Content-Based Readability Assessment: A Study Using A Syllabic Alphabetic Language (Thai)

Abstract. Text readability is typically defined in terms of "grade level"; the expected educational level of the reader at which the text is directed. Mechanisms for measuring readability in English documents are well established; however this is not in case in many other languages, such as syllabic alphabetic languages. In this paper seven different mechanisms for assessing the readability of syllabic alphabetic language texts are proposed and compared. The mechanism are grouped under three headings: (i) graph ranking, (ii) document ranking, and (iii) hybrid. The presented comparison was conducted using the Thai language with respect to the reading age associated with secondary school, high school, and undergraduate students in the context of scientific abstract.

Keywords: readability metrics·reading level assessment·alphasyllabary· readability assessment

1 Introduction

The ease whereby written texts can be read can be measured using mathematical *readability measures*. Readability measures are generally based on lexical features such as: number of words in a sentence, number of letters per word, number of syllables per word and the "difficulty" of words and sentences. The concept of readability formulas have been studied since the 1920s (Kitson 1921). Some well known and commonly used measures include: (i) the Flesch Reading Ease formula (Flesch 1948), (ii) the Flesch-Kincaid Grade Level (Kincaid et al. 1975), (iii) the Gunning Fog Index (Gunning 1952), (iv) the SMOG Index (McLaughlin 1969), (v) the Coleman-Liau Index (Coleman and Liau 1975), and (vi) the Automated Readability Index (ARI) (Senter and Smith 1967). The objective of these readability measures is primarily to find the readability level for a person who needs to read and understand a text. More recently, readability measures have been utilised by insurance agencies, political parties, financial institutes, hospitals, and non-profit organisations with respect to the nature of the documented material that they put out. Readability measures also provide writers with information on how best to reach their intended audiences. However, readability measures will only assess the readability of the text, not of the reader. Moreover, readability may not necessarily be an appropriate measure for the understandability, context, required prior knowledge, interest level, difficulty of concepts, coherence or quality of a text. Work on measuring the level of text readability is not only a research issue in English language texts, but also in other languages (Rabin et al. 1988). To the best knowledge of the authors, Thai and other syllabic alphabetic languages are still in the early stages of such research. There is thus a requirement for readability measures directed at syllabic alphabetic languages.

In this paper, we propose a series of techniques for assessing the text readability of related documents, so-called Content-based Readability Assessment (CRA). Note that existing measures are typically directed at categorising single documents, the work presented in this paper is directed at categorising groups of documents. We focus on syllabic alphabetic languages because, as noted above, little research work has been conducted with respect to readability measures for such languages. More specifically we focus on the Thai language (an exemplar of the syllabic alphabetic languages group) and particularly scientific project abstracts written in Thai. Syllabic alphabetic languages have no explicit boundary markers for words or sentences, this thus provides an additional challenge. Seven different CRA ranking mechanisms are proposed in this paper, grouped under the following headings: Graph, Document and Hybrid. The proposed CRA technique is suitable for ranking all kinds of related texts, but especially directed at syllabic alphabetic language texts.

2 Related Work

Traditional readability formulas are mainly based on properties of words, sentences, and documents. The Flesch Reading Ease Formula (FREF) (Flesch 1948) used sentence length and syllable count to generate a single statistic using a 0 to 100 scale to give a rating from "very difficult" to "very easy". This formula has been widely used with respect to English texts. However, the way that this formula is defined is ambiguous because terms such as sentence, word, and vowels are interpreted in a non-standard manner. The concept of syllables is also not well suited to syllabic alphabetic language texts. With respect to the work presented in this paper the Coleman-Liau Index (CLI) proposed in (Coleman and Liau 1975) and the Automated Readability Index (ARI) proposed in (Senter and Smith 1967) were used for evaluation purposes.

The CLI readability "grade level" is determined according to the average length of words and sentences as follows:

$$CLI = 0.0588L - 0.296S - 15.8, \tag{1}$$

where L and S are the average number of characters per 100 words, and the average number of sentences per 100 words, respectively. The equation has been shown (Coleman and Liau 1975) to be equivalent to:

$$CLI = \frac{(5.89 \times LC) - (30.0 \times SC)}{WC} - 15.8, \tag{2}$$

where LC, SC, and WC are the number of characters, number of sentences and number of words, respectively.

