from u to v prior to any reinforcement then,

$$p_T(d_i,d_j) = (1-\lambda).p^*(d_j) + \lambda.\frac{p_0(d_i,d_j).N_T(d_j)}{D_T(d_i)} \tag{12}$$

where $N_T(d_j)$ is the number of times the walk has visited d_j up to time T and,

$$D_T(d_i) = \sum_{d_j \in V} p_0(d_i,d_j)N_T(d_j) \tag{13}$$

DivRank was originally proposed in a query independent context, thus it is not directly applicable to diversification of search results. We introduce a query dependent prior and thus utilize DivRank into a query dependent ranking schema. In our setting, we use documents that are in the initial retrieval set N for a given query q, create the citation network between those documents and apply DivRank algorithm to select top-k divers documents in S.

Algorithm 6 Produce diverse set of results with DivRank

```
Input: Set of candidate results N, size of diverse set k
Output: Set of diverse results $S \subseteq N, |S| = k$
  $S = \emptyset$
  for $u, v \in N$ do
      $B(u,v) = A_{u,v}$        ▷ connectivity matrix is based on citation network adjacency matrix
  end for
  $p = f_{powermethod}(B)$        ▷ Calculate stationary distribution of Eq. 12. (Omitted for clarity)
  while $|S| < k$ do
      Find $u = argmax_{x_{i \in N}} p(x_i)$        ▷ Sort and select $top - k$ documents
      Set $S = S \cup \{u\}$
      Set $N = N \setminus \{u\}$
  end while
```

- **Grasshopper:** A similar with DivRank ranking algorithm, is described in [5]. This model starts with a regular time-homogeneous random walk and in each step the vertex with the highest weight is set as an absorbing state.

$$p_T(d_i,d_j) = (1-\lambda).p^*(d_j) + \lambda.\frac{p_0(d_i,d_j).N_T(d_j)}{D_T(d_i)} \tag{14}$$

Table 1. Parameters tested in the experiments

Parameter	Range
algoritmhs tested	MMR, MaxMin, MaxSum, Mono, LexRank, BiasedLexRank, DivRank, GrassHopper
tradeoff l values	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
candidate set size n = $\|N\|$	100
result set size k = $\|S\|$	5, 10, 20, 30
# of sample queries	298

where $N_T(d_j)$ is the number of times the walk has visited d_j up to time T and,

$$D_T(d_i) = \sum_{d_j \in V} p_0(d_i, d_j) N_T(d_j) \tag{15}$$

Since Grasshopper and DivRank utilize a similar approach and will ultimately present rather similar results we utilized Grasshopper distinctively from DivRank. In particularly, instead of creating the citation network of documents belonging to the initial result set, we form the adjacency matrix based on document similarity, as previously explained in LexRank Algorithm 5.

4. Experimental Setup

In this section, we describe the legal corpus we use, the set of query topics and the respective methodology for subjectively annotating with relevance judgments for each query, as well as the metrics employed for the evaluation assessment. Finally, we provide the results along with a short discussion.

4.1. Legal Corpus

Our corpus contains 3.890 Australian legal cases from the Federal Court of Australia[8]. The cases were originally downloaded from AustLII[9] and were used in [47] to experiment with automatic summarization and citation analysis. The legal corpus contains all cases from the Federal Court of Australia spanning from 2006 up to 2009. From the cases, we extracted all needed text and citation links for our diversification framework. Our index was built using standard stop word removal and porter stemming, with log based $tf - idf$ indexing technique, resulting in a total of 3.890 documents, 9.782.911 terms and 53.791 unique terms.

Table 1 summarizes testing parameters and their corresponding ranges. To obtain the candidate set N, for each query sample we keep the $top - n$ elements using cosine similarity and a log based $tf - idf$ indexing schema. Our experimental studies are performed in a two-fold strategy: i) qualitative analysis in terms of diversification and precision of each employed method with respect to the optimal result set and ii) scalability analysis of diversification methods when increasing the query parameters.

4.2. Evaluation Metrics

As the authors of [48] claim that "there is no evaluation metric that seems to be universally accepted as the best for measuring the performance of algorithms that aim to obtain diverse rankings", we have chosen to evaluate diversification methods using various metrics employed in TREC Diversity Tasks[10]. In particular, we report:

[8] http://www.fedcourt.gov.au
[9] http://www.austlii.edu.au
[10] http://trec.nist.gov/data/web10.html

Table 2. West Law Digest Topics as user queries

1:	Abandoned and Lost Property	3:	Abortion and Birth Control
24:	Aliens Immigration and Citizenship	31:	Antitrust and Trade Regulation
61:	Breach of Marriage Promise	84:	Commodity Futures Trading Regulation
88:	Compromise and Settlement	199:	Implied and Constructive Contracts
291:	Privileged Communications and Confidentiality	363:	Threats Stalking and Harassment

- **a-nDCG:** a-Normalized Discounted Cumulative Gain [49] metric quantifies the amount of unique aspects of the query q that are covered by the $top - k$ ranked documents. We use $a = 0.5$, as typical in TREC evaluation.
- **ERR-IA:** Expected Reciprocal Rank - Intent Aware [50] is based on inter-dependent ranking. The contribution of each document is based on the relevance of documents ranked above it. The discount function is therefore not just dependent on the rank but also on the relevance of previously ranked documents.
- **S-Recall:** Subtopic-Recall [51] is the number of unique aspects covered by the $top - k$ results, divided by the total number of aspect. It measures the aspect coverage for a given result list at depth k.

4.3. Relevance Judjements

Evaluation of diversification requires a data corpus, a set of query topics and a set of relevance judgments, preferably assessed by domain experts for each query. One of the difficulties in evaluating methods designed to introduce diversity in the legal document ranking process is the lack of standard testing data. While TREC added a diversity task to the Web track in 2009, this dataset was designed assuming a general web search, and so it not possible to adapt it to our setting. Having only the document corpus, we need to define (a) the query topics, (b) a method to derive the subtopics for each topic, and, (c) a method to subjectively annotate the corpus for each topic. In the absence of a standard dataset specifically tailored for this purpose, we looked for an subjective way to evaluate and assess the performances of various diversification methods on our corpus.[11]

To this end, we have employed an subjective way to annotate our corpus with relevance judgments for each query:

User Profiles/ Queries. We used the West Law Digest Topics[12] as candidate user queries. In other words, each topic was issued as candidate query to our retrieval system. Outlier queries, whether too specific/rare or too general, where removed using the interquartile range, below or above values $Q1$ and $Q3$, sequentially in terms of number of hits in the result set and score distribution for the hits, demanding in parallel a minimum cover of $min|N|$ results. In total, we kept 289 queries. Table 2 provides a sample of the topics we further consider as user queries.

Query assessments and ground-truth. For each topic/ query we kept the $top - n$ results. An LDA [52] topic model, using an open source implementation[13], was trained on the $top - n$ results for each query. Topic modeling gives us a way to infer the latent structure behind a collection of documents. Based on the resulting topic distribution, with an acceptance threshold of 20%, we infer whether a document is relevant for an topic/ aspect. Thus, using LDA we create our ground-truth data consisting aspect assessments for each query.

[11] We do acknowledge the fact that the process of automatic query generation is at best an imperfect approximation of what a real person would do.

