EMITTER International Journal of Engineering Technology
Vol. 6, No. 1, June 2018
ISSN: 2443-1168

Automatic Summarization Task for News Article

Afrida Helen

Politeknik Elektronika Negeri Surabaya
Kampus PENS, jln Raya ITS, Sukolilo Surabaya
helen@pens.ac.id

Abstract

Understanding the contents of numerous articles requires strenuous effort. While manually reading the summary or abstract is one way, automatic summarization offers more efficient way in doing so. The current research in automatic summarization focuses on the statistical method and the Natural Processing Language (NLP) method. Statistical method produces Extractive Summary that the summaries consist of independent sentences considered important content of document. Unfortunately, the coherence of the summary is poor. Besides that, the Natural Processing Language expected can produces summary where sentences in summary should not be taken from sentences in the document, but come from the person making the summary. So, the summaries closed to human-summary, coherent and well structured. This research proposed Extractive summarization for news article about Coruption in Indonesia. We use five classes of important word/ phrase and make them in one sentence as summary. We find that there are still opportunities to develop better outcomes that are better coherence and better accuracy

Keywords: Abstractive, Extractive, Statistic, Natural Processing Language, News Article.

1. INTRODUCTION

People are now depending on the Internet in searching for News Article. By using existing tools, such as Google, Yahoo, Bing and so forth, the documents are easy to obtain, in large quantities, and coming from various sources. To understand the content of the documents quickly, the readers should read the summary (or abstract) of the documents. For summary usually contain only the important sentences in short form, and represent the contents of the news article as a whole.

The summary system was built by Luhn [1] in the late 1950s. He built abstracts of scientific articles, which are placed on the top position in the scientific paper format, so that the readers have a choice to understand the contents of a scientific article through the Abstract. The Abstract is written by

the author manually, has few amount sentences and contains only important sentence from the paper.

The problem occours when the amount of the articles is numerous. The readers obviously need more times to read and understand, despite of the summary system in each document. Therefore, the need of more advance summary system increases.

An automated summary system is one if the advance. It is generated from a collection of important sentences from a paper, whereas important sentences are constructed from important word or phrase. The system starts from marking the important words with the degree of occurrence of words or phrase in entire of articles. Sentences are ranked based on the frequency of word occurrences. Top rank of the sentence list becomes summary sentence. Then the technique to choose important sentence evolve, the characteristic of sentence involves the position of the sentence, the length of the sentence and so on. These characteristics make the accuracy of summary increase. The common processes of summarization as shown in Fig. 1. However, this technique is low in coherence because the content of each sentence in the summary has no relation.

This weakness triggers the researchers to develop new method like human-summaries. The main characteristic of human-summary sentences are having good relationship with each other, good sentence structure, and using rhetorical move from the beginning to the end of the summary. Researchers developed the abstractive concepts that resemble human-summary. This study looks at new opportunities to make summaries results more coherent. Several tasks will be discussed in Section 2, followed by the developed method in developing Summary in Section 3. Section 4 will discuss Experiment and Result. Finally, the last Section will explain the conclusion of this study and about the opportunity of the research development.

