Text summarization from legal documents: a survey

3 authors, including:

Ambedkar Kanapala
Indian Institute of Technology (ISM) Dhanbad
6 PUBLICATIONS   **16** CITATIONS

SEE PROFILE

Sukomal Pal
Indian Institute of Technology (Banaras Hindu University) Varanasi
19 PUBLICATIONS   **60** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Sub-document level search View project

Project   TExt Processing in IL View project

Noname manuscript No.
(will be inserted by the editor)

Text Summarization from Legal Documents: A Survey

Ambedkar Kanapala · Sukomal Pal ·
Rajendra Pamula

the date of receipt and acceptance should be inserted later

Abstract Enormous amount of online information, available in legal domain, has made legal text processing an important area of research. In this paper, we attempt to survey different text summarization techniques that have taken place in the recent past. We put special emphasis on the issue of legal text summarization, as it is one of the most important areas in legal domain. We start with general introduction to text summarization, briefly touch the recent advances in single and multi-document summarization, and then delve into extraction based legal text summarization. We discuss different datasets and metrics used in summarization and compare performances of different approaches, first in general and then focused to legal text. we also mentioned highlights of different summarization techniques. We briefly cover a few software tools used in legal text summarization. We finally conclude with some future research directions.

Keywords Summarization from legal documents · Single-document summarization · Multi-document summarization · Legal domain

Ambedkar Kanapala
Department of Computer Science and Engineering,
Indian Institute of Technology (ISM), Dhanbad, India
E-mail: ambedkar.kanapala@live.com

Sukomal Pal
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU), Varanasi, India
E-mail: sukomalpal@gmail.com

Rajendra Pamula
Department of Computer Science and Engineering,
Indian Institute of Technology (ISM), Dhanbad, India
E-mail: rajendrapamula@gmail.com

1 Introduction

Ever-increasing volume of legal information in the public domain has necessitated matching effort in the field of automatic processing and access of relevant information from juridical science as per user's need. Obtaining relevant and useful legal information from a huge repository is important for its different stakeholders like scholars, professionals and ordinary citizen. When the documents are long, having a quick summary is often useful to the users. Text summarization reduces the content of a document without compromising on its essence to cut down users' time and cognitive effort. This is particularly important in the field of law. Legal document summarization is an emerging subtopic of text summarization. At present, lot of effort goes into preparing manual drafting of case summaries. This process is slow, labor-intensive and expensive. Lawyers and judges forward certain cases to legal editors for making summaries of them. The courts rely on a group of specialized staffs whose task is to summarize cases. In order to find a solution to a legal problem, lawyers must search through previous judgments to support their cases. On the other hand, novice users often want to get a feel whether there is any past evidences of similar court cases (summaries). Thus, automatic text summarization can become an important tool to assist different stakeholders. On one hand, it can help the human summarizers to find important points that qualify to be included in the case summaries. Legal practitioners often need to find out a set of feasible arguments to answer questions related to the case and support their claims. They have to rely on the human-generated summaries for this. It causes delay and uncalled-for dependence when the requirement is immediate and/or cursory. With automatic summarizers, instead, they can independently find the documents and decide which documents may be useful based on their domain knowledge. They can focus more on legal problems rather than on the process of finding documents. Automatic text summarizer is particularly useful for novice users or ordinary citizen. Since legal documents are increasingly available in the public domain, people can easily access them. But legal documents are often long and full of legal jargons which play the role of major obstacles to have a first rough idea of cases. If we have techniques that automate the process of drafting the case summaries, the user's liberty towards consulting legal information is also greatly improved. The user can select cases based on his choice without interference of lawyers, who can hide many cases from the user. Automatic text summarization can thus help connect them to legal domain. The important challenge in legal text summarization is identifying the informatory part while avoiding the irrelevant one. The issues are:

- what content should the summaries have?
- how to present to the user in the best possible way?
- how to automatically extract this content from the legal documents? etc.

Although the issues are present in text summarization in general, legal documents have certain specialties.

1.1 Legal Text Differs

Legal text has different characteristics than newspaper articles or scientific text [91]. Here we discuss some of the differences.

– Size: The size of the legal documents tend to be longer than documents in other domains because many domains still depend on collections of abstracts rather than the full text of the documents.
– Structure: Legal documents show a different internal structure. They have status and administrative codes and follows hierarchical structure.
– Vocabulary: The vocabulary of legal texts is different. Legal language uses a number of domain specific terminology besides the standard language.
– Ambiguity: Legal text may be ambiguous as there can be multiple different meaning for the same term, phrase, statement. The same text could be interpreted differently if it occurred in high court opinion than if it occurred in the opinion issued by a district courts.
– Citations: Citations play a prominent role in legal domain than they do in other domains and generally indicate main issues of the case.

These differences play a major role in legal text summarization. For example, there is little or no hierarchy in a general document of, say, news genre. The summarization here primarily focuses on the content words rather than on the structure. But in legal text, the hierarchy of the structure is important. Presence of the same word at different hierarchy will contribute differently. Source of a ruling (whether it is from District court/ State Court / Supreme or Federal court) will determine the importance of the words therein. Also, we can, in general, ignore references/citations in text summarization, but that may not be possible in case of legal texts.

Hence legal text summarization needs special treatment and therefore, demands a study on its own, different from general text summarization. Quite a few legal text summarization techniques are reported at different reputed venues in the recent past. However no systematic review has been conducted till date. We believe that a systematic review improves understanding of the literature, identification of the specific issues of the field and presents directions for future research. We, therefore, attempt to offer a comprehensive overview of the methods and techniques used in summarization with a focus on legal text summarization. We start with some seminal papers of text summarization to give necessary background of text summarization and then to elicit its difference and specialties legal text summarization. The literature we considered mostly range from 2003 till 2014. In this survey, we generally aim to explore how experimental methods have been used and their reported scores. We could not do comprehensive performance evaluation and comparison as the datasets differ.

The rest of the paper is organized as follows. We discuss what motivates us to take up this work in Section 2. Section 3 sets the premise of text summarization in general and then Section 4 progresses to discuss the area of legal text summarization in particular. Section 5 describes different software tools

in legal text summarization. Finally, we conclude stating direction to future work in Section 6.

2 Motivation

Very large volume of information in the legal domain is generated by the legal institutions across the world. For example, a country like India having 24 high courts [97] (provincial courts) and 600 district courts [96] put the legal proceedings in the public domain. This is of paramount importance as a huge number of cases are pending in different courts of India (as of 2011, High Courts $4,217,903$, Supreme Court of India $57,179$) [77]). Also this massive quantity of data available in the legal domain has not been properly explored for providing legal information to general users. Since most court judgments are generally long documents, headnotes are generated which act as summary of the judgments. However headnotes are generated by experienced legal professionals only. This is an extremely time-consuming task which involves considerable human involvement and cognitive effort. Automatic processing of the legal documents can help legal professionals in quickly identifying important sentences from the documents and thus reduce human effort. Text summarization can be applied in extracting important sentences and thus in generating headnotes.

3 Text Summarization

3.1 Introduction to Text Summarization

Automatic summarization is the technique of reducing a text document with a system in order to create a summary that retains the most important information of the original document. With increasing amount of electronic documents, use of such systems become progressively pertinent and inevitable.

Substantial amount of research has been directed to explore different types of summaries, methods to create them and also the methods to evaluate them. Das et al. [13] surveyed some of the approaches both in single and multiple document summarization, giving importance to empirical methods and extractive techniques. The authors discussed issues of automatic summarization, some techniques, and their evaluation that were attempted during 1991-2007.

Nenkova et al. [73] provided an overview of the most prominent methods of automatic text summarization. The authors discussed how representation, sentence scoring or summary selection strategies alter the overall performance of the summarizer. They also pointed out challenges to machine learning approaches for the problem. The paper surveyed the body of works published during 1995-2010.

Lloret and Palomar [65] surveyed a general state-of the-art summarization approaches, describing the important concepts, appropriate international forums and automatic evaluation of summaries from 1995-2010.

Sparck Jones et al. [50] distinguished the summarization activities based on three classes of context factors, namely input, purpose and output factors.

− *Input factors*: text length, genre, single vs. multiple documents
− *Purpose factors*: who is the user and the purpose of summarization
− *Output factors*: running text or headed text etc.

Goldstein enumerated different dimensions to summarization [31]:

− *Construct*: A natural language created summary, abstract, produced by the use of a semantic representation that symbolizes the structure and main points of the text, whereas an extract summary contains pieces of the original text such as key words, phrases, paragraphs, full sentences, and/or condensed sentences.
− *Type*: A summary is a brief account giving the main points of document's content. This includes a set of keywords, a headline, title, abstract, extract, goal-focused abstract, index or table of contents.
− *Purpose*: A generic or overview summary presents a complete feel of the document. On the other hand a query related summary provides the content related to user query or user model, and a goal focused presents information related to specific objective.
− *Number of summarized documents*: A single document summary provides overview of one document and multi document summary renders many functionality.
− *Document length*: The length of single documents often will indicate the degree of redundancy that may be present. For example, newswire articles are usually intended to be summaries of an event and therefore contain minimal amounts of redundancy. Nevertheless, legal documents are generally written to present a point, extend on the point and repeat it in the conclusion.
− *User goal*: When searching for particular information, the objectives would be fulfilled with a well constructed informative summary, take out the need to refer any original documents.
− *Genre*: The information contained in the genres of documents can provide linguistic and structural information useful for summary creation. There are different genres like news documents, opinion pieces, letters and memos, email, scientific documents, books, web pages, legal judgments and speech transcripts (including monologues and dialogues).
− *Presentation*: The summary can be presented in a text-only format, or text with hyperlink references to the original text(s). It can be keywords, phrases, sentences, paragraphs. Also, with the help of suitable user interfaces, the summary length can be expanded or contracted along with provision of adding more contextual sentences around each summary sentence or component in case of text extractive summaries.
− *Source language*: The emergence of translingual (or cross-lingual) and multilingual information retrieval made portions of the input document set can be in different language than the output language.

- *Quality*: As in any presentation, the summary must be coherent, cohesive and if it contains sentences, grammatical as well. It should be readable and accurately reflect the original text, i.e., not contain false implications based on missing references or poor constructions.

The summaries can be classified into extracts and abstracts. The summary produced by reformulating sentences from input text is abstract and extract is produced by extracting sentences from source text. The process of summarization can also be divided as either generic or query oriented. A query based summary presents results which is most relevant to user queries, where as a generic summary provides an overall sense of the document content.

3.2 Single Document Summarization

When a document is too long to go through in detail, and/or the user in a hurry to have a quick overview of its content, a summary of the document concerned seems immensely helpful. Single document summarization is, therefore, an important research issue. In extractive summarization traditional approaches deal with sentence level identification of the important content. Several techniques have been applied to select important sentences from a document. In the following we attempt to categorize some of the recent approaches. This classification is not a comprehensive list, for which one can refer to [13] [73] [65]. Instead we include the papers which were published after those surveys and are not therefore part of the surveys.

