



NLP-Based Quranic Verse Retrieval In Islamic Education: A Development Model And Preliminary Evaluation Of Computational Thinking

Irsyad Fauzan Nurdin*, Lala Septem Riza, Rani Megasari

Computer Science Education, Faculty of Math and Science,

Universitas Pendidikan Indonesia.

*Corresponding Author. Email: irsyadfauzan28@upi.edu

Abstract: This study aims to develop a Natural Language Processing (NLP)-based Quranic verse retrieval model to support Quranic verse exploration and Computational Thinking (CT)-oriented learning in Islamic Education. The study employed the ADDIE development framework, and the model was implemented with Grade XI students and Islamic Education teachers at a private high school. Data were collected using a mixed-methods approach, including pre- and post-tests, perception questionnaires, expert validation, and teacher interviews. Quantitative data were analyzed using descriptive statistics, the Shapiro–Wilk test, the Wilcoxon signed-rank test, N-gain analysis, and descriptive percentage analysis of questionnaire responses, while qualitative data were analyzed through thematic analysis. The results demonstrated a significant improvement in students' mean scores, increasing from 74.40 to 96.40 ($p^* < 0.001$), with a high average N-gain of 0.882. Student and teacher responses reached 90.19% and 77.50%, respectively, indicating positive acceptance of the proposed model. Although the one-group pretest–posttest design limits causal inference and CT was operationalized through learning-process indicators and user perceptions rather than comprehensive performance-based assessment, the findings suggest that the model has considerable potential as a technology-enhanced medium for exploratory learning in Islamic Education. Future research should focus on expanding the Quranic corpus, integrating *tafsir* resources, validating retrieval performance using standard information retrieval metrics, and incorporating performance-based assessments to measure Computational Thinking more comprehensively.

Article History

Received: 18-04-2026

Revised: 30-05-2026

Accepted: 15-06-2026

Published: 20-07-2026

Key Words:

Natural Language

Processing;

Quranic Verse Retrieval;

Computational Thinking;

Islamic Education;

Merdeka Curriculum.

How to Cite: Nurdin, I. F., Riza, L. S., & Megasari, R. (2026). NLP-Based Quranic Verse Retrieval In Islamic Education: A Development Model And Preliminary Evaluation Of Computational Thinking. *Jurnal Paedagogy*, 13(3), 1200-1208. <https://doi.org/10.33394/jp.v13i3.20755>



<https://doi.org/10.33394/jp.v13i3.20755>

This is an open-access article under the [CC-BY-SA License](#).



Introduction

Digital transformation has changed the way students access religious knowledge. Islamic Education is no longer mediated only through textbooks and teacher explanations but also through search engines, mobile applications, and digital religious resources (Untoro, 2022). This shift creates an important instructional challenge: technology must not only provide fast access to Quranic information, but also help students explore verses thematically, critically, and responsibly. Keyword-based retrieval may be insufficient because a single Islamic theme can be expressed through different lexical forms and semantic contexts. Therefore, Islamic Education requires learning media that support thematic verse exploration while preserving the teacher's interpretive role.

Natural Language Processing (NLP) provides a relevant technological approach for this need because it enables computers to process human language, represent textual data, and support information retrieval (Jackson & Moulinier, 2007). In Quranic studies, NLP has been applied to semantic search, thematic classification, concept mapping, and meaning-relation



analysis, although challenges remain regarding Arabic linguistic complexity, corpus quality, and source validation (Bashir et al., 2023). Semantic search studies indicate that Quranic retrieval can move beyond literal word matching and support users who search using translation or non-Arabic input (Mohamed & Shokry, 2022; Rajagede et al., 2023).

The pedagogical value of an NLP-based retrieval system emerges when the system is embedded in learning activities that require students to formulate themes, select keywords, compare retrieved verses, filter relevance, and construct reflections. These activities are related to Computational Thinking (CT), which involves decomposition, pattern recognition, abstraction, and algorithmic reasoning (Beecher, 2017). CT has increasingly been framed as a cross-disciplinary competence rather than a skill limited to computer science. For example, CT has been integrated into humanities learning through design principles for language arts, social studies, and arts (Caskurlu et al., 2022), assessed in social studies classrooms (Manfra, 2022), and discussed as a way of using computational tools to reason with concepts in the humanities (Christensen, 2023). In Islamic Education, this means that students can decompose broad religious themes into searchable concepts, recognize patterns across verses, abstract core meanings, and organize inquiry steps coherently.

