Analysing Features of Lecture Slides and past Exam Paper Materials
Towards Automatic Associating E-materials for Self-revision

Petch Sajjacholapunt and Mike Joy
Department of Computer Science, University of Warwick, Coventry, U.K.

Keywords: Information Retrieval, Technical Terms Extraction, Technology Enhanced Learning.

Abstract: Digital materials not only provide opportunities as enablers of e-learning development, but also create a new challenge. The current e-materials provided on a course website are individually designed for learning in classrooms rather than for revision. In order to enable the capability of e-materials to support a students revision, we need an efficient system to associate related pieces of different e-materials. In this case, the features of each item of e-material, including the structure and the technical terms they contain, need to be studied and applied in order to calculate the similarity between relevant e-materials. Even though difficulties regarding technical term extraction and the similarities between two text documents have been widely discussed, empirical experiments for particular types of e-learning materials (for instance, lecture slides and past exam papers) are still rare. In this paper, we propose a framework and relatedness model for associating lecture slides and past exam paper materials to support revision based on Natural Language Processing (NLP) techniques. We compare and evaluate the efficiency of different combinations of three weighted schemes, term frequency (TF), inverse document frequency (IDF), and term location (TL), for calculating the relatedness score. The experiments were conducted on 30 lectures ($\sim$ 900 slides) and 3 past exam papers (12 pages) of a data structures course at the authors' institution. The findings indicate the appropriate features for calculating the relatedness score between lecture slides and past exam papers.

1 INTRODUCTION

E-materials that are normally provided on a course website comprise lecture slides, past exam papers, homework and assignments, and a list of related textbooks. They are likely to be just supplementary materials for the class, and are not intended specifically for self-study. Although it is difficult to deny that attending lectures must be the first priority of a student, rather than paying attention only to self-study, students still need self-study to recall knowledge before an examination.

In order to succeed in preparing materials before examinations, students need to understand the course materials that are delivered by the lecturers. These e-course materials, however, have not yet maximised the advantage of being digital media. Most e-materials on a course website are closed materials in that they are independent with no possibility of free linking and combining features (Krnel and Barbra, 2009). Many studies illustrate the difficulties of using online e-materials. For example Mertens et al. (2006) states that searching through a pile of digital content is time-consuming and requires skill and knowledge of using computers and computer applications. Moreover, students sometimes suffer from inadequate lecture note content or the poor linking of key concepts between course materials.

The main goal of our research is to address the difficulties of using e-materials by associating relevant content among different types of e-materials. In order to achieve this goal, the structure of each item of material, has to be studied and determined, using techniques including natural language processing (NLP) for technical term extraction. In this paper, we construct a framework for associating e-materials based on existing NLP techniques. We also experiment with comparing potential features of e-materials that are appropriate when constructing a weighting scheme for calculating the relatedness scores between different e-materials. At this preliminary stage, we focus our experiment on the structure of lecture slides and past exam papers, as they are the primary set of materials that students tend to review (Sajjacholapunt and Joy, 2014). Additional e-materials will be examined in the near future. The detail of the framework, tech-

Sajjacholapunt P. and Joy M..
Analysing Features of Lecture Slides and past Exam Paper Materials - Towards Automatic Associating E-materials for Self-revision.
DOI: 10.5220/0005371201690176
In *Proceedings of the 7th International Conference on Computer Supported Education* (CSEDU-2015), pages 169-176
ISBN: 978-989-758-107-6

niques and experimental results are discussed in the following subsections.

2 RELATED WORK

Lecture slides are a common material that is used for presentation of lectures in class, and many studies have focused on the use of lecture slide in a classroom context. For example, Frey and Birnbaum (2002) stated that instructors do not wish to make PowerPoint slides as a substitute for lecturing. They aim at presenting only main ideas, but not a summary of the lecture. Holmes (2004) also stated that a presentation can serve as a guide for listeners or readers, but it can never be a medium that is capable of replacing a skilled teacher. What is possible is that it can be used to conceal poor-quality teaching by providing validity, albeit without gains in terms of the results of learning (Pros et al., 2013).