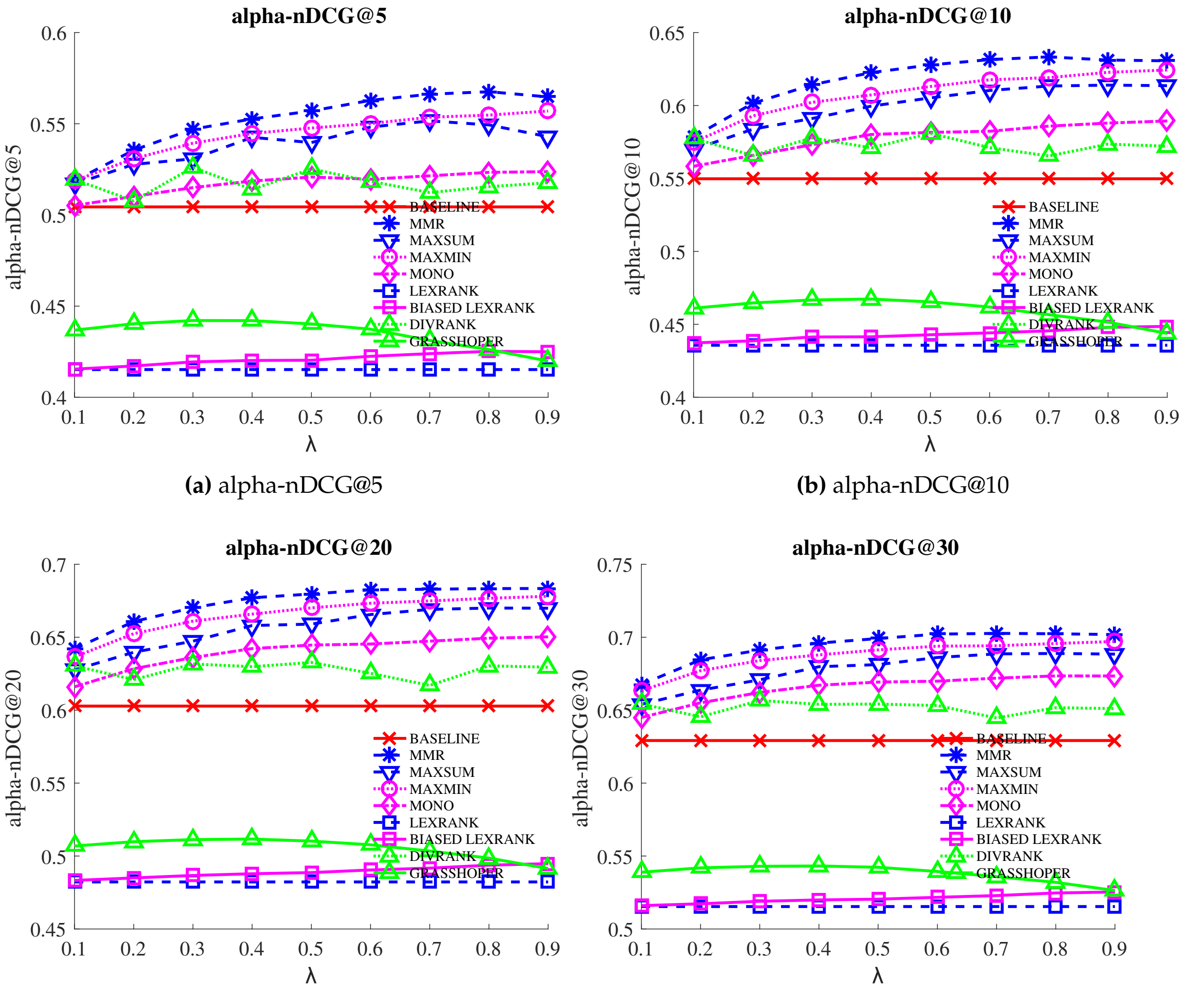
[12] The West American Digest System is a taxonomy of identifying points of law from reported cases and organizing them by topic and key number. It is used to organize the entire body of American law

[13] http://mallet.cs.umass.edu/

We have made available our complete dataset, ground-truth data, queries and relevance assessments in standard qrel format, as to enhance collaboration and contribution in respect to diversification issues in legal IR. [14].

4.4. Results

As a baseline to compare diversification methods, we consider the simple ranking produced by cosine similarity and log based $tf - idf$ indexing schema. For each query, our initial set N contains the $top - n$ query results. The interpolation parameter $\lambda \in [0..1]$ is tuned in 0.1 steps separately for each method. We present the evaluation results for the methods employed, using the aforementioned evaluation metrics, at cut-off values of 5, 10, 20 and 30, as typical in TREC evaluations. Results are presented with fixed parameter n = $|N|$. Note that each of the diversification variations, is applied in combination with each of the diversification algorithms and for each user query.



(a) alpha-nDCG@5

(b) alpha-nDCG@10

(c) alpha-nDCG@20

(d) alpha-nDCG@30

Figure 2. alpha-nDCG at various levels @5, @10, @20, @30 for baseline, MMR, MAXSUM, MAXMIN, MONO, LEXRANK, BIASEDLEXRANK, DIVRANK and GRASSHOPPER methods. [Best viewed in color]

[14] https://github.com/mkoniari/LegalDivEval

Figure 2 shows the a-Normalized Discounted Cumulative Gain (a-nDCG) of each method for different values of λ. Interestingly, web search result diversification methods (MMR, MaxSum, MaxMin and Mono) outperformed the baseline ranking, while text summarization methods (LexRank, Biased LexRank and GrassHopper, as it was utilized without a network citation graph) failed to improve the baseline ranking performing lower than the baseline ranking at all levels across all metrics. Graph-based methods (DivRank) results vary across the different values of λ. We attribute this finding to the extreme sparse network of citations since our dataset covers a short time period (3 years).

The trending behavior of MMR, MaxMin, and MaxSum is very similar especially at levels @10, and @20, while at level @5 MaxMin and MaxSum presented nearly identical a-nDCG values in many λ values (e.g., 0.1, 0.2, 0.4, 0.6, 0.7). Finally, MMR constantly achieves better results in respect to the rest methods, following by MaxMin and MaxSum. MONO despite the fact that performs better than the baseline in all λ values, still always presents the lower performance when compared to MMR, MaxMin, and MaxSum. It is clear that web search result diversification approaches (MMR, MaxSum, MaxMin and Mono) tend to perform better than the selected baseline ranking method. Moreover, as λ increases, preference to diversity as well as a-nDCG accuracy increases for all tested methods.