The Merdeka Curriculum provides an appropriate context for this integration because it emphasizes active, contextual, and student-centered learning. Islamic Education and Character Education materials also emphasize the internalization of Islamic values in everyday life, which requires learning to move beyond memorization toward exploration, discussion, and reflection (Untoro, 2022). Accordingly, an NLP-based Quranic verse retrieval model can be positioned as a bridge between digital literacy, Islamic Education learning, and CT-oriented inquiry.

Several prior studies have addressed Quranic information retrieval, cross-language retrieval, and semantic search (Agustian & Wulandari, 2013; Taan et al., 2021; Rajagede et al., 2023). Other studies have examined NLP in educational contexts, including collaborative learning analysis and misconception detection (Sullivan & Keith, 2019; K  kver et al., 2024). However, studies that explicitly position an NLP-based Quranic verse retrieval system as a classroom-based Islamic Education learning medium, map its use onto CT indicators, and evaluate it through learning outcomes and user responses remain limited. The novelty of this study lies in the pedagogical integration of a retrieval prototype, Islamic Education learning activities, and CT indicators within a development model, rather than in a claim of algorithmic superiority. This study aims to develop and preliminarily evaluate an NLP-based Quranic verse retrieval model that supports thematic verse exploration and CT-oriented learning in Islamic Education.

Research Method

This study used a research-and-development approach based on the ADDIE framework, which consists of analysis, design, development, implementation, and evaluation. ADDIE was selected because the study aimed not only to produce a technological prototype, but also to position the prototype within a learning design. In the analysis stage, the researcher identified the needs of Islamic Education learning, the characteristics of Grade XI students, teachers' needs, and the limitations of ordinary keyword-based verse retrieval. In the design stage, the system workflow, learning activities, and evaluation instruments were developed. In the development stage, the retrieval prototype was built. In the implementation stage, the system was used in Islamic Education learning. In the evaluation stage, the system was assessed using quantitative and qualitative data.



The study involved Grade XI students and Islamic Education teachers at a private high school. Learning-outcome data were analyzed based on 25 complete pre-test and post-test scores. Student questionnaire data were collected from 27 respondents; some students completed the user-experience evaluation, although not all had complete learning-outcome data. This difference in sample size was separated during analysis so that each conclusion corresponded to the relevant dataset. Teachers were involved as learning facilitators, pedagogical evaluators, and interview informants.

The prototype used a corpus consisting of Quranic translation texts and Islamic Education materials relevant to the Merdeka Curriculum. The retrieval pipeline consisted of preprocessing, TF-IDF text representation, and cosine similarity. The preprocessing stages included lowercasing, punctuation cleaning, tokenization, stop-word removal, and Indonesian stemming. The system displayed a list of verses or materials with the highest similarity scores to the user's query. At this preliminary stage, retrieval performance was not evaluated using information-retrieval metrics such as precision, recall, mean reciprocal rank, or nDCG. Instead, result relevance was validated formatively through functional testing and teacher/expert input to determine whether the system was sufficient to support classroom activities.

The research instruments consisted of a pre-test, post-test, student questionnaire, teacher questionnaire, expert judgment sheet, and teacher interview guide. The pre-test and post-test measured students' understanding of Islamic Education materials based on verses and thematic relationships. The student questionnaire used a 1–4 scale to assess ease of use, support for understanding verse relationships, independent exploration, learning engagement, and perceptions of structured thinking processes. The teacher questionnaire assessed system relevance, ease of integration, feasibility as a learning medium, and development needs. Expert judgment was used to assess content feasibility, system functionality, and implementation potential (Darvishi et al., 2022).

In this study, CT was operationalized as learning-process indicators and user perceptions rather than a complete performance-based CT assessment. Four indicators were mapped onto the learning process: decomposition, when students broke down a theme into keywords; pattern recognition, when students compared retrieval results; abstraction, when students extracted the core message of verses; and algorithmic reasoning, when students organized search and reflection steps. Quantitative analysis began with descriptive statistics, including minimum, maximum, mean, and standard deviation. The Shapiro-Wilk test was used to assess normality because of the relatively small sample size. Since the data were not normally distributed, the Wilcoxon Signed-Rank test was used to compare pre-test and post-test scores. Learning improvement was calculated using N-Gain so that the score change was not interpreted only through statistical significance but also through students' room for improvement (Hake, 1998). Questionnaire data were analyzed using percentages and interpretation categories. Interview data were analyzed thematically through data reduction, data display, and conclusion drawing (Miles et al., 2014).