Many studies emphasise improving presentation video based on presentation slides such as synchronising a speech transcript and presentation slides (Chen and Heng, 2003), annotating each video segment with the related presentation slide(Sack and Waitelonis, 2006) and improving search performance in video recording presentation by indexing the presentation slides(Vinciarelli and Odobez, 2006; Le et al., 2008).

There are also other related works that aim at improving presentation slides by adding their own information. Hayama and Kunifuji (2011) describes a method to extract related pieces of information from presentation slides and display them properly as preview information. Hill (2011) purposes the idea of checking and grading presentation file assignments by comparing student documents with a correct version of assignment. Yuanyuan and Kazutoshi (2011) concentrates on improving browsing performance of presentation slides by recommending other related presentation slides and identifying their relationship.

Many research projects have analysed and reused content in presentation slides for summarising slides content for a quick overview, building an index for quick access to other digital media and for linking related slides together. These projects have not yet concentrated on improving presentation slides based on other e-course materials.

At this stage, we considered to use past exam paper as a supplementary material for enriching lecture slides content. This is because past exam paper is also a commonly used resource during revision. Most students use a past exam paper to help them get familiar with terminologies and exam style questions that will be used in the actual exam (du Boulay, 2011). They also are used to help them work out time required for each question as well as to identify subject areas to focus on in their revision. Enhancing e-lecture slide based on past exam paper is a challenge problem. We need to understand the structure of both materials including their features.

3 METHODOLOGY

In order to implement a system for automatic association between lecture slides and past exam papers, we need a framework as a guideline for designing the system. The framework for associating e-lecture slides with related past exam paper materials is illustrated in Figure 1. This framework was designed based on common NLP techniques presented in (Baeza-Yates and Ribeiro-Neto, 2011). This framework starts by determining the potential features of the selected e-materials, extracting candidate terms and calculating the similarity table of these two documents.

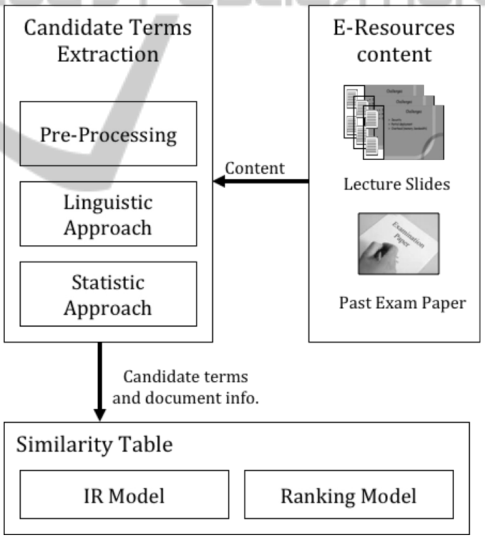

Figure 1: The framework for associating e-lecture slides with related past exam paper materials.

3.1 Determining Potential Features

Determining key features of individual e-materials is a challenging issue because their structures are normally different. By their nature, lecture slides and exam paper materials contain a lot fewer terminologies and content than does an e-book. In order to create a link between these related materials, it is necessary to understand their structure. The following section explores lecture slide and exam paper structure, as well as identifying potential common features.

3.1.1 Lecture Slides

The format of lecture slides usually contains two major parts — the title and the content as presented in Figure 2. The former offers an overview of the content that is in the body, while the latter contains key information related to the title. The content is mostly presented as a bullet list which allows students to quickly obtain information.

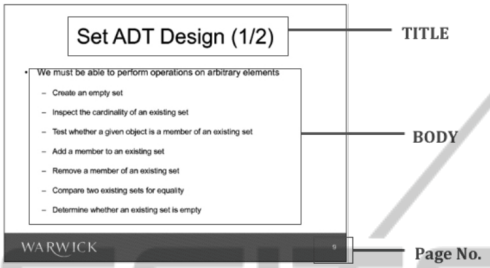

Figure 2: Lecture Slides Structure.

