Test-driven summarization: combining formative assessment with teaching document summarization

Test-driven summarization: combining formative assessment with teaching document summarization

Luca Cagliero, Laura Farinetti, Elena Baralis
Politecnico di Torino
Torino (Italy)
Email: {*name*}.{*surname*}*@polito.it*

***Abstract*—The diffusion of learning technologies has fostered the use of mobile and Web-based applications to assess the knowledge level of learners. In parallel, an increasing research interest has been devoted to studying new learning analytics tools able to summarize the content of large sets of learning documents. To bridge the gap between formative assessment tools and document summarization systems, this paper addresses the problem of recommending short summaries of large sets of learning documents based on the outcomes of multiple-choice tests. Specifically, it presents a new methodology for integrating formative assessment through mobile applications and summarization of learning documents in textual form. The content of the multiple-choice tests is exploited to drive the generation of document summaries tailored to specific topics. Furthermore, the outcomes of the tests are used to automatically recommend the generated summaries to learners based on their actual needs. As a case study, we performed an evaluation experience of students' progresses, which was conducted in the context of a university-level course. The achieved results show the applicability of the proposed methodology.**

1. Introduction

The advent of electronic devices such as laptops, tablets, and mobile phones has radically changed learning paradigms [1]. The interaction between learners and teachers has become extremely simplified thanks to the diffusion of
(i) *E-learning platforms*, which allow teachers to share electronic learning materials through Web-based platforms.
(ii) *Learning analytics systems*, which allow teachers to extract significant information hidden in large learning datasets and to support learners in various activities.
(iii) *Formative assessment tools*, which allow teachers to monitor students' progresses and to get an early feedback on the most critical shortcomings or misunderstandings.

E-learning platforms allow teachers to share lecture notes, books, reports, articles, or scientific papers in digital form with their students for learning purposes. In this work we will focus on the analysis of learning materials in textual form (i.e., documents), which represent the most widespread content type [2]. These digital documents can be explored through multiple devices, such as laptops, tablets, and smartphones [1]. However, the amount of learning documents retrievable from the Web or shared through learning platforms is becoming so large that their manual inspection may become practically unfeasible.

Learning analytics entails using learning materials and learner-produced data to discover information and social connections useful for predicting and advising people's learning [3]. In this field, several efforts have been made to design and develop text mining solutions aimed to analyze digital learning documents (e.g. [4], [5], [6], [7], [8], [9]). Some of them (e.g. [4], [7], [8], [9]) addressed the problem of summarizing learning materials to ease document exploration or to improve content accessibility through portable devices with limited bandwidth or resolution. The importance of text summaries in learning activities has been confirmed by previous studies [10]. Automatic summary generation is particularly useful in the learning context to quickly identify salient concepts for study and review without the need for manually inspecting large sets of learning documents, possibly written in different languages.

The aim of this work is to recommend short summaries of the learning documents to learners based on their actual needs. To identify learners' needs, teachers may take advantage of formative assessment tools (e.g. [11], [12], [13]) whose main focus is to monitor students' progress towards learning objectives. Formative assessment entails conducting formal or informal procedures to improve students' performance [14]. Learners are typically asked to undergo quick tests, such as solving short exercises or answering to multiple-choice questions. Teachers may constructively exploit the outcome of the tests to highlight frequent mistakes, to adapt the content of future lessons to the audience's level, or to plan tutoring sessions. Most of the available formative assessment tools, such as Kahoot! (getkahoot.com) or VoxVote (www.voxvote.com), provide Web-based and/or mobile interfaces supporting interesting features, among which the creation and personalization of simple tests, surveys, or games, the collection of tests' outcomes, and the computation of basic statistics. The advantages of using these instruments to assess the ongoing knowledge level of the students are many-fold. On the one hand, learners may improve the retention of the concepts learnt and the awareness of their shortcomings/misunderstandings. On the other hand, teachers get an early feedback on students' progresses. Based on the tests' outcomes, teachers may

(i) recall concepts unclear to the majority of the students,

(ii) adapt the detail level of the oral lessons to the knowledge level of the audience, and

(iii) provide learners with additional materials to revise unclear concepts or to deepen their studies.