Results and Discussion

The results are presented in three parts: the development of the NLP-based Quranic verse retrieval model, changes in students' learning outcomes, and user responses from students and teachers. This structure was used to show the model not only as a technical prototype, but also as a learning medium implemented in Islamic Education.

The developed model positioned the NLP-based retrieval system as a medium for thematic Quranic verse exploration. In the learning process, students were guided to formulate Islamic themes, transform them into searchable keywords, examine the retrieved verses or materials, compare their relevance, and construct reflections under teacher guidance. Figure 1 presents the conceptual model that connects Islamic Education learning needs, the Quranic and learning-material corpus, the retrieval process, Computational Thinking-oriented activities, and learning outcomes. The model emphasizes that the system functions as a learning support tool, not as a replacement for the teacher or tafsir authority.

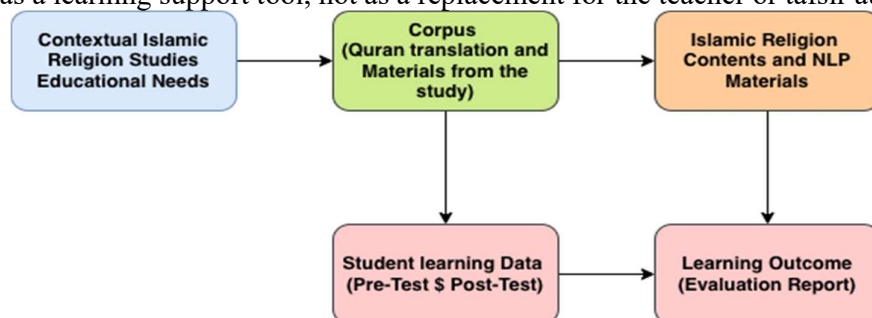


Figure 1. Conceptual model integrating the NLP system, PAI learning, and computational thinking

Students' learning outcomes showed an increase after the implementation of the model. The mean pre-test score was 74.40 with a standard deviation of 20.43, while the mean post-test score increased to 96.40 with a standard deviation of 12.54. The minimum score also increased from 20 in the pre-test to 50 in the post-test. This result indicates that students' achievement improved after the learning activities supported by the NLP-based retrieval system. However, because the study used a one-group pre-test-post-test design, this increase should be interpreted as a preliminary indication of learning improvement rather than as direct causal evidence.

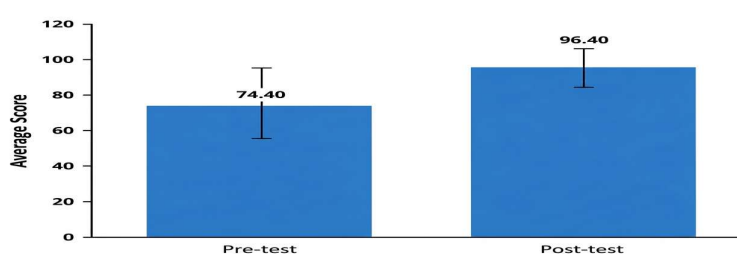


Figure 2. Comparison of students' pre-test and post-test scores. Bars indicate means, and error lines indicate standard deviations

The Shapiro-Wilk test showed that the pre-test, post-test, and score-difference data were not normally distributed. The p -values for the pre-test, post-test, and score difference were 0.0063, < 0.001, and 0.0128, respectively. Therefore, the Wilcoxon Signed-Rank test was used to examine the difference between pre-test and post-test scores. The test produced a statistic of 1.0 with $p < 0.001$, indicating a statistically significant difference between students' scores before and after the implementation of the model.

The N-Gain analysis showed an average score of 0.882 among 19 valid students, which falls into the high category. Six students were not included in the N-Gain calculation because their pre-test scores had already reached 100, leaving no room for improvement

under the N-Gain formula. Of the 19 valid students, 17 were categorized as high and 2 were categorized as low. This result suggests that most students who still had room for improvement experienced substantial learning gains during the implementation.

