Set of Frequent Word Sequence (SFWS) as Document Model for Feature Based Document Clustering

Gusti Ayu Putri Saptawati

School of Electrical Engineering & Informatics Institut Teknologi Bandung, Indonesia putri.saptawati@stei.itb.ac.id

Abstract: Sequence of word sequence has been considered as an appropriate text representation since text reveal inherent sequential nature. Those representations are Frequent Word Sequence (FWS), Set of Frequent Word Sequence (SFWS) and Frequent Word Itemsets (FWI). Moreover, Maximal Frequent Sequence (MFS) is text feature that exploiting sequential property of textual data. In this paper, we proposed SFWS as the best text representation for document clustering. SFWS considers document as set of sentences in which sentence is the language highest grammatical hierarchy, conveying a complete thought. Consequently, document clustering would have accurate results. The main contribution of this work is the data pre-processing, feature extraction and selection based on SFW. Since SFWS works based on sentence, we need to construct sequence sentences of all document into sequence database for sentences. Then, sequential pattern mining was applied to extract set of frequent sentence sequence. And finally, we select features with maximal set of frequent sequence (MSFS). We conducted experiments on Twenty News Group Text Data (TNTD). To do so, we developed Feature based clustering (FBC) algorithm with MSFS as text feature based on SFWS representation. The experimental results showed that document clustering based on SFWS had the highest accuracy, compared with FWS and FWI.

Keywords: Frequent Word Sequence (FWS), Set of Frequent Word Sequence (SFWS), Frequent Word Itemset (FWI), Maximal Frequent Sequence (MFS), document clustering, Feature Base Clustering.

1. Introduction

Nowadays, ubiquitous information technology has been contributing the availability of digital text document corpora. As the volume of text corpora continues to grow rapidly, automated document management become crucial for document management in term of document search, access, and use. Document clustering is text mining technique that support automated document management by grouping individual documents into categories, each of which includes documents relating to a theme or topic [Zhao and Karypis 2005]. Recently, various text clustering methods has been proposed including distance-based clustering algorithms, word & phrase-based clustering, probabilistic document clustering & topic model, online clustering with text stream, clustering text in networks, semi-supervised clustering [Aggarwal and Zhai 2012; Li et al.2008; Zhao and Karypis 2005].

Text is unstructured, so text document necessitates an appropriate feature representation for many applications of data mining and information retrieval. Most document clustering methods use vector space model, n-grams, bag of words to represent document in document space [Thanh and Yamada 2011; Xu et al.2012]. However, the meaning of natural language depends on words sequences and accordingly text document representation should preserve the sequential relationship among words in document [Duocet 2005; Li et al 2008; Rachmawati et al.2015]. Clustering based-on Frequent Word Sequence (CFWS) algorithm works based on frequent word sequence to reduce the high dimensionality of document and to measure the closeness between documents [Li et al 2008]. It implemented association rule miner to get frequent 2-word sequence in which all lengths of frequent word sequence were constructed.

Received: December 10th, 2018. Accepted: December 2nd, 2019 DOI: 10.15676/ijeei.2019.11.4.13 Meanwhile, maximal frequent sequences (MFS) was proposed for text feature as it guarantees longest possible frequent sequences [Duocet 2005]. MFS was considered as an effient way to account for sequential property of textual data. To do so, sequential pattern mining was recommended for extracting word sequences since it outperformed association rule mining [Duocet 2005, Jaliet et al.2008]. Sequential pattern mining to extract MFS for document clustering has been implemented using FBC algorithm [Rahmawati et.al.2015]. Initially, FBC was developed to overcome computational complexity of hierarchical (UPGMA) and partitioning (K-mediod) algorithms since these algorithms must compute the pairwise of similarity between all sequences [Guralnik and Karypis 2001]. Accordingly, FBC took advantage computationally efficient scheme of K-means so it is scalable for large data set. To cluster documents with FBC, MFS was generated from set of FWS features which were extracted based on sequential pattern.

It has been showed that MFS as document features based on FWS as text representation is a promising method to reduce dimension of vector space in FBC [Rahmawati et.al. 2015]. Yet, with FWS representation, it ignores grammatical hierarchy on document. It means that FWS ignores the presence of sentence in document, as it treats document as one sentence. Meanwhile, in term of language, sentences are considered as language highest grammatical hierarchy since they convey a complete thought. Accordingly, the best way to represent topic in document is by expressing it in sentences [Schönhofen and András 1995].