In this paper we aim at supporting teachers in achieving Objective (iii) by proposing a new methodology, namely TEST-driven SUMmarization (TestSumm). TestSumm automatically generates summaries of additional learning materials based on the outcome of multiple-choice tests. Summaries consist of a selection of representative document sentences (separated by punctuation marks) and they can be shared to learners through e-learning platforms. Our goal is to trigger the automatic generation of short summaries of large document sets tailored to specific topics. To customize summaries to the actual learner's needs, the summary topics are chosen based on the content of multiple-choice tests. Specifically, the key terms occurring in the multiple-choice questions and answers are exploited to drive the summarization process and to select only the sentences that are most pertinent to the topic covered by the tests. In such a way, multiple summaries, covering different topics, are generated and automatically recommended to learners based on their outcome in the test. To allow learners to easily move to the original teaching content for further readings, the summary sentences are linked to the corresponding occurrence in the original documents.

To test the proposed methodology on real learning data, we summarized documents using a state-of-the-art summarization system, i.e., the Multilingual Weighted Itemset-based Summarizer (MWISum) [15] relying on an established data mining technique (i.e., frequent itemset mining [16]). The summarizer picks the sentences covering the largest number of combinations of frequently co-occurring terms. To drive the summarization process according to the topic of the test, we extended the summarizer by integrating a dictionary of on-topic terms in the summarization process. Specifically, the co-occurrence of terms in the dictionary (i.e., the terms that are peculiar to the topic under analysis) are rewarded so that sentences ranging over the same topic are most likely to be included in the summary.

As a case study, we investigated the applicability of the proposed approach in a real learning context, a B.S. course on databases held in our university. Specifically, we conducted tests at the end of each course lesson and collected the outcomes of the tests. Then, we summarized the reference course textbooks by using dictionaries related to the topics covered by the tests.

The rest of paper is organized as follows. Section 2 describes the proposed methodology, while Section 3 summarizes the results of the evaluation experience in a real learning context. Finally, Sections 4 discuss challenges and future perspectives of this work and draw conclusions, respectively.

2. The TEST-driven SUMmarization methodology

Documents in textual form are the most common digital learning materials available on e-learning platforms [17]. These documents may be, for instance, lecture notes, e-books, scientific articles, or technical reports.

This section presents the proposed methodology to support learners in their activities. The goal is to identify topics on which learners would benefit from additional learning materials and provide them with on-topic textual summaries.

It consist of the three main steps, which are briefly summarized below. More detailed descriptions are given in the following sections.

- *Formative assessment.* Learners undergo multiple-choice tests through a dedicated platform (see Section 2.1).
- *Text preparation.* Tests and learning documents in textual form are prepared for the subsequent summarization process (see Section 2.2).
- *Summarization.* A set of document summaries is generated. Each summary contains the most salient sentences of the documents on a specific topic, among those covered by the test (see Section 2.3).
- *Summary recommendation.* Based on the tests' outcomes, on-topic summaries are recommended to learners (see Section 2.4).

2.1. Formative assessment

Formative assessment tools are mobile or Web-based applications that can be used to monitor students' progress towards learning objectives [14]. They provide ongoing feedback that may be exploited by teachers to improve the quality of their learning and by learners to understand their strengths and weaknesses as well as to tailor their future studies on specific subjects. In literature, formative assessment conceptually differs from summative assessment because tests are planned during lectures and not at the end of courses or teaching units [14].

The diffusion of mobile and Web-based platforms for formative assessment, such as Kahoot! (http://kahoot.it), has simplified the interaction between learners and teachers thus allowing a more extensive use of these learning instruments. Kahoot! is a free game-based formative assessment tool. It provides a social learning environment, accessible through mobile devices or Web browsers, through which learners may undergo surveys, quizzes, or questionnaires during or immediately after the lessons. The application is designed for social learning, with learners gathered around a common screen and equipped with an electronic device (e.g. smartphone, tablet, laptop). During tests, the questions and up to four multiple-choice answers are displayed on the main screen. Every answer corresponds to a distinctive color and shape. On the screen of the learners' devices, there are at most four rectangles with the color and the shape on each, and the learner needs to click or tap on the rectangle representing the correct answer. The game design is such that the players are required to frequently look up from their devices, enabling social interaction with the teacher and their peers.