The ARI (Senter and Smith 1967) measure produces an estimate of the readability of a document, derived from the number of letters per word and the number of words per sentence, expressed in terms of a minimum "age level" required for a given document to be understood. ARI is calculated as follows:

$$ARI = 4.71 \times (\frac{C}{W}) + 0.5 \times (\frac{W}{S}) - 21.43, \tag{3}$$

where C is the total number of letters, digits and punctuation marks; W is the number of spaces; and S is the number of sentences.

Note that both CLI and ARI used the number of characters per word as opposed to the number syllables per word (as in the case of the FREF measure). The advantage of this is that the number of characters is more readily and accurately counted by computer programs than syllables. The CLI and ARI measures are also more suited to syllabic alphabetic languages (but see below).

The above "traditional" readability metrics are mainly "tuned" for use with English language texts. They are thus argued to be inapplicable for use with other languages such as Arabic (Al-Khalifa and Al-Ajlan 2010), Thai (Daowadung and Chen 2011), and Bengali (Sinha et al. 2012). A syllabic alphabetic language is a language whose writing system consists of consonant and inherent vowels as single units. Each unit is based on a consonant letter; an inherent vowel can be changed to another vowel or muted by means of diacritics (Ager 2014). Vowel diacritics may appear above, below, to the left, to the right, or around the consonant. This writing system, also called Abugida (Daniels 1990) or Alphasylabary (Bright 2000), is widely used in South and Southeast Asia. Note that there are no spaces between words in some of these languages where spaces are mainly used to indicate the end of a clause or sentence. Clause or sentence length varies in relation to the rhythm of the writer's speech. Also note that the number of syllables in a word can be counted in different and inconsistent ways where both are considered correct. These factors need to be taken into consideration when developing readability measures for use with syllabic alphabetic language texts.

3 Proposed Content-Based Readability Assessment Framework

The generic CRA framework is presented in Fig. 1. The framework comprises four stages. Each stage is discussed further in this section. During the preprocessing stage word segmentation and part-of-speech (POS) tagging is applied to the document collection. Note that in the context of syllabic alphabetic languages there tend not to be explicit boundary markers for words and sentences as in the case of languages such as English. In the case of the Thai language there are no sentence boundary markers while a space has several usages: (i) as an intrinsic part of a word or (ii) as a word, phrase or sentence delimiter; depending on context and writer's style. For the work presented in this paper, we applied the bigram word segmentation technique presented in (Meknavin et al. 1997) to identify words and associate POS tags with the identified words. At the end of

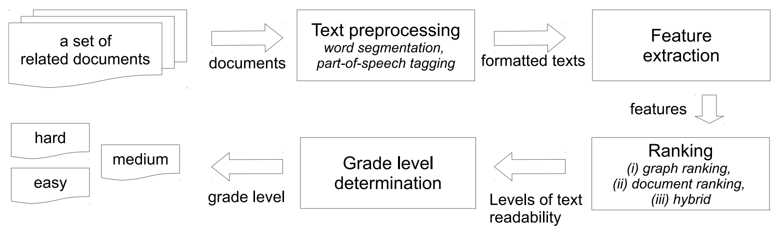


Fig. 1. A Framework for Content-Based Readability Assessment

this stage documents are represented in terms of sequences of words each with an associated POS tag.

Text readability measures generally use features such as the number of characters, syllables,words or sentences. Some of them exploit some special features such as the number of easy words (defined as two syllables or less), the number of hard words (defined as three syllables or more), the number of a single word (words with hyphens) or the number of proper nouns. However, syllable counts have been shown to be unreliable (Coleman and Liau 1975, Senter and Smith 1967) and do not lend themselves to use with syllabic alphabetic languages. For the work described in this paper the different proposed ranking mechanisms use different features. During the ranking stage documents are ranked according to their readability level. This is achieved by allocating a readability score to each document, a number between 0 and 1. How this is achieved depends on the ranking mechanism used and this will be discussed further in the following section, Section 4). The final stage is where the readability scores used for the ranking are translated into a grade level measure.

4 Ranking Mechanism

From the introduction it was noted that seven different ranking mechanisms, for application to syllabic alphabetic language texts, are proposed in this paper and that these can be categorised under the following headings: (i) Graph, (ii) Document and (iii) Hybrid. Each is discussed in further detail in the following three sub-sections.