3.2.1 Linguistic feature based approaches

Wang et al. [95] proposed a summarization algorithm based on LSA (Latent Semantic Analysis) which blends term description with sentence description for each topic. They select three sentences at most for each topic and the sentences selected not only have the best representation of the topic but also include the terms that can best represent this topic. As with Gong et al. [32], a concept can be depicted by the sentence that has the largest index value in the corresponding right singular vector. They made another assumption that a concept can also be represented by a few of terms, and these terms should have the largest index values in the corresponding left singular vector. They describe these two concepts as sentence description and term description. They also bring up the concept of neighbor weight and propose a novel way that tries to utilize the mutual reinforcement between neighbor sentences to create the term-sentence matrix.

Pal et al. [74] proposed a single-document summarization based on simplified Lesk algorithm with an online semantic dictionary WordNet. A list of distinct sentences is first created from the input text. The WordNet is used to extract all the meaningful words of glosses (dictionary definitions). The intersection is carried out between input text and glosses. The summation of all intersected sentences gives the sentence weight. The sentences are arranged in

descending order of their obtained weights. Top subset of sentences are chosen based on the specific percentage of summarization.

3.2.2 Statistical feature based approaches

Vodolazova et al. [93] described the interaction between a set of statistical and semantic features and their impact on the process of extractive text summarization. They used features like term frequency (TF), inverse term frequencies (ITF), inverse sentence frequencies (ISF), word sense disambiguation (WSD), anaphora resolution (AR), textual entailment (TE), stopwords filtering, standard stopword list (SSW), standard and extended stopwordlist (ASW), not using the stopwords filtering (NOSW) etc. The obtained results have shown that semantic-based method involving AR, TE and WSD benefit the redundancy detection stage. Once the redundant information is detected and discarded, the statistical methods, such as TF and ISF offer a better mechanism to select the most representative sentences to be included in the final summary.

Sharma et al. [85] proposed a web-based text summarization tool called Abstractor. The data and meta-data are retrieved from HTML DOM tree. They implemented a four-fold model, which calculates the scores based on term frequency, font semantics, proper nouns and signal words. Selection of important text depends on aggregated output of the above four scores given to each sentence.

Batcha et al. [4] proposed the use of Conditional Random Field (CRF) and Non-negative Matrix Factorization (NMF) in Automatic Text Summarization. Although NMF is a good technique for summarization which uses a set of semantic features and hidden variables, taking appropriate initialization plays crucial role in convergence of NMF. The authors used CRF to identify and extract the correct features by classifying or grouping the sentences based on the identified patterns in the domain specific corpus. Once classification or grouping is done, they identify contributing terms to that segment or topics. CRF is used to model the sequence labelling problem. Their proposed method showed effective extraction of semantic features with improvement over the existing techniques.

3.2.3 Language-independent approaches

Cabral and Luciano [8] described a platform for language independent summarization (PLIS) that combines techniques for language identification, content translation and summarization. The language identification task was performed using CALIM [59] language identification algorithm followed by translation into English with Microsoft API. The summarization process is done with three extractive features such as word frequency, sentence length, and sentence position. Each approach computes values for the sentences of the text. These values are aggregated and ranked. Top scoring sentences are selected for the summary according to the threshold provided by user, which

may be the sentence quantity or percentage of the text size (e.g. 20 sentences or 30%).

Gupta [39] proposed an algorithm for language independent hybrid text summarization. The author considered language independent features from four different papers proposed by Krishna et al. [56], Fattah and Ren [21], Bun and Ishizuka [7] and Lee and Kim [62]. The text summarizer uses seven features as such as words form similarity with title line, n-gram similarity with title, normalized term and proportional sentence frequency, position feature, relative length feature, extraction of number data, user specified domain specific keywords feature.

3.2.4 Evolutionary Computing based approaches

Abuobieda et al. [1] introduced Differential Evolution (DE) algorithm to optimize the process of sentence clustering. Using five features, namely title, sentence length, sentence position, presence of numeric data and thematic word, they compared three different distance metrics for text clustering, namely Jaccard measure (JM), Normalized Google Distance [10] and cosine similarity. The JM outperformed the other two. Proper selection of similarity measure plays an important role in determining the quality of the summary.

Abuobieda and Albaraa [2] proposed Opposition Differential Evolution (ODL) based method for text summarization. They implemented opposition-based learning (OBL) on DE algorithm to enhance the text summarization. They focus only on the initial population of the DE algorithm which is enhanced using the OBL machine learning approach.

Garcia et al. [28] proposed use of genetic algorithm (GA) for automatic single extractive text summarization. Starting from preprocessing, they described chromosome encoding, initial population, fitness function, parent selection, crossover and mutation step. They considered a document of n-sentences as a chromosome. Some subset of these n sentences can be present in the summary which are represented as 1 and the rest as 0. GA starts with a population of random solutions (initial population step) that are evaluated according to the objective function to optimize (fitness function step). Fitness function is set as a product of expressivity and relevance of the sentence based on its position. Two best solution are chosen which provide high fitness value (parents selection step). In the cross-over step, two best solutions are mixed satisfying the constraint of keeping a fixed number of words in the summary. In the mutation step, they considered the probability of 0.1. The new population is evaluated again and the process is repeated until a satisfactory solution is reached or until some arbitrary stop-criteria is reached (stop condition).

Mendoza and Martha [69] proposed a method of extractive single-document summarization based on genetic operators and guided local search, called MA-SingleDocSum. MA-SingleDocSum based on the approach presented by Hao et.al [47]. A memetic algorithm is used to integrate the own-population-based search of evolutionary algorithms with a guided local search strategy. The proposed algorithm consists of selecting the parents of a new offspring. In this

setting, the father is selected by rank selection strategy and the mother by Roulette wheel selection based on Sivanandam and Deepa [86]. The generation of offspring is done by one-point crossover strategy. The mutation technique applied is related to multi-bit strategy. The local optimization algorithm used in MA-SingleDocSum is guided by local search which maintains an exploitation strategy directed by the information of the problem. The objective function defined for MA-SingleDocSum method is constituted by features such as position, relationship with the title, length, cohesion, and coverage.

Ghalehtaki et al. [30] used cellular learning automata (CLA) for calculating similarity of sentences in Particle swarm optimization (PSO) and Fuzzy logic. They used PSO to assign weights to the features in terms of their importance and then fuzzy logic for scoring sentences. The CLA method concentrates on reducing the redundancy problem but PSO and fuzzy logic methods concentrate on the scoring technique of the sentences. They proposed two methods of text summarization: the first one is based on CLA only and the second one based on combination of fuzzy, PSO and CLA. They used a linear combination of features such as word feature, sentence length, sentence position for selecting the important sentences. Text summarization based on fuzzy PSO CLA provided better performance than text summarization based on CLA only.

3.2.5 Graph based approaches

Hirao and Tsutomu [48] presented a single document summarization based on the Tree Knapsack Problem (TKP). The process is two fold. First, they propose rules for altering a rhetorical structure theory (RST) based discourse tree into a dependency-based discourse tree (DEP-DT), which permits to take a tree trimming approach for summarization. Next, they specified the problem of trimming a DEP-DT as a TKP, then resolve it with integer linear programming (ILP).

Kikuchi and Yuta [52] described single document summarization with a nested tree structure. They represent each document as a nested tree, which is composed of a document tree and a sentence tree. The document tree has sentences as nodes and head modifier relationships between sentences obtained by RST as edges. The relationship between words is obtained by the dependency parser. The nested tree was built by regarding each node of the document tree as a sentence tree. Summarization was achieved by trimming of the nested tree by seeing the problem as that combinatorial optimization using ILP.

Miranda-Jimenez et al. [72] proposed a method for single-document abstractive summarization, based on conceptual graphs as the underlying text representation of Sowa et al. [88]. They focused on ranking nodes and selecting the most important nodes according to Hyperlink Induced Topic Search algorithm of Kleinberg et al. [55] over weighted conceptual graphs and using other heuristics based on semantic patterns of VerbNet Kipper et al. [54]. The summary at semantic level is the resulting structure of selected nodes. The authors have carried out experiments by creating three groups of documents of sentence length (sen) 2, 3, 4 from news dataset.

Ferreira and Rafael [22] describe a graph model based on four dimensions (similarity, semantic similarity, co-reference resolution, discourse relation). Similarity measures the overlap in content between pairs of sentences. Semantic similarity applies ontologic conceptual relations such as synonyms, hyponym, and hypernym. The authors represent each sentence as a vector of terms and calculated the semantic similarity between each pair of terms using WordNet. Co-reference resolution links up sentences that relate to the same subject. The developed prototype provides named, nominal, and pronominal co-reference resolutions. Discourse relations highlight the related relationships in the text. The TextRank algorithm with four dimensional graph model attains better precision, recall and F-measure (quantitative) and also shows better qualitative results.

Ledeneva et al. [60] presented single extractive text summarization using graph based ranking with Maximal Frequent Sequences (MFS). The nodes of the graph in term selection step are considered as MFS, which are then ranked in term weighting step using graph based algorithm Text Rank.

Hamid et al. [46] introduced an unsupervised graph based ranking model for text summarization. A graph is built by collecting words and their lexical relationships in the document. They implemented semantic information (definition, sentimental polarity) of words to improve edge weights (interconnectivity) between nodes (words). The polarity based ranking algorithm applied over the graph and subset of high-ranked and low-ranked collected words named as keywords. They extract sentences based on the high rank which is defined by rank vector of keywords.

3.3 Multi document summarization

Multi-document summarization, as the name suggests, automatically creates a summary from multiple texts written about the same topic. The summary helps the user to quickly get familiarized with the information contained in a larger cluster of documents. Generally summaries so created are both concise and comprehensive. Unlike single document counterpart, multi-document summarization is more complex and difficult due to thematic diversity within a large number of documents. However there have been several attempts to meet the challenge. In the following, we mention a few recent ones.

3.3.1 Linguistic feature based approaches

Chen et al. [9] designed a multi-document summarization system based on common fact detection. First important terms are extracted from the citation sentences occurring within the given set of documents. They construct a term co-occurrence base from a collection of 18514 scientific abstracts in the domain of computational linguistics. For each extracted term, the co-occurrence base is consulted to expand the term which helps in detecting common facts in citations. The citation sentences are clustered based on the common facts.

Summaries are generated by selecting top-few sentences from the cluster after removing redundancy.

Gross et al. [33] proposed an 'Association Mixture Text Summarization' method which is unsupervised and language-independent. The summaries are generated based on a mixture of two underlying assumptions: one, association between two terms are relevant if they co-occur in a sentence with a higher probability than that when they are mutually independent, and two, their association is characteristic of the particular document if they co-occur in the document more frequently than in the overall corpus. The association between a pair of terms is weighted with the help of log-likelihood ratio test. Sentences with strong word pair associations are selected to generate the summary.

Ma et al. [66] described summarization based on combination of multiple features such as n-gram (unigram, bigram, skip-bigram), dependency word pair (DWP) co-occurrence and global TF∗ IDF. They combined weights of all the above features to find the significance of the text and extracted the relevant sentences by greedy algorithm based on the combined score. Cosine similarity is used to detect duplicate sentences and only unique sentences with high combined score are used to generate a summary.

3.3.2 Evolutionary Computing based approaches

Lee et al. [61] describes a multi-document summarization method using using both a topic model and a fuzzy method. The Latent Dirichlet Allocation (LDA) model of Blei et al. [5] is used for extracting the topic words and each sentence in the input documents is scored by using the topic words. They used fuzzy technique to extract the important sentences in the documents. They evaluated the generated summaries using Kullback Leibler (KL) divergence, Jensen Shannon divergence, cosine similarity, proportion of topic terms, unigram and multinomial probabilities between summary and input documents.