The preliminary step of our methodology is the assessment of the knowledge level of the learners on the subjects covered in each lesson through multiple-choice tests. Multiple-choice tests are among the most popular objective assessment forms, because of their simplicity in educational assessment [1]. Given a question, learners are asked to select one or more correct answers out of the choices from a list. We consider multiple-choice tests because questions and answers typically consist of short phrases containing key terms, which recall the topics covered by the test. Thus, they can be exploited to drive the summarization process on learning documents (see Section 2.3). To implement our methodology we adopted Kahoot! as reference mobile formative assessment application. However, since the proposed methodology is general, different formative assessment tools can be easily integrated as well.

2.2. Text preparation

Let $\mathbf{D}$ be the set of textual documents d_1, d_2, …, d_n considered in our analyses. In this study we disregard non-textual content such as pictures and references in textbooks, slides, videos, highlights, and annotations.

Each document can be modeled as a set of sentences (i.e., portions of text separated by periods, question marks, or exclamation marks). Let s^i_j be the j-th sentence of document $d_i \in \mathbf{D}$. The goal is to generate a summary consisting of a selection of sentences s^i_j in $\mathbf{D}$, which represent the most salient content of the document set.

To adapt the document set to the summarization process, we apply the following two established text preprocessing steps.

Stopword elimination aims at discarding the sentence words having little semantic content, such as prepositions, articles, or conjunctions, because their presence would bias the quality of the following data mining phase.

Stemming aims at reducing the sentence words to their root form (i.e., the stem). This step, which can be enabled or disabled according to analyst's preferences, reduces the variance of the textual content to a more compact set of word roots. This step is particularly useful when the text summarization process relies on frequency-based term evaluation metrics.

Note that stopword and stemming algorithms are currently available for a large variety of languages. Hence, the proposed methodology is portable to documents written in different languages.

The output of the text preparation phase is a document set $\mathbf{D}_p$, where each sentence is a bag-of-word (BOW), i.e., an unordered set of word stems.

To select on-topic sentences we will exploit a dictionary to drive the summarization process. Let $\mathbf{T}$ be the set of multiple choice tests. For each test $t_z \in \mathbf{T}$ let diz(t_z) be a dictionary consisting of all the stems that occur in the corresponding questions or answers.

To each stem in the BOW of the document set $\mathbf{D}_p$ we assign a relevance score, which is a variant of the term frequency-document frequency (tf-df) statistics introduced in [15]. It considers three main factors:

(i) The frequency of the stem in each document (hereafter denoted as term frequency).

(ii) The number of documents in which the stem occurs at least once (denoted as document frequency).

(iii) The presence/absence of the stem in the dictionary (denoted as term rewarding/penalty score).

Specifically, the relevance score rs_{zi} of stem s_{zi} is computed as follows: $rs_{zi} = \Delta(s_{zi}) \cdot \frac{o_{zi}}{|d_i|} \cdot \frac{|\{d_i \in \mathbf{D}_p \,:\, s_{zi} \in d_i\}|}{|\mathbf{D}_p|}$, where o_{zi} is the number of occurrences of the z-th stem s_{zi} in the i-th document d_i, $\mathbf{D}_p$ is the document set under analysis, $|d_i|$ is the number of stems that are contained in the i-th document d_i, $\frac{|\{d_i \in \mathbf{D}_p \,:\, s_{zi} \in d_i\}|}{|\mathbf{D}_p|}$ represents the document frequency of the stem s_{zi} in the whole document set, and $\Delta(s_{zi})$ is a boolean function that returns a user-specified penalty score $\delta \in [0, 1]$ if stem s_{zi} is not present in the dictionary or 1 (no penalty) otherwise.