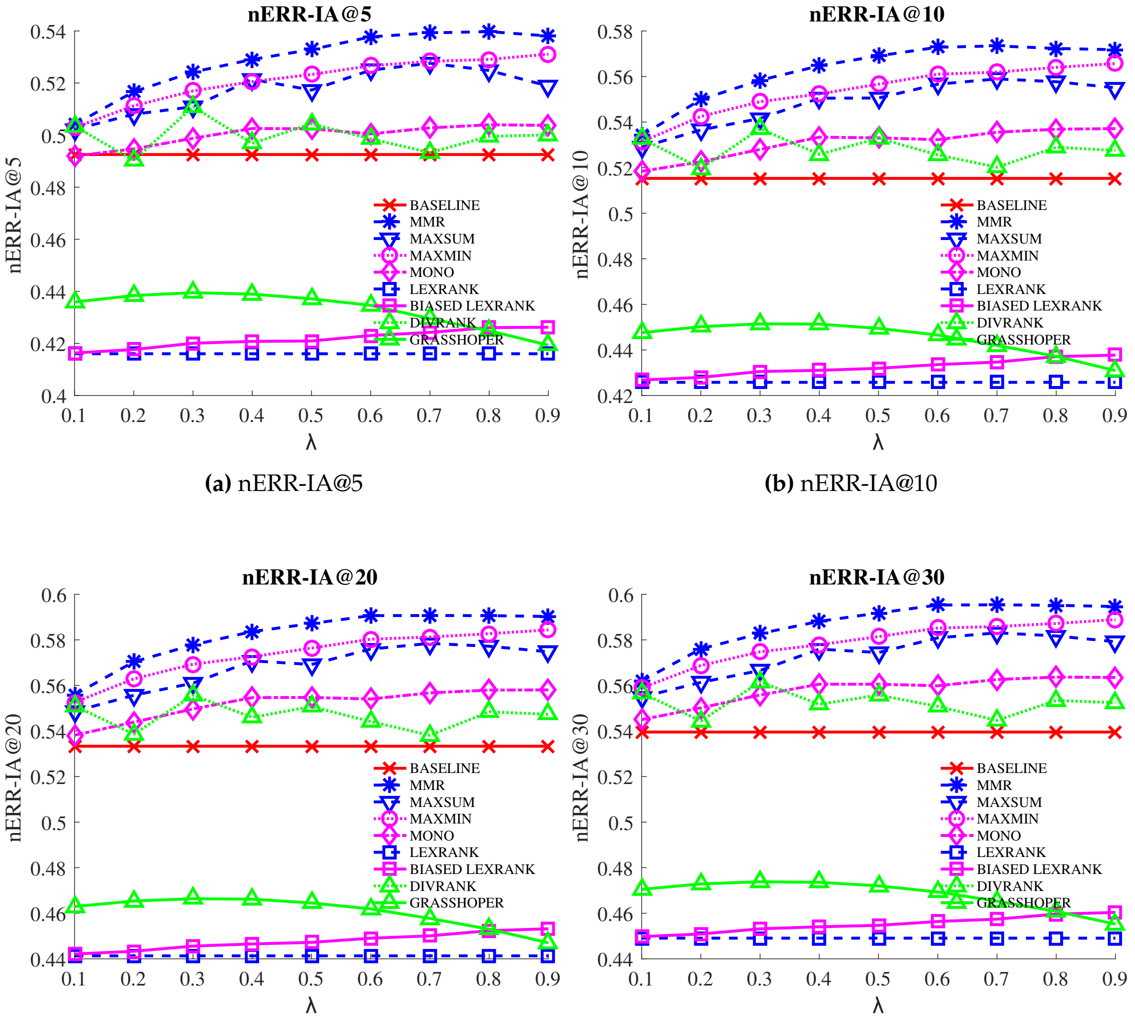
Figure 3 depicts the normalised Expected Reciprocal Rank - Intent Aware (nERR-IA) plots for each method in respect to different values of λ. It is clear that web search result diversification approaches (MMR, MaxSum, MaxMin and Mono) tend to perform better than the selected baseline ranking method. Moreover, as λ increases, preference to diversity as well as nERR-IA accuracy increases for all tested methods. Text summarization methods (LexRank, Biased LexRank) and GrassHopper, once again failed to improve the baseline ranking at all levels across all metrics, while as in a-nDCG plots DivRank results vary across the different values of λ. MMR constantly achieves better results in respect to the rest methods. We also observed that MaxMin tends to perform better than MaxSum. There were few cases where both methods presented nearly similar performance especially in lower recall levels (e.g., for nERR-IA@5 when λ equals to 0.1, 0.4, 0.6, 0.7). Once again, MONO presents the lower performance when compared to MMR, MaxMin, and MaxSum for nERR-IA metric for all λ values applied.

Figure 4 shows the Subtopic-Recall at various levels @5, @10, @20, @30 of each method for different values of λ. It is clear the web search result diversification methods (MMR, MaxSum, MaxMin and Mono) tend to perform better than the baseline ranking. As λ increases, preference to diversity increases for all methods except MMR. Subtopic-Recall accuracy of all methods, except MMR, increases when increasing λ. For lower levels (e.g., @5, @10) MMR clearly outperforms other methods, while for upper levels (e.g., @20, @30) MMR and MAXMIN scores are comparable. We also observe that MAXMIN tends to perform better than MAXSUM, which in turn constantly achieves better results than MONO. Finally, LexRank, Biased LexRank and GrassHopper approaches fail to improve the baseline ranking at all levels across all metrics. Overall, we noticed a similar trending behavior with the ones discussed for Figure 2 and Figure 3.

In summary, among all the results, we note that the trends in the graphs look very similar. Clearly enough, the utilized web search diversification methods (MMR, MAXSUM, MAXMIN, MONO) statistically significantly[15] outperform the baseline method, offering legislation stakeholders broader insights in respect to their information needs. Furthermore, trends across the evaluation metric graphs, highlight balance boundaries for legal IR systems between reinforcing relevant documents or sampling the information space around the legal query.

Table 3 summarizes average results of the diversification methods. Statistically significant values, using the paired two-sided t-test with $p_{value} < 0.05$ are denoted with $^\circ$ and with $p_{value} < 0.01$ with *.

[15] Statistical significance with the paired two-sided t-test ($p - value < 0.05$ and $p_{value} < 0.01$)



(a) nERR-IA@5

(b) nERR-IA@10

(c) nERR-IA@20

(d) nERR-IA@30

Figure 3. nERR-IA at various levels @5, @10, @20, @30 for baseline, MMR, MAXSUM, MAXMIN, MONO, LEXRANK, BIASEDLEXRANK, DIVRANK and GRASSHOPPER methods. [Best viewed in color]

Table 3. Retrieval Performance of the tested algorithms with interpolation parameter $\lambda \in [0..1]$ tuned in 0.1 steps for $N = 100$ and $k = 30$. Highest scores are shown in bold. Statistically significant values, using the paired two-sided t-test with $p_{value} < 0.05$ are denoted with $^\circ$ and with $p_{value} < 0.01$ with *

	a-nDCG				nERR-IA				ST Recall			
Method	@5	@10	@20	@30	@5	@10	@20	@30	@5	@10	@20	@30
$\lambda = 0.1$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	0.5187	**0.5785***	**0.642***	**0.6676***	**0.5041**	**0.5341**	**0.5559**$^\circ$	**0.562**$^\circ$	0.6145°	**0.7875***	**0.9135***	**0.9543***
MaxSum	0.5170	0.5699°	0.6276*	0.6541*	0.5022	0.5290	0.5486	0.5549	0.6083	0.7626°	0.8851*	0.9294*
MaxMin	0.5188	0.5749*	0.6365*	0.6633*	0.5029	0.5313	0.5526°	0.5589°	0.6173°	0.7820*	0.8990*	0.9481*
MonoObjective	0.5052	0.5584	0.6160	0.6450	0.4919	0.5184	0.5382	0.5450	0.5889	0.7543	0.8740°	0.9273*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4155*	0.4373*	0.4833*	0.5160*	0.4163*	0.4268*	0.4421*	0.4498*	0.4228*	0.5370*	0.6734*	0.7654*
DivRank	**0.5195**	0.5774*	0.6304*	0.6543*	0.5035	0.5328	0.5511	0.5567	**0.6208**$^\circ$	0.7820*	0.8976*	0.9384*
GrassHopper	0.4368*	0.4611*	0.5069*	0.5389*	0.4359*	0.4476*	0.4630*	0.4705*	0.4567*	0.5758*	0.7059*	0.7931*

continued ...