Kumar and Yogan Jaya [58] described multi document summarization based on news components using fuzzy cross-document relations. There are three main phases which include component sentence extraction, cross-document relation (CST relation) identification and sentence scoring using fuzzy reasoning. From the set of input documents D, component sentences are extracted using the gazetteer list and named entity recognition. CST relation is identified using Genetic-Case Base Reasoning (CBR) model with five different features, namely Cosine similarity, Word overlap, Length type, noun phrase similarity, verb phrase similarity for each sentence pair taken from a document cluster. The fuzzy reasoning model is used for sentence scoring. The sentence selection is based on sentence scores after removing duplicates.

3.3.3 Graph based approaches

Samei and Borhan [81] proposed summarization based on the combination of graph ranking and minimum distortion. The input documents were transferred into a directed weighted graph by adding a vertex for each sentence. Each two

sentences were then examined by a distortion measure representing the semantic relation between them, and an edge was added between two sentences if the distortion was below a predefined threshold. This distortion measure, based on normalized squared difference of term frequencies between two sentences x and y is used to represent the semantic distance between the nodes as the weight of the edges. After the graph is built, they used (Brin et al. [6]) PageRank algorithm to select the important sentences.

Ferreira and Rafael [23] proposed sentence clustering algorithm to deal with the redundancy and information diversity problems based on graph model that uses statistical similarities and linguistic features. The proposed algorithm uses the text representation proposed in Ferreira et al. [24] to convert the text into a graph model. It identifies the main sentences from the graph using TextRank method of Mihalcea and Tarau [70] and groups the sentences based on the similarity between them. The authors have generated summaries with 200 and 400 words (W) for each collection of documents.

Alliheedi et al. [3] proposed automated detection of rhetorical figures for improving the performance of the *MEAD*[1] summarizer [79].

They used *JANTOR* [29], a computational annotation tool for detecting and classifying rhetorical figures. They considered four figures, namely antimetabole, epanalepsis, isocolon, and polyptoton. The rhetorical figure value RF_j feature ($RF_j = \frac{n}{N}$, j is a type of rhetorical figure such as polyptoton or isocolon, and n is the total number of occurrences of j in the document and the total number of all rhetorical figures denoted by N, which contains all four figures considered that exist in the document) is added to *JANTOR-MEAD* system along with rhetorical figures. The rhetorical value is incorporated along with other features in calculating sentence score. High scoring sentences are chosen to generate a summary. The JANTOR-MEAD using all four figures provides better results than MEAD in every ROUGE method.

3.4 Evaluation of Text Summarization

Once the summaries are generated, it is imperative to judge their quality and thereby performance of the summarizer. Evaluating the performance of different search activities is a crucial issue that drives to the future research. TREC[2] (Text REtrieval Conference), MUC[3] (Message Understanding Conference), DUC[4] (Document Understanding Conference), TAC[5] (Text Analysis Conference) and FIRE[6] (Forum for Information Retrieval Evaluation) have created a benchmark data and established baselines for performance study.

---

[1] MEAD is an open-source summarization system which allows researchers to experiment with different features and methods for the single and multi-document summarization.

[2] http://www.trec.nist.gov

[3] http://www-nlpir.nist.gov/related_projects/muc/

[4] http://duc.nist.gov/

[5] http://www.nist.gov/tac/

[6] http://www.isical.ac.in/ fire/

Since summarization is a subjective issue, human satisfaction plays a very crucial role. Evaluation therefore involves human intervention. System generated summaries are compared against human generated ones, considered as gold standard. Evaluation using a reasonably large-scale data was started by MUC followed by DUC and TAC from NIST (National Institute of Standard and Technology), USA. Along with providing benchmark data, these forums also offer standard metrics to quantitatively evaluate the performance of single and multi document summarization. Some of the benchmark data used are listed in the Table 1.

Conference	#Documents	Source Genre
MUC-1	10 messages	Military reports
MUC-2	105 messages	Military reports
MUC-3	1300 messages	News reports
MUC-4	1500 messages	News reports
MUC-5	1200 to 1600 documents	News reports
MUC-6	100 articles	News reports
MUC-7	100 articles	News reports
DUC 2001	30 sets of 10 documents each	Newswire/paper stories
DUC 2002	30 sets of 10 documents each	Newswire/paper stories
DUC 2003	30 TDT clusters-298 doc, 30 TREC clusters-326 doc, 30 TREC Novelty clusters-734 doc	Newswire/paper stories
DUC 2004	50 and 24 clusters of 10 documents each	English/Arabic Newswire
DUC 2005	50 DUC topics and each topic 25-50 documents	English News
DUC 2006	50 DUC topics and each topic 25-50 documents	English News
DUC 2007	45 DUC topics and each topic 25-50 documents	English News
TAC-08	48 topics contains 20 documents per topic	English News/blog
TAC-09	44 topics contains 20 documents per topic	English News
TAC-10	46 topics contains 20 documents per topic	English News
TAC-11	44 topics contains 20 documents per topic	English News

Table 1: **Summarization conferences and their collections**

3.4.1 Evaluation Metrics

ROUGE [63] means Recall-Oriented Understudy for Gisting Evaluation. It automatically measures the quality of human generated summary by using its measures based on n-gram co-occurrence statistics. ROUGE includes five variants like ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S and ROUGE-SU. These variants reveal different performance aspects of a summarization system. We briefly describe various versions of ROUGE below.

- *ROUGE-N:* It measures the n-gram units common between a candidate summary and a collection of reference summaries. It is computed as follows.

$$ROUGE - N$$
$$= \frac{\sum_{S\epsilon\{ReferenceSummaries\}} \sum_{gram_n\epsilon\{S\}} Count_{match}\left(gram_n\right)}{\sum_{S\epsilon\{ReferenceSummaries\}} \sum_{gram_n\epsilon\{S\}} Count(gram_n)}$$

  Where S is a sentence, n stands for the length of the n-gram, $gram_n$ and $count_{match}$ is the maximum number of $n - grams$ co-occurring in a candidate summary and a set of reference summaries. Here N is the n-gram's length (E.g. ROUGE-1 (R-1) is unigram, ROUGE-2 (R-2) is bigram, ROUGE-3 (R-3) is trigram, ROUGE-4 (R-4) is 4-gram). The number of n-grams in the denominator of ROUGE-N increases as we add more references. Therefore multiple references can be easily integrated into this metric. The numerator sums over all reference summaries. This effectively gives more weight to matching n-grams occurring in multiple references. Therefore a candidate summary that contains words shared by more references is favored by the ROUGE-N measure.
- *ROUGE-L:* It computes Longest Common Subsequence (LCS) metric. The longer LCS between the system and human summaries shows more similarity and therefore higher quality of the system summary. LCS does not require consecutive matches but in-sequence matches that reflect sentence level word order as n-grams. It automatically includes longest in-sequence common n-grams, therefore no predefined n-gram length is necessary. In order to overcome some of the shortcomings of ROUGE-N, more precisely the fact that the measure may be based on too small sequences of text, ROUGE-L takes into account the LCS between two sequences of text divided by the the length of one of the text. Even if this method is more flexible than the previous one, it still suffers from the fact that all n-grams have to be continuous.
- *ROUGE-W:* It is the weighted longest common subsequence metric. One problem with ROUGE-L is that all LCS with same lengths are rewarded equally. The LCS can be either related to a consecutive set of words or a long sequence with many gaps. While ROUGE-L treats all sequence matches equally, it makes sense that sequences with many gaps receive lower scores in comparison with consecutive matches. ROUGE-W considers an additional weighting function that awards consecutive matches more than non-consecutive ones. ROUGE-W introduces a weighting factor of 1.2 to better score contiguous common subsequences.
- *ROUGE-S:* It is the skip-bigram co-occurrence statistics metric. It measures the number of overlapping skip-bigrams in the evaluation. Skip-bigram is any word pair in the sentence order with random gaps. ROUGE-S4 measures any bigram with a distance less than 4.
- *ROUGE-SU:* It measures skip-bigram plus unigram-based co-occurrence statistics. This measure considers both unigrams and skip-bigrams in the evaluation. ROUGE-S does not give any credit to a system generated sen-

tence if the sentence does not have any word pair co-occurring in the reference sentence. To solve this problem, ROUGE-SU was proposed which is an extension of ROUGE-S that also considers unigram matches between the two summaries. ROUGE-SU4 and ROUGE-SU6 are used to measure any bigram with the distances less than 4 and 6 respectively.

Even though ROUGE has been widely used for evaluation of summaries but it suffers from few limitations.

- The automatic summaries have to be evaluated by comparing with human summaries that involve expensive human effort [94].
- ROUGE scores donot take into consideration linguistic qualities such as human readability [52].
- It solely relies on lexical overlaps (n-gram and sequence overlap) between the terms and phrases in the sentences. Therefore, in cases of terminology variations and paraphrasing, ROUGE is not very effective [11].
- ROUGE metrics seem to be ignoring redundant information [15].
- Sequential comparison of a system summary with a human summary is not robust with respect to the number of sentences and their order [15].
- It depends on the length of the system summaries (i.e., the longer is the system with respect to the human, the higher are expected to be the ROUGE scores) [76]

There are two other most frequently used measures, namely precision and recall coming from information retrieval domain.

Precision (P) is the fraction of retrieved documents that are relevant

$$\text{P} = \frac{\#\text{relevant items retrieved}}{\#\text{retrieved items}}$$

Recall (R) is the fraction of relevant documents that are retrieved

$$\text{R} = \frac{\#\text{relevant items retrieved}}{\#\text{relevant items}}$$

F-measure (popularly used F_1 score) is the harmonic mean of precision and recall.

$$F_1 = 2 \cdot \frac{\text{P} \cdot \text{R}}{\text{P} + \text{R}}$$

Ferreira et al. [24] performed a quantitative and qualitative assessment of 15 extractive text summarization algorithms for sentence scoring. They considered different features such as word frequency, TF/IDF, upper case, proper noun, word co-occurrence, lexical similarity, cue-phrase, inclusion of numerical data, sentence length, sentence position, sentence centrality, resemblance to the title, aggregate similarity, TextRank score and Bushy path. Comparative performances of single document summarization techniques on different datasets are grouped based on corpus and described in Table 2. Performances of multi-document summarization techniques are grouped according to corpus and summarized in Table 3.