4.1 Graph Ranking

As the name suggests, graph ranking is based on the idea of representing a document collection in terms of a graph. The process commences with a graph generation process after which the desired ranking are calculated using the graph represented data. We refer to such graphs as *ranking graphs*. A ranking graph is a directed graph where each vertex is a document "item" of some form and each directed edge is a relation between the preceding and succeeding item. Note that any specific item can only appear once in the graph and the directed edges collectively represent the sequence in which the items appear in the given

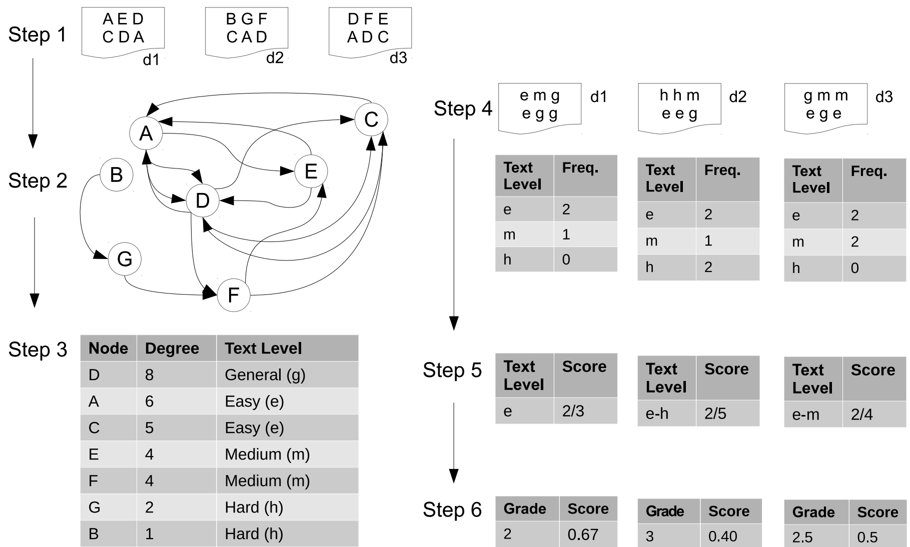


Fig. 2. Graph Ranking Process

set documents. Four alternative ranking graphs are considered according to the nature of the "item" encapsulated at the nodes:

Word Ranking Graph: Each vertex is a word and each edge is a relation between preceding and succeeding words.

POS Ranking Graph: Each vertex is a POS tag, and each edge is a relation between preceding and succeeding POS tags.

Word and POS Ranking Graph: Each vertex is a word with its POS tag, and each edge is a relation between preceding and succeeding word and POS pairs.

Number of Characters Ranking Graph: Each vertex is a number of characters of a word, and each edge is a relation between preceding and succeeding number of characters in a word.

The graph ranking approach operates as shown in Figure 2 shows the process for generating a ranking graph. The process comprises six steps: (i) item set identification, (ii) ranking graph generation, (iii) vertex ranking and readability level allocation, (iv) text readability level frequency calculation, (v) text readability score calculation and (vi) conversion of text readability score to a point on a n point readability scale. During Step 1 the items of interest (words, POS tags, word and POS tag pairings or sizes of word), input document set, are identified. In the figure the set $\{A, B, C, D, E, F\}$ indicates the global set of items that might appear in the individual documents in the given document collection.The desired ranking graph, describing the entire document collection, is then generated in Step 2 where each vertex is an item, and each edge is a relation between the preceding and succeeding items. For example, A $\rightarrow$ E $\rightarrow$ D

Data: total degree of a node d
Result: text level of a node t ($\{hard, medium, easy, general\}$)
$gend$: The number documents in the document collection $\times$ 2;
$mind$: The minimum total degree in the given ranking graph;
$maxd$: The maximum total degree of the given ranking graph (which must be less than or equal to $gend$);
if $mind \leq d \leq (mind + maxd) \times 1/3$ **then**
  $t = hard$;
else if $(mind + maxd) \times 1/3 < d \leq (mind + maxd) \times 2/3$ **then**
  $t = medium$;
else if $(mind + maxd) \times 2/3 < d \leq maxd$ **then**
  $t = easy$;
else
  $t = general$;
end

Algorithm 1: Text readability level determination for the Graph Ranking approach

means A followed by E followed by D. In Step 3, the degree of each vertex in the ranking graph is extracted and the text readability level assigned to each vertex. The total degree of each vertex is the summation of the "indegree" (the number of head endpoints adjacent to a vertex) and the "outdegree" (the number of tail endpoints adjacent to a node) for exam vertex. The total degree of vertices are then sorted in descending order. The text readability level for each vertex is calculated using Algorithm 1. Note that a four point scale is used where the "General" level represents the situation where the vertex total degree is more than twice the size of document collection (the total number of documents in the collection). Note that using this algorithm the greater the total degree of a given vertex the more commonly the associated item occurs in the given document collection and consequently the "easier" the assigned readability level. The general category is therefore applied to very frequently occurring items.