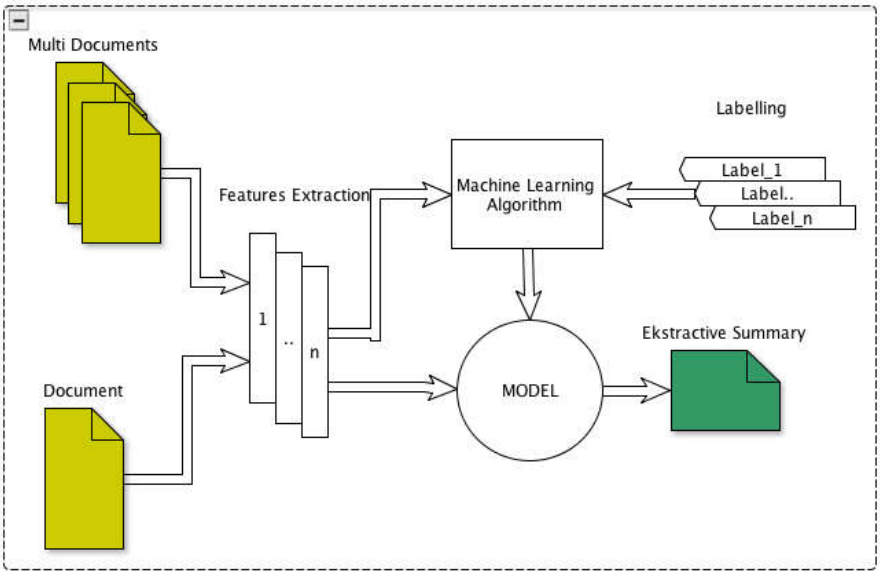


Fig 1. The Process for Generating Summarization Automatically

2. ABSTRACTIVE SUMMARIZATION

Technology evolves along with needs. Similarly, the need for better summary results has led researchers to investigate the possibilities for automatic summary. Summaries made by humans have abstraction values, where sentences in summary are new sentences that essentially convey what is contained in the original article. The automatically generated abstractive summaries are expected to be better and closer to the human-built summary. Abstractive summary requires a deeper analysis of the text and the ability to generate new sentences, which provide an obvious advantage to improve the accuracy of a summary, reducing its redundancy and keeping a good compression rate. The tasks for generating abstractive summaries [2] are 1) sentence Compression that removes peripheral information from a sentence to shorten summary, 2) sentence Fusion, that merge information from multiple sentences and reduces redundancy in summary and 3) reorganization. Sentences to make the summary coherent. Most Abstractive summary task uses Natural Language Processing. The natural language is a task for parsing of sentences in the text. Parsing is identification the Parts of Speech of each word and the grammatical relations between the words in the sentence. The output of parsing is usually a tree of grammatical relations and dependencies between the words in the text. These trees are called dependency trees.

2.1 Sentence Compression

Sentence compression can be broadly described as the task of creating a grammatical summary of a single sentence with minimal information loss. The task requires a quantity called inter-phrase dependency strength. In the training process, original sentences are parsed. The number of tokens is counted for each pair of phrases, and connected with each other by a dependency path of certain length. The statistics is used to estimate the interphrase dependency strength required in the sentence compression process.

Prior tasks mostly use Decision Tree algorithm [3], random forest, and gradient boosting algorithms to analyze the data. Madanapali's {1} research used Intersection Algorithm to align paired sentences and a swallow parser to combine the sentences. The sentence was mapped into the structure of the predicate, get news content of phrase, and compare the predicate. Phrases that contain general information were selected, sorted, and added with some entities. Those were combined and arranged to generate the Summary. The results could be concluded using a tree that can improve the quality of the language of the summary significantly, and simultaneously minimize repetition. Unfortunately this approach did not involve the context of text when exchange between sentences.

Furthermore, Genest and G. Lapalme [2] proposed a short summary and a good abstract from several articles in the same topic. This scheme was the extraction of information based on heuristic content selection and one or

more rules to generate sentences. Each abstraction is tailored to the topic or sub category. When the rule was raised, multiple verbs and nouns that have the same meaning were omitted and the position of the rule was identified. The extraction process found several candidate rules on each topic of each category. Based on information extraction module, the content module selected the best rule candidate from each category and passed it to the summary generation module. The best candidates of the content were selected using a single rule that extracted and adapted to one or two categories of rule, and then updated the rule to construct a summary sentence. This study had the potential to make a summary with more information. However, the method was purely using a rule written manually that takes a lot of time.

Another research was conducted by Trevor C and Mirella L [4]. The task was simply deleting word and rewrite sentence using additional such as, recording, insertion. They developed grammar rule for given source to set of possible output. Applying a series of grammar rules created each rule. Where each rule match a fragment of the source and creates a fragment of the target tree. A rule in the grammar consisted of a pair of elementary trees and a mapping between the variables in both trees. A derivation was a sequence of rules yielding a target tree with no remaining variables. Each grammar rule could assign a weigh. These weigh are learnt in discriminative training, find to set of related sentences target for given source sentence and create output.

2.2 Sentence Fusion

We cannot separate the task of compression and fusion when constructing a summary. Both of them have to be done in order to generate a summary.

When a sentence has been compressed which generates two or more sentences that have the same intent, the fusion technique is required to combine the two sentences or more and then discard the repetitive words. The heuristic technique used is very closed to a number of rules because the most appropriate approach is to find a number of rules based on existing examples. Similarly, the fusion task, the rule is extracted using the language pattern, then is used to identify the piece of text. The Dependency trees are used in this task. They play a prominent role in sentence fusion. Sentences, which will be fused, are represented at first as dependency trees. These trees are merged in to a tree and converted to a sentence that is known as the fused sentence.