Table 3 – *Continued from previous page*

	a-nDCG				nERR-IA				ST Recall			
Method	@5	@10	@20	@30	@5	@10	@20	@30	@5	@10	@20	@30
$\lambda = 0.2$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	**0.5356***	**0.6015***	**0.6607***	**0.6845***	**0.5167**°	**0.5499***	**0.5704***	**0.576***	**0.6547***	**0.8388***	**0.9322***	**0.9696***
MaxSum	0.5277°	0.5838*	0.6397*	0.6637*	0.5080	0.5366°	0.5557°	0.5613°	0.6422*	0.7993*	0.9017*	0.9398*
MaxMin	0.5309*	0.5929*	0.6524*	0.6771*	0.5113	0.5425*	0.5629*	0.5687*	0.6533*	0.8187*	0.9246*	0.9640*
MonoObjective	0.5102	0.5658	0.6284*	0.6550*	0.4947	0.5226	0.5439	0.5502	0.6035	0.7654*	0.8941*	0.9398*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4172*	0.4387*	0.4850*	0.5173*	0.4176*	0.4280*	0.4433*	0.4509*	0.4277*	0.5391*	0.6761*	0.7668*
DivRank	0.5077	0.5657	0.6209°	0.6454°	0.4902	0.5196	0.5386	0.5444	0.6159°	0.7931*	0.9100*	0.9453*
GrassHopper	0.4403*	0.4647*	0.5099*	0.5419*	0.4384*	0.4502*	0.4654*	0.4729*	0.4657*	0.5799*	0.7100*	0.7965*
$\lambda = 0.3$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	**0.547***	**0.6142***	**0.6702***	**0.6912***	**0.5242***	**0.5584***	**0.5778***	**0.5828***	**0.6955***	**0.8581***	**0.9439***	**0.9682***
MaxSum	0.5308*	0.5911*	0.6473*	0.6708*	0.5109	0.5416*	0.5610*	0.5666*	0.6512*	0.8111*	0.9093*	0.9460*
MaxMin	0.5394*	0.6022*	0.6610*	0.6840*	0.5170°	0.5490*	0.5693*	0.5748*	0.6775*	0.8339*	0.9343*	0.9675*
MonoObjective	0.5150	0.5731°	0.6361*	0.6621*	0.4988	0.5280	0.5497	0.5558	0.6159°	0.7779*	0.9059*	0.9481*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4194*	0.4414*	0.4867*	0.5190*	0.4201*	0.4305*	0.4456*	0.4532*	0.4298*	0.5433*	0.6754*	0.7675*
DivRank	0.5261°	0.5779*	0.6316*	0.6566*	0.5109	0.5371°	0.5557°	0.5616°	0.6394*	0.7882*	0.8886*	0.9356*
GrassHopper	0.4421*	0.4667*	0.5113*	0.5430*	0.4395*	0.4514*	0.4664*	0.4738*	0.4713*	0.5848*	0.7121*	0.7965*
$\lambda = 0.4$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	**0.5527***	**0.6226***	**0.677***	**0.696***	**0.5291***	**0.5647***	**0.5836***	**0.5881***	**0.7093***	**0.8727***	**0.9481***	0.9696*
MaxSum	0.5425*	0.5996*	0.6580*	0.6798*	0.5214*	0.5505*	0.5708*	0.5759*	0.6740*	0.8187*	0.9183*	0.9522*
MaxMin	0.5447*	0.6073*	0.6659*	0.6879*	0.5206*	0.5524*	0.5726*	0.5778*	0.6962*	0.8422*	0.9405*	**0.9723***
MonoObjective	0.5186	0.5802*	0.6422*	0.6671*	0.5025	0.5334	0.5546°	0.5605°	0.6208°	0.7903*	0.9114*	0.9543*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4202*	0.4415*	0.4879*	0.5199*	0.4208*	0.4310*	0.4465*	0.4541*	0.4298*	0.5412*	0.6768*	0.7682*
DivRank	0.5140	0.5710°	0.6298*	0.6540*	0.4968	0.5258	0.5460	0.5517	0.6111	0.7785*	0.9045*	0.9439*
GrassHopper	0.4421*	0.4673*	0.5117*	0.5432*	0.4389*	0.4513*	0.4662*	0.4736*	0.4740*	0.5896*	0.7142*	0.7979*
$\lambda = 0.5$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	**0.557***	**0.6278***	**0.6796***	**0.6991***	**0.5329***	**0.5691***	**0.5872***	**0.5918***	**0.7218***	**0.8844***	**0.9495***	**0.9737***
MaxSum	0.5397*	0.6052*	0.6590*	0.6812*	0.5173°	0.5506*	0.5692*	0.5744*	0.6824*	0.8381*	0.9211*	0.9571*
MaxMin	0.5477*	0.6130*	0.6701*	0.6913*	0.5233*	0.5567*	0.5764*	0.5814*	0.7024*	0.8554*	0.9433*	**0.9737***
MonoObjective	0.5208	0.5816*	0.6446*	0.6693*	0.5024	0.5331	0.5547°	0.5605°	0.6318*	0.7965*	0.9183*	0.9571*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4203*	0.4429*	0.4887*	0.5204*	0.4209*	0.4319*	0.4472*	0.4547*	0.4291*	0.5446*	0.6782*	0.7675*
DivRank	0.5252°	0.5810*	0.6327*	0.6542*	0.5044	0.5329	0.5507	0.5558	0.6450*	0.8000*	0.8976*	0.9280*
GrassHopper	0.4402*	0.4654*	0.5103*	0.5423*	0.4371*	0.4494*	0.4645*	0.4720*	0.4713*	0.5869*	0.7128*	0.7986*
$\lambda = 0.6$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	**0.5628***	**0.6315***	**0.6825***	**0.7021***	**0.5377***	**0.5729***	**0.5906***	**0.5953***	**0.7363***	**0.8872***	**0.9516***	**0.9744***
MaxSum	0.5482*	0.6103*	0.6656*	0.6861*	0.5250*	0.5566*	0.5760*	0.5809*	0.7024*	0.8422*	0.9260*	0.9557*
MaxMin	0.5501*	0.6176*	0.6733*	0.6939*	0.5267*	0.5610*	0.5803*	0.5852*	0.7038*	0.8602*	0.9467*	0.9723*
MonoObjective	0.5196	0.5824*	0.6454*	0.6699*	0.5005	0.5323	0.5540°	0.5598°	0.6325*	0.8014*	0.9218*	0.9606*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4225*	0.4442*	0.4905*	0.5217*	0.4230*	0.4336*	0.4490*	0.4564*	0.4332*	0.5446*	0.6817*	0.7675*
DivRank	0.5185	0.5711°	0.6253*	0.6532*	0.4986	0.5256	0.5442	0.5508	0.6401*	0.7945*	0.9017*	0.9467*
GrassHopper	0.4374*	0.4619*	0.5077*	0.5394*	0.4346*	0.4465*	0.4619*	0.4693*	0.4657*	0.5806*	0.7107*	0.7958*
$\lambda = 0.7$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	**0.5662***	**0.6333***	**0.6829***	**0.7026***	**0.5393***	**0.5734***	**0.5907***	**0.5954***	**0.7467***	**0.8893***	**0.9516***	**0.9744***

continued . . .