Therefore in this paper, we propose SFWS (set of frequent word sequence) as sentencedbased text representations to solve problem of grammatical hierarchy in FWS. SFWS models document as set of sentence sequences by which, each consists of sequence of words. Consequently, document topic is extracted based on SFWS (set of sentences). Then, for document clustering with FBC, we need to extract text features based on maximal set frequent sequence (MSFS) which is developed based on MFS principle. For feature extraction, we apply an efficient sequential pattern mining, i.e. PrefixSpan. Contrast with data pre-processing in [Rachmawati et al.2015], we need to construct sequence database consists of sequence sentences. These sentences consist of sequence of words. The result of PrefixSpan is a set of frequent sentences which then need a post-process in order to select MSFS feature.

This paper is organised as follow. In Section II we provide literature overview of word sequence, frequent word sequence, text representations based on frequent word sequence, and Feature Based Clustering (FBC). In Section III we describe evaluation of FWS, SFWS and FWI to express meanings in natural languages. In section IV we propose an architecture of document clustering based on FBC, three text representations (FWS, SFWS, FWI) and MFS as feature selection. Finally, in section V, we present the experimental results conducted on Twenty News Group Text Data. Finally, conclusion of this work is presented in section VI.

2. Literature Review

A. Word Sequence & Frequent Word Sequence

A word sequence S is an ordered sequence of two or more words, and denoted as $\langle w_1, w_2, \cdots \rangle$. Meanwhile a frequent word sequence FWS/FS is denoted $\langle w_1, w_2, w_3, w_4, w_5 \rangle$ which means that w_4 comes after w_1 , w_2 , and w_3 in text document [Agrawal et.al.2001; Guralnik and Karypis 2001]. Unlike phrase in [Lent et.al.1997] that does not allow gap between words, FWS relaxes the rule by allowing other words appear between w_1 , w_2 , w_3 , w_4 , w_5 as long as they are not frequent. Accor. For example, the following is a set of sentences: [H.A-Myka 2005].

- 1. The Congress subcommittee backed away from mandating specific retaliation against foreign countries for unfair foreign trade practices.
- 2. He urged Congress to reject provision that would mandate US retaliation against foreign unfair trade practices.
- 3. Washington charged France West Germany the U.K. Spain and the EC Commissions with unfair practices on behalf of Airbus.

Thus, a sequence *«unfair practices»* can be found in all of three sentences [Myka 2005]

Definition 1: a sequence $p = a_1, ..., a_k$ is a subsequence of a sequence q if all the items $a_i, 1 \le i \le k$ occur in q and they occur in the same order as in p. If a sequence p is a subsequence of a sequence q, it is called that p occurs in q. For example is the sequence *<unfair practices*> with respect to all of three sentences.

Definition 2: A sequence p is frequent in S if p is a subsequence of at least ∂_{\min} sentences in S where ∂_{\min} is a given minimum frequency threshold. For example: 2 is the minimum frequency threshold, accordingly there two frequent sequences. i.e. *<congress retaliation against foreign unfair trade practices>* and *<unfair practices>*.

Definition 3: A sequence p is maximal frequent sequence in S if there is not exist any sequence p' in S such that p is subsequence of p' and p' is frequent in S. For example, *«unfair practices»* is not maximal since it is subsequence of *«congress retaliation against foreign unfair trade practices»* which is frequent.

B. Text Representation Based on Word Sequence

Beside FWS as text representation, there are other two representations based on word sequence namely SFWS (set of frequent words sequence) and FWI (frequent word itemsets) [Agrawal et.al.2001]. First representation (FWS) consider document an ordered list of FWSs in which each FWS holds in ordered set of words according to their occurrence in document. Second representation (SFWS) takes concept of sentence into account document. It considers document as sets of ordered FWS based on its occurrence in document. On the other hand, third representation (FWI) considers document as a ordered list of sequential patterns, each is considered as an unordered set of words. Yet, FWI is different from 'bag of word' which is solely set of words without considering association between words, while FWI still considers the order of different set of words. Consequently, FWI will treat all FWS with same words but different order, as one FWI. For example, we have three different FWS < *congress unfair trade practice* >, *< unfair trade congress practise* >, *< unfair practise congress trade* >. These FWS will be regarded as one FWI, or in other word, FWI is loosely FWS. Table 1 summarizes these three representations.