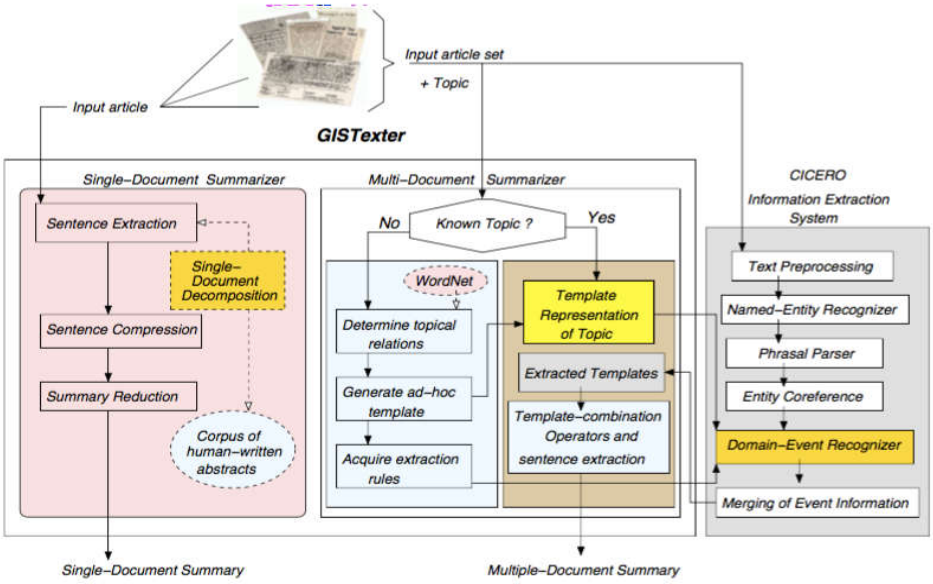


Fig 2. The architecture of Gistexter [5]

However, this study has not yielded satisfactory results. S. M. Harabagiu and F. Lacatusu, [5] built a summary of several news documents with the same topic and call GISTEXTER. They used Text snippets to generate information coherence using a summary algorithm. Significant result of the research was the high level of coherence of the summary result. On the other hand, this research could not work with a large number of documents, only in one single source.

Figure 2 show the architecture of Gistexter that is developed by Harabagieu et al. This figure is devided into three tasks. The first task is a common process for building Abstract by Human. This Abstract will be a corpus and will be used for learning. Second task is multi document summarization process. The process can be distinguished, 1) if the input document has not a template yet in corpus, and 2) if the input document has a template in corpus. The process continues to extract the sentences to build a summary. Otherwise, the System automatically generated an ad-hoc template to acquire rules. The third task combines the rule of linguist extraction pattern with co-reference knowledge to produce a good quality of summary. Texts snipped are used to generate information coherence using a summary algorithm. Significantly development of this research is the coherence highly enough. On the other hand, this research only works in a single source document, could not handle if the source in multi document.

Furthermore, Barzilay [6] presented research on the automatic fusion of same document topic. The method for summarizing was a specific type of input: 1) news articles presenting different descriptions of the same event, 2) a content planner selects and orders propositions from an underlying knowledge base to form text content, 3) a sentence planner determines how to combine propositions into a single sentence, and a sentence generator

realizes each set of combined propositions as a sentence, mapping and building syntactic structure. The content planner found an intersection of phrases by comparing the predicate-argument structures. This process selected the phrases that able to mention the common information of the topic, order them, and augment them with information needed for clarification. The next step of generating sentence begun with phrases. The task was to produce fluent sentences that combine these phrases by arranging them in new contexts. In this process, new grammatical constraints may be imposed and paraphrasing may be needed. Redundant statement in a summary is selected by one sentence from the set of similar sentences. Therefore, need to intersect the topic sentences to identify the common phrases and then generate a new sentence. Phrases, which were produced by topic intersection, will form the content of the generated summary. Then, matching the fact was done to identify similarities between phrases instead of identifying words. If paraphrasing rules are known, the predicate-argument structure of the sentences can be compared and common parts are found. Paraphrasing pattern is obtained from studying corpus then used for intersection algorithm.

2.3 Reorganization and Revision

Reorganizing and Revising a sentence needs to be done for the abstractive summary to obtain coherency. Jing and McKeown (7) found that human summarization can be traced back to cut-and-paste operations of a text and proposed a revision method consisting of sentence reduction and combination modules with a sentence extraction part.