Table 3 – *Continued from previous page*

	a-nDCG				nERR-IA				ST Recall			
Method	@5	@10	@20	@30	@5	@10	@20	@30	@5	@10	@20	@30
MaxSum	0.5516*	0.6134*	0.6690*	0.6886*	0.5276*	0.5590*	0.5784*	0.5831*	0.7073*	0.8450*	0.9280*	0.9543*
MaxMin	0.5536*	0.6191*	0.6749*	0.6941*	0.5283*	0.5619*	0.5812*	0.5858*	0.7093*	0.8623*	0.9481*	0.9702*
MonoObjective	0.5215	0.5859*	0.6473*	0.6720*	0.5028	0.5355°	0.5567°	0.5626°	0.6353*	0.8083*	0.9190*	0.9599*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4239*	0.4456*	0.4917*	0.5229*	0.4242*	0.4347*	0.4501*	0.4574*	0.4353*	0.5467*	0.6837*	0.7689*
DivRank	0.5123	0.5654	0.6170	0.6446	0.4933	0.5202	0.5381	0.5446	0.6353*	0.7896*	0.8865*	0.9391*
GrassHopper	0.4312*	0.4566*	0.5034*	0.5359*	0.4295*	0.4420*	0.4577*	0.4653*	0.4533*	0.5702*	0.7038*	0.7945*
$\lambda = 0.8$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	$\mathbf{0.5676}^*$	$\mathbf{0.6312}^*$	$\mathbf{0.6834}^*$	$\mathbf{0.7024}^*$	$\mathbf{0.5397}^*$	$\mathbf{0.5723}^*$	$\mathbf{0.5906}^*$	$\mathbf{0.5952}^*$	$\mathbf{0.7502}^*$	$\mathbf{0.881}^*$	$\mathbf{0.9516}^*$	$\mathbf{0.9737}^*$
MaxSum	0.5494*	0.6140*	0.6700*	0.6889*	0.5248*	0.5577*	0.5772*	0.5817*	0.7093*	0.8512*	0.9343*	0.9571*
MaxMin	0.5547*	0.6228*	0.6767*	0.6957*	0.5291*	0.5640*	0.5827*	0.5872*	0.7156*	0.8706*	0.9509*	0.9716*
MonoObjective	0.5234	0.5880*	0.6493*	0.6735*	0.5040	0.5369°	0.5580*	0.5636*	0.6443*	0.8118*	0.9232*	0.9626*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4252*	0.4479*	0.4937*	0.5247*	0.4261*	0.4371*	0.4524*	0.4596*	0.4346*	0.5488*	0.6858*	0.7709*
DivRank	0.5155	0.5735*	0.6304*	0.6517*	0.4995	0.5289	0.5484	0.5534	0.6090	0.7806*	0.8976*	0.9280*
GrassHopper	0.4262*	0.4515*	0.4986*	0.5320*	0.4249*	0.4372*	0.4530*	0.4608*	0.4457*	0.5647*	0.6969*	0.7931*
$\lambda = 0.9$												
baseline	0.5044	0.5498	0.6028	0.6292	0.4925	0.5153	0.5333	0.5395	0.5827	0.7260	0.8464	0.9010
MMR	$\mathbf{0.5647}^*$	$\mathbf{0.6306}^*$	$\mathbf{0.6834}^*$	$\mathbf{0.7018}^*$	$\mathbf{0.5381}^*$	$\mathbf{0.5718}^*$	$\mathbf{0.5902}^*$	$\mathbf{0.5946}^*$	$\mathbf{0.7439}^*$	$\mathbf{0.8817}^*$	$\mathbf{0.9529}^*$	$\mathbf{0.9737}^*$
MaxSum	0.5429*	0.6136*	0.6699*	0.6884*	0.5188*	0.5551*	0.5748*	0.5792*	0.6997*	0.8554*	0.9419*	0.9626*
MaxMin	0.5570*	0.6244*	0.6781*	0.6971*	0.5311*	0.5657*	0.5844*	0.5889*	0.7211*	0.8727*	0.9495*	0.9716*
MonoObjective	0.5238	0.5894*	0.6502*	0.6734*	0.5037	0.5371°	0.5580*	0.5635*	0.6471*	0.8166*	0.9260*	0.9619*
LexRank	0.4152*	0.4357*	0.4823*	0.5154*	0.4160*	0.4258*	0.4413*	0.4491*	0.4228*	0.5329*	0.6713*	0.7647*
BiasedLexRank	0.4250*	0.4486*	0.4948*	0.5254*	0.4262*	0.4377*	0.4532*	0.4604*	0.4325*	0.5502*	0.6872*	0.7702*
DivRank	0.5177	0.5721°	0.6295*	0.6511*	0.5001	0.5275	0.5473	0.5524	0.6187°	0.7785*	0.8969*	0.9280*
GrassHopper	0.4199*	0.4438*	0.4915*	0.5264*	0.4193*	0.4309*	0.4470*	0.4552*	0.4332*	0.5495*	0.6851*	0.7882*

The effectiveness of diversification methods is also depicted in Table 4 which illustrates the result sets for three example queries, using our case law dataset ($|S| = 30$ and $N = 100$) with $\lambda = 0$ (no diversification), $\lambda = 0.1$ (light diversification), $\lambda = 0.5$ (moderate diversification) and $\lambda = 0.9$ (high diversification). Only MMR results are shown since, in almost all variations, it outperforms other approaches. Due to space limitations, we show the case title for each entry hyper linked to the full text for that entry. When $\lambda = 0$ the result set contains the top-5 elements of S ranked with the sim scoring function. The result sets with no diversification contain several almost duplicate elements, defined by terms in the case title. As λ increases, less "duplicates" are found in the result set, and the elements in the result set "cover" many more subjects again as defined by terms in the case title. We note that the result set with high diversification contains elements that have almost all of the query terms, as well as other terms indicating that the case is related to different subjects among the other cases in the result set.

Table 4. Result sets (document titles for three example queries, using the dataset ($|S| = 30$ and $N = 100$) with $\lambda = 0$ (no diversification), $\lambda = 0.1$ (light diversification), $\lambda = 0.5$ (moderate diversification) and $\lambda = 0.9$ (high diversification)

	Baseline ($\lambda = 0$)	**MMR**: light diversity ($\lambda = .1$)	**MMR**: moderate diversity ($\lambda = .5$)	**MMR**: high diversity ($\lambda = .9$)
		query 24: Aliens Immigration and Citizenship		
1	Virgin Holdings SA v Commissioner of Taxation [2008] FCA 1503 (10 October 2008)	Virgin Holdings SA v Commissioner of Taxation [2008] FCA 1503 (10 October 2008)	Virgin Holdings SA v Commissioner of Taxation [2008] FCA 1503 (10 October 2008)	Virgin Holdings SA v Commissioner of Taxation [2008] FCA 1503 (10 October 2008)