	FWS			
structure	A list of ordered FWS	Sets of ordered FWS	List of FWI	
	< <u>(w1 w2</u>) (<u>w3 w4 w5</u>),> FWSi FWSi+1	<<(<u>w1 w2</u>) (w3 w4),> <(<u>w5 w6</u>)>> FWSi FWSi+1 FWSj	< (w1 w2) (w3 w4 w5),> FWi FWi+1	
		Sentencei sentencei+1		
	FWS is ordered based on its	Senteces are ordered based on their	FWS is ordered based on its	
	occurrence in document	occurrence	occurrence in document	
	FWS's elements are ordered	FWS is ordered based on its	The order of FWS's	
	based on their occurrence in	occurrence in document	elements are not matter	
	document	FWS's elements are ordered based on		
		their occurrence in document		

Table 1 Text Representation based on word sequence [Agrawal et.al.1999]

C. Feature Based Clustering (FBC)

Originally, FBC was developed for clustering data having an inherent sequential nature [Guralnik and Karypis 2001]. FBC integrates K-means algorithm as main clustering method since it is near linear complexity in term of number of sequences. The essential part of FBC, is extracting a set of features which describe the sequence nature in data. These sequence-based features need to be projected into new space having them as dimensions. Finally, with respect to vector space model, K-means would cluster data-sequence in the new dimensions.

FBC requires those features must satisfy particular properties, i.e.: [Guralnik and Karypis 2001, Rahmawati et.al. 2015]

1. The features must describe the sequential pattern among item which are naturally embedded in the data-sequence. This property is significant since FBC would cluster the data-sequence heavily based on the similarity of the features.

- 2. The features should be nontrivial with respect to user-specified number, since rare features are not strong enough for clustering data-sequence.
- 3. The features must be complete, means that all features have been transformed as dimensions in the new space.

Seq.Id	Sequences
S1	AQVHGHKKSVDAM
S2	AQHKKSGSDGLP
S 3	AQVHAHVAQIVAKDP
S4	AQVHDALGPHKKS
S5	DALGPAQMHVHKKS
S6	AQIKDDLGPAQP
S 7	KKSPQIKDQVG
S8	DIKDALGMAQVHPL

a. Sequence database

Pattern	Support
ALG	4
AQV	6
DAL	4
HKK	4
IKD	4
KKS	5
QIK	4
QVH	6
AQVH	6
DALG	4
HKKS	4
QIKD	4

b. Sequential pattern with support 50%, and length 3, 4

Seq.ID	ALG	KKS	AQVH	DALG	HKKS	QIKD
S 1			Х		Х	
S2		Х	Х			
S 3			Х			Х
S4			Х	Х	Х	
S5				Х		Х
S6				Х		Х
S 7		Х				X
S 8	Х		Х			Х

c. Feature Selection of Resulted Sequential Patterns Figure 1. Example of Local-based Feature Selection [Guralnik and Karypis 2001]

The second step after extracting features is selecting a set of independent features in which two features are independent if they are supported by non-overlapping segments in the same data-sequence. This approach (called 'local approach') is outperformed 'global approach' in term of time [Guralnik and Karypis 2001]. Figure 1 describes an example of local-based feature selection. Figure 1(a) is sequence database of amino-acid, and Figure 1(b) is amino-acid sequential patterns whose support are 50% and each length are 3 or 4. Figure 1(c) is the selected features based on local approach

The final step of FBC is clustering the data-sequence with K-means, started by projecting the features into vector space model. It means that all sequential features are dimensions of the vector space model, and data-sequence will be represented as M (number of feature) dimensional vector. Then, cosine of two data-sequence, is calculated to identify the similarity between them. Moreover, each feature is scaled up based on inverse-document-frequency.

Based on all explanations, it is clear that we need to extract text features, select them locally, and projecting them into new vector space model in order to apply FBC for document clustering. Figure 2 below describes the process of FBC for document clustering using MFS based as text feature.