Since all documents in the analyzed set are assumed to cover the same subject, we exploit a relevance score that gives higher importance to word stems that frequently occur both locally (within a document) and globally (in the document set), as they are deemed as the best representatives of the documents' content. To tailor summaries to the content of the tests, we reward word stems occurring in the dictionary, while penalizing the others. Penalty score δ is set by the domain expert in the range [0,1]. The lower δ the more focused the summaries will be on the dictionary content, because the penalization of terms not in the dictionary becomes more relevant. If δ is set to zero only the word stems in the dictionary get non-zero relevance scores. Oppositely, if the penalty score δ is set to one, dictionary content is not rewarded at all (i.e., dictionary stems are as important as all the others) thus the resulting summaries are less likely to be tailored to the corresponding topic.

2.3. Summarization

This step entails generating multiple summaries of the document set according to the given dictionaries. For each test t_z in $\mathbf{T}$ we generate a summary $S(t_i)$ tailored to the corresponding dictionary diz(t_z). Each summary consists of a selection of the most representative sentences in the document set, where sentence relevance is evaluated according to the content of both document set and dictionary.

To test the proposed methodology on real learning data, we summarized textual documents using a state-of-the-art summarization system, i.e., the Multilingual Weighted Itemset-based Summarizer (MWISum) [15]. The MWISum summarizer relies on the following steps: (i) frequent itemset mining and (ii) sentence selection and ranking. The key idea behind the algorithm is to pick the sentences covering the largest number of combinations of frequently co-occurring word stems. At step (i) a model consisting of weighted frequent itemsets [18], i.e., sets of word stems characterized by fairly high importance, is generated from transactional representation of the source data [19]. Then at step (ii) a subset of sentences are selected and included in the output summary. A sentence *covers* an itemset if it contains the corresponding combination of word stems. Since itemsets represent the most significant underlying correlations among words, the number of covered itemsets per sentence is exploited as the evaluation criterion of sentence relevance in the document set.

2.4. Summary recommendation

To support learners in study and revision, summaries are recommended to learners based on the outcome of the multiple-choice tests. For each question in the test, the summary tailored to the corresponding content is recommended to all the students who gave wrong answers. Exploring short summaries tailored to their shortcomings, instead of the entire document set, allows learners to focus their attention on the most critical aspects. Furthermore, to allow learners to easily move to the original teaching content for further readings, the summary sentences are linked to the corresponding occurrence in the original documents. Hyperlinks to the Uniform Resource Locator of the documents allow learners to easily move from the summary to the original teaching materials.

Summaries can be shared to learners through the e-leaning platform. Since the document summarizer considered in our study is applicable to document sets written in different languages and generates short textual summaries that are easily accessible through mobile devices, we envision the use of an integrated mobile learning application capable of online processing the test's results, triggering summary generation, and recommending targeted summaries to students based on their outcome.

Summary generation is currently performed offline and required tens of seconds on the tested documents. A tight integration between formative assessment and document summarization tools allows learners to explore the summaries through their smartphone immediately after the end of the test thus getting an early feedback on their current level of knowledge.

3. Case study

We analyzed the applicability of the proposed methodology in a real learning context, i.e., a B.S. course held in our university. Specifically, we conducted an experience of formative assessment and learning document summarization by involving the students of a B.S. course on databases. At the end of 6 lessons, students were invited to login to the Kahoot! mobile app with a nickname and to undergo a test in anonymous form. Each test consists of a set of multiple-choice questions ranging over the main topic covered by the lesson. The number of questions per test, the number of participants, and the topics covered by each test are summarized in Table 1, while in Table 3 we report the outcome of the performed tests, in terms of number of participants and the average percentages of correct answers computed over all the participants. The interface of the Kahoot! environment contains an example of question shown by the teacher in the classroom (on the left-hand size of the image) and the interface used by the students to answer via a smartphone (on the right-hand side).

The aim of this learning activity was twofold.
(i) Assess the students' level of attention and comprehension of the concepts taught in the lesson.
(ii) Suggest students targeted readings taken from one of the recommended course textbooks [20].