Researcher(s), year, Reference	Technique	Corpus	Metrics	Results
Wang et al. (2013) [95]	LSA	DUC2002, DUC2004	ROUGE-1, 2, SU4, L	R-1-0.47, R-2-0.26, SU4-0.17, R-L-0.42, R-1-0.38, R-2-0.09, SU4-0.13, R-L-0.35
Vodolazova et al. (2013) [93]	Statistical and semantic features	DUC2002	ROUGE-1	SSW AR WSD TE : TF-R-1-P-0.45 NOSW ISF: TF-R-1-R-0.42 ASW AR WSD TE: TF-R-1-R-0.43
Abuobieda et al. (2013) [1]	DE, JM, Real-to-integer modulator, Feature-based approach	DUC2002	ROUGE-1, 2, L, Precision, Recall, F_1-score	R-1-AVG-P-0.50, R-1-AVG-R-0.44, R-1-AVG-F-0.46, R-2-AVG-P-0.25, R-2-AVG-R-0.22, R-2-AVG-F-0.23, R-L-AVG-P-0.48, R-L-AVG-R-0.48, R-2-AVG-F-0.48
Abuobieda and Albaraa (2013) [2]	ODL	DUC2002	ROUGE-1, 2, L	R-1-AVG-P-0.50, R-1-AVG-R-0.44, R-1-AVG- F-0.46, R-2-AVG-P-0.25, R-2-AVG-R-0.22, R-2-AVG-F-0.23, R-L-AVG-P-0.46, R-L-AVG-R-0.40, R-L-AVG-F-0.42
Garcia et al. (2013) [28]	GA	DUC2002	ROUGE	P-0.48, R-0.48, F-0.48
Mendoza and Martha (2014) [69]	MA-SingleDocSum	DUC2001, DUC2002	ROUGE-1, 2	R-1-0.47, R-2-0.20, R-1-0.48, R-2-0.22
Ghalehtaki et al. (2014) [30]	Fuzzy logic, PSO, CLA	DUC2002	ROUGE-1, 2, L	R-1-AVG-P-0.49, R-1-AVG-R-0.43, R-1-AVG-F-0.46, R-2-AVG-P-0.22, R-2-AVG-R-0.19, R-2-AVG-F-0.20, R-L-AVG-P-0.45, R-L-AVG-R-0.40, R-L-AVG-F-0.43
Ferreira and Rafael (2013) [24]	15 Sentence scoring methods	CNN News, Cosmic varience and Internet Explorer Blog, SUMMAC	ROUGE	AVG-P-0.35, AVG-R-0.73, AVG-F-0.46, AVG-P-0.62, AVG-R-0.76, AVG-F-0.68, AVG-P-0.28, AVG-R-0.36, AVG-F-0.29

Cabral and Luciano (2014) [8]	CALIM, Microsoft API, Word Frequency, Sentence Length, Sentence Position	CNN (English, Spanish), TeMrio-Portuguese (Brazilian)	ROUGE	AVG-P-0.29, AVG-R-0.71, AVG-F-0.41, AVG-P-0.11, AVG-R-0.72, AVG-F-0.20, AVG-R-0.77
Ferreira and Rafael (2013) [22]	Similarity, Semantic similarity, Coreference resolution, Discourse relations	CNN News	ROUGE	Quantitative: AVG-P-0,40, AVG-R-0,60, AVG-F-0,48, Qualitative-34,87%
Pal et al. (2014) [74]	Simplified Lesk algorithm and WordNet	News paper articles	Precision, Recall, F_1-score	50% close to human summaries
Ledeneva et al. (2014) [60]	MFS	DUC News	ROUGE	F-48.65
Gupta (2014) [39]	Statistical based features	Wikipedia (Punjabi), The Tribune, Hindustan Times, Indian Express (English)	Intrinsic and extrinsic	AVG-F-87.56 (top 20% sentences)
Sharma et al. (2014) [85]	Abstractor web based tool	Wikipedia	Precision, Recall, F_1-score	90% close to human summaries
Hirao and Tsutomu (2013) [48]	DEP-DT	RST-DTB	ROUGE-1, 2	R1-R-0.32, R1-F-0.31, R2-R-0.11, R2-F-0.10
Kikuchi and Yuta (2014) [52]	Nested Trees	RST-DTB	ROUGE-1	R1-0.35
Miranda-Jimenez et al. (2013) [72]	Conceptual Graphs	DUC 2003	Precision, Recall, F_1-score	2-Sen-P-0.50, 2-Sen-R-0.50, 2-Sen-F-0.50, 3-Sen-P-0.25, 3-Sen-R -0.22, 3-Sen-F-0.23, 4-Sen-P-0.50, 4-Sen-R-0.50, 4-Sen-F-0.50

Table 2: **Single Document Text Summarization**

3.4.2 Discussions

Summarization techniques used can be classified into two categories as described above. Here we would discuss the performance of different techniques based on the categories.

Single document Summarization

The authors used different summarization approaches on variety of datasets like DUC2002, news data set (English and other languages), RST-Discourse Treebank (DTB) and Wikipedia (English, Punjabi). Even when the same datasets are used, different metrics are reported, and, therefore it is diffi-

cult to compare their performances straightway and make specific comments. Nevertheless, we can make some general observations from the reported scores.

Wang et al. [95], Vodolazova et al. [93], Abuobieda et al. [1], Abuobieda and Albaraa [2], Garcia et al. [28], Mendoza and Martha [69], Ghalehtaki et al. [30] implemented different methodologies on DUC2002 dataset. In all the above, the approaches of Abuobieda et al. [1] DE with JM, real-to-integer modulator, feature such as title, sentence length, sentence position, numerical data, thematic words performed better in comparison to other ones.

Ferreira and Rafael [24], Cabral and Luciano [8], Ferreira and Rafael [22], Pal et al.(2014) [74], Ledeneva et al. [60] used various techniques on news dataset. However, the best performance was reported on almost all metrics by Ferreira and Rafael [24] with 15 different sentence scoring methods.

Papers by Gupta [39], Sharma et al. [85] respectively implemented different approaches on Wikipedia data, but they reported diverse metrics.

Hirao and Tsutomu [48], Kikuchi and Yuta [52] adopted different methods on RST-DTB data set. Among the three, the best score is achieved by Kikuchi and Yuta [52] with nested tree method. Below we discuss some of the advantages of the different methodologies in single document summarization.

- LSA (Wang et al. 2013 [95]): This method selects the sentences and terms which have the best representation of the topic.
- Statistical and semantic features (Vodolazova et al. 2013 [93]): These methods comprising AR, TE and WSD eliminates redundancy in summarization. The statistical methods like TF and ISF select the most representative sentences to be included in the final summary.
- DE (Abuobieda et al. (2013) [1]): This method optimizes the allocation of sentence to a group. JM improves the coverage and topic diversity of summary. The feature-based approach captures the full relationship between a sentence and other sentences in a cluster or a document. Therefore, best representative sentence are selected from each cluster to represent the topic.
- ODL (Abuobieda and Albaraa (2013) [2]): A method name ODL generates more optimal solution then classical DE by allocating each sentence to a group.
- GA (Garcia et al. (2013) [28]): This approach optimizes the sentence selection process based on frequency of the terms so that important sentences are selected in summary.
- MA-SingleDocSum (Mendoza and Martha (2014) [69]): This algorithm addresses the generation of summaries as a binary optimization problem indicates presence or absence of the sentence in the summary instead of sentence belongs to a group. Memetic algorithm redirects the search towards a best solution. Multi-bit mutation encourages information diversity.
- Fuzzy logic, PSO, CLA (Ghalehtaki et al. (2014) [30]): Fuzzy logic generate the score for every sentence based on features and PSO assigns suitable weights to each feature to select important sentences. The CLA reduces the information redundancy.

– Sentence scoring methods (Ferreira and Rafael (2013) [24]): The sentence position feature extracts important phrases at the beginning and end of document from the news data set. The sentence length features identifies small text from blog data set. The Resemblance to the title feature extracts title text in scientific papers.
– PLIS (Cabral and Luciano (2014) [8]): The proposed approach is a language independent summarizer works for 25 different languages.
– Graph model (Ferreira and Rafael (2013) [22]): This approach covers different dimension in identifies the relation between sentence.
– Simplified Lesk (Pal et al. (2014) [74]): The method with WordNet extracts the relevant sentences based on semantic information of the text.
– MFS (Ledeneva et al. (2014) [60]): This method identifies most important information in text without the need of deep linguistic knowledge, nor domain or language specific annotated corpora, which makes it highly portable to other domains, genres, and languages.
– Statistical based features (Gupta (2014) [39]): These techniques are independent of any language and summarizes the text different language. So, these techniques do not need any additional linguistic knowledge or complex linguistic processing.
– Abstractor tool (Sharma et al. (2014) [85]): This methodology retrieves data as well as structural details of the document from HTML content. Another important advantages is Font semantics, determined by considering features like italicizing and underlining the text which points out certain important portions of the document.
– DEP-DT (Hirao and Tsutomu (2013) [48]):This method defines parent-child relationships between textual units at higher positions in a tree so that summary is a the optimal.
– Nested trees (Kikuchi and Yuta (2014) [52]): This method takes account of both relations between sentences and relations between words without losing important content in the source document.
– Conceptual graph (Miranda-Jimenez et al. (2013) [72]): This approach represents complete semantic relations among the text units so that meaningful summary can be generated.

Multi-document Summarization

Kumar and Yogan Jaya [58], Samei and Borhan [81], Ferreira and Rafael [23] implemented different approaches on DUC2002 data set. The performance of Kumar and Yogan Jaya [58] is better with Genetic-CBR and fuzzy reasoning model. Gross et al. [33], Ma et al. [66], Lee et al. [61] adopted different methods on news data set. The association mixture method of Gross et al. [33] exhibited the best performance.

Researcher(s), year, Reference	Technique	Corpus	Metrics	Results
Kumar and Yogan Jaya (2014) [58]	Genetic-CBR Model, Fuzzy Reasoning Model	DUC2002	ROUGE-1, 2, S, SU	AVG-P-0.34, AVG-R-0.33, AVG-F-0.33, AVG-P-0.12, AVG-R-0.12, AVG-F-0.12, AVG-P-0.09, AVG-R-0.09, AVG-F-0.09, AVG-P-0.10, AVG-R-0.09, AVG-F-0.10
Samei and Borhan (2014) [81]	Minimum Distortion	DUC2002	ROUGE-1	R-1-0.35
Ferreira and Rafael(2014) [23]	Statistical and linguistic features	DUC2002	ROUGE	400W-F-25,4% 200W-F-30%
Gross et al. (2014) [33]	Association mixture method	DUC2007	ROUGE-1, 2, L	R-1-F-0.42, R-2-F-0.10, R-3-F-0.03, R-L-F-0.38
Ma et al. (2014) [66]	Tf*idf, Unigram, Bigram, Skip-bigram, DWP	TAC-2010	ROUGE-2, SU4	R-2-0.10, R-SU4-0.13
Lee et al. (2013) [61]	LDA topic modeling, Fuzzy technique	Newsblaster systems	SIMetricx-v1	KL-1.85
Chen et al. (2014) [9]	Common Fact Detection	ACL Anthology Network	ROUGE-1, 2	P-0.57, R-0.66, F-0.60, P-0.39, R-0.45, F-0.43
Alliheedi et al. (2014) [3]	Rhetorical Figuration	American Presidency Project	ROUGE-1, L, S4, SU4, ROUGE with Porter Stemmer	R-1-0.71, R-L-0.70, R-S4-0.51, R-SU4-0.51, R-1-0.74, R-L-0.73, R-S4-0.56, R-SU4-0.56

Table 3: **Multi Document Text Summarization**

Below we discuss some of the advantages of the different methodologies in muti-document summarization.

– Genetic-CBR model (Kumar and Yogan Jaya(2014) [58]): This approach identifies the relations between sentences directly from un-annotated documents instead of manual annotation by human experts and fuzzy reasoning model is designed to rank sentence based on the type of relation it holds and not solely on the total number of relations.
– Minimum Distortion (Samei and Borhan(2014) [81]): This method represents the semantic difference between two sentences which covers important content of the document with minimum redundancy.
– Statistical and linguistic features (Ferreira and Rafael(2014) [23]): This methodology deals with redundancy and information diversity. As it is completely unsupervised method, it does not need annotated corpus.

