

Ask4Summary: A Summary Generation Moodle Plugin Using Natural Language Processing Techniques

Mohammed SALEH, Maria F. IRIARTE & Maiga CHANG*

Faculty of Science and Technology, Athabasca University, Canada
*maiga.chang@gmail.com

Abstract: Online learning management systems (e.g., Moodle) provide students with a variety of materials in a course. If learning materials include large amount of text, students may find it difficult to gain insights. Natural Language Processing (NLP) uses computational language models to obtain knowledge from text and speech. This study develops a Moodle plugin, Ask4Summary plugin, which summarizes relevant course content for the questions asked by students in a discussion forum. First, the system parses all course text-based content documents to extract and store the information. As students ask in a discussion forum, the system parses the questions and creates responses for them. The response is a customized summary generated via matching and retrieving from the stored information. Ask4Summary plugin is an open-source project and has been approved by Moodle. The tool is available for users to download from the Moodle plugins directory. For the next step, the research team has an evaluation plan to implement the plugin in a graduate level academic writing English course in collaboration with an Asian university. The questionnaire covers the perceived usability of the plugin and teacher's and student's perceptions. It is expected that evaluation results demonstrate that the plugin generates an adequate summary for each question. The plugin might extend beyond the use in online learning (e.g., medical records, legal documents).

Keywords: Language Learning, N-gram, Part of Speech, Cosine Similarity, Discussion Forum

1. Introduction

Due to the pandemic, on-site education programs moved to the online environment to ensure education delivery (Dhawan, 2020). Nowadays, online learning resources in educational systems keeps growth (García-Morales et al., 2021; Setiawan et al., 2021). Nevertheless, information overload caused by technological development might influence learning performance. Feroz et al. (2021) conclude that an excess of information hinders students' knowledge acquisition.

This study addresses students' difficulties in handling extensive course materials across of different document types. The research team develops a Natural Language Processing (NLP) plugin for Moodle, a popular open-source learning management system. The main research contribution is the generation of a summary based on a student's question about a topic covered in the course materials.

The following sections are structured as follows: Section 2 elaborates on the related research background. Section 3 introduces the Ask4Summary Moodle plugin and explains the summary generation process. Section 4 discusses the document scanning workflow. Section 5 describes the technical issues of the summary generation. Section 6 proposes an evaluation plan for assessing the usability and satisfaction on the generated summary. Section 7 makes a conclusion and suggests future research.

2. Background

Natural Language Processing focuses on obtaining knowledge from the speech and text data with computational language models. Over time, the introduction of technology in language learning has resulted in favorable outcomes (Ahmadi, 2018). Research shows the advantage of NLP feedback systems for language learning over traditional technologies (Dodigovic, 2007; Nagata, 1993) and

suggest evolving from corrective feedback to interactive tutoring systems (Slavuj et al., 2015). Recent research confirms that automated personalized feedback provided by Intelligent Tutoring Systems (ITSs) promotes student knowledge acquisition (Al-Bastami & Naser, 2017; Kochmar et al., 2020). This study meets new trends in Technology Enhanced Language Learning (TELL) by generating a customized summary for each student's question.

N-gram is an NLP model based on a Markov chain that only accounts for the sequence of n components of a text (Russell & Norvig, 2022). To reduce the ambiguity inherent in natural language, the n-grams can be grammatically classified and labelled the words according to their part of speech (PoS) tags (Wilks & Stevenson, 1998). Ask4Summary plugin relies on two NLP open access web services: (1) *N-gram & PoS Identifier and Verifier* to select valid n-gram and PoS tags and (2) *Automatic Answering Service for Coronavirus Question*. The latter adopts both of NLP techniques and cosine similarity method (Pal, Chang, & Iriarte, 2022). Cosine similarity is a common metric that compares two vectors via the cosine of the angle they form, being 1 complete similarity and -1 total dissimilarity.

Moodle is a widespread open-source learning management system used in diverse educational options (Al-Ajlan & Zedan, 2008). Research highlights the system's comprehensive range of capabilities (Lisnani et al., 2020). Teachers can design a course and select the features that best fit their needs.