The format of lecture slides can be altered based on the style of the author. Two major styles that the author normally chooses are media selection and media format. Media selection is a process that is concerned with what media should be used in the lecture slide, such as figure, table or text information. Media format is a process of selecting the property of the media that is to be presented on the slide. For example, the text information can be adjusted in terms of its font size, colour or location.

3.1.2 Past Exam Paper

Generally, the format of exam papers contains two major parts — cover page and question page. The former, as presented in Figure3, consists of general information with regard to course information, date and time, exam time and length, as well as instructional information. The information on the cover page usually does not contain any subject regarding the course content.

The latter, as presented in Figure 4, consists of a set of main questions and sub-questions. Sub-questions are not only defined in the form of an interrogative sentence, but sometimes are also defined in the form of an affirmative sentence or equation following a core question in order to measure abilities in terms of explanation and discussion. A mark for each sub-question is commonly provided to indicate the level of difficulty.

3.2 Candidate Term Extraction

The following subsections explain the processes and approaches for extracting candidate terms, including techniques that are considered appropriate in the research.

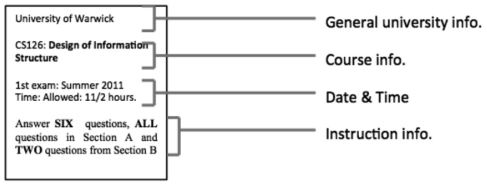

Figure 3: Cover page of exam paper structure.

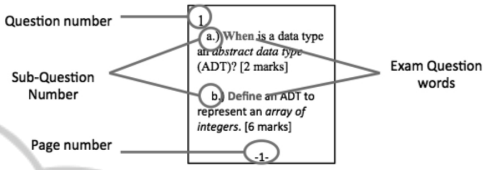

Figure 4: Question page of exam paper structure.

3.2.1 Data Pre-Processing

In order to convert the current format of e-materials to plain text, we consider applying the six common pre-processing techniques presented in Table 1. The iText (Lowagie, 2007), a Free Java-PDF library 5.5.0 API, is used as a library for converting PDF E-Materials to text format.

Table 1: Common pre-processing tasks.

Converted Document to Plain Text
Sentence Segmentation
Tokenisation
Part-of-Speech Tagging
Stemming and Lemmatisation
Stop-words Filtering

Converting E-Lecture Slides and Exam Paper to Plain Text

Converting a document to plain text is a process of transforming a non-plain text document (e.g., JPEG, .PDF, .PPT) to a plain text document (Foo, 2012). Much research has considered use of e-lecture slides in PowerPoint (PPT) format. In the real world, however, most of the provided e-lecture slides and past exam papers are presented in PDF format for accessibility and security reasons. We thus mainly focus on converting the PDF document to plain text. The iText (Lowagie, 2007), a Free Java-PDF library 5.5.0 API, is used as a library for transforming PDF-to-TEXT. The Apache OpenNLP [1] was selected as a

[1] http://opennlp.apache.org

tool to perform sentence segmentation, tokenisation, part-of-speech tagging, stemming and lemmatisation in this research because it is an open-source tool and its popularity for using in term extraction system. Finally, the 517-stop words list built by Salton (1971) for the experimental SMART information retrieval system at Cornell University was used in stop-word filtering process.

3.2.2 Linguistic and Statistic Approaches for Term Extraction

Techniques for extracting terminologies can mainly be classified into two approaches (Pazienza et al., 2005; Conrado et al., 2013), which are (1) linguistic approaches and (2) statistical approaches. The former approach is to deal with pure linguistic properties to extract terminologies, such as part-of-speech patterns and words related to the stem. The latter approach applies statistics to measure the degree of termhood of candidate terms to decide whether to choose the term. For example, basic term frequencies and word length counts. The details of these two approaches are shown in the following section.