Figure 2. FBC for Document Clustering

3. Propositions of FWS, SFWS, AND FWI

Those three representation will be evaluated based on 3 important criteria i.e.: (i) syntax structure, (ii) representation for expressing meaning of topic with respect to natural language and grammar [Jones and Douglas 2007; Schönhofen and András 1995; UCL 1998]. Accordingly, sentences are claimed as the most powerful language unit for expressing a complete thought and consequently they constitute semantical meaning to support topic of document.

Based on the above explanation, text feature with FWS naturally ignores grammatical hierarchy since all FWS's are considered equally in one sequence. With structure as shown in Table 1 Representation based on word sequence [Agrawal et.al, 1999], all FWS would express the meaning of document topics based on list of ordered words' occurrence. In other word, a specific topic is represented by a specific ordered of word occurrence, or different ordered word occurrence would express different meaning. For example: <(*play*) (*tennis*) (*afternoon*)> has different meaning with <(*afternoon*) (*tennis*) (*play*)>. Yet, FWS/MFS is an appropriate feature for document clustering based on FBC method as claimed in [Rahmawati et.al. 2015].

SFWS on the other hand, considers sentences as grammatical structure in expressing thought or topic. As shown in Table 1 Representation based on word sequence [Agrawal et.al, 1999] each FWS is grouped based on its occurrence in same sentences. It means that a sentence can be represent by a SFWS whose FWS's occur with the same order as sequence of words in the sentence. For example, document with structure as follow:

"Obesity presents numerous problems for the child. Obesity increase the risk of diabetes in adulthood"

Consequently, representation SFWS would not include FWS such as (child, risk, diabetes) since the word 'child' occurred in the first sentence, whilst the word 'risk' and 'diabetes' occurred in the second sentence. So, like sentences, SFWS is considered better/stronger than FWS in expressing complete semantic of topic since it preserves the occurrence of FWS in sentences as well as FWS's elements. Moreover, in term of features, it is clear that number of SFWS will be fewer than of FWS.

Like FWS, FWI representation does not take into account the grammatical hierarchy by which it regards document as one sequence whose element is FWI [Maylawati, Saptawati 2017]. It expresses the semantic of meaning based on occurrence of itemsets. In term of sequential pattern, sequence of itemset occurrence is matter whilst sequence of items is not. For example, two words with sequence 'chicken eat' and 'eat chicken' are regarded as the same

itemsets. It is obvious FWI losses meaning information since in fact those two word sequences have different meaning. In the first sequence, 'chicken' is subject while in the second sentence, 'chicken' is object. Therefore, FWI is the most lossless representation compared with FWS and SFWS [Maylawati, Saptawati 2017].

4. Proposed Architecture

Figure 3 describes the architecture of document clustering based on FBC and sequential pattern. The architecture was developed based on the previous one developed by Rahmawati [Rahmawati et.al 2015]. We modified some modules, as well as added new modules, in order to support FWS, SFWS, and FWI.



Figure 3. Architecture of FBC for Document with MFS Feature

Below are the descriptions for modules being modified:

- Module for constructing data sequence based on sentences. As SFWS work based on sentences in text document, firstly we need to identify sentences in text. Then, we construct text sequence with regard to those sentences. Consequently, we develop a sentence sequence which consists of words sequence. Figure 4 is an example of those sequences.
- Module for MFS-based Feature Extraction for SFWS and FWI. Unlike FWS which are the result of Prefix-Span, we need to add process for constructing groups of FWS contained in sentences (named as Maximal Set of Frequent Sequence/MSFS). To do so, follow are the process based on Prefix-Span result:
 - a) Select FWSs according to user requirement (such as length and gap)
 - b) Scan those FWSs in set of sentences. A group of FWS consists of one or more FWSs occur in one sentence, and it is candidate feature.
 - c) If a FWS overlap with another, select FWS with higher frequency since we assume that it is more likely represent document dependency. If both FWS have the same frequency, we choose the first FWS.