For these purposes, we considered the chapters of the textbook as separate input documents for the summarization process, disregarding exercises, bibliography, footnotes, indices, tables, figures, captions, and all non-textual content. Note that even if each chapter of the book analyzes a different aspect of database systems, chapters often cover related topics and recall theoretical concepts mentioned in other ones. For this reason, we have chosen to use to content of the entire textbook as the input document set of the summarization process. We deemed the summarization algorithm as suitable for summarizing the content of the entire textbook (see Section 2.3). For the sake of simplicity, in our experiments we considered only the English-written version of the textbook. However, the adopted summarization algorithm can handle documents written in different languages as well [15]. For running the MWI-Sum algorithm we used the default configuration setting for English-written text corpora (Wminsup = 0.8%).

To adapt summaries to the topics covered by multiple-choice tests, we first generated on-topic dictionaries, one for each test, and then we extracted targeted summaries of the analyzed documents by exploiting the summarization algorithm driven by the dictionaries (see Section 2.3). Each dictionary consists of a selection of word stems that best characterize the topic covered by the test.

In the current implementation of the proposed methodology, dictionaries were generated by applying the following semi-automatic procedure. First, the text preparation steps described in Section 2.2 were applied. Specifically, we performed stopword elimination and stemming on the content of the tests (including both questions and answers) by applying the Wordnet algorithms for English-written documents (https://wordnet.princeton.edu/) and we computed the relevance score of each word stem. Stem relevance was evaluated according to the variant of the tf-df statistics described in Section 2.2, where we set the value of the penalty score δ to 0.3 (meaning that the occurrences of word stems occurring in the dictionary are awarded, on average, by 70%). Then, word stems with low relevance score in the document set are pruned, because they are less likely to represent interesting information. In particular, we considered only the word stems in the first two quartiles according to the distribution of the term relevance score in the test set. The resulting set of word stems has been validated by a domain expert prior to running the summa-

TABLE 1. CHARACTERISTICS OF MULTIPLE-CHOICE TESTS.

Test Id	Num. of questions	Topics	Dictionary
1	8	Principles of the relational model	*Relation*, *Cardinality*, *Primary*, *Foreign*, *Key*, *Integrity*, *Constraint*, *Tuple*, *Domain*, *Uniqueness*, *Attribute*, *Record*, *Schema*, *Instance*, *Reference*
2	10	Properties of transactions in the relational model	*Transaction*, *Commit*, *Rollback*, *Consistency*, *Durability*, *Atomicity*, *Reliability*, *ACID*, *State*, *Failure*, *Start*, *Automatic*
3	16	Principles of the Entity-Relationship model	*Entity*, *Association*, *Cardinality*, *Attribute*, *Composite*, *Hierarchy*, *Transaction*, *Inheritance*, *Conceptual*, *Design*, *Normal*, *BCNF*, *Schema*
4	14	Principles of the SQL language	*Declarative*, *Language*, *Relation*, *Join*, *Instruction*, *Clause*, *Table*, *DBMS*, *Definition*, *Command*
5	12	Syntax of the SQL language	*Select*, *Where*, *From*, *Join*, *Check*, *Like*, *Instruction*, *Clause*, *Operator*, *Not*, *Unique*, *Union*, *Create*, *Order*
6	9	DBMS and client-server architectures	*DBMS*, *Client*, *Server*, *Application*, *Architecture*, *Tier*, *API*, *Call*, *Interface*, *Connection*, *JDBC*, *Layer*

TABLE 2. TEST 2. TOP-5 SENTENCES IN THE SUMMARY.