Hideki Tanaka et al [8] method did not use the coreference relation of noun phrases (NPs), but rather insertion and substitution of the phrases to modify the same head chunk in lead and other sentences. It addressed the problem of revising the lead sentence in a news text to increase the amount of the key information. For analyzing, the method suggested to devise the lead sentence revision algorithm and present the outline. The syntactical analyzed are, 1) Trigger search for the same chunks in the lead and body sentences, 2) Phrase alignment that identify the maximum phrases of each trigger of which phrases are aligned according to a similarity metric, 3) Substitution if a body phrase has a corresponding phrase in the lead. The body phrase was richer in information, so the method substituted the body phrase for the lead phrase, and finally 4) Insertion if a body phrase has no counterpart in the lead that the phrase is floating.

The method inserts and substitutes any type of phrase that modifies the trigger and therefore had no limitation in syntactic type. Although NP elaborates, there are other useful syntactic types for revision. Khodra.M.L et al [9] evoked a summary with 15 templates for sentences from one scientific paper, and extracted seven features using statistical methods. They proposed summary coherence by substituting certain words such as subject, active verb, passive verb, phrase substitution, and discarding unimportant phrases

called Surface Repair. Manually evaluated, readers are quite satisfied with the results of the summary.

3. EXTRACTIVE SUMMARIZATION

Extractiive Summarization is the summary that contains a collection of importance sentence from an article. While the content of summary based on reader requirement, the collection of sentences should be accordance with the reader request. It is very important to know the characteristic of the sentences that fill in the summary. The characteristic of the sentence called Feature. This technique simpler and more dependent on its Features.

3.1 Our Proposed Method

Some summarization researchs have developed by researcher. Many methods have done to increase the precision. Starting from using a sentence summary strategy by ranking the words that often appear. Then another technique develops, adding features that can recognize these important sentences. There are three categories of feature that common used, First, Entity feature is a feature that finds the correct representation of a sentence that is required. Second, Lexical feature is a feature that involves lexical (word term) provides better performance than other types of features. This feature is rich in words that involve every word term. Unfortunately the words will lose meaning when the feature come from the phrase beheading. Third, Syntax feature is a feature that can keep the syntax of sentences consisting of a collection of words. Word/ phrase requires special features to be recognized. This study proposed Extractive summarization using two categories of feature i.e. Entity Feature and Lexical Feature. The list of our features shown in Table 1.

Table 1. List of the Feature

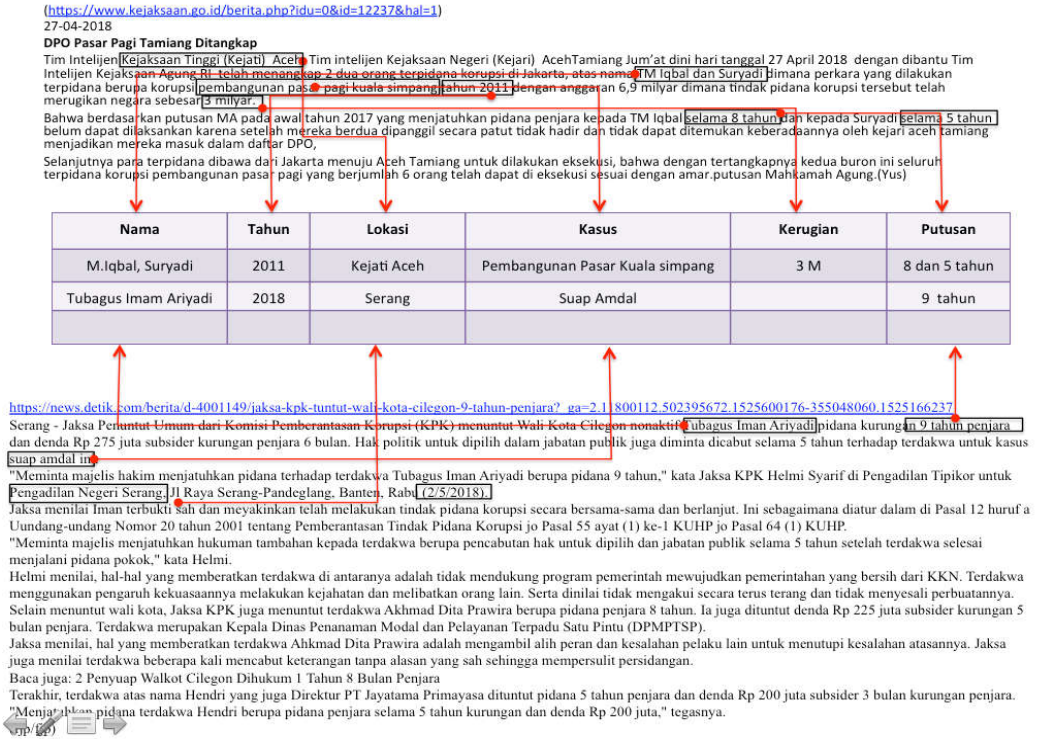
No		Feature	Description	Value
1	Entity	isNumeric	There is numbers in the sentence.	0, 1
2		isCapital	There is conjunction word egg. “karena” ,”sebab” ,	0, 1
3		isCurrency	Rupiah, Rp, Dolar, $.	0, 1
4		isMonth	The name of the Month	0, 1
5		isCity,	There is a word or sequence word beginning with	0, 1
6		indicateW	‘Korupsi’, ‘suap’, ‘gratifikasi’, ‘OTT ‘, ‘kasus’,	0, 1
7		isYear	There is preposition word eg. "di", "pada", "saat",	0, 1
8		isName	The order of words that started with capital letter	0, 1
9		isPrepositi	di, dari, ke.	0, 1
10	Lexica	Lexical	Term of words	Numeri

