

Diversifying the Legal Order

Marios Koniaris, Ioannis Anagnostopoulos, Yannis Vassiliou

► **To cite this version:**

Marios Koniaris, Ioannis Anagnostopoulos, Yannis Vassiliou. Diversifying the Legal Order. 12th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2016, Thessaloniki, Greece. pp.499-509, 10.1007/978-3-319-44944-9_44 . hal-01557597

HAL Id: hal-01557597

<https://hal.inria.fr/hal-01557597>

Submitted on 6 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Diversifying the Legal Order

Marios Koniaris¹, Ioannis Anagnostopoulos², and Yannis Vassiliou¹

¹ KDBS Lab, School of ECE, National Technical University of Athens, Greece

² Department of Computer Science & Biomedical Inf., University of Thessaly, Greece

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 world³, 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 domain. 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 reveal that users receive broader insights across the results they get from a legal information retrieval system.

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-Lex⁴, the European Union’s database of regulations, the on-line version of the United States Code⁵, United Kingdom⁶, and the Australian⁷, just to mention a few, serve as 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 his/ 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

³ <http://www.worldlii.org/worldlii/declaration/>

⁴ <http://eur-lex.europa.eu/>

⁵ <http://uscode.house.gov/>

⁶ <http://www.legislation.gov.uk/>

⁷ <https://www.comlaw.gov.au/>

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 [3].

The objective of this paper is to define and evaluate the potential of results diversification in the field of legal information retrieval. To this end, we adopt various methods from the literature that are introduced for search result diversification [MMR [5], Max-Sum [12], Max-Min [12] and MonoObjective [12]]. 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 information retrieval and evaluated using diversity-aware evaluation metrics.

Our findings reveal that, 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. Furthermore our qualitative 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 and in the field of legal text retrieval. 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 firstly present related work on query result diversification and then we focus on same issues in legal text retrieval techniques.

In order to satisfy a wide range of users, query results diversification has attracted a lot of attention in the field of text mining. The published literature on search result diversification is reviewed in [8]. The maximal marginal relevance criterion (MMR), presented in [5], 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 [12] 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 that is 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. [18] 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. [1] propose 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. [14] 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.

In respect to legal text retrieval that traditionally relies on external knowledge sources, such as thesauri and classification schemes, various techniques are presented in [17]. Several supervised learning methods that have been proposed to classify sources of law according to legal concepts can be found in [4],[15],[13]. Legal document summarization techniques that scope to make the content of the legal documents, notably cases, more easily accessible are described in [9], [10], [16].

Finally, a similar approach with our work is described in [2], 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.

3 Legal Document ranking using diversification

Here, we firstly provide an overview of general diversification processes focusing in the problem we address. Then, we define the ranking features and describe the diversification algorithms employed in this work.

3.1 Diversification Overview

Initially, the user submits his/ her query as a way to express an information need and receives relevant documents. Diversification aims at finding a subset of those documents that maximize an objective function that quantifies the diversity of documents in S . 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 .*

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

Typicaly, 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, each document u can be represented as a term vector $U = (is_{w_1u}, is_{w_2u}, \dots, is_{w_mu})^T$, where w_1, w_2, \dots, 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.

- **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:

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

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

$$d(u, v) = 1 - sim(u, v) \quad (3)$$

- **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

$$r(q, u) = \cos(q, u) \quad (4)$$

3.2 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 [5], 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

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

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. The set S is usually 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 [12] aims at maximizing the sum of the relevance and diversity in the final result set. This is achieved by a greedy approximation algorithm that selects a pair of documents that maximizes Eq.6 in each iteration.

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

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 .

- **MaxMin:** The Max-Min diversification objective function [12] aims at maximizing the minimum relevance and dissimilarity of the selected set. This is achieved by a greedy approximation algorithm that select a document that maximizes Eq. 7 in each iteration.

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

where $\min_{v \in S} d(u, v)$ is the minimum distance of u to the already selected documents in S .

- **MonoObjective:** MonoObjective[12] combines the relevance and the similarity values into a single value for each document. It is defined as:

$$f_{MONO}(u, q) = r(u, q) + \frac{\lambda}{|N| - 1} \sum_{v \in N} d(u, v) \quad (8)$$

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 Australia⁸. The cases were originally downloaded from AustLII⁹ and were used in [11] 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 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.

Table 1: Parameters Tested in the Experiments

Parameter	Range
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
# of sample queries	298

4.2 Evaluation Metrics

We evaluate diversification methods using metrics employed in TREC Diversity Tasks¹⁰. In particular we report:

- **a-nDCG**: a -Normalized Discounted Cumulative Gain [7] 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.