Rank	Sentence	Selected by domain experts (YES/NO)
TEST ID 1 Principles of the relational model		
1	In a database there is a part that is invariant in time, called the schema of the database, made up of the characteristics of the data, and a part that changes with time, called the instance or state of the database, made up of the actual values.	Yes
2	However, a single relation is not usually sufficient for this purpose: a database is generally made up of several relations, whose tuples contain common values where this is necessary in order to establish correspondences.	No
3	For this purpose, the concept of integrity constraint was introduced, as a property that must be satisfied by all correct database instances.	Yes
4	The database shows one of the fundamental characteristics of the relational model, which is often expressed by saying that it is 'value-based': the references between data in different relations are represented by means of the values of the domains that appear in the tuples.	Yes
5	In practice, we adopt a simple solution, which makes it possible to guarantee the unambiguous identification of each tuple and refer to it from within other relations: null values are forbidden on one of the keys (called the primary key) and usually (that is, unless specified otherwise) allowed on the others.	Yes
TEST ID 2 Properties of transactions in the relational model		
1	Transactions must possess particular properties: atomicity, consistency, isolation and durability.	Yes
2	A transaction identifies an elementary unit of work carried out by an application, to which we wish to allocate particular characteristics of reliability and isolation.	Yes
3	A transaction can be defined syntactically: each transaction, irrespective of the language in which it is written, is enclosed within two commands: begin transaction (abbreviated to bot) and end transaction (abbreviated to eot).	No
4	Before executing the commit of its atomic unit, any failure will cause the elimination of all the effects of the transaction, whose original state is recreated.	Yes
5	Consistency demands that the carrying out of the transaction does not violate any of the integrity constraints defined on the database.	Yes
TEST ID 3 Principles of the Entity-Relationship model		
1	Other systems provide tools to carry out the reverse operation: reconstructing a conceptual schema based on an existing relational schema.	No
2	The generalization is transformed into two one-to-one relationships that link the parent entity E with the child entities E_1 and E_2.	Yes
3	The aim of logical design is to construct a logical schema that correctly and efficiently represents all of the information described by an Entity- Relationship schema produced during the conceptual design phase.	Yes
4	Remember that entities identified externally always participate in the relationship with a minimum and maximum cardinality of one; this type of translation is valid independently of the cardinality with which the other entities participate in the relationship.	Yes
5	Normalization allows the non-normalized schemas to be transformed into new schemas for which the satisfaction of a normal form is guaranteed.	Yes

TABLE 3. TEST OUTCOMES.

Test Id	Num. of participants	Avg. perc. of correct answers
1	42	63%
2	47	74%
3	42	67%
4	34	47%
5	41	72%
6	49	54%

rization process to prune irrelevant words, which accidentally occured in the dictionaries. In our tests, approximately between 10% and 15% of the word stems selected by the procedure described above were filtered out, because they represent redundant or out-of-topic information. Note that to avoid manual dictionary validation an alternative would have been to integrate topic detection algorithms (e.g. [21]) into the text preparation phase. However, since dictionaries are typically small (they contain from 10 to 30 word stems) they can be easily explored by domain experts through manual inspection. Furthermore, validated dictionaries can be reused to summarize multiple document sets acquired from different sources or collected in different periods. The dictionaries generated in our experiments are summarized in the right-hand column of Table 1, where, for the sake of readability, we reported entire words instead of the corresponding stems.

4. Conclusions and future perspectives

This paper presented a methodology for supporting learners' activities through the automatic generation of short summaries of potentially large sets of learning documents in textual form. The main innovation provided by the proposed methodology is in the tight integration between formative assessment and document summarization tools, which allows, on the one hand, teachers to get an early feedback on the learners' level of comprehension of the lesson taught and learners to explore succinct summaries of learning materials customized to their needs.

To apply the proposed methodology to real learning data, we implemented a preliminary version of our method integrating a mobile formative assessment application and a state-of-the-art summarization algorithm. The summarizer has been modified to adapt the generated summaries to the outcome of the multiple-choice tests. The experiments, conducted in the context of a B.S. database course, show the applicability of the proposed approach in a real case study.

To improve the manageability and accessibility of the results, we plan to integrate portable formative assessment procedures and summarization tools into a new integrated mobile application, which allows learners to explore summaries during or immediately after the test. Since mobile devices often have limited bandwidth and low resolution, accessing short summaries instead of large documents or complex models can be a more effective way to support learner's activities. Furthermore, we will investigate the extension of the proposed method to cope with Big document sets. Specifically, we aim at extending the current implementation of the summarization algorithm, which currently relies on an itemset-based model, to cope with Big datasets. Instead of generating one summary per topic, a scalable learning system may generate and recommend one summary per question tailored to the learner's needs. After giving each answer, learners may immediately verify its correctness and explore the corresponding summary to review the related concepts.