– Association mixture method (Gross et al.(2014) [33]): This approach selects most salient information based on relative association between words instead of using hand-crafted linguistic resources.
– N-gram and DWP (Ma et al.(2014) [66]): The unigram indicates the common keywords of topic and bigram, skip-bigram describes the sequential relationships in the sentences. DWP describes the syntactic relationships between words. TF*IDF discovers important keywords in a document.
– LDA topic modeling and fuzzy method (Lee et al.(2013) [61]): These methods reduces the error rates of divergence between the input document and the summary.
– Common Fact Detection (Chen et al.(2014) [9]): The summary created using common facts helps researchers who want a brief description of a group citation about a topic.
– Rhetorical Figuration (Alliheedi et al.(2014) [3]): This method automatically detects patterns of persuasive language, generally at the sentence level, can provide linguistic knowledge and expected to highlight significant sentences preserved in a summary.

4 Legal Text Summarization

Legal text summarization is a process of generating summaries from court judgments. The summarization here is different from that of other genres. The legal text, specifically court judgments, are structurally distinct and not comparable to general text. Some of salient specialties are discussed in Sec. 1.1. In legal text summarization we generate a headnote (summary) from a court judgment which necessarily contains Article No., regulations and/or some other statutory wordings. On the other hand, in general text summarization, important information is extracted without the constraints. Legal documents are much longer than office memos or newspaper articles or magazine articles. They exhibit wide range of internal structure (sections, articles and paragraphs in statutes, sections and sub-sections in regulations). The importance of individual documents is determined, to a great extent, by their origin. The same text would be interpreted differently if it occurs in a higher court opinion than in the opinion issued by a lower court. Citations play a major and crucial role in legal documents which indicate important information about the case. Due to the reasons, legal text summarization demands special attention and straightway application of successful techniques of general text summarization may not be effective here. Below we discuss a few classes of summarization techniques that have been tried till date. Although they may have similar groupings as with general text categorization for a few cases, there are differences at finer level of application.

4.1 Feature based approaches

Galgani et al. [26] described summarization of legal documents, by applying knowledge base (KB) to combining different summarization techniques. The creation of KB of rules are based on ripple down rules of Compton and Jansen [12]. These rules describes the selection of important sentences as candidate catchphrases. They developed a tool that assists in checking of legal dataset, rule creation, selection, specification of features based on present case context and using different information in different situations. They used data set of AustLII (Australasian Legal Information Institute) and evaluated using ROUGE-1 with threshold of 0.5 and 0.7 with average number of extracted sentences per document (SpD). The KB outperformed other methods in precision, followed by KB+Citations (CIT), although KB+Citations achieved higher recall.

Galgani et al. [25] described citation-based approach to summarization. They generated cathphrases from citation text or using citation to extract sentences from the document. The choice of best citations as candidates extraction was based on class of centroid or centrality-based summarisation of Erkan and Guneset [14], Radev et al. [79]. They calculated centrality scores based on average value of similarity and similarity of link between two citances/citphrases. The corpus was collected from AustLII and evaluated with ROUGE-1, SU6 and W. The CpSent (citphrases are used to rank sentences of the target document) method gave the best performance in precision and recall.

Kumar et al. [57] described an approach to generate a short summary from the given legal judgment using the topics obtained from LDA. The documents were passed to LDA as bag of words and based on the probabilistic model different topics were generated. They made an assumption that, the number of topics obtained from LDA is equal to seven that represents different rhetorical role of each judgment as described in Saravanan et al. [83]. They developed an algorithm to calculate sentence scores based on the probability of occurrence of each word with respect to each topic. The final summary is generated based on the topics from LDA. The data set consists of 116 documents from 5 different sub-domains (Income Tax (I), Rent control Act (R), Motor Act (M), Negotiable Instrument Act (N), Sales Tax(S)) belonging to civil case in India collected from http://www.keralawyer.com/.

In the general text summarization, the feature based methods involving anaphora resolution, textual entailment and word sense disambiguation help identify semantics of text, while term frequency and inverse sentence frequency select the most representative sentences to be included in the final summary. But, in the legal text summarization, TF-IDF and other methods extract catchphrases and presents the case depending on the context. Absence of freely available legal dictionary or ontology hinders automatic understanding of legal text and therefore semantic analysis is a far cry. The LSA method used in general text summarization selects the set of sentences and terms that best represent the topic. On the contrary, the number of topics in legal text is

limited and therefore application of LDA is effective. The variety of feature-based of techniques that have been successfully applied in text summarization in general, can certainly be tried here as well, but with thoughtful customization.

4.2 Graph based approaches

Kim et al. [53] describe a graph based algorithm for extractive summarization on legal judgments. They consider sentences as nodes, and form a directed graph of sentences. A directed edge is added between two nodes when probability of a sentence being embedded in the second crosses a pre-defined threshold. The connected component in the graph represent a summary topic. However, representative sentences are chosen from the connected components based on the key-word strength of the sentences over another threshold of key-value. Representative sentences are supplemented in the summary by other supporting sentences in the connected component if there is direct link from the representative sentence to the supporting sentence. The corpus used is a judgments of the House of Lords (HOLJ) and evaluation done by using precision, recall and F-measure.

Schilder et al. [84] show the influence of the repetition of legal phrases in the text by using graph-based approach. The graphical representation of legal text is based on a similarity function between sentences. The similarity function as well as the voting algorithm on the derived graph representation is different from other graph based approaches (e.g. LexRank). For legal text, the authors hypothesize that some paragraphs summarize the whole text or some part of the text. Identification of such kind of paragraphs computes inter-paragraph similarity scores and selects the best match for every paragraph. The system works like a voting system where each paragraph assigns vote for another paragraph (its best match). The top paragraphs with the most votes were selected as the summary. The process of vote casting can be seen as a similarity function which is based on phrase similarity. The phrase similarity is calculated by co-occurrence of phrases in two paragraphs. The longer the matched phrase is the higher the score.

The graph based methods in legal domain use repetition of legal phrases, find similarity between sentences or paragraphs, vote according to the best matches between them and rank them. Here sentences, even paragraphs sometimes, represent nodes in the graph. But in general text summarization, important sentences are extracted by ranking them. The sentences are taken as nodes and functions of incident edges and weights decide importance of the nodes. Semantic and linguistic aspects between the sentences or terms are determined using dictionary. Even sentimental polarity of words is sometimes used to find interconnectivity between words. But as stated earlier, non-availability of legal text resources, causes difficulty in semantic or sentiment study for legal text.

4.3 Rhetorical role based approaches

Rhetorical roles are used to specify a group of sentences under labels. A document is divided into paragraphs under a particular label. Grover and Claire [35] described progress of an automatic text summarization system for judicial proceedings from HOLJ. The authors show primary annotation scheme of seven rhetorical roles *fact*, *proceedings*, *background*, *proximation*, *distancing*, *framing*, *disposal* assigning a label specifying the argumentative role of each sentence in a fragment of the corpus. They used methodology of Teufel and Moens ([89], [90]) for automatic summarization. NLP techniques were used to distinguish main, subordinate clauses and find the tense (past tense (past), present tense (pres)), aspect features of the main verb in every sentence. They performed linguistic analysis by processing the data through XML-based tools from the Edinburgh Language Technology Group(LTG)[7] Text Tokenisation Toolkit(TTT) [38] and *LT XML* tool-sets.

The main LT TTT program is *ltpos*, a statistical combined part-of-speech (POS) tagger and sentence identifier [71]. The method used for chunking is another use of *fsgmatch* [34], availing hand-written rule which set for noun and verb groups. Once verb groups have been identified they used *fsgmatch grammar* to analyse the verb groups and encoded information about tense, aspect, voice and modality. The main method clause structure identifies main verb and tense of the sentence. They used a probabilistic clause identifier from sections 15-18 of the Penn Treebank [68].

In a follow-up work, Grover and Hachey [36] enriched the corpus HOLJ. It contained a header with structured information, succeed by a sequence of Law Lord's judgments consisting of free-running text. The document contains information such as the respondent, appellant and the date of the hearing. They experimented with five classifiers such as decision trees (C4.5) [78], Naive Bayes (NB) [49], support vector machines (SVM) [75], Littlestone's algorithm for mistake-driven learning of a linear separator (Winnow) [64]. They added new features like *location* (L), *thematic words* (T), *sentence length* (S), *quotation* (Q), *named entities*(E), *cue phrases* (C). The preliminary experiments using C4.5 with location features gave encouraging results.

Grover et al. [37] proposed argumentative roles and defined sub-categories to different categories of rhetorical roles. The *background* category is subdivided into 2 sub-categories: *precedent* and *law*; category *case* into *event* and *lower court decision*; category *own* as *judgment*, *interpretation* and *argument*. The classification of sentences are done based on their argumentative roles.

Farzindar et al. [20] described an approach to summarize the legal proceedings of federal courts in Canada and presenting it as a table-style summary in the legal domain. The important sentences extracted based on the identification of thematic structure of the document and determination of argumentative themes of the textual units in the judgment [18]. The summary is generated in four themes *introduction*, *context*, *juridical analysis* and *conclusion*. Based

---

[7] https://www.ltg.ed.ac.uk/software/lt-ttt2/

on the experimental work of judge Mailhot [67] they divide the legal decisions into thematic segments.

Farzindar et al. [18] extended the work by presenting the summary with an additional thematic structure *decision data* on top of *introduction*, *context*, *juridical analysis*, *conclusion*. The implementation of the approach is a system called LetSum (Legal text Summarizer). The summary is generated in four steps: thematic segmentation to identify the structures of document, filtering to remove insignificant quotations and noises, building best candidate units through selection and finally table style summary production. The process of thematic segmentation is done based on the particular knowledge of the legal field. Each thematic segment can be associated with an argumentative role in the judgment based on the presence of significant section titles and certain linguistic markers.

Ravindran and Saravanan [80] [82] proposed a novel idea for applying probabilistic graphical models for automatic text summarization task related to a legal domain. They identified seven rhetorical roles *identifying the case*, *establishing facts of the case*, *arguing the case*, *history of the case*, *arguments*, *ratio decidendi*, and *final decision* present in the legal document. They performed text segmentation of a given legal document into seven roles using CRF. A linear chain of CRF with parameters $C = C_1, C_2..$ defines a conditional probability for a label sequence $l = l_1,...l_W$ given an observed input sequence $s = s1,...s_W$ to be

$$P_C(l|s) = \frac{1}{Z_s} exp \left[\sum_{t=1}^{W} \sum_{k=1}^{m} C_k f_k(l_{t-1}, l_t, s, t) \right]$$

where Z_s is a normalized factor and $f_k(l_{t-1}, l_t, s, t)$ is one of the m features function and C_k is a learned weight associated with the feature function.

They created a collection features like *cue phrase*, *named entity recognition*, *local features* and *layout features*, *state transition features*, *legal vocabulary features* etc to identify the labels. The term distribution model Saravanan et al. [83] used assigns probabilistic weights and normalize the occurrence of the terms so that it selects important sentences from a legal document. They used the legal judgments of three sub-domains (Rent control, Income tax, Sales tax) from www.kerelawyer.com and evaluated using F-measure and Mean Average Precision (MAP) and ROUGE.