In Step 4 the frequency with which each text readability level ({hard, medium, easy, general}) occurs is calculated with respect to each document (note that in the figure h equates to 'Hard', m to 'Medium' and so on), the items in each document from the collection will be allocated respect to their text levels. Note that the General text readability level is excluded from this process as such items occur so frequently that it is assumed that they play no role in assessing the overall readability of the documents in the given document collection. In Step 5 the text levels from Step 4 which have the highest frequency with respect to each document is selected as the text readability level for the document in question and a readability score derived. The readability score is the maximum frequency of the selected text readability level divided by the total frequency of all the readability levels in the document. Note that in Step 5 it is possible to have more than one text readability level if there exists more than one level that features a maximum frequency. In Step 6 we calculate a *readability grade* using the conversion table presented in Table 1. For ease of understanding we

Table 1. Text Readability Level to Readability Grade Conversion

Text Readability Level	Readability Grade
Hard	4
Medium-Hard	3.5
Medium, Easy-Hard, Easy-Medium-Hard	3
Easy-Medium	2.5
Easy	2
General	1

Table 2. Mappings Between Grade Values and Readability Levels

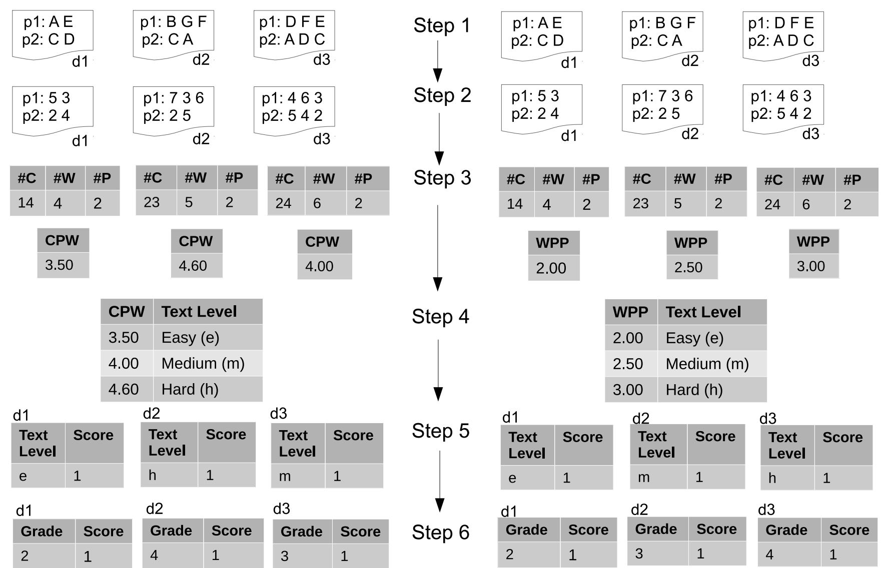
Readabilty Grade Values	5 Point Readability Scale	3 Point Readability scale
3.5 - 4.0	Very Hard	Hard
3.0 - 3.5	Hard	Hard
2.5 - 3.0	Medium	Medium
2.0 - 2.5	Easy	Medium
0.0 - 2.0	Very Easy	Easy

can convert the numeric readability grades to either a five points or a three point readability scale. Note that inspection of the table indicates that we can map the grade values using either a three or a five point readability scale. These scales were chosen because they are the most commonly used scale as reported in the literature.

4.2 Document Ranking

In the case of the document ranking methods occurrence count statistics are used. Two approaches are considered: (i) Characters Per Word (CPW) Document Ranking where documents are ranked according to average number of characters per word, and (ii) Words Per Paragraph (WPP) Document Ranking where documents are ranked according to average number of words per paragraph. The general process for determining document ranking using CPW is as shown in Figure 3(a) while that for WPP is as shown in Figure 3(b). In the figure the set $\{A, B, C, D, E, F, G\}$ represents the global set of words that may appear in the individual documents in the document collection, while the identifiers $p1$ and $p2$ indicate paragraphs (note that for illustrative purposes it is assumed that each document comprises two paragraphs). From the figure it can again be seen that we have a six step process: (i) item set identification, (ii) number of characters in word calculation, (iii) characters, word, and paragraph occurrence calculation, (iv) document ranking and readability level allocation, (v) text readability score calculation and (vi) conversion of text readability score to a point on a readability scale.