Moodle allows researchers to extend the standard version by creating plugins in various areas. New plugins require submission for reviewing and approval to become available to all Moodle sites. The research team designs Ask4Summary plugin as a block plugin. "Blocks" are a common plugin type that host functionalities in boxes distributed on the sides of Moodle webpages.

3. Ask4Summary Moodle Plugin

Ask4Summary plugin is a free tool approved by Moodle and available to users in the Moodle plugins directory (https://moodle.org/plugins/block_ask4summary). More documentation, workflow charts and tutorials can be found on the project website (<https://ask4summary.vipresearch.ca/>) and the research team YouTube channel (<https://youtube.vipresearchgroup.ca>). Tutorials cover installation, configuration, teacher use and student use including examples and results. Teachers can configure and add the block to their courses and have access to the service with a brief setup. Once the Ask4Summary Moodle plugin is installed and added to a course, teachers can choose from a variety of settings to control the summary generation service and the learning database.

The general process of Ask4Summary is to utilize the built-in Moodle tables to access the posts in a certain forum and use the course content to produce a response. Figure 1 shows the overall process flow. Once students have been told which forum to use, they can simply ask their questions with a "Helper Name" – a name that allows Ask4Summary to distinguish which post is a question. On a pre-set interval, a scheduled "Forum Scanning" task is queued. The task checks within the specific discussion forum in the Moodle database for the student's posts that contains the "Helper Name".

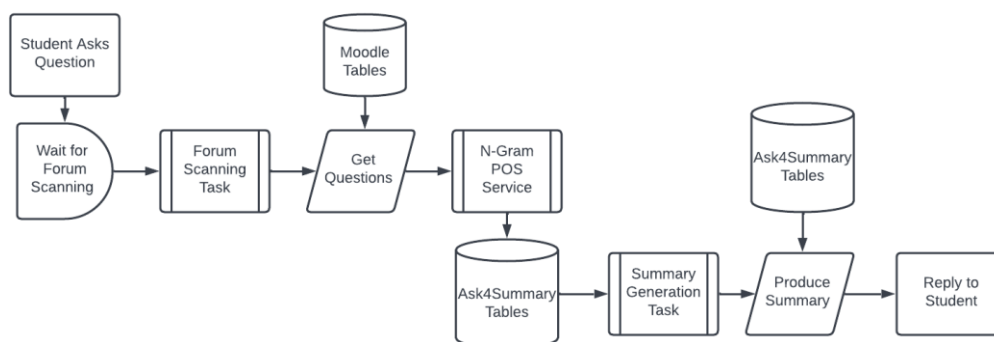


Figure 1. Ask4Summary Process.

With any post identified to be a student’s question, another “N-Gram POS” service gathers and stores the n-grams and PoS tags from the post into Ask4Summary tables, by using the N-Gram & PoS Identifier and Verifier. From there, Ask4Summary has the “Summary Generation” ad-hoc task queued for producing a summary as the response of the student’s post with the n-grams and PoS tags generated from course content. The ad-hoc task then uses the generated summary to respond the student’s post in the designated forum.

There are settings that handle many of these processes and teachers can configure according to their needs. In the “Block Configuration” there are three general sections, for forum scanning, document scanning, and summary generation. Teachers can choose to enable/disable the plugin for responding questions, and how (i.e., the Helper Name) and where (i.e., a chosen discussion forum) students can ask their question as Figure 2 shows. They can also choose which file types they would like the plugin to consider as the base of its learning content database, and how many of these resources should be considered for responses as Figure 3 displays.

▼ Forum Settings

Ask4Summary Helper Name

Enable Ask4Summary When enabled, Ask4Summary will parse forums and respond to questions.

Ask4Summary Response Method

Select the forum (if applicable)

Automatically generated form name (if applicable)

Figure 2. Forum Scanning.

▼ Course Learning Object Scan Settings

Enable URL Parsing When enabled, Ask4Summary will scan through course URL links and get their content.

URL Crawling Depth

Enable PDF Parsing When enabled, Ask4Summary will scan PDF modules and get their content. ⓘ
For more information on why PDF parsing is disabled by default, see [the documentation page](#).