⁸ <http://www.fedcourt.gov.au>

⁹ <http://www.austlii.edu.au>

¹⁰ <http://trec.nist.gov/data/web10.html>

- **ERR-IA:** Expected Reciprocal Rank - Intent Aware [6] 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 [19] quantifies the amount of unique aspects of the query q that are covered by the $top - k$ ranked documents

4.3 Relevance Judgements

As mentioned above, the evaluation of diversification requires a data corpus, a set of query topics and a set of relevance judgments, preferably made by human assessors for each query. In the absence of a standard dataset and since it was not feasible to involve legal experts in this study, we have employed an subjective way to annotate our corpus with relevance judgments for each query. To this end, we employed the following method:

User Profiles/ Queries We used the West Law Digest Topics¹¹ 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, were 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.

Table 2: West Law Digest Topics as user queries

1: Abandoned and Lost Property	3: Abortion and Birth Control
24: Aliens Immigration and Citizenship	88: Compromise and Settlement
291: Privileged Communications and Confidentiality	363: Threats Stalking and Harassment

Query assessments and ground-truth. For each topic/ query we kept the $top - n$ results. An LDA topic model, using an open source implementation¹², was trained on the $top - n$ results for each query. Based on the resulting topic distribution and with an acceptance threshold of 20%, we can infer whether a document is relevant for an aspect. 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 IR¹³.

¹¹ 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

¹² <http://mallet.cs.umass.edu/>

¹³ <http://www.dbnet.ntua.gr/~mkoniari/LegalDiv>

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. The interpolation parameter $\lambda \in [0..1]$ is tuned in 0.1 steps separately for each method. 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.

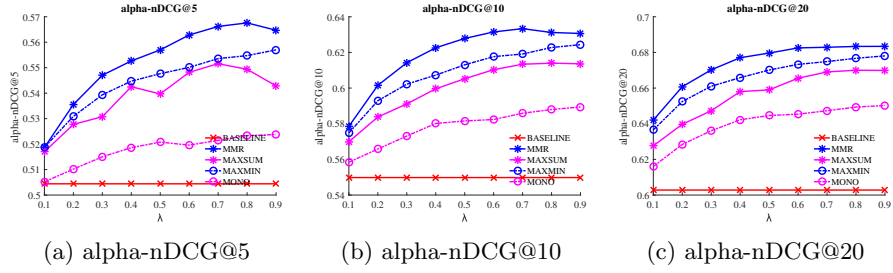


Fig. 1: alpha-nDCG at various levels @5, @10, @20 for baseline, MMR, MAXSUM, MAXMIN and MONO methods [Best viewed in color]

Figure 1 shows the a-nDCG of each method for different values of λ . Interestingly, all methods (MMR, MaxSum, MaxMin and Mono) outperformed the baseline ranking, while as λ increases, preference to diversity also increases for all methods. 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, while MaxMin and MaxSum follow. 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.

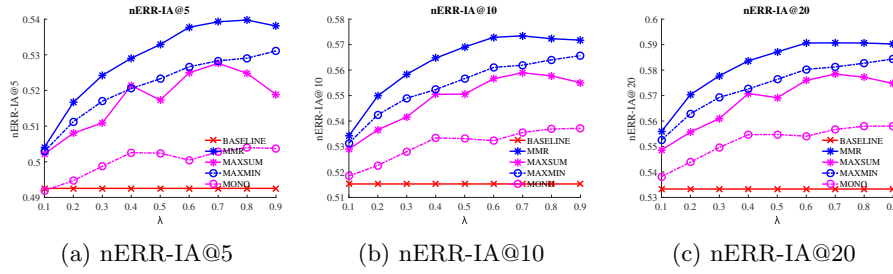


Fig. 2: nERR-IA at various levels @5, @10, @20 for baseline, MMR, MAXSUM, MAXMIN and MONO methods. [Best viewed in color]

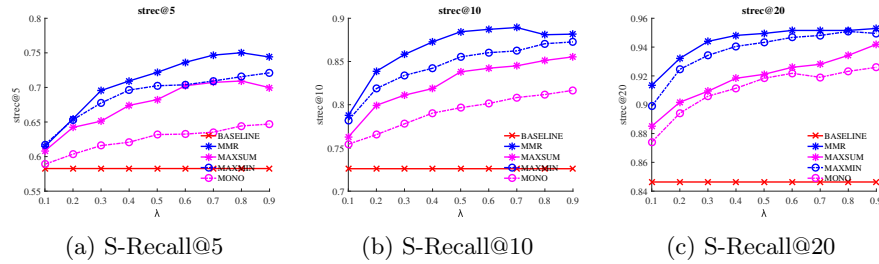


Fig. 3: SubTopic Recall at various levels @5, @10, @20 for baseline, MMR, MAXSUM, MAXMIN and MONO methods. [Best viewed in color]

Figure 2 depicts the ERR-IA plots for each method in respect to different values of λ , while similarly Figure 3 shows the Subtopic-Recall plots. It is clear that all of the 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 Subtopic-Recall accuracy increases for all tested methods. We noticed a Similar trending behavior with the one discussed for Figure 1. 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, and for S-Recall@5 when λ equals to 0.1, 0.2, 0.6, 0.7, 0.8). Once again, MONO presents the lower performance when compared to MMR, MaxMin, and MaxSum for both nERR-IA and S-Recall metric for all λ values applied.