For FWI, we need to merge FWSs consist of same words since FWI ignores the words occurrence. To do so, we need to

- i. Identify FWSs which contain same words.
- ii. Merge those FWSs as one FWI.
- iii. Calculate the frequency of each FWI on document collection based on singular principle.
- Module for Feature Selection. This module is developed in order to select text features MFS and MSFS based on local approach as described in Section II.3.

obes, present, number, problem,child>, <add, adulthood,<br="" increase,="" obes,="" risk,="">ildhood, obes, lead,caus, vari, diseas, pediatr, hyperten, associate, pediatr, pertren,associate, type, diabet, mellitus, increase, high, risk, coronoari, art,diseas, increase, stress, weight, bear,joint, low, self, esteem, affect, ationshiop, peer>></add,>
<obes, child="" numer,="" present,="" problem,=""></obes,>
n Anthe Rel Octa CA
Sentence 1
add, increase, risk, obes, adulthood,childhood, obes, lead, caus, vari, diseas,, pediatr, hypertren, associate,type, diabet, mellitus, increase, high, risk, coronary, heart, diseas, increase,stress, weight, bear, joint, low, self,
esteem, affect, relationship, peer >
Sentence 2

Figure 4. Example of Sequential Data in Sentence

5. Experiment Results

The experiment was intended to evaluate FWS, SFWS, and FWI performance for document clustering regarding accuracy. The experiment was conducted for 5 datasets extracted from Twenty News Group Text Data (TNTD), and on PrefixSpan* library to obtain sequential patterns. Due to limitation of PrefixSpan*, datasets were designed as follows:

No.	Dataset_id	Description	Total files	Original class		
				Class_ label	# of files	
1.	dataset_1	5 different classes, size of each file: 4-5 kb		alt.atheism rec.autos rec.motorcycles sci.med talk.religion.misc	3 6 3 4 5	
	dataset_2	7 different classes, size of each file: 3-6 kb		comp.graphics comp.windows.x sci.med sci.space rec.sport.baseball talk.politics.guns talk.politics.misc	7 9 8 8 10 9	
	dataset_3	3 different classes with high similarity within one class, size of each file: 3-6 kb		sci.electronics sci.meds talk.politics.misc	9 10 5	
	dataset_4	6 different classes with high similarity within one class, size of each file: 3-6 kb		comp.graphics sci.crypt sci.electronics sci.med talk.politics.mideast talk.politics.misc	8 7 12 12 9 7	
	dataset_5	10 different classes, size of each file: 3-6 kb		alt.atheism comp.sys.ibm.pc.hardware comp.sys.mac. hardware comp.windows.x rec.sport.baseball rec.sport.hockey sci.crypt soc.religion.christian talk.politics.mideast talk.religion.misc	10 10 10 10 10 10 10 10 10 10	



Figure 5. Figure 8 describes the evaluation of document clustering conducted based on each data set.

Figure 5 Accuracy of Document Clustering from Data set 1

The experiments were conducted with the following procedures:

- For each dataset, we conducted 9 experiments to cluster documents based on FWS, SFWS, and FWI (each 3 experiments).
- For each FWS, SFWS, or FWI, we set 3 4 different parameter of min sp length, min sp frequency, min gap for MFS/MSFS, feature selection method
- K-value (number of cluster) was the same as the number of class documents in dataset
- The quality ('goodness') of document clustering was evaluated based on F-measures as implemented in [Steinbach, et.al, 2000] since the document groups would be compared with the known category. This evaluation is called external quality measures.



Figure 6. Accuracy of Document Clustering from Data set 2



Figure 7. Accuracy of Document Clustering from Data set 3

The experimental results showed that text representation SFWS had the highest accuracy for datasgeet_1, dataset_3, and dataset_4 with average accuracy rate 0.8698, 0.865, and 0.8285 respectively. For text representation FWS, the average accuracy rates for those 3 datasets were 0.7397, 0.8155, and 0.802. Meanwhile for FWI, the average accuracy rates were 0.7149, 0.7812, and 0.79.

For dataset_4, text representation FWI had the highest accuracy rate with average accuracy rate 0.542. These dataset was extracted randomly from Twenty Newsgroup Text Data which does not have appropriate language structures [Hammouda & Mohamed. 2004].



Figure 8. Accuracy of Document Clustering from Data set 4

6. Conclusions

Experimental results show that the highest accuracy of document clustering is based on text representation SFWS, while the lowest accuracy in based on FWI. Consequently, it is clear that representation SFWS would better maintain semantical aspect of document, compared with representation FWS and FWI. Although in some cases the highest accuracy is based on FWS, this is merely due to the fact that email in dataset were less structured. Moreover, it also concludes that text representation SFWS is appropriate for structured/formal documents such as academic documents, scientific documents, reports, and publications. On the other hand, text representation FWI could be appropriate for least structured documents such as tweet, slang, etc. Finally, we propose future research on slang/tweet clustering based on FWI.