Enable DOCX Parsing When enabled, Ask4Summary will scan DOCX modules and get their content.

Enable PPTX Parsing When enabled, Ask4Summary will scan PPTX modules and get their content.

Enable Moodle Page Parsing When enabled, Ask4Summary will scan Moodle Pages and get their content.

Figure 3. Document Scanning Settings.

Outside of the settings and in the block itself, there are two links that only the teachers can see: “Documentation” and “Logistics”. The “Documentation” page has a variety of information on the plugin itself, and its range of functionality. The “Logistics” page shows a table for the scanning progress of each enabled file type, including the amount of unscanned and scanned course modules. It provides the names of each resource that were considered for the learning database and showcases all the questions that Ask4Summary was either able or unable to answer. The page also displays the average time that the plugin took to answer, and each question with who asked it and when.

4. Reading Content with Natural Language Processing

Before a summary can be generated, Ask4Summary needs a content base to choose sentences from. This content base is built from the course’s material, where the plugin attempts to scan and parse course modules – activities and resources a course has – for their text content. Once the sentences in the

selected resources – teachers can choose which type(s) of materials should be scanned by the plugin as Figure 3 shows earlier – are retrieved, the plugin then uses the *N-Gram & PoS Identifier and Verifier* to find the valid n-grams based on their part of speech. Afterwards, this content is stored into Ask4Summary’s learning tables. The document scanning and other system workflow scan be found in the documentation section of the project website (Vip Research Group, 2022).

During later hours of the day, Ask4Summary looks through the course materials that have not already considered to scan. Currently, the plugin supports the scanning of Word and PDF documents, PowerPoint presentation slides, URL web pages, and those built-in Moodle page modules. Depending on the filetype, Ask4Summary handles the parsing differently. For example, Word and PDF documents are already considered to be in sentence format, whereas PowerPoint slides are mostly sentence fragments. Regardless of type, Ask4Summary attempts to gather text content from these course modules.

Once the sentences are retrieved, Ask4Summary begins to parse the sentences for their n-grams and PoS tags. This is done by packaging a query that includes a sentence and the preferred settings for sending to the *N-Gram & PoS Identifier and Verifier* to get the top valid N-Grams. These n-grams and PoS tags and the corresponding sentence are stored in a variety of created XML based tables the plugin has. The tables form the content base.

5. Summary Generation

Forum scanning and summary generation work off one another to produce a response – summary generation in Ask4Summary can never commence if a question is not sent to the answering service. Once Ask4Summary begins its “Forum Scanning” task, it retrieves all the course’s forum posts that contain the specified “Helper Name”. Then, following the same process it reads and processes the course materials, the plugin goes through each forum post, breaks the question into sentences, and sends to the online service to get the valid n-grams and their part of speech tags.

Prior to starting the answering service, Ask4Summary compares the n-grams to its “response” table to see if similar question has the same identified n-gram set before was already answered. If so, it is not adding that post to the queue for summary generation but simply replies to the forum post with the previously generated summary. Once all the questions in the course were considered, if any questions were not already answered, Ask4Summary queues the background task “Question Answering”.

The summary generation workflow is as follows. The “Summary Generation” task looks at all the forum posts that are considered to be valid. It gets the count of each n-gram the question has and creates a vector to represent the question. It also goes through each course learning object (e.g., a Moodle course page, a reading article in PDF, a teaching material in PowerPoint PPTX, an exercise document in DOCX, a supplemental webpage, etc.) and creates frequency vectors based on the same n-gram set for the learning object.

With the vectors, the task performs Cosine Similarity Calculation to find the top “X” documents, where the value of “X” can also be predefined and set by the teacher of the course. Now, these top documents are analyzed at the sentence level where all sentences from the documents are reviewed and the correspondent vectors for the same n-gram set are created. The Cosine Similarity Calculation is done again to compare how similar each sentence vector is to the original question vector. The top “Y” sentences will be considered, where “Y” value is also predefined and set by the teacher. With the identified “Y” sentences Ask4Summary can generate a summary to the student’s question by replying his or her forum post. Figure 4 shows an example of a question asked by a student in the forum and Ask4Summary generates a summary to respond the student’s question.