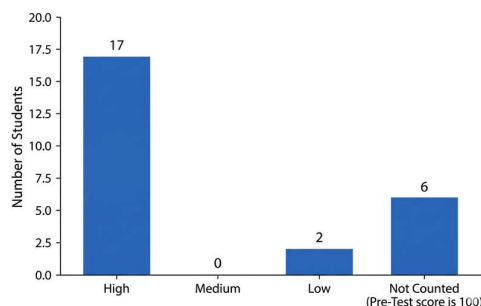


Figure 3. Distribution of N-Gain categories. Six students were not included because their pre-test scores had already reached 100

Student responses were in the very good category, with an average of 90.19%. The highest item was ease of application use (94.44%), followed by the system's ability to support searching by theme or meaning (91.67%) and to understand relationships among verses (90.74%) systematically. Items related to CT were also high, namely support for logical and structured thinking (87.96%) and the development of skills in identifying patterns, selecting important information, and arranging search steps (88.89%). These data support the conclusion that students perceived the system as useful, while still representing perception data.

Teacher responses were in the good category, with an average of 77.50%. Teachers considered the system easy to use and relevant as a supporting medium, but not yet sufficiently robust for more specific topics. Teacher interviews indicated that the system helped with initial searching, but needed to be differentiated more clearly from general search engines. Teachers also emphasized the need to include Arabic verses, concise tafsir, thematic metadata, and more in-depth reference sources. Figure 4 shows the difference between average student and teacher responses; this difference is important because students assessed user experience, while teachers assessed pedagogical feasibility and implementation sustainability.

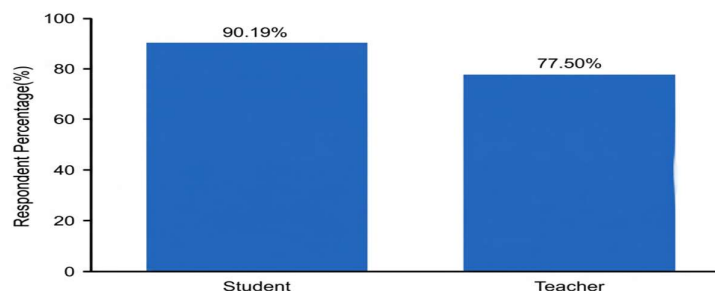


Figure 4. Average Percentage of Students and Teacher Responses

Table 1 summarizes the quantitative and qualitative findings. The table shows that the strongest evidence lies in score changes and student acceptance, while the weaker aspects are retrieval-performance validation, performance-based CT measurement, and database depth. Thus, the data not only support the system's preliminary benefits but also indicate an agenda for system revision.

**Table 1. Summary of evaluation results for the NLP-based verse retrieval system**

Evaluation aspect	Main result	Cautious interpretation	Implication
Descriptive statistics	Pre-test M = 74.40; SD = 20.43. Post-test M = 96.40; SD = 12.54.	There was an increase within one group.	A control group is needed for causal inference.
Normality	Shapiro-Wilk pre-test p = 0.0063; post-test p = 0.0000; score difference p = 0.0128.	The data were not normal.	A non-parametric test was used.
Difference test	Wilcoxon Signed-Rank: statistic = 1.0; p = 0.000177.	A significant pre-post difference was found.	Supports a preliminary indication of effectiveness.
N-Gain	19 valid students; mean = 0.882; 17 high; 2 low; 6 not calculated because pre-test = 100.	Most students with room for improvement showed strong development.	Learning differentiation remains necessary.
Student responses	Average 90.19%, very good category.	Positive perception data, not evidence of CT performance.	Students easily accepted the system.
Teacher responses	Average 77.50%, good category.	Teachers considered the system feasible but requiring deeper development.	Add Arabic verses, tafsir, metadata, and relevance validation.
Teacher interview	The system supported initial searching and should be distinguished from general search engines.	The system is a supporting medium, not a final source of tafsir.	The teacher's role remains central.

Note. M = mean; SD = standard deviation. The table presents the main results and a critical interpretation to ensure that the claims are consistent with the research design.

Discussion

The findings indicate that integrating an NLP-based retrieval model into Islamic Education learning can support a more exploratory learning process. The increase from the pre-test to the post-test suggests that students were able to use the learning sequence to improve their understanding of thematic verse relationships. However, because the study used a single-group design, the improvement should be interpreted as a preliminary indication rather than definitive causal evidence. Other learning factors, including teacher explanation, peer discussion, practice, and test familiarity, may also have contributed to the score increase. This cautious interpretation is consistent with methodological guidance for pre-experimental designs, which require further control or comparison groups before causal claims can be made (Campbell & Stanley, 1963).