Use of the linguistic approaches or statistical approach alone does not provide an effective result (Pazienza et al., 2005). There are only a few works that use only the statistical method without touching any of the linguistic approaches (Jones et al., 1990; Salton et al., 1975) . In this research, we therefore chose to use a hybrid approach for the lecture slide materials. The open-source JATEtoolkit [2] were used because it can perform both approaches. However,we do not apply the statistical approach to extract candidate terms in past exam paper because the nature of exam paper contains low term frequencies for which a statistical approach may not be useful. For the statistical approach, the C-value algorithm was tested first because it is outperformed among other algorithms and it does not need reference corpus as a training data (Pazienza et al., 2005).

3.3 Similarity Calculation

Identifying a lecture slide that relates to relevant past exam papers is another challenge. The degree of similarity between each item of material needs to be calculated. Having analysed the structure of both the lecture slides and exam papers (in the subsection 3.1), we considered three potential features, for testing the effectiveness of association between lecture slides and past exam papers, which are the following.

- **Terms frequency (TF)**: is a measure of frequency of a candidate term that appears in the target document. The following equations show how normalised TF is computed (Salton and Buckley, 1988).

  For the candidate term t_i in the document d_j, the normalised term frequency $tf(i,j)$ derives from a fraction of the raw frequency $freq(i,j)$ and the maximum of the raw frequency $\max_{i,j} freq(i,j)$ of term t_i over all terms mentioned in any documents d_j.

$$tf(i,j) = \frac{freq(i,j)}{\max_{i,j} freq(i,j)}. \quad (1)$$

- **Inverse Document frequency (IDF)**: Sometime there is a term that has a high frequency in the document but it does not represent the document because it also has a high frequency in another document. In this case, the statistical term extraction technique called TF-IDF value is applied. The TF-IDF is computed by the following equation.

  Let N be the total number of documents and n_i be number of the document that a candidate term appears. The inverse document frequency of term t_i is computed by (Salton and Buckley, 1988).

$$idf(i) = log\frac{N}{n_i} \quad (2)$$

  The common term weighting score $tf-idf(i,j)$ is computed by (Salton and Buckley, 1988).

$$tf-idf(i,j) = tf(i,j) \times idf(i) \quad (3)$$

- **Terms Location (TL)**: This is a location where the term is displayed in the document. The significance of the term can be changed based on the location. For example, in the lecture slides, terms that appear in titles should be more important than terms that appear on in body position.

As we mentioned earlier, candidate terms that present in past exam paper likely be a subject area that the student must know before the exam. We then can assume that term that appears in the lecture slide and also appears in the exam paper is considered as a key for linking these two documents. Simple Boolean matching (of candidate terms) as presented in Figure 5 thus, is a promising technique for identifying the lecture slides that relates to past exam paper.

In this research, we chose to investigate and compare the three mentioned features to identify the best

[2] https://code.google.com/p/jatetoolkit/

Figure 5: Technical term association.

features for scoring and ranking relatedness between lecture slides and exam papers. The experiment was done using the four following cases.

Denote :
$w^{(\cdot)}(i,j)$ = Weight of term t_i in slide S_j,
$freq(i,j)$ = Number of term t_i in slide S_j,
H_j := Header of slide S_j,
W_H := Weight of Header (vary from range 2 - 10).

- **Case 1**: Term Frequency (TF) feature. Calculate the TF score of candidate terms that appear in both slide S_j and past exam paper P_k.

$$w^1(i,j) = \frac{freq(i,j)}{\max_{i,j} freq(i,j)}. \quad (4)$$

- **Case 2**: Term Frequency (TF) feature with term location (TL) adjustment. In addition to Case 1, candidate terms located in the header of the slide S_j have more weight than candidate terms that appear on the body slide.

$$w^2(i,j) = \frac{\hat{freq}(i,j)}{\max_{i,j} \hat{freq}(i,j)}, \quad (5)$$

where

$$\hat{freq}(i,j) = \begin{cases} freq(i,j) & \text{if } i \notin H_j, \\ freq(i,j) \cdot W_H & \text{if } i \in H_j. \end{cases}$$

- **Case 3**: Term Frequency and Inverse Document Frequency (TF-IDF) feature. Calculate the TF-IDF score of candidate terms that appear in both slide S_j and past exam paper P_k.