News article is written in unstructured format that force reader reading whole of article to understand the content.

Figure 3. Google response when we are looking for Corruption News Article.

Even though, the important content sometimes only represented in 4 or 5 sentences only. This paper proposed the summarization technique for Corruption news article in one sentence. So many cases of corruption occurred in Indonesia. Many links that appear when we look for news corruption in online media, as shown in Fig. 3. Google found 11.300.000 links in 0.32 second. It is too much. We need big space to collect them all. The reader takes a lot of time to select the desired article. So we proposed to represent each article in one sentence.

The step of the research:
1). Extracting six important word classes related to corruption. There are Year, Name of Corruptor, Location, Case, Amount of many, Status, as shown in Fig. 4.


(https://www.kejaksaan.go.id/berita.php?idu=0&id=12237&hal=1)
27-04-2018
DPO Pasar Pagi Tamiang Ditangkap
Tim Intelijen Kejaksaan Tinggi (Kejati) Aceh Tim intelijen Kejaksaan Negeri (Kejari) AcehTamiang Jum'at dini hari tanggal 27 April 2018 dengan dibantu Tim Intelijen Kejaksaan Agung RI telah menangkap 2 dua orang terpidana korupsi di Jakarta, atas nama TM Iqbal dan Suryadi dimana perkara yang dilakukan terpidana berupa korupsi pembangunan pasar pagi kuala simpang tahun 2011 dengan anggaran 6,9 milyar dimana tindak pidana korupsi tersebut telah merugikan negara sebesar 3 milyar.
Bahwa berdasarkan putusan MA pada awal tahun 2017 yang menjatuhkan pidana penjara kepada TM Iqbal selama 8 tahun dan kepada Suryadi selama 5 tahun belum dapat dilaksankan karena setelah mereka berdua dipanggil secara patut tidak hadir dan tidak dapat ditemukan keberadaannya oleh kejari aceh tamiang menjadikan mereka masuk dalam daftar DPO,
Selanjutnya para terpidana dibawa dari Jakarta menuju Aceh Tamiang untuk dilakukan eksekusi, bahwa dengan tertangkapnya kedua buron ini seluruh terpidana korupsi pembangunan pasar pagi yang berjumlah 6 orang telah dapat di eksekusi sesuai dengan amar.putusan Mahkamah Agung.(Yus)

Nama	Tahun	Lokasi	Kasus	Kerugian	Putusan
M.Iqbal, Suryadi	2011	Kejati Aceh	Pembangunan Pasar Kuala simpang	3 M	8 dan 5 tahun
Tubagus Imam Ariyadi	2018	Serang	Suap Amdal		9 tahun