The high N-Gain value suggests that students who still had room for improvement experienced meaningful development during the learning sequence. In educational evaluation, N-Gain is useful because it considers the possible improvement range rather than only the absolute score difference (Hake, 1998). In this study, the N-Gain result supports the view that technology-enhanced exploration can help students move from initial understanding to more systematic comprehension of lesson-related verses. Nevertheless, the six students who had already reached a perfect pre-test score also show that future implementation should



include differentiated tasks so that high-performing students can engage in more advanced inquiry.

The model's contribution to CT-oriented learning can be understood through the structure of the learning activities. Students were not merely asked to type keywords and accept results. They were guided to decompose broad Islamic themes into searchable concepts, recognize patterns among retrieved verses, abstract the core messages, and organize their search-reflection steps. This interpretation aligns with cross-disciplinary CT research showing that CT can be operationalized beyond computer science when learners model problems, identify patterns, and reason systematically within disciplinary content (Caskurlu et al., 2022; Manfra, 2022; Christensen, 2023). Therefore, the deconstruction of religious themes into keywords and conceptual patterns can reasonably be treated as a CT-oriented learning process, although it was not yet measured with a full performance rubric in this study.

Compared with conventional verse-search activities, the NLP-based model provides a more structured path for inquiry. Conventional learning often depends on teacher-selected verses or literal keyword searches. Such methods remain important, especially for maintaining interpretive accuracy, but they may limit students' opportunity to explore related verses independently. NLP-based retrieval expands access to possible verse connections and can stimulate discussion about relevance, meaning, and context. This role is consistent with research on Quranic semantic retrieval, which shows that retrieval systems can help users move beyond exact word matching toward meaning-oriented search (Mohamed & Shokry, 2022; Rajagede et al., 2023). However, the teacher's role remains central because retrieval output still requires pedagogical framing and tafsir guidance.

The positive student responses indicate that the model was usable and perceived as beneficial. Ease of use was the highest-rated item, suggesting that the system did not create a major technical barrier for students. This is important because educational technology can only support learning when learners are able to use it without excessive cognitive load. Student perceptions of logical and structured thinking also suggest that the system supported the intended CT-oriented process. Similar educational NLP studies have shown that NLP can be used to analyze learning discussions and misconceptions, thereby supporting more data-informed learning processes (Sullivan & Keith, 2019; Kökver et al., 2024). The present study extends this line of work into Islamic Education by using NLP as a medium for thematic Quranic exploration.

Teacher responses were positive but more cautious than student responses. This difference is pedagogically meaningful. Students mainly assessed usability and perceived benefit, while teachers assessed feasibility, content adequacy, and long-term implementation needs. Teachers' suggestions to add Arabic verses, concise tafsir, thematic metadata, and relevance validation show that the system must develop beyond a simple retrieval prototype if it is to become a mature learning medium. These findings are consistent with the broader literature on trustworthy AI-supported learning systems, which emphasizes the need for transparency, expert validation, and responsible integration into teaching practice (Darvishi et al., 2022).

Overall, the study demonstrates the potential of combining NLP, Islamic Education, and CT-oriented pedagogy. The model does not replace teachers or tafsir, but it creates an inquiry space in which students can search, compare, discuss, and reflect. The strongest evidence lies in learning-outcome changes and positive user acceptance. The weakest areas are retrieval-performance validation, objective CT measurement, and corpus depth.



Therefore, the model should be viewed as a promising preliminary development that requires stronger validation in future research.

Conclusion

This study developed an Islamic Education learning model based on NLP-driven Quranic verse retrieval within the ADDIE framework. The model positions the technological system as an exploration medium that supports students in formulating themes, entering keywords, comparing retrieval results, and developing reflections under teacher guidance, rather than replacing the teacher or serving as a final source of tafsir. The preliminary evaluation showed an improvement in learning outcomes within one group. The mean pre-test score increased from 74.40 to 96.40 in the post-test. The Wilcoxon Signed-Rank test indicated a significant difference ($p < 0.001$), and the mean N-Gain of 0.882 was in the high category. Student responses were very good (90.19%), while teacher responses were good (77.50%). These findings indicate that the model has potential to support exploratory, technology-enhanced Islamic Education learning.