$$w^3(i,j) = \frac{freq \times idf(i,j)}{\max_{i,j} freq \times idf(i,j)}. \quad (6)$$

- **Case 4**: Term Frequency and Inverse Document Frequency (TF-IDF) feature with term location (TL) adjustment. In addition to Case 3, candidate terms located in the header of the slide S_j have more weight than candidate terms that appear only in the body slide.

$$w^4(i,j) = \frac{\hat{freq} \times idf(i,j)}{\max_{i,j} \hat{freq} \times idf(i,j)}, \quad (7)$$

where

$$\hat{freq} \times idf(i,j) = \begin{cases} freq \times idf(i,j) & \text{if } i \notin H_j, \\ freq \times idf(i,j) \cdot W_H & \text{if } i \in H_j. \end{cases}$$

Relatedness, $RScore^{(\cdot)}(j,k)$, is a score of similarity between lecture slide S_j and past exam paper P_k where term i is presented in both S_j and P_k . The *RScore* ranking model is calculated as the following equation.

$$RScore^{(\cdot)}(j,k) = \begin{cases} \frac{\sum_{i \in j} w^{(\cdot)}(i,j)}{\max_j(\sum_{i \in j} w^{(\cdot)}(i,j))} & \text{if } i \in k, \\ 0 \quad \text{otherwise.} \end{cases} \quad (8)$$

A higher relatedness score implies that more information related to a past exam paper is contained in a lecture slide. The appropriate relatedness threshold, therefore, has to be defined. This can be achieved by using different relatedness thresholds to retrieve documents and measure their effectiveness. The results will be compared with the answer set to evaluate the accuracy of different cases. The expected results are a combination of features that can provide the most fitting retrieved results including a similarity threshold and weight of terms located in the header of lecture slides that can produce the most accurate result.

$$\text{Precision} = \frac{|\text{answer set} \cap \text{retrieved document}|}{\text{retrieved document}} \quad (9)$$

$$\text{Recall} = \frac{|\text{answer set} \cap \text{retrieved document}|}{\text{answer set}} \quad (10)$$

$$\text{F-Score} = 2 \cdot \frac{\text{Precision}}{\text{Recall}} \quad (11)$$

4 EXPERIMENTS

In order to perform the experiment, we designed a system based on the above-mentioned framework. The four cases weighting scheme will be investigated to determine the appropriate weighting scheme for the proposed framework.

4.1 Datasets for Training

The datasets for training the system were obtained from a corpus of undergraduate lecture slides at the Department of Computer Science, University of Warwick. The course CS126 Design of Information Structures was selected because it provided sufficient materials such as lecture slides, past exam papers as well as a recommended textbook for the purposes of our study.

Past exam paper in the module were available for 3 years. All three past exam papers (12 pages) and 30 lectures ($\sim$ 900 lecture slide pages) were selected for the experiment. The answer set were manually defined by matching lecture slide pages with related past exam papers based on the rule that content in a lecture slide page must contains technical terms that appear in a past exam paper.

Result of related documents in different cases of calculating similarity table would be retrieved to compare with the answer set based on different similarity threshold. The precision (Eq.9), recall (Eq. 10), and F-score (Eq.11) were used as the evaluation results for comparison.

4.2 Evaluation Results

As part of the evaluation for determining the effective feature for the proposed framework, we studied the effectiveness of TF and TF-IDF weighting schemes by adjusting the weight based on the term location. Bar chart is used to represent the precision, recall, f-score value at different similarity threshold. Ten bars at each similarity threshold in the bar chart represent different adjusted weight (1-10) from left to right.

4.2.1 Weighting Approach of TF Feature

At any similarity threshold in Figure 6, increasing the weight for the terms in the header location does not greatly affect the average precision score. This preliminary result suggests that the term in the header location is relatively unimportant compared to the relatedness score. The highest precision score is at 27.59% with a weight >than 4 and the relatedness threshold 0.25. It is improved by only 1.84% from the highest precision score associated with using the TF feature with no adjusted weight.