https://news.detik.com/berita/d-4001149/jaksa-kpk-tuntut-wali-kota-cilegon-9-tahun-penjara?_ga=2.1.800112.502395672.1525600176-355048060.1525166237
Serang - Jaksa Penuntut Umum dari Komisi Pemberantasan Korupsi (KPK) menuntut Wali Kota Cilegon nonaktif Tubagus Iman Ariyadi pidana kurungan 9 tahun penjara dan denda Rp 275 juta subsider kurungan penjara 6 bulan. Hak politik untuk dipilih dalam jabatan publik juga diminta dicabut selama 5 tahun terhadap terdakwa untuk kasus suap amdal ini.
"Meminta majelis hakim menjatuhkan pidana terhadap terdakwa Tubagus Iman Ariyadi berupa pidana 9 tahun," kata Jaksa KPK Helmi Syarif di Pengadilan Tipikor untuk Pengadilan Negeri Serang, Jl Raya Serang-Pandeglang, Banten, Rabu (2/5/2018).
Jaksa menilai Iman terbukti sah dan meyakinkan telah melakukan tindak pidana korupsi secara bersama-sama dan berlanjut. Ini sebagaimana diatur dalam di Pasal 12 huruf a Uundang-undang Nomor 20 tahun 2001 tentang Pemberantasan Tindak Pidana Korupsi jo Pasal 55 ayat (1) ke-1 KUHP jo Pasal 64 (1) KUHP.
"Meminta majelis menjatuhkan hukuman tambahan kepada terdakwa berupa pencabutan hak untuk dipilih dan jabatan publik selama 5 tahun setelah terdakwa selesai menjalani pidana pokok," kata Helmi.
Helmi menilai, hal-hal yang memberatkan terdakwa di antaranya adalah tidak mendukung program pemerintah mewujudkan pemerintahan yang bersih dari KKN. Terdakwa menggunakan pengaruh kekuasaannya melakukan kejahatan dan melibatkan orang lain. Serta dinilai tidak mengakui secara terus terang dan tidak menyesali perbuatannya.
Selain menuntut wali kota, Jaksa KPK juga menuntut terdakwa Akhmad Dita Prawira berupa pidana penjara 8 tahun. Ia juga dituntut denda Rp 225 juta subsider kurungan 5 bulan penjara. Terdakwa merupakan Kepala Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP).
Jaksa menilai, hal yang memberatkan terdakwa Ahkmad Dita Prawira adalah mengambil alih peran dan kesalahan pelaku lain untuk menutupi kesalahan atasannya. Jaksa juga menilai terdakwa beberapa kali mencabut keterangan tanpa alasan yang sah sehingga mempersulit persidangan.
Baca juga: 2 Penyuap Walkot Cilegon Dihukum 1 Tahun 8 Bulan Penjara
Terakhir, terdakwa atas nama Hendri yang juga Direktur PT Jayatama Primayasa dituntut pidana 5 tahun penjara dan denda Rp 200 juta subsider 3 bulan kurungan penjara.
"Menjatuhkan pidana terdakwa Hendri berupa pidana penjara selama 5 tahun kurungan dan denda Rp 200 juta," tegasnya.
(rip/gp)

Figure 4. Extracting the important words or phrase related to Corruption article

2) Collect extracted important words into a defined Format. We propose a format for the summary sentence. We add words, comma, and conjunction in this format. There are 'Tahun', ',', 'tersangka', 'di', 'sebesar', 'dihukum') between important word to construct the summary sentence, as shows in Fig. 5.

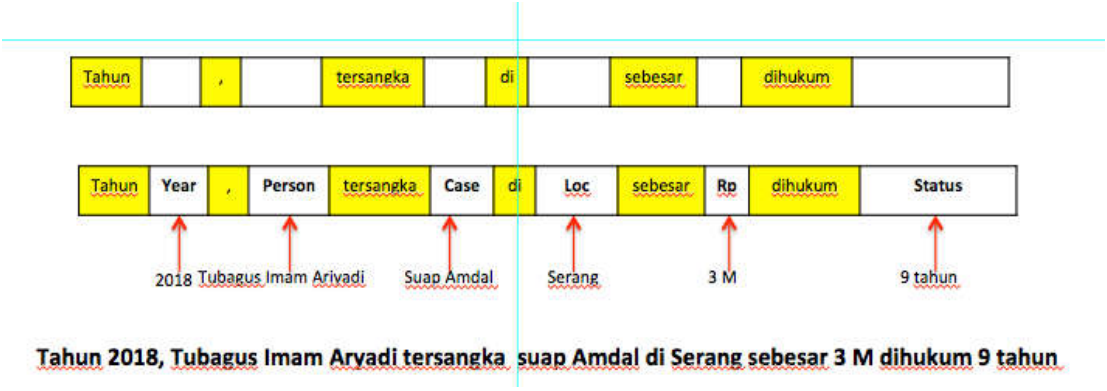


Figure 5. Format for constructing the summary sentence.