Kavila and Selvani [51] described a hybrid system based on key phrase/key word matching and case based technique. The author proposed rhetorical roles *appeal no*, *year*, *case*, *judges*, *petitioner*, *respondent*, *counsel for the appellant*, *counsel for the respondent*, *judgments by*, *sections*, *facts*, *positive* & *negative*, *judgment*. The identification of labels was done using different kind of features like structure identification, abbreviated words, length of word, etc.

4.4 Classification based approaches

Hachey et al. [40] described a classifier which determines the rhetorical status
of sentences in texts from a HOLJ corpus. The sentences extracted based
on the feature set of Teufel and Moens [90]. The rhetorical roles identified
are *fact*, *proceedings*, *background*, *framing*, *disposal*, *textual*, *others*. They ran
experiments with five classifiers: C4.5, NB, SVM, Winnow using the same
features. C4.5 yielded better results, in terms of micro-averaged F-score, (65.4)
with location features only and SVMs performed second best (60.6) with all
features. NB is next (51.8) with all but thematic word features. Winnow gave
poor performance with all the features. The authors have reported individual
(I) and cumulative scores (C) for the different features.

Hachey et al. [41–45] performed series of experiments using machine learn-
ing approaches with same features set for rhetorical status classification. They
used $p(y = yes \mid \overrightarrow{x})$ for ranking the sentences. They applied point-biserial
correlation coefficient for quantitative evaluation with NB and ME. They ex-
perimented with ME and sequence modeling (SEQ), previous labels but no
search (PL). Here, the conditional probability of a tag sequence $y_1 \ldots y_n$ given
a lord's speech $s_1 \ldots s_n$ is approximated as:

$$p(y_1..y_n|s_1..s_n) \approx \prod_{i=1}^{n} p(y_i/x_i)$$

where $p(y_i|x_i)$ is normalized probability at sentence i of a tag y_i given the
context x_i. Also, $p(y_i|x_i)$ has the following log linear form.

$$p(y_i|x_i) = \frac{1}{Z(x_i)} exp(\sum_j \lambda_i f_j(x_i, y_i))$$

where f_j include the features. They included lemmatised (L) token and hy-
pernym (H) cue phrase features gave a better results compared with normal
cue phrases. Among all the methods SVM and ME sequence models showed
good performance.

Yousfi-Monod et al. [98] described sentence classification with NB classifier
using set of surface (Sur), emphasis (Em), and content features (vocabulary
(Voc)). They define four sections or themes: *introduction*, *context*, *reasoning*
and *conclusion* for summarization and named it as *PRODSUM* (PRobabilistic
Decision SUMmarizer). The authors have taken two sub-domains immigration
(I) and tax (T) for English (E) and French (F) languages.

Here, we present a summary of the reported results in different text process-
ing tasks including summarization of legal documents and these are grouped
based on corpus (Table 4).

Researcher(s), year, Reference	Technique	Corpus	Metrics	Results
Grover and Claire (2003) [35]	POS tagging, Noun and verb group chunking, Clause boundary identification, Main verb and tense identification.	HOLJ	Precision, Recall, F-Measure	Tense identification Past-P-97.78, Past-R-88.00, Past-F-92.63, Pres-P-81.58, Pres-R-93.93, Pres-F-87.32, Main verb group identification P-90.80, R-84.04, F-87.29
Grover and Hachey (2004) [36]	C4.5, NB, Winnow, SVM	HOLJ	Human judgments	C4.5-L-AVG-F-65.4
Hachey et al.(2004) [40]	C4.5, NB, Winnow, SVM	HOLJ	Human judgments	C4.5-L-I-AVG-F-65.4, C4.5-L-C-AVG-F-59.7, NB-Q-C–AVG-F-51.8, Winnow-T-C-AVG-F-41.4, SVM-T-C-AVG-F-60.6
Hachey et al. (2004) [41]	ME, PL, SEQ	HOLJ	Human judgments	ME-T-C-57.5, SEQ-Q-C-61.2, PL-T-C-58.1
Hachey et al. (2005) [43]	NB, ME	HOLJ	Precision, Recall, F-Measure	NB-S-F-29.8, ME-Q-F-28.0, Point-biserial correlation NB-S-C-0.23, ME-T-C-0.22
Hachey et al. (2005) [44]	ME sequence model	HOLJ	Precision, Recall, F-Measure	Mic AVG-P-61.4, Mic AVG-R-60.9, Mic AVG-F-61.2
Hachey et al.(2005) [42]	NB, ME	HOLJ	Precision, Recall, F-Measure	NB-T-F-31.2, ME-T-F-27.3, Mic AVG-P-51.4, Mic AVG-R-17.6, Mic AVG-F-26.3, NB-C-L-P-38.9, NB-C-L-R-20.1, NB-C-L-F-26.5, ME-C-L-P-63.2, ME-C-L-R-13.0, ME-C-L-F-21.6, NB-C-H-P-23.6 NB-C-H-R-32.7, NB-C-H-F-27.4, ME-C-H-P-60.3, ME-C-H-R-13.4, ME-C-H-F-22.0
Hachey et al. (2006) [45]	NB, ME	HOLJ	Precision, Recall, F-Measure	Mic AVG-P(ME)-51.4, Mic AVG-R(ME)-17.6, Mic AVG-F(ME)-26.3
Kim et al. (2013) [53]	Graph based method	HOLJ	Precision, Recall, F-Measure	P-31.3, R-36.4, F-33.7
Schilder et al. (2006) [84]	Graph based method	FSupp	ROUGE-2, ROUGE-SU4	R-2-0.90, R-4-0.93
Farzindar et al. (2004) [20]	LetSum tool	Federal Court of Canada	ROUGE	R-1-0.61

				Sur-P-0.46, Sur+Em+Voc-R-0.57, Sur+Em+Voc-F-0.50, PRODSUM-E-I-P-0.44,
Yousfi-Monod et al. (2010) [98]	NB with Surface Emphasis Content	Immigration, Tax, Intellectual property	Precision Recall, F-Measure, ROUGE-2, SU4	PRODSUM-E-I-R- 0.57, PRODSUM-E-I-F-0.50, PRODSUM-E-T-F-0.44, PRODSUM-F-I-F-0.48, PRODSUM-F-T-F-0.36, PRODSUM-E-I-R-2-0.63, PRODSUM-E-I-R-SU-0.64, PRODSUM-E-T-R-2-0.66, PRODSUM-E-I-R-SU-0.66
Galgani et al. (2012) [26]	KB	AustLII	ROUGE-1, Precision, Recall, F-Measure	KB-SpD-0.5-P-0.87, KB+CIT- SPD-0.5-R-0.66, KB-SpD-0.7-P-0.69, KB+CIT- SPD-0.7-R-0.36
Galgani et al. (2012) [25]	Citation based method	AustLII	ROUGE-1, SU6,W-1.2, Precision, Recall, F-Measure	CpSent-P-0.82, CpSent-R-0.70, CpSent-F-0.75, CpSent-R-1-P-0.18, CpSent-R-1-R-0.46, CpSent-R-1-F-0.24, CpSent-R-SU6-P-0.06, CpSent-R-SU6-R-0.18, CpSent-R-SU6-F-0.08, CpSent-R-W-1.2-P-0.12, CpSent-R-W-1.2-R-0.22, CpSent-R-W-1.2-F-0.14
Ravindran and Saravanan M (2010) [80] [82]	CRF, Features, Term distribution	keralawyer.com	MAP, F-Measure, ROUGE-1, 2	MAP-0.64, F-0.65, R-1-0.68, R-2-0.41
Kumar et al. (2012) [57]	LDA	keralawyer.com	Precision, Recall, F-Measure	I-P-0.60, I-R- 0.58, I-F-0.59, R-P-0.58, R-R-0.56, R-F-0.57, M-P-0.55, M-R-0.54, M-F-0.54, N-P-0.52, N-R-0.51, N-F-0.51, S-P-0.53, S-R-0.55, S-F-0.54

Table 4: **Legal Text Summarization**

4.4.1 Discussions

Grover and Claire [35] used a number of NLP techniques for legal text processing: POS tagging, noun and verb group chunking, clause boundary, main verb and tense identification on the HOLJ corpus. The performance of main verb and tense identification is good in all the metrics, however they did not do summarization.

Grover and Hachey [36, 40] used classifiers like C4.5, NB, Winnow, SVM with different features to classify sentences of HOLJ corpus for summarization.

The authors evaluated performance of the summarizer with human generated baselines.

Hachey et al. in a series of work [41–45] used different classifiers like NB, ME, PL, SEQ for sentence classifications. In [42] they performed experiments with NB and ME on HOLJ corpus and evaluated with point-biserial correlation coefficients also. In all the above, ME sequence model with rhetorical categorization gave better results.

Farzindar et al. [20] used LetSum tool on Federal Court of Canada corpus. Evaluation with ROUGE-1 gave high score.

Galgani et al. [25,26] used KB and citation based methods on AustLII data set. The citation based methods gave best scores.

Ravindran and Saravanan [80,82], and Kumar et al. [57] implemented CRF with numerous features including term distribution on court judgments from keralawyer.com. The methods of Ravindran and Saravanan [80, 82] showed better performance compared to [57].

4.5 Highlights of Different Techniques

Below we discuss some of the advantages of the different methodologies in legal text summarization.

- C4.5 (Grover and Hachey (2004) [36,40]):This classifies sentences with good accuracy and assigns the sentences to appropriate rhetorical roles.
- ME sequence model (Hachey et al. (2004-2005) [41–45]): This method takes sequence of unlabeled observations and anticipate their corresponding labels.
- Graph based method (Kim et al. (2013) [53]): This method determines compression rate automatically to every document as each individual document compression rate can be different. The graph for a document is an unconnected graph (set of connected graphs) ensures diversity of topic so that summaries have high cohesion.
- Graph based method (Schilder et al. (2013) [84]): This method produces an ordered list of paragraphs according to their importance and also extracts the sentences that is most similar to query.
- Letsum (Farzindar et al. (2004) [20]): This tool uses thematic structure which improves coherency and readability of the text and linguistic markers helps to identify important sentences in a judgment.
- NB with surface, emphasis, content features (Yousfi-Monod et al.(2010) [98]): The surface feature extracts relevant legal content. The emphasis feature identifies some of the leading words in a sentences. The content feature identifies some of the important words, which are generally relevant in the summary.
- KB (Galgani et al. (2012) [26]): This method automatically extracts catch-phrases that specifies which sentences should be extracted from text.

– Citation based method (Galgani et al. (2012) [25]): This approach uses
cited and citing cases to extract catch phrases as these phrases present
important legal point of a case.
– CRF (Ravindran and Saravanan (2010) [80] [82]): This method is used for
text segmentation of legal judgments and term distribution model identifies
term patterns and frequencies of the terms so that important sentence are
extracted.
– LDA (Kumar et al. (2012) [57]): This method generates set of topics and
these topic are used as a basic elements for summarization by extracting
topic related sentences from the given document.