In Step 1 the items of interests (words in this case) are identified. In step 2 the number of characters per word in each document is determined. In Step 3 either the average CPW per document or the average WPP per document is calculated. CPW is the total number of characters in a document divided by the



(a) CPW Document Ranking (b) WPP Document Ranking

Fig. 3. Document Ranking Process

total number of words in a document. WPP is the total number of words in a document divided by the total number of paragraphs in the document. In the figure the labals "#C", "#W" and "#P" indicate the number of characters, the number of words, and the number of paragraphs in the document, respectively. In Step 4 from all documents in the collection is sorted in ascending order according to average CPW (WPP) and the text readability levels assigned to each the document. The text readability level of a documents is determined as shown in Algorithm 2. The intuition here is that the smaller the average number of characters per words (words per paragraph) the easier the text is to read.

For compatibility with the graph ranking approach in Step 5 a readability score of 1 is allocated to each document (because it is impossible to have more than one text readability level in a document using the document ranking approach). However, we still need to convert the text readability level to the readability grade, This is done in Step 6 using the conversion table presented in Table 1. Note that the readability grade and its score from the document ranking will be used in the hybrid ranking approach discussed in the next Sub-section.

4.3 Hybrid Ranking

In the hybrid approach all of the above six ranking methods are combined to produce a single "readability index" which we have called the CR Index (CRI). The CRI is calculated as follows.

$$\text{CRI} = \frac{1}{n} \times \sum_{i=1}^{n} (GL_i \times Sc_i), \tag{4}$$

where n is the number of ranking methods, GL is the readability grade (numerical value) for method i (from Steps 6 above) , and Sc is the score associated with the readability grade from the ranking technique. In this paper, $n = 6$. We can interpret the CRI value as a readability point as in the case graph and document ranking approaches.

Data: dr the document ranking value
Result: text level of document t $\{easy, medium, \text{or } hard\}$
$mindr$: the minimum value of the document ranking method;
$maxdr$: the maximum value of the document ranking method;
if $mindr \leq dr \leq (mindr + maxdr) \times 1/3$ **then**
  $t = easy$;
else if $(mindr + maxdr) \times 1/3 < dr \leq (mindr + maxdr) \times 2/3$ **then**
  $t = medium$;
else
  $t = hard$;
end

Algorithm 2: Text level determination in Document Ranking

5 Experimental Settings and Results

From the above we have seven alternative ranking mechanisms: (i) Word Ranking Graph, (ii) POS Ranking Graph, (iii) Word and POS Ranking Graph, (iv) Number of Characters Ranking Graph, (v) CPW Document Ranking, (vi) WPP Document Ranking and (vii) CRI Hybrid Ranking. In this section, the evaluation of the proposed ranking techniques is presented. The objectives of the evaluation were as follows:

- **Individual Classification**: Which ranking method is most appropriate for classifying documents with respect to each of the individual identified class labels: secondary school (S), high school (H) and undergraduate students (U).
- **Overall Classification**: Which ranking method produces the best overall classification result.
- **Graph Ranking Comparison**: Which of the four proposed graph ranking methods produced the best performance.
- **Document Ranking Comparison**: Which of the two proposed document ranking methods produced the best performance.
- **Hybrid Ranking Comparison**: Whether the propose CRI method provided an advantage or not, and why.

In the case of the first and second objectives the operation of the proposed ranking mechanism was compared with two baseline methods: (i) the Coleman-Liau Index (CLI) (Coleman and Liau 1975) and (ii) the Automated Readability Index (ARI) (Senter and Smith 1967). Both CLI and ARI were discussed in the previous work section above (Section 2). These baseline methods were chosen

Table 3. Mapping the output of CLI and ARI to the dataset

Text Readability Level	Abstract Level	CLI (Grade)	ARI (Age)
Hard	U	13+	20+
Medium	H	10-12	17-19
Easy	S	7-9	14-16

Table 4. Dataset Statistics

Text Readability Level	Abstract Level	# Documents	# Words	#Characters
Hard	Undergraduate (U)	47	13,675	53,164
Medium	High school (H)	44	13,559	47,783
Easy	Secondary school (S)	44	14,704	51,907
	total	135	41,938	152,854

because they are commonly referred to in the literature and because they are well suited to syllabic alphabetic languages such as the Thai language (the formulae used to calculate CLI and ARI do not require syllable counts as required by some established measures). However, so as to achieve compatibility with our proposed measures, we have mapped CLI "grade levels" and ARI "age levels" to our three reading grade levels ($\{Hard, Medium, Easy\}$) as shown in Table 3.