Continued on next page

Table 4 – *Continued from previous page*

	Baseline $\lambda = 0$	**MMR**: light diversity ($\lambda = .1$)	**MMR**: moderate diversity ($\lambda = .5$)	**MMR**: high diversity ($\lambda = .9$)
2	Undershaft (No 1) Limited v Commissioner of Taxation [2009] FCA 41 (3 February 2009)	Fowler v Commissioner of Taxation [2008] FCA 528 (21 April 2008)	Fowler v Commissioner of Taxation [2008] FCA 528 (21 April 2008)	Soh v Commonwealth of Australia [2008] FCA 520 (18 April 2008)
3	Fowler v Commissioner of Taxation [2008] FCA 528 (21 April 2008)	Wight v Honourable Chris Pearce, MP, Parliamentary Secretary to the Treasurer [2007] FCA 26 (29 January 2007)	Coleman v Minister for Immigration & Citizenship [2007] FCA 1500 (27 September 2007)	SZJDI v Minister for Immigration & Citizenship (No. 2) [2008] FCA 813 (16 May 2008)
4	Wight v Honourable Chris Pearce, MP, Parliamentary Secretary to the Treasurer [2007] FCA 26 (29 January 2007)	Undershaft (No 1) Limited v Commissioner of Taxation [2009] FCA 41 (3 February 2009)	Charlie v Minister for Immigration and Citizenship [2008] FCA 1025 (10 July 2008)	Charlie v Minister for Immigration and Citizenship [2008] FCA 1025 (10 July 2008)
5	Coleman v Minister for Immigration & Citizenship [2007] FCA 1500 (27 September 2007)	Coleman v Minister for Immigration & Citizenship [2007] FCA 1500 (27 September 2007)	VSAB v Minister for Immigration and Multicultural and Indigenous Affairs [2006] FCA 239 (17 March 2006)	VSAB v Minister for Immigration and Multicultural and Indigenous Affairs [2006] FCA 239 (17 March 2006)
	query 84: Commodity Futures Trading Regulation			
1	BHP Billiton Iron Ore Pty Ltd v The National Competition Council [2006] FCA 1764 (18 December 2006)	BHP Billiton Iron Ore Pty Ltd v The National Competition Council [2006] FCA 1764 (18 December 2006)	BHP Billiton Iron Ore Pty Ltd v The National Competition Council [2006] FCA 1764 (18 December 2006)	BHP Billiton Iron Ore Pty Ltd v The National Competition Council [2006] FCA 1764 (18 December 2006)
2	Australian Securities & Investments Commission v Lee [2007] FCA 918 (15 June 2007)	Australian Securities & Investments Commission v Lee [2007] FCA 918 (15 June 2007)	Australian Securities & Investments Commission v Lee [2007] FCA 918 (15 June 2007)	Australian Competition and Consumer Commission v Dally M Publishing and Research Pty Limited [2007] FCA 1220 (10 August 2007)
3	Woodside Energy Ltd (ABN 63 005 482 986) v Commissioner of Taxation (No 2) [2007] FCA 1961 (10 December 2007)	Woodside Energy Ltd (ABN 63 005 482 986) v Commissioner of Taxation (No 2) [2007] FCA 1961 (10 December 2007)	Woodside Energy Ltd (ABN 63 005 482 986) v Commissioner of Taxation (No 2) [2007] FCA 1961 (10 December 2007)	Heritage Clothing Pty Ltd trading as Peter Jackson Australia v Mens Suit Warehouse Direct Pty Ltd trading as Walter Withers [2008] FCA 1775 (28 November 2008)
4	BHP Billiton Iron Ore Pty Ltd v National Competition Council (No 2) [2007] FCA 557 (19 April 2007)	Keynes v Rural Directions Pty Ltd (No 2) (includes Corrigendum dated 16 July 2009) [2009] FCA 567 (3 June 2009)	Keynes v Rural Directions Pty Ltd (No 2) (includes Corrigendum dated 16 July 2009) [2009] FCA 567 (3 June 2009)	Travelex Limited v Commissioner of Taxation (Corrigendum dated 4 February 2009) [2008] FCA 1961 (19 December 2008)
5	Keynes v Rural Directions Pty Ltd (No 2) (includes Corrigendum dated 16 July 2009) [2009] FCA 567 (3 June 2009)	Queanbeyan City Council v ACTEW Corporation Limited [2009] FCA 943 (24 August 2009)	Heritage Clothing Pty Ltd trading as Peter Jackson Australia v Mens Suit Warehouse Direct Pty Ltd trading as Walter Withers [2008] FCA 1775 (28 November 2008)	Ashwick (Qld) No 127 Pty Ltd (ACN 010 577 456) v Commissioner of Taxation [2009] FCA 1388 (26 November 2009)
	query 291:Privileged Communications and Confidentiality			
1	Siam Polyethylene Co Ltd v Minister of State for Home Affairs (No 3) [2009] FCA 839 (7 August 2009)	Siam Polyethylene Co Ltd v Minister of State for Home Affairs (No 3) [2009] FCA 839 (7 August 2009)	Siam Polyethylene Co Ltd v Minister of State for Home Affairs (No 3) [2009] FCA 839 (7 August 2009)	Siam Polyethylene Co Ltd v Minister of State for Home Affairs (No 3) [2009] FCA 839 (7 August 2009)
2	AWB Limited v Australian Securities and Investments Commission [2008] FCA 1877 (11 December 2008)	AWB Limited v Australian Securities and Investments Commission [2008] FCA 1877 (11 December 2008)	AWB Limited v Australian Securities and Investments Commission [2008] FCA 1877 (11 December 2008)	Krueger Transport Equipment Pty Ltd v Glen Cameron Storage [2008] FCA 803 (30 May 2008)
3	Brookfield Multiplex Limited v International Litigation Funding Partners Pte Ltd (No 2) [2009] FCA 449 (6 May 2009)	Brookfield Multiplex Limited v International Litigation Funding Partners Pte Ltd (No 2) [2009] FCA 449 (6 May 2009)	Autodata Limited v Boyce's Automotive Data Pty Limited [2007] FCA 1517 (4 October 2007)	Futuretronics.com.au Pty Limited v Graphix Labels Pty Ltd [2007] FCA 1621 (29 October 2007)
4	Cadbury Schweppes Pty Ltd (ACN 004 551 473) v Amcor Limited (ACN 000 017 372) [2008] FCA 88 (19 February 2008)	Barrett Property Group Pty Ltd v Carlisle Homes Pty Ltd (No 2) [2008] FCA 930 (17 June 2008)	Barrett Property Group Pty Ltd v Carlisle Homes Pty Ltd (No 2) [2008] FCA 930 (17 June 2008)	Australian Competition & Consumer Commission v Visy Industries [2006] FCA 136 (23 February 2006)

Continued on next page

Table 4 – *Continued from previous page*

	Baseline $\lambda = 0$	**MMR**: light diversity ($\lambda = .1$)	**MMR**: moderate diversity ($\lambda = .5$)	**MMR**: high diversity ($\lambda = .9$)
5	Barrett Property Group Pty Ltd v Carlisle Homes Pty Ltd (No 2) [2008] FCA 930 (17 June 2008)	Cadbury Schweppes Pty Ltd (ACN 004 551 473) v Amcor Limited (ACN 000 017 372) [2008] FCA 88 (19 February 2008)	Optus Networks Ltd v Telstra Corporation Ltd (No. 2) (includes Corrigendum dated 7 July 2009) [2009] FCA 422 (9 July 2009)	IO Group Inc v Prestige Club Australasia Pty Ltd (No 2) [2008] FCA 1237 (11 August 2008)

5. Conclusions

In this paper, we studied the problem of of diversifying results in legal documents. We adopted and compared the performance of several state of the art methods from the web search, network analysis and text summarization domains as to handle the problems challenges. We evaluated all the methods using a real data set from the Common Law domain that we subjectively annotated with relevance judgments for this purpose. Our findings reveal that diversification methods offer notable improvements and enrich search results around the legal query space. In parallel, we demonstrated that that web search diversification techniques outperform other approaches e.g., summarization-based, graph-based methods, in the context of legal diversification. Finally, we provide valuable insights for legislation stakeholders though diversification, as well as by offering balance boundaries between reinforcing relevant documents or information space sampling around legal queries.

A challenge we faced in this work was the lack of ground-truth. We hope on an increase of the size of truth-labeled data set in the future, which would enable us to draw further conclusions about the diversification techniques. To this end, our complete dataset is publicly available in open and editable format, along with ground-truth data, queries and relevance assessments.