4.6 Multi-document summarization

Objective of the multi-document summarization can be to generate a summary
from multiple documents that cover similar information. Legal judgments are
complex in nature. To the best of our knowledge, there is no literature on
multi-document summarization in legal domain. One of the possible reasons
could be difficulty of tracking the presence of topics. Identification of similar
topics can best be done manually, otherwise need to be found using automatic
clustering. Absence of such data in the legal domain is the reason of absence
of any work on multi-document summarization.

5 Legal Text Summarization Tools

A number of software tools have been developed by different academic or
corporate research groups for legal text processing. Below are some automatic
summarizers implementing different models and approaches discussed earlier.

LetSum: Farzindar et al. [19] describe LetSum system, which decides the
thematic structure of a judgment in four themes *introduction*, *context*, *juridical
analysis* and *conclusion* and identifies the relevant sentences for each theme.
The purpose of the tool is to summarize the legal record of the proceedings
of federal courts in Canada and presenting as table-style summary for the re-
quirement of lawyers and experts in the legal domain. The most important
units extraction is based on the identification of the thematic structure within
the document and determining the argumentative themes in the judgment.
(Farzindar et al. [20]). The summary is generated using four steps : thematic
segmentation is to detect legal document structure, filtering to eliminate unim-
portant quotations and noises, selection of the candidate units and production
of structured summary.

FLEXICON: The FLEXICON project developed by Smith et al. [87]
generate a summary of legal cases by using IR based on location heuristics,
occurrence frequency of index terms, and the use of indicator phrases. A term
extraction module identifies concepts, case citations, statute citations, and fact
phrases which guides to the generation of a document profile. This project was
developed for the decision reports of Canadian courts.

SALOMON: The SALOMON project developed by Uyttendaele et al. [92] for the automatic processing of legal texts. The main aim is to summarize the Belgian criminal cases automatically which can help access large number of cases. These techniques are developed to reduce the work of lawyers. The SALOMON developed with dual methodology: one, applying additional knowledge and features; and the other, using occurrence statistics of index terms. It attains initial categorization and structuring of the cases and afterwards most relevant text units from the alleged offences and the opinion of the court are extracted.

HAUSS (Hybrid AUtomatic Summarization System): Galgani et al. [27] describe a summarizer which combines different methods into a single approach. The relevance term extraction is based on Cue Terms, Frequent Pattern, Legal Terms and Fcfound score. They build rules that combine different type of features to extract sentences. They consider term level attributes and sentence level attributes to build rules. The term level attributes used are Term frequency, AvgOcc, Documentfrequency, TFIDF, CpOcc, Fcfound, CitSen, CitCp, CitAct, pos tag, proper noun, Cue terms, Legal terms. The sentence level attributes are Find, Position, HasCitCase, CommonCitCp, CommonCitSen, HasCitLaw, CommonCitLaw, AnyPattern, Patterns. Term level attributes specify conditions which pertain to single terms and the sentences which are made of such terms and satisfy such conditions are extracted. Sentence level attributes are conditions based on a single constraint on the entire sentence. The created rules blend frequency, centrality, citation and linguistic information in a context-dependent way. The creation of these rules gives an incremental knowledge acquisition framework, utilizing a training corpus to guide rule acquisition, and create a KB precise to the domain. They created a KB for catchphrase extraction in legal text. HAUSS KB on set of 5-sentence obtains highest precision of 0.765, 0.486 with threshold 0.5 and 0.7 respectively. HAUSS single condition KB on set of 5-sentence obtains highest recall and f-measure of 0.579 ,0.624 with threshold 0.5 and 0.305, 0.342 with threshold 0.7 respectively.

Other than the above initiatives from academic community, a corporate organization called 'NLP Technologies' [16] has also developed a number of tools for the judicial domain. The company is engaged in research, development and provides software tools and services. The company came up with summarization tool called DecisionExpress.

DecisionExpress: Based on the work of Farzindar [17] NLP Technologies has developed a summarization system for the legal domain based on a thematic segmentation of the text. They distinguish four themes *introduction*, *context*, *reasoning*, and *conclusion*. The tool divides the legal decisions into thematic segments based on the work of judge Mailhotet al. [67]. The system describes information like name of the judge who has signed the judgment and what kind of tribunal he/she pertains to, which domain the law belongs to and what is the subject of information (for example, immigration and application for permanent residence), a brief description of the litigated point, conclusion of the judge's (allowed or dismissed), hyperlinks to the summary

and the original judgment. The judgments are automatically translated into Canadian languages (French or English) allows legal professionals to work in their preferred language irrespective of the language in which the decision was published. Some of the legal summarization tools listed in the Table 5.

Tools	Approach	Features
LetSum	Linguistic features	Table style summary
SALOMON	Covering clustering algorithm, K-mediod mehtod	Strctured represenataion
HAUSS	KB	Diversity, Context tailored, Purpose specific
DecisionExpress	Specific knowledge of the legal field	Consistency, Cost reduction, Bilingual information exaction

Table 5: **Legal text summarization tools**

6 Conclusion

Text summarization has been one of the active areas of research for about last two decades. Even though substantial amount of work has been done in the area of extraction based summarization, automatic generation of abstractive summaries from text documents is comparatively a new area which is not explored in much detail and depth. As far as legal text is concerned, often the documents are quite long and different from text of other genres. The need for automatic summarization of legal documents and other text processing has been felt for long, but focus of computer scientists has come only very recently. In this survey, we have attempted to give an account of different text summarization techniques giving special emphasis on the legal text summarization. We started with general definition of text summarization with a brief account of the recent works in the domain so that an unfamiliar reader can readily relate to the techniques that are used in legal text summarization. Gradually we discussed different legal text summarization tools used in the domain. Specifically, we covered some state of the art summarization techniques, the datasets and metrics they used, their performances and a comparative study. The same treatment was then followed for legal text summarization with techniques based on different categories like features, graph, rhetorical roles and classification. The summarization of legal text has some domain-specific issues like the structure of the documents, different terminologies and evaluation criteria for the task. Even though sometimes scores achieved in the legal text summarization are comparable and competitive to those in general domain counterpart, they are not consistent across the datasets. Unless some more research is done, it is difficult to have a comparative analysis. Also, the summarization techniques we were able to find in the literature are only extraction-based, coming from single documents only. It is imperative to

explore whether abstractive summarization is possible or not in the legal domain. Also, although it is a far cry, there is certainly a need to have automatic categorization of the similar court cases and their verdicts. Multi-document summarization from the similar cases can provide the legal practitioners a brief yet holistic view of a particular type of court-cases or a quick chronological account of important milestones in a single case. There are a number of areas along with a plenty of issues therein where information science community can explore as far as legal text summarization is concerned.

References

1. Abuobieda, A., Salim, N., Kumar, Y.J., Osman, A.H.: An improved evolutionary algorithm for extractive text summarization. In: Intelligent information and database systems, pp. 78–89. Springer (2013)
2. Abuobieda, A., Salim, N., Kumar, Y.J., Osman, A.H.: Opposition differential evolution based method for text summarization. In: Intelligent Information and Database Systems, pp. 487–496. Springer (2013)
3. Alliheedi, M., Di Marco, C.: Rhetorical figuration as a metric in text summarization. In: Advances in Artificial Intelligence, pp. 13–22. Springer (2014)
4. Batcha, N.K., Aziz, N.A., Shafie, S.I.: Crf based feature extraction applied for supervised automatic text summarization. Procedia Technology **11**, 426–436 (2013)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research **3**, 993–1022 (2003)
6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems **30**(1), 107–117 (1998)
7. Bun, K.K., Ishizuka, M.: Topic extraction from news archive using tf* pdf algorithm. In: Web Information Systems Engineering, International Conference on, pp. 73–73. IEEE Computer Society (2002)
8. Cabral, L.d.S., Lins, R.D., Mello, R.F., Freitas, F., Ávila, B., Simske, S., Riss, M.: A platform for language independent summarization. In: Proceedings of the 2014 ACM symposium on Document engineering, pp. 203–206. ACM (2014)
9. Chen, J., Zhuge, H.: Summarization of scientific documents by detecting common facts in citations. Future Generation Computer Systems **32**, 246–252 (2014)
10. Cilibrasi, R.L., Vitanyi, P.: The google similarity distance. Knowledge and Data Engineering, IEEE Transactions on **19**(3), 370–383 (2007)
11. Cohan, A., Goharian, N.: Revisiting summarization evaluation for scientific articles. arXiv preprint arXiv:1604.00400 (2016)
12. Compton, P., Jansen, R.: Knowledge in context: A strategy for expert system maintenance. Springer (1990)
13. Das, D., Martins, A.F.: A survey on automatic text summarization. Literature Survey for the Language and Statistics II course at CMU **4**, 192–195 (2007)
14. Erkan, G., Radev, D.R.: Lexrank: graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research pp. 457–479 (2004)
15. Ermakova, L.: Automatic summary evaluation. rouge modifications. In: VI   (RuSSIR2012) (2012)
16. Farzindar, Hosseiny, M.: Nlptechnologies, url=http://www.nlptechnologies.ca/en/nlp-technologies-services-ans-solutions, urldate=2016-08-17
17. Farzindar, A.: Résumé automatique de textes juridiques
18. 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. Citeseer (2004)
19. Farzindar, A., Lapalme, G.: Letsum, an automatic legal text summarizing system. Legal knowledge and information systems, JURIX pp. 11–18 (2004)

20. Farzindar, A., Lapalme, G.: The use of thematic structure and concept identification for legal text summarization. Computational Linguistics in the North-East (CLiNE2002) pp. 67–71 (2004)
21. Fattah, M.A., Ren, F.: Automatic text summarization. World Academy of Science, Engineering and Technology **37**, 2008 (2008)
22. Ferreira, R., Freitas, F., de Souza Cabral, L., Dueire Lins, R., Lima, R., França, G., Simskez, S.J., Favaro, L.: A four dimension graph model for automatic text summarization. In: Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on, vol. 1, pp. 389–396. IEEE (2013)
23. Ferreira, R., de Souza Cabral, L., Freitas, F., Lins, R.D., de França Silva, G., Simske, S.J., Favaro, L.: A multi-document summarization system based on statistics and linguistic treatment. Expert Systems with Applications **41**(13), 5780–5787 (2014)
24. Ferreira, R., de Souza Cabral, L., Lins, R.D., e Silva, G.P., Freitas, F., Cavalcanti, G.D., Lima, R., Simske, S.J., Favaro, L.: Assessing sentence scoring techniques for extractive text summarization. Expert systems with applications **40**(14), 5755–5764 (2013)
25. Galgani, F., Compton, P., Hoffmann, A.: Citation based summarisation of legal texts. In: PRICAI 2012: Trends in Artificial Intelligence, pp. 40–52. Springer (2012)
26. 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. Association for Computational Linguistics (2012)
27. Galgani, F., Compton, P., Hoffmann, A.: Hauss: Incrementally building a summarizer combining multiple techniques. International Journal of Human-Computer Studies **72**(7), 584–605 (2014)
28. García-Hernández, R.A., Ledeneva, Y.: Single extractive text summarization based on a genetic algorithm. In: Pattern recognition, pp. 374–383. Springer (2013)
29. Gawryjolek, J.J.: Automated annotation and visualization of rhetorical figures (2009)
30. Ghalehtaki, R.A., Khotanlou, H., Esmaeilpour, M.: A combinational method of fuzzy, particle swarm optimization and cellular learning automata for text summarization. In: Intelligent Systems (ICIS), 2014 Iranian Conference on, pp. 1–6. IEEE (2014)
31. Goldstein, J.: Automatic text summarization of multiple documents (1999)
32. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 19–25. ACM (2001)
33. Gross, O., Doucet, A., Toivonen, H.: Document summarization based on word associations. In: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, pp. 1023–1026. ACM (2014)
34. Group, E.L.T.: fsgmatch, url=https://files.ifi.uzh.ch/cl/broder/tttdoc/c385.htm, urldate=2016-08-17
35. Grover, C., Hachey, B., Hughson, I., Korycinski, C.: Automatic summarisation of legal documents. In: Proceedings of the 9th international conference on Artificial intelligence and law, pp. 243–251. ACM (2003)
36. Grover, C., Hachey, B., Hughson, I., et al.: The holj corpus: Supporting summarisation of legal texts. In: Proceedings of the 5th international workshop on linguistically interpreted corpora (LINC-04) (2004)
37. Grover, C., Hachey, B., Korycinski, C.: Summarising legal texts: Sentential tense and argumentative roles. In: Proceedings of the HLT-NAACL 03 on Text summarization workshop-Volume 5, pp. 33–40. Association for Computational Linguistics (2003)
38. Grover, C., Matheson, C., Mikheev, A., Moens, M.: Lt ttt-a flexible tokenisation tool. In: LREC (2000)
39. Gupta, V.: A language independent hybrid approach for text summarization. In: Emerging Trends in Computing and Communication, pp. 71–77. Springer (2014)
40. Hachey, B., Grover, C.: A rhetorical status classifier for legal text summarisation. In: Proceedings of the ACL-2004 Text Summarization Branches Out Workshop (2004)
41. Hachey, B., Grover, C.: Sentence classification experiments for legal text summarisation. In: Proceedings of the 17th Annual Conference on Legal Knowledge and Information Systems (Jurix) (2004)