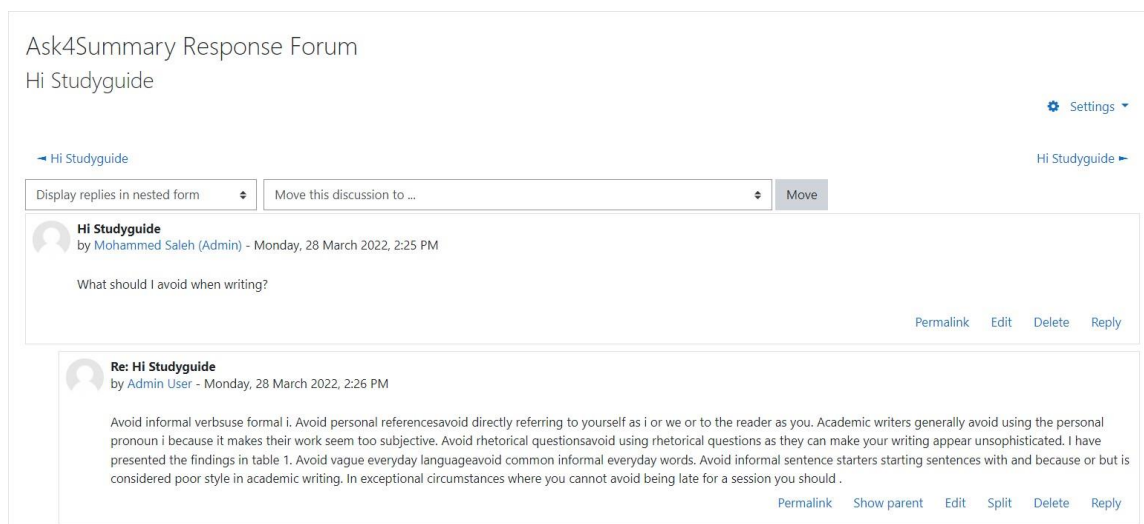


Figure 4. Student’s question and the correspondent summary generated by Ask4Summary.

6. Evaluation Plan

To verify the usability and effectiveness of Ask4Summary, the research team plans to have the Ask4Summary plugin adopted and used in a graduate-level Academic English course through a partnership with an Asian university. The post-questionnaire focuses on several aspects of the system: usability, perceived relevance and satisfaction, and is preceded by an introduction and instructions.

The data analysis starts a review of the asked questions and the correspondent summary generated by Ask4Summary to check the teacher’s and students’ perceived relevance and satisfaction degree. The research team is going to study the relevance and satisfaction degrees (in score 1 to 10) with the means of descriptive statistical analysis. The evaluation also foresees consultation with the teacher on discrepancies between perceptions on the same summary and opportunities for system improvement.

For the perceived relevance and satisfaction, the analysis will devote special attention to:

- Low relevance and low satisfaction scores: they require a detailed review on the question and the summary to identify possible reasons for the low score and system refinements.
- Relevance and satisfaction scores are not correlated: in general, the two scores on a question and its summary generated should not far apart. A relevant summary may generate satisfaction and vice versa. For this reason, inconsistent scores lead to investigation of the causes.

7. Conclusion

This study addresses the difficulty for students to extract information from the online learning materials in a course. The research team develops Ask4Summary, an NLP plugin for the open-source online learning management system Moodle. The plugin responds student questions posed in a forum by providing customized summaries from the text-based course materials (e.g., PDF, DOCX, PPTX, webpages, video transcripts).

There are further aspects of the study that have not yet been undertaken:

- The Ask4Summary plugin evaluation is planned on a Graduated Academic English course via a partnership with an Asian university.
- Influence of study on the improvement of students' knowledge acquisition and its possible contribution to learning success.
- Reduction of the summary generation time by applying other NLP techniques (e.g., distributed representations)

The research team has also observed that quality of a summary declines when generated from a presentation file (e.g., PowerPoint slides in PPTX format). This finding is understandable because the

presentation document might structure the content in tables and bullet points instead of sentences. Analysis of the collected data may provide more solid evidence to verify the observation.