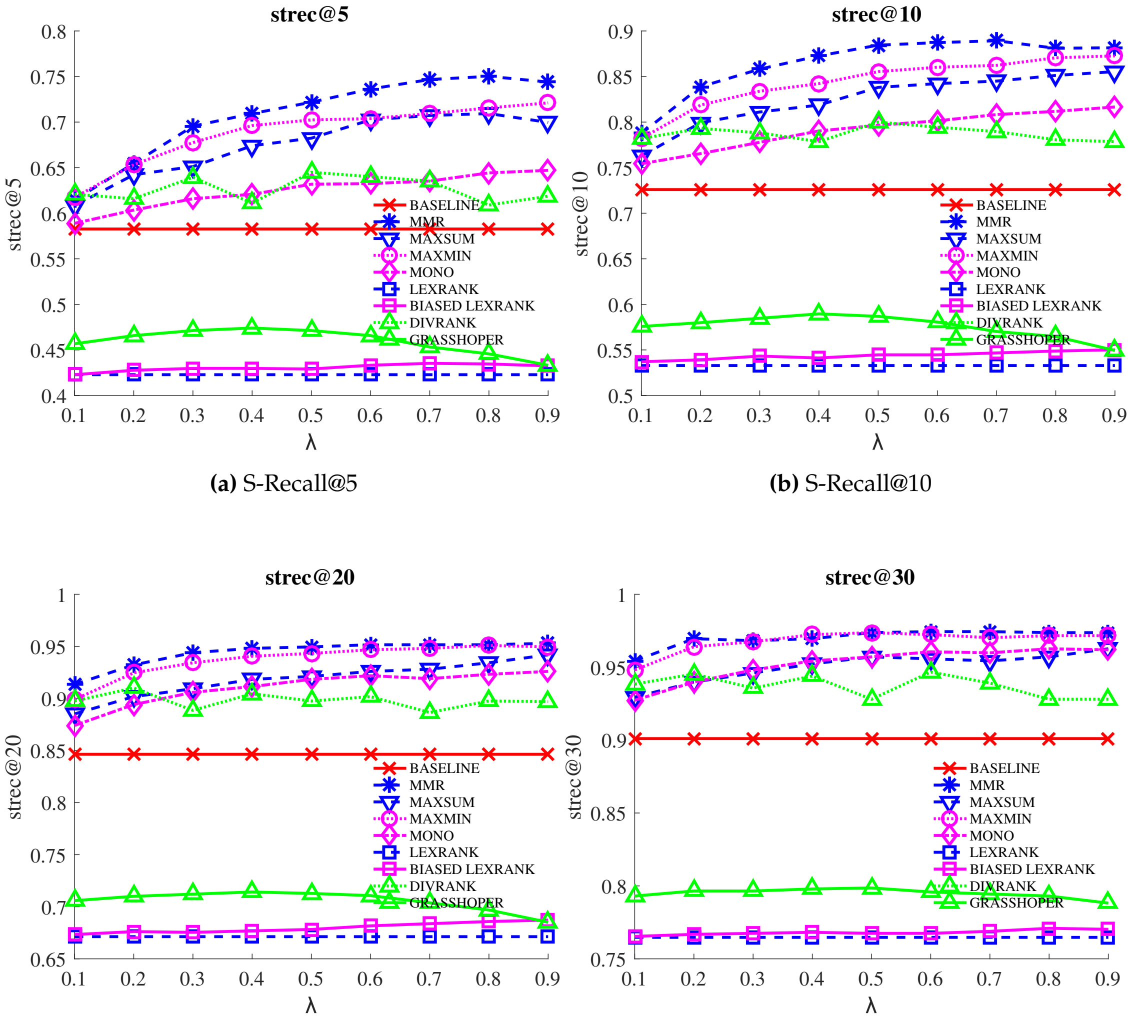
In future work, we plan to further study the interaction of relevance and redundancy, in historical legal queries. While access to legislation generally retrieves the current legislation on a topic, point-in-time legislation systems address a different problem, namely that lawyers, judges and anyone else considering the legal implications of past events need to know what the legislation stated at some point in the past when a transaction occurred, or events occurred which have led to a dispute and perhaps to litigation [53].

Author Contributions: Marios Koniaris conceived the idea, designed and performed the experiments, analyzed the results, drafted the initial manuscript and revised the manuscript; Ioannis Anagnostopoulos analyzed the results, helped to draft the initial manuscript and revised the final version; Yannis Vassiliou provided feedback and revised the final manuscript

Conflicts of Interest: The authors declare no conflict of interest

Bibliography

1. Alces, K.A. Legal diversification. *Columbia Law Review* **2013**, pp. 1977–2038.
2. Erkan, G.; Radev, D.R. LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research* **2004**, *22*, 457–479.
3. Otterbacher, J.; Erkan, G.; Radev, D.R. Biased LexRank: Passage retrieval using random walks with question-based priors. *Information Processing & Management* **2009**, *45*, 42–54.
4. Mei, Q.; Guo, J.; Radev, D. Divrank: the interplay of prestige and diversity in information networks. Proceedings of KDD'10. Association for Computing Machinery (ACM), 2010, pp. 1009–1018.
5. Zhu, X.; Goldberg, A.B.; Van Gael, J.; Andrzejewski, D. Improving Diversity in Ranking using Absorbing Random Walks. HLT-NAACL, 2007, pp. 97–104.
6. Carbonell, J.; Goldstein, J. The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of SIGIR '98. Association for Computing Machinery (ACM), 1998, pp. 335–336.
7. Gollapudi, S.; Sharma, A. An Axiomatic Approach for Result Diversification. Proceedings of WWW '09. Association for Computing Machinery (ACM), 2009, pp. 381–390.



(a) S-Recall@5

(b) S-Recall@10

(c) S-Recall@20

(d) S-Recall@30

Figure 4. SubTopic Recall at various levels @5, @10, @20, @30 for baseline, MMR, MAXSUM, MAXMIN, MONO, LEXRANK, BIASEDLEXRANK, DIVRANK and GRASSHOPPER methods. [Best viewed in color]

8. Singh, J.; Nejdl, W.; Anand, A. History by Diversity: Helping Historians Search News Archives. Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval; ACM: New York, NY, USA, 2016; CHIIR '16, pp. 183–192.

9. Giannopoulos, G.; Koniaris, M.; Weber, I.; Jaimes, A.; Sellis, T. Algorithms and criteria for diversification of news article comments. *Journal of Intelligent Information Systems* **2015**, *44*, 1–47.

10. Cheng, S.; Arvanitis, A.; Chrobak, M.; Hristidis, V. Multi-Query Diversification in Microblogging Posts. EDBT, 2014, pp. 133–144.

11. Koniaris, M.; Giannopoulos, G.; Sellis, T.; Vasileiou, Y. Diversifying microblog posts. International Conference on Web Information Systems Engineering. Springer, Springer Science Business Media, 2014, pp. 189–198.

12. Song, K.; Tian, Y.; Gao, W.; Huang, T. Diversifying the image retrieval results. Proceedings of the 14th ACM international conference on Multimedia. ACM, Association for Computing Machinery (ACM), 2006, pp. 707–710.

13. Ziegler, C.N.; McNee, S.M.; Konstan, J.A.; Lausen, G. Improving recommendation lists through topic diversification. Proceedings of the 14th international conference on World Wide Web. ACM, Association for Computing Machinery (ACM), 2005, pp. 22–32.