42. Hachey, B., Grover, C.: Automatic legal text summarisation: experiments with summary structuring. In: Proceedings of the 10th international conference on Artificial intelligence and law, pp. 75–84. ACM (2005)

43. Hachey, B., Grover, C.: Sentence extraction for legal text summarisation. In: International Joint Conference on Artificial Intelligence, vol. 19, p. 1686. Lawrence Erlbaum Associates Ltd (2005)

44. Hachey, B., Grover, C.: Sequence modelling for sentence classification in a legal summarisation system. In: Proceedings of the 2005 ACM symposium on Applied computing, pp. 292–296. ACM (2005)

45. Hachey, B., Grover, C.: Extractive summarisation of legal texts. Artificial Intelligence and Law **14**(4), 305–345 (2006)

46. Hamid, F., Tarau, P.: Text summarization as an assistive technology. In: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments, p. 60. ACM (2014)

47. Hao, J.K.: Memetic algorithms in discrete optimization. In: Handbook of memetic algorithms, pp. 73–94. Springer (2012)

48. Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., Nagata, M.: Single-document summarization as a tree knapsack problem. In: EMNLP, pp. 1515–1520 (2013)

49. John, G.H., Langley, P.: Estimating continuous distributions in bayesian classifiers. In: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pp. 338–345. Morgan Kaufmann Publishers Inc. (1995)

50. Jones, K.S., et al.: Automatic summarizing: factors and directions. Advances in automatic text summarization pp. 1–12 (1999)

51. Kavila, S.D., Puli, V., Raju, G.P., Bandaru, R.: An automatic legal document summarization and search using hybrid system. In: Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), pp. 229–236. Springer (2013)

52. Kikuchi, Y., Hirao, T., Takamura, H., Okumura, M., Nagata, M.: Single document summarization based on nested tree structure. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 2, pp. 315–320 (2014)

53. Kim, M.Y., Xu, Y., Goebel, R.: Summarization of legal texts with high cohesion and automatic compression rate. In: New Frontiers in Artificial Intelligence, pp. 190–204. Springer (2013)

54. Kipper, K., Dang, H.T., Palmer, M., et al.: Class-based construction of a verb lexicon. In: AAAI/IAAI, pp. 691–696 (2000)

55. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM) **46**(5), 604–632 (1999)

56. Krishna, R., Kumar, S.P., Reddy, C.S.: A hybrid method for query based automatic summarization system. Int J Comput Appl Comput Appl **68**, 39–43 (2013)

57. Kumar, R., Raghuveer, K.: Legal document summarization using latent dirichlet allocation

58. Kumar, Y.J., Salim, N., Abuobieda, A., Albaham, A.T.: Multi document summarization based on news components using fuzzy cross-document relations. Applied Soft Computing **21**, 265–279 (2014)

59. L. Cabral, R.L., Lima, R., Ferreira, R., Freitas, F., Silva, G., G.Cavalcanti, e, S.S., Favaro, L.: A hybrid algorithm for automatic language detection on web and text documents,. In: 11th IAPR International Workshop on Document Analysis Systems,Tours-Loire Valley, France (2014)

60. Ledeneva, Y., García-Hernández, R.A., Gelbukh, A.: Graph ranking on maximal frequent sequences for single extractive text summarization. In: Computational Linguistics and Intelligent Text Processing, pp. 466–480. Springer (2014)

61. Lee, S., Belkasim, S., Zhang, Y.: Multi-document text summarization using topic model and fuzzy logic. In: Machine Learning and Data Mining in Pattern Recognition, pp. 159–168. Springer (2013)

62. Lee, S., Kim, H.j.: News keyword extraction for topic tracking. In: Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on, vol. 2, pp. 554–559. IEEE (2008)

63. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Text summarization branches out: Proceedings of the ACL-04 workshop, vol. 8. Barcelona, Spain (2004)

64. Littlestone, N.: Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. In: Foundations of Computer Science, 1987., 28th Annual Symposium on, pp. 68–77. IEEE (1987)

65. Lloret, E., Palomar, M.: Text summarisation in progress: a literature review. Artificial Intelligence Review **37**(1), 1–41 (2012)

66. Ma, Y., Wu, J.: Combining n-gram and dependency word pair for multi-document summarization. In: Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on, pp. 27–31. IEEE (2014)

67. Mailhot, L., Carnwath, J.D.: Decisions, Decisions–: A Handbook for Judicial Writing. Cowansville, Québec: Éditions Y. Blais (1998)

68. Marcus, M.P., Marcinkiewicz, M.A., Santorini, B.: Building a large annotated corpus of english: The penn treebank. Computational linguistics **19**(2), 313–330 (1993)

69. Mendoza, M., Bonilla, S., Noguera, C., Cobos, C., León, E.: Extractive single-document summarization based on genetic operators and guided local search. Expert Systems with Applications **41**(9), 4158–4169 (2014)

70. Mihalcea, R., Tarau, P.: Textrank: Bringing order into texts. Association for Computational Linguistics (2004)

71. Mikheev, A.: Automatic rule induction for unknown-word guessing. Computational Linguistics **23**(3), 405–423 (1997)

72. Miranda-Jiménez, S., Gelbukh, A., Sidorov, G.: Summarizing conceptual graphs for automatic summarization task. In: Conceptual Structures for STEM Research and Education, pp. 245–253. Springer (2013)

73. Nenkova, A., McKeown, K.: A survey of text summarization techniques. In: Mining Text Data, pp. 43–76. Springer (2012)

74. Pal, A.R., Saha, D.: An approach to automatic text summarization using wordnet. In: Advance Computing Conference (IACC), 2014 IEEE International, pp. 1169–1173. IEEE (2014)

75. Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines (1998)

76. Plaza, L.: Comparing different knowledge sources for the automatic summarization of biomedical literature. Journal of biomedical informatics **52**, 319–328 (2014)

77. Press Information Bureau, G.o.I.: Cases Pending in High Courts and Supreme Court, url=http://pib.nic.in/newsite/erelease.aspx?relid=73624, urldate=2015-07-10

78. Quinlan, J.R.: C4. 5: programs for machine learning. Elsevier (2014)

79. Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Celebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., et al.: Mead-a platform for multidocument multilingual text summarization (2004)

80. Ravindran, B., et al.: Automatic identification of rhetorical roles using conditional random fields for legal document summarization (2008)

81. Samei, B., Samei, B., Estiagh, M., Eshtiagh, M., Keshtkar, F., Hashemi, S., Hashemi, S.: Multi-document summarization using graph-based iterative ranking algorithms and information theoretical distortion measures. In: The Twenty-Seventh International Flairs Conference (2014)

82. Saravanan, M., Ravindran, B.: Identification of rhetorical roles for segmentation and summarization of a legal judgment. Artificial Intelligence and Law **18**(1), 45–76 (2010)

83. Saravanan, M., Ravindran, B., Raman, S.: Improving legal document summarization using graphical models. Frontiers in Artificial Intelligence and Applications **152**, 51 (2006)

84. Schilder, F., Molina-Salgado, H.: Evaluating a summarizer for legal text with a large text collection. In: 3 rd Midwestern Computational Linguistics Colloquium (MCLC). Citeseer (2006)

85. Sharma, A.D., Deep, S.: Too long-didn't read a practical web based approach towards text summarization. In: Applied Algorithms, pp. 198–208. Springer (2014)

86. Sivanandam, S., Deepa, S.: Introduction to genetic algorithms. Springer Science & Business Media (2007)

87. Smith, J., Deedman, C.: The application of expert systems technology to case-based law. In: ICAIL, vol. 87, pp. 84–93 (1987)
88. Sowa, J.F.: Conceptual structures: information processing in mind and machine (1983)
89. Teufel, S., Moens, M.: Sentence extraction as a classification task. In: Proceedings of the ACL, vol. 97, pp. 58–65 (1997)
90. Teufel, S., Moens, M.: Summarizing scientific articles: experiments with relevance and rhetorical status. Computational linguistics **28**(4), 409–445 (2002)
91. Turtle, H.: Text retrieval in the legal world. Artificial Intelligence and Law **3**(1-2), 5–54 (1995)
92. Uyttendaele, C., Moens, M.F., Dumortier, J.: Salomon: automatic abstracting of legal cases for effective access to court decisions. Artificial Intelligence and Law **6**(1), 59–79 (1998)
93. Vodolazova, T., Lloret, E., Muñoz, R., Palomar, M.: The role of statistical and semantic features in single-document extractive summarization. Artificial Intelligence Research **2**(3), p35 (2013)
94. Wang, T., Chen, P., Simovici, D.: A new evaluation measure using compression dissimilarity on text summarization. Applied Intelligence **45**(1), 127–134 (2016)
95. Wang, Y., Ma, J.: A comprehensive method for text summarization based on latent semantic analysis. In: Natural Language Processing and Chinese Computing, pp. 394–401. Springer (2013)
96. wikipedia: district_courts, Legal Domain, url=http://en.wikipedia.org/wiki/List_of _district_courts_of_India, urldate=2015-07-10
97. wikipedia: High_Courts, Legal Domain, url= http://en.wikipedia.org/wiki/List_of _High_Courts_of_India, urldate=2015-07-10
98. Yousfi-Monod, M., Farzindar, A., Lapalme, G.: Supervised machine learning for summarizing legal documents. In: Advances in Artificial Intelligence, pp. 51–62. Springer (2010)