Recommendation

For teachers, the NLP-based Quranic verse retrieval model should be used as a guided exploration medium, not as an independent source of interpretation. Teachers need to prepare themes, guide students in selecting appropriate keywords, help students evaluate the relevance of retrieved verses, and connect the results with valid Islamic references and classroom discussion. Teachers are also advised to combine the system with reflective worksheets that explicitly ask students to identify themes, compare verses, abstract core messages, and explain their reasoning steps so that CT indicators become visible during learning.

For future researchers, the model should be tested using stronger research designs, such as a quasi-experimental design with a comparison group. Future studies should also develop performance-based CT rubrics and collect search-activity logs to measure students' CT processes more objectively. From a technical perspective, future work should enrich the corpus with Arabic verses, validated translations, concise tafsir, thematic metadata, and expert-reviewed relevance labels. Retrieval performance should be evaluated using standard information-retrieval metrics such as precision, recall, mean reciprocal rank, and nDCG. These improvements are necessary to strengthen both the pedagogical and technical validity of the model.

References

- Agustian, S., & Wulandari, I. S. (2013). Sistem Quran retrieval terjemahan Bahasa Indonesia berbasis web dengan reorganisasi korpus. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 1–6.
- Bashir, M. H., Azmi, A. M., Nawaz, H., Zaghouani, W., Diab, M., Al-Fuqaha, A., & Qadir, J. (2023). Arabic natural language processing for Qur'anic research: A systematic review. *Artificial Intelligence Review*, 56(7), 6801–6854. <https://doi.org/10.1007/s10462-022-10313-2>
- Beecher, K. (2017). *Computational thinking: A beginner's guide to problem-solving and programming*. BCS Learning & Development.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.



- Caskurlu, S., Hu, A. D., Yadav, A., & Santo, R. (2022). Computational thinking integration design principles in humanities. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2* (p. 1174). Association for Computing Machinery. <https://doi.org/10.1145/3478432.3499071>
- Christensen, I.-M. F. (2023). Computing with concepts using tangible, computational tools: A 21st century competency for teachers and students in the humanities. *Læring og Medier*, 15(27). <https://doi.org/10.7146/lom.v15i27.134149>
- Darvishi, A., Khosravi, H., Sadiq, S., & Gašević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, 53(4), 844–875. <https://doi.org/10.1111/bjet.13233>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Jackson, P., & Moulinier, I. (2007). *Natural language processing for online applications: Text retrieval, extraction and categorization*. John Benjamins Publishing Company.
- Kökver, Y., Pektaş, H. M., & Çelik, H. (2024). Artificial intelligence applications in education: Natural language processing in detecting misconceptions. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12919-1>
- Manfra, M. M., Hammond, T. C., & Coven, R. M. (2022). Assessing computational thinking in the social studies. *Theory & Research in Social Education*, 50(2), 255–296. <https://doi.org/10.1080/00933104.2021.2003276>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Mohamed, E. H., & Shokry, E. M. (2022). QSST: A Quranic semantic search tool based on word embedding. *Journal of King Saud University - Computer and Information Sciences*, 34(3), 934–945. <https://doi.org/10.1016/j.jksuci.2020.01.004>
- Rajagede, R. A., Haryono, K., & Qardafil, R. (2023). Semantic retrieval for Indonesian Quran autocompletion. *Jordanian Journal of Computers and Information Technology*, 9(2), 94–106. <https://doi.org/10.5455/jjcit.71-1668279800>
- Sullivan, F. R., & Keith, P. K. (2019). Exploring the potential of natural language processing to support microgenetic analysis of collaborative learning discussions. *British Journal of Educational Technology*, 50(6), 3047–3063. <https://doi.org/10.1111/bjet.12875>
- Taan, A. A., Khan, S. U. R., Raza, A., Hanif, A. M., & Anwar, H. (2021). Comparative analysis of information retrieval models on Quran dataset in cross-language information retrieval systems. *IEEE Access*, 9, 169056–169067. <https://doi.org/10.1109/ACCESS.2021.3126168>
- Untoro, R. C. (2022). *Pendidikan Agama Islam dan budi pekerti untuk SMA/SMK/MA kelas XI*. Pusat Perbukuan, Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.