The evaluation criteria used were precision, recall and the F-score. Precision indicates the percentage of detected documents that matches a target readability level. Recall indicates the total number of target documents in the dataset that are retrieved. The F-score (or F-measure) gives an overall performance measure. The remainder of this section is organised as follows. The data set used for the evaluation is presented in Sub-section 5.1. The results are presented in Sub-section 5.2. A discussion, in the context of the above objectives is then given in Sub-section 5.3.

5.1 Evaluation Data

To evaluate the proposed approach we collected scientific project abstracts[1] for secondary school (S), high school (H) and undergraduate (U) levels of education. These abstracts summarised scientific projects undertaken during tuition. In total 135 abstracts were collected. Some statistics concerning these abstracts are presented in Table 4.

5.2 Results

The experimental results obtained are shown in Tables 5, 6 and 7 with respect to precision, recall, and the F-score. Best results are highlighted in bold font.

5.3 Discussion

In this sub-section the results presented in the foregoing sub-section are discussed with respect to the five evaluation objectives identified above.

[1] http://www.vcharkarn.com/project/

Table 5. Experimental Results (Precision)

Readability Technique	Precision (%)			
	U (Hard)	H (Medium)	S (Easy)	Average
Word Ranking Graph	34.81	0.00	0.00	11.60
POS Ranking Graph	17.39	35.00	29.17	27.19
Word and POS Ranking Graph	34.81	0.00	0.00	11.60
Number of Characters Ranking Graph	30.77	20.51	33.73	28.34
CPW Document Ranking	**60.00**	35.56	**44.44**	**46.67**
WPP Document Ranking	24.44	**40.90**	26.67	30.67
CRI Hybrid Ranking	35.56	36.96	31.11	34.54
CLI (baseline)	0.00	0.00	19.51	6.50
ARI (baseline)	34.81	0.00	0.00	11.60

Table 6. Experimental Results (Recall)

Readability Technique	Recall (%)			
	U (Hard)	H (Medium)	S (Easy)	Average
Word Ranking Graph	**100.00**	0.00	0.00	33.33
POS Ranking Graph	8.51	31.82	47.73	29.35
Word and POS Ranking Graph	**100.00**	0.00	0.00	33.33
Number of Characters Ranking Graph	8.51	18.18	**63.63**	30.11
CPW Document Ranking	57.45	36.36	45.45	**46.42**
WPP Document Ranking	23.40	**40.90**	27.27	30.52
CRI Hybrid Ranking	34.04	38.64	31.82	34.83
CLI (baseline)	0.00	0.00	18.19	6.06
ARI (baseline)	**100.00**	0.00	0.00	33.33

Individual Classification. The first objective of the evaluation was to determine the most appropriate ranking method with respect to each of the individual identified class labels. From Table 7 it can be seen that the CPW Document Ranking is best suitable for class U and S (accuracies of 58.7% and 44.94%), while the WPP Document Ranking is best suited to class H (40.90%). The results indicate that students in higher education tend to write scientific abstracts using longer words and write more words per paragraph.

Overall Classification. The second objective was directed at the identification of the best overall ranking method. The CPW Document Ranking gives the highest F-score (46.53%), followed by the CRI hybrid ranking (34.67%), and the WPP Document Ranking (30.59%). Thus the average number of characters per word plays an important role in classifing the text readability level of the Thai language abstracts. Moreover, it can also be concluded that the variety of words and frequency of POS tags used in the texts are also useful to differentiate between levels of text readability.