4. EXPERIMENT RESULT

Research data consist of 80-corruption article. We collected them from Indoenesia online media. There are detik.com, infokorupsi.com, antikorupsi.org, kpk.go.id, ti.or.id, polri.go.id, kejaksaan.go.id, and kemenkumham.go.id. Some of articles shown in Table 2.

Table 2. The collection of corruption article

No	Title
1	Kasus Bank Century - Never Ending Story
2	Segera Adili Setya Novanto
3	In-Depth Analysis: KPK Tetapkan PT. DGI Sebagai Tersangka Korporasi Pertama
4	Mengurai Proyek Mangkrak di PLN
5	KPK Tetapkan Bupati Mojokerto dan 3 Pihak Lainnya sebagai Tersangka Suap dan Gratifikasi
6	Aktivis Mahasiswa dan Perilaku Korupsi
7	8 Pejabat Bersiasat Sembunyikan Uang Hasil Korupsi
8	Sambil Menangis, Eks Pejabat Bakamla Akui Terima Duit Suap Proyek
9	Gubernur Jabar Serahkan SK Pemberhentian Bupati Subang
10	Korupsi DED PLTA, Dirut PT KPIJ Didakwa Perkaya Diri Rp 5 Miliar
11	Periksa Banyak Saksi Kasus UPS, Komjen Buwas: Kami Hati-hati
12	Korupsi Sitaan Korupsi
13	Staf Jaksa Korupsi Uang Korupsi, Vonis 15 Tahun Penjara Tepat
14	Nasib Proyek Hambalang
15	Hambalang, Proyek 'Maling' yang Kini Dimaling
16	Anton Bambang Akui Beri Uang Rp 100 Juta ke Jaksa Sistoyo
17	Terlibat Korupsi, Eks Petinggi PLN Dituntut 10 Tahun Bui
18	Anggota DPRD Sumbar Terdakwa Korupsi Minta Sumbangan
19	Tersangka Korupsi Perumahan di Kabanjahe Ditahan
20	Kejagung Ciduk Koruptor Dana Aparatur Desa
21	KPK Sita Rubicon Milik Pejabat Kemenkeu Terkait Kasus Suap
22	Kadis di Lampung Tengah Didakwa Jadi Perantara Suap Bupati Mustafa
23	Dalami Gratifikasi Proyek di Mojokerto, KPK Periksa 11 Kontraktor
24	Amin Santono jadi Tersangka Suap APBN-P, PD Merasa Kecolongan
25	Koleksi Mobil Mewah Bupati Abdul Latif yang Disita KPK
26	Buron Korupsi Dana Desa Riau Ditangkap Jualan Kopi di Jakarta
27	KPK Geledah 2 Lokasi di Jakarta Terkait Kasus Bupati Mojokerto
28	Ditangkap di Kalibata City, Ini Tampang Koruptor Rp 1,6 Miliar

The number of sentences is 1,580. Each sentence is labeled according to its class. We used Naïve Bayes algorithm for learning. Our experiments show quite encouraging results for initial research. The Precision and Recall shown in Table 3.

The average of precision for six classes is 73.5 %, and recall average is 73.7 %. Some of articles have no important word/ phrase, so the data is not enough to be trained.

Table 3. The Precision and Recall for six classes of word/ phrase

Precision	0.67	0.79	0.82	0.66	0.67	0.8
Recall	0.82	0.5	0.78	0.75	0.9	0.67

In addition, the information extracted to fill the summary sentence in phrase form. For example, the person's name, at least two words and has capital letters at the beginning of each word. The problem is that corruption reports mention a lot of names, need another characteristic to determine who's the witnesses, the police or the officers who have the authority as shown in Fig. 6. It is still difficult to determine whether Setya Novanto or Irvanto Hendra Pramudi Cahyo as a suspect.