7. Acknowledgments

I acknowledge my team, Rindhu Astry Nalistia, Dini Rachmawati, Riza Ramadan, Kania Adityarani, Yudhie Hatmadji, and Dian Sa'adillah Maylawati working on "*Features Based Clustering for Document Clustering*". This work has been carried out in Database Laboratory, Informatic Engineering Study Program, School of Electrical & Informatics Engineering, Institut Teknologi Bandung, Indonesia.

8. References

- [1] Aggarwal and C.Zhai, " A Survey of Text Clustering Algorithms", Chapter in Book 'Mining Text Data', Book, Springer, 2012.
- [2] R. Agrawal, R. Srikant, B.S.Lent, "Mapping words, phrases using sequential pattern to find user specific trends in text database", United States Patent Number 6,006,223, 1999.
- [3] A. Doucet, "Advanced document descriptor: A sequential approach", Academic Dissertation, Faculty of Science, University of Helsinki, 2005.
- [4] V. Guralnik and G. Karypis. "A scalable algorithm for sequential data", University of Minessota, Department of Computer Science and Army HPC Research, 2001
- [5] S. Jaliet, A. Laurent, M.Teisseire, "Sequential pattern for text categorization", Intelligent Data Analysis, 2006.
- [6] N.M. Jones, J.K.M. Douglas, "Representing word meaning and information order in composite holographic lexicon", University of Colorado, USA, 2007.

- [7] B. Lent, R. Agrawal, R. Srikant, "Discovering trends in text databases", in Proc. of the 3rd International Conference on Knowledge Discovery and Data Mining, 1997.
- [8] Y.Li, S.M.Chung, and D.J. Holt. "Text document clustering based on frequent word meaning sequences", *Data & Knowledge Engineering*, 64(1), 381-404, 2008
- [9] D.S.Maylawati, G.A.P.Saptawati, 'Sequence of Frequent Word Item sets As Feature Representation for Text with Indonesian Slang', *Journal of Physics: Conference Series* 801 012066, 2017.
- [10] H.A-Myka, "Mining all maximal frequent word sequences in a set of sentences", ACM *International Conference On Information And Knowledge Management*, 2005.
- [11] D.Rachmawati, G.A.P.Saptawati, Y.Widyani, 'Document Clustering Using Sequential Pattern (SP) : Maximal Sequence Frequent (MFS) as SP Representation", *International Conference on Data & Software Engineering (ICoDSE), Jogyakarta Indonesia*, 2015.
- [12] P.Schönhofen and A.B. András, "Feature selection based on word-sentence relation", *Informatics Laboratory Hungarian Academy of Science*, 1995.
- [13] M.Steinbach, G.Karypis, V.Kumar, 'A Comparison of Document Clustering Techniques', Technical Report, *Department of Computer Science & Engineering*, University of Minnesota, 2000.
- [14] N.C.Thanh and K.Yamada, "Document representation and clustering with WordNet based on similarity rough set model", *International Journal of Computer Science Issues*, 8(5), 2011
- [15] University College London, "The grammatical hierarchy : words, phrases, clauses, and sentences", Survey of English Usage, 1998, <u>http://www.ucl.ac.uk/internet-grammar/clauses/xclau7.htm</u>, date of access 15 May 2017.
- [16] Z.Xu, M.Chen, K.Q. Weinberger, F.Sha, "From eBoW to dCoT marginalized encoder for text representation", CIKM'12 Proceedings of the 21st ACM international conference on information and knowledge management, 2012, pp 1879-1884.
- [17] Y.Zhao and G.Karypis, "Hierarchical clustering algorithms for document datasets", *Data Mining & Knowledge Discovery*, 10(2), pp 141-168, 2005.



Gusti Ayu Putri Saptawati received Bachelor & Doctoral Degree from Department of Informatics Engineering Institut Teknologi Bandung, and Master Degree from Information Systems the University of New South Wales. Since 1993, she has been a lecturer and researcher in the Department of Informatics Engineering Institut Teknologi Bandung. Her current research interests are in the area of data mining, machine learning for complex data structure, urban informatics, and graph-based recommender systems. She has published around 40 papers, participated in 6 R&D Project, primary funded

by Indonesian Government.