This research is part of a larger project where other members in the research group are investigating and designing more sophisticated NLP methods. The research took the n-gram model as a starting point only and was aware of the limitation of the n-gram model and the chosen Cosine similarity method. The future work is aiming to improve the plug-in through the evaluation results and the investigation and trial outcome on the innovative and comprehensive NLP methods.

Acknowledgements

The authors acknowledge the support of the National Science and Engineering Research Council of Canada (NSERC) [RGPIN-2018-05545], NSERC Undergraduate Student Research Awards and Mitacs (Globalink program).

References

- Ahmadi, D. M. R. (2018). The Use of Technology in English Language Learning: A Literature Review. *International Journal of Research in English Education*, 3(2), 115–125. <https://doi.org/10.29252/ijree.3.2.115>
- Al-Ajlan, A., & Zedan, H. (2008). Why Moodle. *2008 12th IEEE International Workshop on Future Trends of Distributed Computing Systems*, 58–64. <https://doi.org/10.1109/FTDCS.2008.22>
- Al-Bastami, B. G., & Naser, S. S. A. (2017). Design and Development of an Intelligent Tutoring System for C# Language. *European Academic Research*, 4(10).
- Dhawan, S. (2020). Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>
- Dodigovic, M. (2007). Artificial Intelligence and Second Language Learning: An Efficient Approach to Error Remediation. *Language Awareness*, 16(2), 99–113. <https://doi.org/10.2167/la416.0>
- Feroz, H. M. B., Zulfikar, S., Noor, S., & Huo, C. (2021). Examining multiple engagements and their impact on students' knowledge acquisition: The moderating role of information overload. *Journal of Applied Research in Higher Education*, 14(1), 366–393. <https://doi.org/10.1108/JARHE-11-2020-0422>
- García-Morales, V. J., Garrido-Moreno, A., & Martín-Rojas, R. (2021). The Transformation of Higher Education After the COVID Disruption: Emerging Challenges in an Online Learning Scenario. *Frontiers in Psychology*, 12, 616059. <https://doi.org/10.3389/fpsyg.2021.616059>
- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2020). Automated Personalized Feedback Improves Learning Gains in An Intelligent Tutoring System. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (pp. 140–146). Springer International Publishing. https://doi.org/10.1007/978-3-030-52240-7_26
- Lisnani, L., Putri, R. I. I., Zulkardi, & Somakim. (2020). Designing Moodle features as e-learning for learning mathematics in COVID-19 pandemic. *Journal of Physics: Conference Series*, 1657(1), 012024. <https://doi.org/10.1088/1742-6596/1657/1/012024>
- Nagata, N. (1993). Intelligent Computer Feedback for Second Language Instruction. *The Modern Language Journal*, 77(3), 330–339. <https://doi.org/10.1111/j.1540-4781.1993.tb01980.x>
- Pal, S., Chang, M., & Iriarte, M. F. (2022). Summary Generation Using Natural Language Processing Techniques and Cosine Similarity. In A. Abraham, N. Gandhi, T. Hanne, T.-P. Hong, T. Nogueira Rios, & W. Ding (Eds.), *Intelligent Systems Design and Applications* (pp. 508–517). Springer International Publishing. https://doi.org/10.1007/978-3-030-96308-8_47
- Russell, S. J., & Norvig, P. (2022). *Artificial intelligence: A modern approach* (Fourth edition, global edition). Pearson.
- Setiawan, A. M., Munzil, & Fitriyah, I. J. (2021). Trend of learning management system (LMS) platforms for science education before-after Covid-19 pandemic. *AIP Conference Proceedings*, 2330(1), 060005. <https://doi.org/10.1063/5.0043196>
- Slavuj, V., Kovačić, B., & Jugo, I. (2015). Intelligent tutoring systems for language learning. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 814–819. <https://doi.org/10.1109/MIPRO.2015.7160383>
- Wilks, Y., & Stevenson, M. (1998). The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation. *Natural Language Engineering*, 4(2), 135–143. <https://doi.org/10.1017/S1351324998001946>