The trend of the average recall (in Figure 7) shows that the TF features decrease when we increase the relatedness threshold from 0.05 to 0.55, then remain stable after that. This is almost exactly the same at any adjusted weight. This also suggests that increasing the weight of terms that are located in the headers of the lecture slides does not impact a great deal on the recall score.

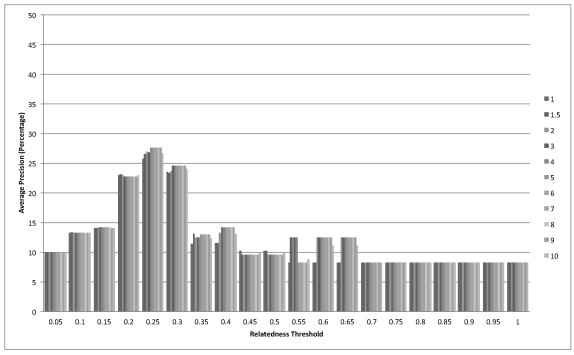

Figure 6: The average precision of TF feature on different weight of terms in the header.

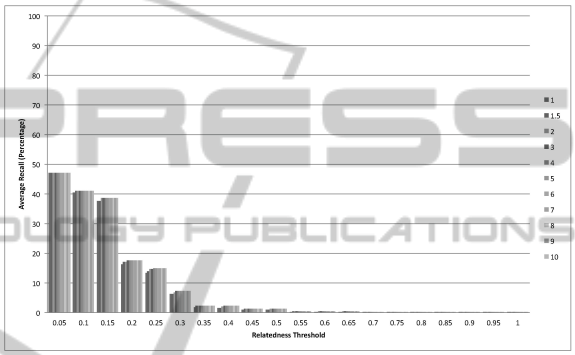

Figure 7: The average recall of TF feature on different weight of terms in the header.

The F-Score, in Figure 8, confirms that giving more weight to the terms located in the headers of the lecture slides does not greatly affect the relatedness score. The highest F-score, presented is 18.92% with a relatedness threshold 0.15 and the weight of the header being 2. It only increases 0.24% from the highest average F-score associated with using the TF feature without weighting.

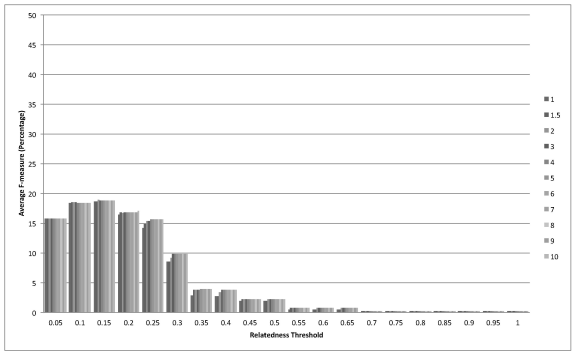

Figure 8: The average F-score of TF feature on different weight of terms in the header.

4.2.2 Weighting Approach of TF-IDF Feature

When applying the IDF score, the trend in terms of the precision score (see Figure 9) fluctuates more than when using only the TF score in Figure 6. From the relatedness threshold 0.25 to the relatedness threshold 0.55, there was a considerable fall in the percentage of the average precision on the TF feature, while the graph of the TF-IDF feature still continues increasing towards the highest average precision which is 34.58% at the relatedness threshold 0.5 , when the weight is 1.5. Thus it can be stated that applying IDF can eliminate some of the non-technical terms extracted from the lecture slides.

At some point of the relatedness threshold, for instance at a relatedness threshold of 0.35, the greater the weighting given to the terms located in the headers, the greater the average precision score. On the other hand, at a relatedness threshold of 0.5, giving more weight can return a lower percentage in terms of average precision.