In summary, among all the results, we note that the trends in the graphs look very similar. Clearly enough, the utilized diversification methods statistically significantly¹⁴ 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.

5 Conclusions

In this paper, we studied the novel problem of diversifying legal search results. We adopted and compared the performance of several state of the art methods from the web search domain as to deal with the challenges in this paradigm. We performed an exhaustive evaluation of all the methods, by using a real data set from the Common Law domain that we subjectively annotated with relevance judgments. Our findings i) reveal that diversification methods offer notable improvements and enrich search results around the legal query space and ii) offer balance boundaries between reinforcing relevant documents or sampling the information space around the legal query.

¹⁴ Statistical significance with the paired two-sided t-test (p - value < 0.05)

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. We also plan to incorporate additional features in our legal search result diversification framework, specifically tailored across the legislation domain. Finally, we aim at investigating the performance of heuristics provided for other domains, e.g. for text summarization and graph diversification.

References

1. Agrawal, R., Gollapudi, S., Halverson, A., Jeong, S.: Diversifying search results. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09. pp. 5–14 (2009)
2. Aktolga, E., Ros, I., Assogba, Y.: Detecting outlier sections in us congressional legislation. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval - SIGIR '11. pp. 235–244 (2011)
3. Alces, K.A.: Legal Diversification. Columbia Law Review pp. 1977–2038 (2013)
4. Biagioli, C., Francesconi, E., Passerini, A., Montemagni, S., Soria, C.: Automatic semantics extraction in law documents. In: Proceedings of the 10th international conference on Artificial intelligence and law - ICAIL '05 (2005)
5. Carbonell, J., Goldstein, J.: The use of mmr, diversity-based reranking for re-ordering documents and producing summaries. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '98. pp. 335–336 (1998)
6. Chapelle, O., Metzler, D., Zhang, Y., Grinspan, P.: Expected reciprocal rank for graded relevance. In: Proceedings of the 18th ACM conference on Information and knowledge management - CIKM '09. pp. 621–630 (2009)
7. Clarke, C.L.A., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '08 (2008)
8. Drosou, M., Pitoura, E.: Search result diversification. ACM SIGMOD Record 39(1), 41 (2010)
9. Farzindar, A., Lapalme, G.: Legal text summarization by exploration of the thematic structures and argumentative roles. In: Text Summarization Branches Out Workshop held in conjunction with ACL. pp. 27–34 (2004)
10. Farzindar, A., Lapalme, G.: Letsum, an automatic legal text summarizing system. Legal knowledge and information systems, JURIX '04 pp. 11–18 (2004)
11. Galgani, F., Compton, P., Hoffmann, A.: Combining different summarization techniques for legal text. In: Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data. pp. 115–123 (2012)
12. Gollapudi, S., Sharma, A.: An axiomatic approach for result diversification. In: Proceedings of the 18th international conference on World wide web - WWW '09. pp. 381–390 (2009)
13. Grabmair, M., Ashley, K.D., Chen, R., Sureshkumar, P., Wang, C., Nyberg, E., Walker, V.R.: Introducing luima. In: Proceedings of the 15th International Conference on Artificial Intelligence and Law - ICAIL '15 (2015)

14. Hu, S., Dou, Z., Wang, X., Sakai, T., Wen, J.R.: Search result diversification based on hierarchical intents. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management - CIKM '15. pp. 63–72 (2015)
15. 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, pp. 50–65. Springer (2008)
16. Moens, M.F.: Summarizing court decisions. *Information Processing & Management* 43(6), 1748–1764 (2007)
17. Moens, M.: Innovative techniques for legal text retrieval. *Artificial Intelligence and Law* pp. 29–57 (2001)
18. Santos, R.L., Macdonald, C., Ounis, I.: Exploiting query reformulations for web search result diversification. In: Proceedings of the 19th international conference on World wide web - WWW '10. pp. 881–890 (2010)
19. Zhai, C.X., Cohen, W.W., Lafferty, J.: Beyond independent relevance. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '03 (2003)