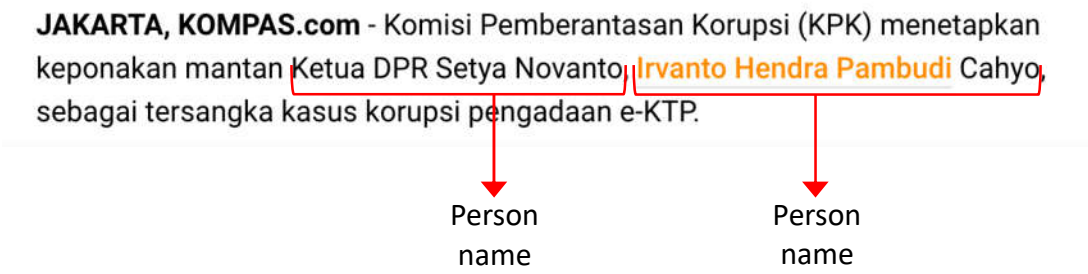


Figure 6. Two-person name in a sentence, need another feature-
To determine who is a suspect.

When we extract the phrase of Case, we also find the problem. The Case is narrated in several words so it must be known the first word and the final word of the phrase Case.

5. CONCLUSION

Automatic text summaries have actually been done since the late fifties, and then become unpopular, because this task difficult enough to do. But since the Internet began to be used to search the desired documents, the number of documents becomes exponentially increasing, needed a system that can summarize the documents becomes more crucial.

The problem of this task, automatically collected articles are not always articles that discuses the act of corruption of a person, but including a seminars article, education article, and general article about the concepts and phenomena of corruption in Indonesia. Besides, the same media, also by another media, repeatedly publishes one case so each case should be clustered first because a person can be a suspect in several cases. For example, Setya Novanto became a suspect of two corruption cases, the case "Papa minta saham" and "e-ktp".

The challenge of this research is to improve the coherence and accuracy of the summaries. The summary content represents the contents of the document, and minimizes the repetition of sentences. Using the Syntaxis feature is more precise than the Lexical feature because the sintaxis feature does not cut a sentence into the word, but consider the syntax of the sentence

References

[1]. Luhn, H.P. (1958). The automatic creation of literature abstracts. IBM Journal of Research and Development.

[2]. PE Genest, G Lapalme. (2011), Framework for Abstractive Summarization using text-to-text generation. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 64–73, Portland, Oregon, 24 June 2011. Association for Computational Linguistics (ACL)

[3]. Madanapalli, Andhra Pradesh, (2008), Knowledge Extraction Using Rule Based Decision Tree Approach. International Journal of Computer Scienceand Network Security IJCSNS, VOL.8 No.7, India.

[4]. Trevor Cohn and Mirella Lapata, (2008), Sentence Compression Beyond Word Deletion Proceedings of the 22nd International Conference on Computational Linguistics (Coling), pages 137–144 Manchester.

[5]. S. M. Harabagiu and F. Lacatusu, [2002], Generating single and multi-document summaries with gistexter," in Document Understanding Conferences (MUC)

[6]. R. Barzilay et al, [2003] ("Information fusion in the context of multi-document summarization," in Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, 1999, pp. 550- 557)

[7]. Jing, H., & McKeown, K. R. (2000). Cut and paste based text summarization. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (pp. 178-185). Association for Computational Linguistics.

[8]. Hideki Tanaka, Akinori Kinoshita, Takeshi Kobayakawa, Tadashi Kumano, and Naoto Kato. (2009). Syntax- driven sentence revision for broadcast news summa- rization. In *Proceedings of the 2009 Workshop on Language Generation and Summarisation*, UCNLG+Sum '09, pages 39–47, Stroudsburg, PA, USA. Association for Computational Linguistics.

[9]. Masayu Leylia Khodra, Dwi Hendratmo Widyantoro, E. Aminudin Aziz Bambang Riyanto Trilaksono, (2012), Automatic Tailored Multi Rhetorical Document Profile and Summary Specification, Journal of ICT Research and Applications. Published by ITB Journal Publisher, Vol. 6, No. 3, 2012, pp 220-239, ISSN: 1978-3086, DOI: 10.5614.

[10]. Afrida Helen, Ayu Purwarianti, Dwi Hendratmo Widyantoro (2014), Extraction and Classification of rhetorical sentences of experimental technical paper based on section class, 2nd International Conference on Information and Communication Technology (ICoICT), 2014, Date of Conference: 28-30 May 2014,
IEEE *Xplore*: DOI: 10.1109/ICoICT.2014.6914099

[11]. Afrida Helen, Ayu Purwarianti, Dwi Hendratmo Widyantoro, (2015), Rhetorical Sentences Classification Based on Section Class and Title of Paper for Experimental Technical Papers, Journal of ICT Research and Applications. Published by ITB Journal Publisher, Vol. 9, No. 3, 2015, pp. 288-310, ISSN: 2337-5787, DOI: 10.5614/