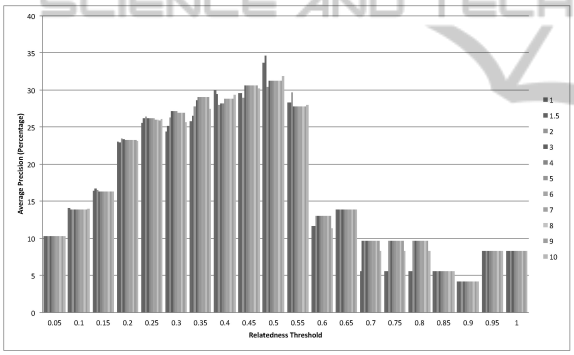

Figure 9: The average precision of TF-IDF feature on different weight of terms in the header.

Bar chart trends (see Figure 10) are similar to the average recall of TF presented in Figure 7. This can confirm that using the TF-IDF feature and giving different weights to the term location cannot increase the number of relevant documents retrieved.

Giving different weights to the term location with the TF-IDF feature does not greatly affect the F-Score. The highest average F-score presented in Figure 11 is 22.93% of a similarity threshold of 0.2 and weight 2. This is just 1% higher than using only the TF-IDF feature.

5 CONCLUSIONS AND FUTURE WORK

In conclusion, the idea of increasing the relatedness threshold is to eliminate non-relevant documents by

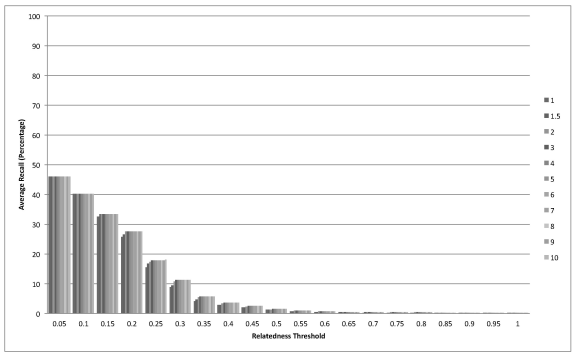

Figure 10: The average recall of TF-IDF feature on different weight of terms in the header.

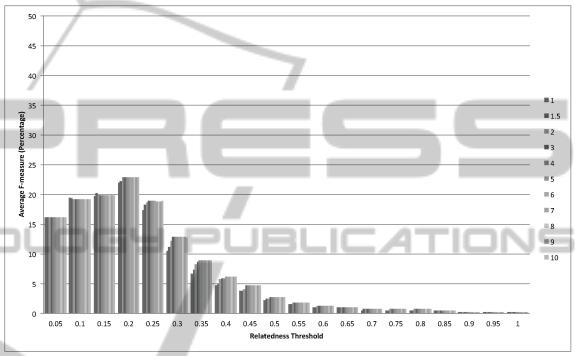

Figure 11: The average F-score of TF-IDF feature on different weight of terms in the header.

retrieving only those documents that have a relatedness score above the threshold. Thus, a target document that has a high relatedness score appears to have a high degree of relevance to the source documents. This is supposed to improve the precision score.

The recall score is initially dependent on the relevant document extraction method. The more relevant the documents (in the answer set) that are retrieved, the greater the recall score. Increasing the relatedness threshold to eliminate the non-relevant documents therefore, cannot improve the recall score. Although increasing the relatedness score does not increase the recall scores, it can still retain or slow down the reduction of the recall score.

In this experiment, however, increasing the relatedness threshold at some point can improve the average precision score but not the recall score. It can be characterized by the fact that, using TF and TF-IDF features, both with and without giving weight to the term location, can eliminate the non-relevant documents from being retrieved. The sharp reduction in recall scores at the initial range of the relatedness threshold between 0.05–0.3 in all cases, implies that some of the relevant documents that were retrieved with a lower relatedness score were elimi-

nated. Thus it can be concluded that the TF-IDF feature is the best feature when it comes to eliminating non-relevant documents and to improving precision among other test features. The term weight location only has a minimal impact on this, and is not significant.

In future, the research will still need to find out in detail why the term location feature is not much help, as well as identifying how to improve a recall score at the beginning of extracting a candidate set of related documents.

REFERENCES

Baeza-Yates, R. and Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition) (ACM Press Books)*. Addison-Wesley Professional, 2 edition.