14. Raman, K.; Shivaswamy, P.; Joachims, T. Online learning to diversify from implicit feedback. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, Association for Computing Machinery (ACM), 2012, pp. 705–713.
15. Makris, C.; Plegas, Y.; Stamatiou, Y.C.; Stavropoulos, E.C.; Tsakalidis, A.K. Reducing Redundant Information in Search Results Employing Approximation Algorithms. International Conference on Database and Expert Systems Applications. Springer, Springer Science Business Media, 2014, pp. 240–247.
16. Zhang, B.; Li, H.; Liu, Y.; Ji, L.; Xi, W.; Fan, W.; Chen, Z.; Ma, W.Y. Improving web search results using affinity graph. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, Association for Computing Machinery (ACM), 2005, pp. 504–511.
17. Chen, H.; Karger, D.R. Less is more: probabilistic models for retrieving fewer relevant documents. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, Association for Computing Machinery (ACM), 2006, pp. 429–436.
18. Cronen-Townsend, S.; Croft, W.B. Quantifying query ambiguity. Proceedings of Human Language Technology Research '02. Association for Computational Linguistics (ACL), 2002.
19. Santos, R.L.T.; Macdonald, C.; Ounis, I. Search Result Diversification. *Foundations and Trends® in Information Retrieval* **2015**, *9*, 1–90.
20. Drosou, M.; Pitoura, E. Search result diversification. *ACM SIGMOD Record* **2010**, *39*, 41.
21. Santos, R.L.; Macdonald, C.; Ounis, I. Exploiting query reformulations for web search result diversification. Proceedings of WWW '10. Association for Computing Machinery (ACM), 2010, pp. 881–890.
22. Agrawal, R.; Gollapudi, S.; Halverson, A.; Ieong, S. Diversifying search results. Proceedings of WSDM '09. Association for Computing Machinery (ACM), 2009, pp. 5–14.
23. Hu, S.; Dou, Z.; Wang, X.; Sakai, T.; Wen, J.R. Search Result Diversification Based on Hierarchical Intents. Proceedings of CIKM '15. Association for Computing Machinery (ACM), 2015, pp. 63–72.
24. Langville, A.N.; Meyer, C.D. A survey of eigenvector methods for web information retrieval. *SIAM review* **2005**, *47*, 135–161.
25. Page, L.; Brin, S.; Motwani, R.; Winograd, T. The PageRank citation ranking: bringing order to the web. **1999**.
26. Moens, M. Innovative techniques for legal text retrieval. *Artificial Intelligence and Law* **2001**, pp. 29–57.
27. Biagioli, C.; Francesconi, E.; Passerini, A.; Montemagni, S.; Soria, C. Automatic semantics extraction in law documents. Proceedings of ICAIL '05. Association for Computing Machinery (ACM), 2005.
28. Mencia, E.L.; Fürnkranz, J. Efficient pairwise multilabel classification for large-scale problems in the legal domain. In *Machine Learning and Knowledge Discovery in Databases*; Springer, 2008; pp. 50–65.
29. Grabmair, M.; Ashley, K.D.; Chen, R.; Sureshkumar, P.; Wang, C.; Nyberg, E.; Walker, V.R. Introducing LUIMA. Proceedings of ICAIL'15. Association for Computing Machinery (ACM), 2015.
30. Saravanan, M.; Ravindran, B.; Raman, S. Improving legal information retrieval using an ontological framework. *Artificial Intelligence and Law* **2009**, *17*, 101–124.
31. Schweighofer, E.; Liebwald, D. Advanced lexical ontologies and hybrid knowledge based systems: First steps to a dynamic legal electronic commentary. *Artificial Intelligence and Law* **2007**, *15*, 103–115.
32. Gangemi, A.; Sagri, M.T.; Tiscornia, D. Metadata for content description in legal information. Procs. of LegOnt Workshop on Legal Ontologies, 2003.
33. Klein, M.C.; Van Steenbergen, W.; Uijttenbroek, E.M.; Lodder, A.R.; van Harmelen, F. Thesaurus-based Retrieval of Case Law. Proceedings of JURIX '06, 2006, Vol. 152, p. 61.
34. Hoekstra, R.; Breuker, J.; di Bello, M.; Boer, A. The LKIF Core ontology of basic legal concepts. Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007), 2007.
35. Farzindar, A.; Lapalme, G. Legal text summarization by exploration of the thematic structures and argumentative roles. Text Summarization Branches Out Workshop held in conjunction with ACL, 2004, pp. 27–34.
36. Farzindar, A.; Lapalme, G. Letsum, an automatic legal text summarizing system. *Legal knowledge and information systems, JURIX* **2004**, pp. 11–18.
37. Moens, M.F. Summarizing court decisions. *Information Processing & Management* **2007**, *43*, 1748–1764.

38. Aktolga, E.; Ros, I.; Assogba, Y. Detecting outlier sections in us congressional legislation. Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval - SIGIR '11. Association for Computing Machinery (ACM), 2011, pp. 235–244.
39. Marx, S.M. Citation networks in the law. *Jurimetrics Journal* **1970**, pp. 121–137.
40. van Opijnen, M. Citation Analysis and Beyond: in Search of Indicators Measuring Case Law Importance. Proceedings of JURIX '12, 2012, pp. 95–104.
41. Fowler, J.H.; Jeon, S. The authority of Supreme Court precedent. *Social Networks* **2008**, *30*, 16–30.
42. Fowler, J.H.; Johnson, T.R.; Spriggs, J.F.; Jeon, S.; Wahlbeck, P.J. Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court. *Political Analysis* **2006**, *15*, 324–346.
43. Galgani, F.; Compton, P.; Hoffmann, A. Citation based summarisation of legal texts. PRICAI 2012: Trends in Artificial Intelligence. Springer Science Business Media, 2012, pp. 40–52.
44. Koniaris, M.; Anagnostopoulos, I.; Vassiliou, Y. Network Analysis in the Legal Domain: A complex model for European Union legal sources. Physics and Society, Cornell University Library, Arxiv, http://arxiv.org/abs/1501.05237, 2015.
45. Lettieri, N.; Altamura, A.; Faggiano, A.; Malandrino, D. A computational approach for the experimental study of EU case law: analysis and implementation. *Social Network Analysis and Mining* **2016**, *6*, 56.
46. Wong, S.M.; Raghavan, V.V. Vector space model of information retrieval: a reevaluation. Proceedings of the 7th annual international ACM SIGIR conference on Research and development in information retrieval. British Computer Society, 1984, pp. 167–185.
47. Galgani, F.; Compton, P.; Hoffmann, A. Combining different summarization techniques for legal text. Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data, 2012, pp. 115–123.
48. Radlinski, F.; Bennett, P.N.; Carterette, B.; Joachims, T. Redundancy, diversity and interdependent document relevance. ACM SIGIR Forum. ACM, Association for Computing Machinery (ACM), 2009, Vol. 43, pp. 46–52.
49. Clarke, C.L.A.; Kolla, M.; Cormack, G.V.; Vechtomova, O.; Ashkan, A.; Büttcher, S.; MacKinnon, I. Novelty and diversity in information retrieval evaluation. Proceedings of SIGIR'08. Association for Computing Machinery (ACM), 2008.
50. Chapelle, O.; Metlzer, D.; Zhang, Y.; Grinspan, P. Expected reciprocal rank for graded relevance. Proceedings of the 18th ACM conference on Information and knowledge management - CIKM '09. Association for Computing Machinery (ACM), 2009, pp. 621–630.
51. Zhai, C.X.; Cohen, W.W.; Lafferty, J. Beyond independent relevance. Proceedings of SIGIR'03. Association for Computing Machinery (ACM), 2003.
52. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *Journal of machine Learning research* **2003**, *3*, 993–1022.
53. Wittfoth, A.; Chung, P.; Greenleaf, G.; Mowbray, A. AustLII's Point-in-Time legislation system: A generic PiT system for presenting legislation'. *Launch of the Point-in-Time legislation system* **2005**, *7*.