Graph Ranking Comparison. The third evaluation objectives was directed at the graph ranking methods. The POS Ranking Graph gives the highest F-score (26.99%) amongst the graph ranking techniques, followed by the Number of Characters Ranking Graph (25.57%). Most of the graph ranking techniques are suited to predicting the hard level. The results imply that the graph ranking technique is useful for ranking the text readability of documents which have

Table 7. Experimental Results (F-score)

Readability Technique	F-score (%)			
	U (Hard)	H (Medium)	S (Easy)	Average
Word Ranking Graph	51.65	0.00	0.00	17.22
POS Ranking Graph	11.42	33.33	36.21	26.99
Word and POS Ranking Graph	51.65	0.00	0.00	17.22
Number of Characters Ranking Graph	13.33	19.28	44.10	25.57
CPW Document Ranking	**58.70**	35.96	**44.94**	**46.53**
WPP Document Ranking	23.91	**40.90**	26.97	30.59
CRI Hybrid Ranking	34.78	37.78	31.46	34.67
CLI (baseline)	0.00	0.00	18.82	6.27
ARI (baseline)	51.65	0.00	0.00	17.22

many words, longer words, and more open-classed words (newly created nouns, verbs, and adjectives).

Document Ranking Comparison. In the case of the two document ranking methods, the CPW Document Ranking achieves the highest F-score (46.53%) of all the techniques considered. Moreover, the maximum F-score obtained from the Document Ranking technique is higher than the maximum F-score obtained from the Graph Ranking technique by 19.54% (46.53 - 26.99). The results show that the average number of characters per word, in general, is more significant when classifying text readability than the average number of words per paragraph.

Hybrid Ranking Comparison. The last evaluation objective was to consider the CRI Hybrid Ranking measure in terms of the other proposed ranking methods and the baseline methods. The CRI hybrid ranking is the second best of all proposed ranking methods (34.67% using the F-score). Note also that our proposed seven ranking techniques give better performance than the traditional baselines. However, the CRI Hybrid Ranking measure can be used as another baseline for ranking the text readability of documents with related content when applying these ranking techniques to other kinds of texts and languages. In this paper, the CPW Document Ranking technique is the only ranking technique that gives an overall nest performance better than CRI. We can conclude that the CPW Document Ranking method is the most suitable method for assessing the text readability of this document collection .

6 Conclusion and Future Work

This work presented in this paper has proposed seven content-based readability measures socially designed for syllabic alphabetic languages. The experimental results show that our technique achieve an overall best performance of up to 46.53% using the F-score measure, which is significantly higher than that obtained using the traditional readability formulas considered. The proposed techniques are also suitable for ranking other kinds of the related texts (not just syllabic alphabetic language texts).

References

Ager, S.: Omniglot - writing systems and languages of the world. http://www.omniglot.com, 2014-05-21

Al-Khalifa, H.S., Al-Ajlan, A.A.: Automatic readability measurements of the arabic text: An exploratory study. The Arabian Journal for Science and Engineering 35(2C), 103–124 (December 2010)

Bright, W.: A matter of typology: Alphasyllabaries and abugidas. Studies in the Linguistic Sciences 30(1), 63–71 (2000)

Coleman, M., Liau, T.L.: A computer readability formula designed for machine scoring. Journal of Applied Psychology 60, 283–284 (1975)

Daniels, P.T.: Fundamentals of grammatology. Journal of the American Oriental Society 119(4), 727–731 (October-December 1990)

Daowadung, P., Chen, Y.H.: Using word segmentation and svm to assess readability of thai text for primary school students. In: The Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE). pp. 170–174 (2011)

Flesch, R.: A new readability yardstick. Journal of Applied Psychology 32, 221–233 (1948)

Gunning, R.: The Technique of Clear Writing. McGraw-Hill (1952)

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., Chissom, B.S.: Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. TN: Naval Air Station, Memphis (February 1975)

Kitson, H.D.: The Mind of the Buyer: A Psychology of Selling. Macmillan, New York (1921)

McLaughlin, G.H.: Smog grading - a new readability formula. Journal of Reading 22, 639–646 (1969)

Meknavin, S., Charoenpornsawat, P., Kijsirikul, B.: Feature-based thai word segmentation. In: Proceedings of the Natural Language Processing Pacific Rim Symposium 1997 (NLPRS'97) (1997)

Rabin, A., Zakaluk, B., Samuels, S.: Readability: Its past, present & future, chap. Determining Difficulty Levels of Text Written in Languages Other Than English, pp. 46–76. Newark DE: International Reading Association (1988)

Senter, R., Smith, E.: Automated Readability Index. Wright-Patterson Air Force Base (November 1967)

Sinha, M., Sharma, S., Dasgupta, T., Basu, A.: New readability measures for bangla and hindi texts. In: Proceedings of COLING 2012: Posters. pp. 1141–1150 (December 2012)