Chen, Y. and Heng, W. J. (2003). Automatic synchronization of speech transcript and slides in presentation. In *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, volume 2, pages II–568–II–571 vol.2.

Conrado, M., Pardo, T. A. S., and Rezende, S. O. (2013). *Exploration of a Rich Feature Set for Automatic Term Extraction*, volume 8265 of *Lecture Notes in Computer Science*, pages 342–354. Springer Berlin Heidelberg.

du Boulay, D. (2011). *Study Skills For Dummies*. John Wiley & Sons, (uk edition) edition.

Foo, J. (2012). *Computational Terminology : Exploring Bilingual and Monolingual Term Extraction*. PhD thesis, Linkoping University.

Frey, B. A. and Birnbaum, D. J. (2002). Learners' perceptions on the value of powerpoint in lectures. (ED467192):10.

Hayama, T. and Kunifuji, S. (2011). Relevant piece of information extraction from presentation slide page for slide information retrieval system. In Theeramunkong, T., Kunifuji, S., Sornlertlamvanich, V., and Nattee, C., editors, *Knowledge, Information, and Creativity Support Systems*, volume 6746 of *Lecture Notes in Computer Science*, pages 22–31. Springer Berlin Heidelberg.

Hill, T. G. (2011). Word grader and powerpoint grader. *ACM Inroads*, 2(2):34–36.

Holmes, W. N. (2004). In defense of powerpoint. *IEEE Computer*, 37(7):98–99.

Jones, L. P., Gassie Jr, E. W., and Radhakrishnan, S. (1990). Index: The statistical basis for an automatic conceptual phrase-indexing system. *Journal of the American Society for Information Science*, 41(2):87–97.

Krnel, D. and Barbra, B. (2009). Learning and e-materials. *Acta Didactica Napocensia*, 2(1):97–108.

Le, H. H., Lertrusdachakul, T., Watanabe, T., and Yokota, H. (2008). Automatic digest generation by extracting important scenes from the content of presentations. In *Database and Expert Systems Application, 2008. DEXA '08. 19th International Workshop on*, pages 590–594.

Lowagie, B. (2007). *IText in Action: Creating and Manipulating PDF*. Manning Pubs Co Series. Manning.

Mertens, R., Farzan, R., and Brusilovsky, P. (2006). Social navigation in web lectures. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, HYPERTEXT '06, pages 41–44, New York, NY, USA. ACM.

Pazienza, M., Pennacchiotti, M., and Zanzotto, F. (2005). Terminology extraction: An analysis of linguistic and statistical approaches. In Sirmakessis, S., editor, *Knowledge Mining*, volume 185 of *Studies in Fuzziness and Soft Computing*, pages 255–279. Springer Berlin Heidelberg.

Pros, R. C., Tarrida, A. C., Martin, M. d. M. B., and Amores, M. d. C. C. (2013). Effects of the powerpoint methodology on content learning. *Intangible Capital*, 9(1):184–198.

Sack, H. and Waitelonis, J. (2006). Automated annotations of synchronized multimedia presentations. In *In Workshop on Mastering the Gap: From Information Extraction to Semantic Representation, CEUR Workshop Proceedings*.

Sajjacholapunt, P. and Joy, M. (2014). Exploring patterns of using learning resources as a guideline to improve self-revision. In *INTED2014 Proceedings*, 8th International Technology, Education and Development Conference, pages 5263–5271. IATED.

Salton, G. (1971). *The SMART Retrieval System–Experiments in Automatic Document Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513 – 523.

Salton, G., Yang, C.-S., and Yu, C. T. (1975). A theory of term importance in automatic text analysis. *Journal of the American society for Information Science*, 26(1):33–44.

Vinciarelli, A. and Odobez, J. (2006). Application of information retrieval technologies to presentation slides. *Multimedia, IEEE Transactions on*, 8(5):981–995.

Yuanyuan, W. and Kazutoshi, S. (2011). A browsing method for presentation slides based on semantic relations and document structure for e-learning. *IPSJ Journal*, 52(12):15p.