References

[1] J. L. Moore, C. Dickson-Deane, and K. Galyen, "E-learning, online learning, and distance learning environments: Are they the same?" *The Internet and Higher Education*, vol. 14, no. 2, pp. 129 – 135, 2011.

[2] M. Hossain, M. Masud, A. Alelaiwi, and A. Alghamdhi, "Aco-based media content adaptation for e-learning environments," in *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, May 2014, pp. 118–123.

[3] R. Ferguson, "Learning analytics: Drivers, developments and challengs," *Int. J. Technol. Enhanc. Learn.*, vol. 4, no. 5/6, pp. 304–317, Jan. 2012.

[4] G. Yang, D. Wen, Kinshuk, N.-S. Chen, and E. Sutinen, "Personalized text content summarizer for mobile learning: An automatic text summarization system with relevance based language model," in *Technology for Education (T4E), 2012 IEEE Fourth International Conference on*, July 2012, pp. 90–97.

[5] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414 – 422, 2015.

[6] Y. H. Lee, G. D. Chen, L. Y. Li, Nurkhamid, C. Y. Fan, and K. H. Chiang, "The effect of utilizing the learning skill of highlighting and constructing a map in a networked hyperlink condition on learning performance," in *2012 IEEE 12th International Conference on Advanced Learning Technologies*, July 2012, pp. 546–548.

[7] S. Saraswathi, M. Hemamalini, S. Janani, and V. Priyadharshini, "Multi-document text summarization in e-learning system for operating system domain," in *Advances in Computing and Communications*, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2011, vol. 193, pp. 175–186.

[8] E. Baralis, L. Cagliero, and L. Farinetti, "Generation and evaluation of summaries of academic teaching materials," in *39th IEEE Annual Computer Software and Applications Conference, COMPSAC 2015, Taichung, Taiwan, July 1-5, 2015. Volume 2*, 2015, pp. 881–886.

[9] E. Baralis and L. Cagliero, "Learning from summaries: Supporting e-learning activities by means of document summarization," *IEEE Trans. Emerging Topics Comput.*, vol. 4, no. 3, pp. 416–428, 2016.

[10] S. Elliott, *Educational Psychology: Effective Teaching, Effective Learning*. McGraw-Hill, 2000.

[11] M. Caeiro-Rodriguez, M. Llamas-Nistal, and F. Mikic-Fonte, "Introducing bea into self-regulated learning to provide formative assessment support," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–4.

[12] L. Zhang, "Formative assessment in english for specific purposes," in *2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, March 2016, pp. 315–318.

[13] F. Ali and A. George, "Impact of a formative e-assessment on learning outcomes: A pilot study on a social and behavioural sciences course, college of health sciences, university of bahrain," in *Fifth International Conference on e-Learning*, Oct 2015, pp. 408–412.

[14] Y. Song, Z. Hu, Y. Guo, and E. F. Gehringer, "An experiment with separate formative and summative rubrics in educational peer assessment," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–7.

[15] E. Baralis, L. Cagliero, A. Fiori, and P. Garza, "Mwi-sum: A multilingual summarizer based on frequent weighted itemsets," *ACM Trans. Inf. Syst.*, vol. 34, no. 1, p. 5, 2015.

[16] R. Agrawal, T. Imielinski, and Swami, "Mining association rules between sets of items in large databases," in *ACM SIGMOD 1993*, 1993, pp. 207–216.

[17] F. Grünewald and C. Meinel, "Implementation and evaluation of digital e-lecture annotation in learning groups to foster active learning," *TLT*, vol. 8, no. 3, pp. 286–298, 2015.

[18] W. Wang, J. Yang, and P. S. Yu, "Efficient mining of weighted association rules (war)," in *In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. Press, 2000, pp. 270–274.

[19] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[20] P. Atzeni, S. Ceri, S. Paraboschi, and R. Torlone, *Database Systems - Concepts, Languages and Architectures*. McGraw-Hill Book Company, 1999.

[21] H. Lin, B. Sun, J. Wu, and H. Xiong, "Topic detection from short text: A term-based consensus clustering method," in *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, June 2016, pp. 1–6.