

---

*Original Article*

# Latent Dirichlet allocation for generating concise three-word central themes from research conference titles

Tristan Jay P. Calaguas<sup>1</sup>, Frances Frangelico S. Friginal<sup>1\*</sup>, and Erlinda B. Gonzalez<sup>2</sup>

<sup>1</sup>Information Technology Department, Emilio Aguinaldo College-Manila, Manila Philippines

<sup>2</sup>Library Department, Emilio Aguinaldo College-Manila, Manila Philippines

\*Corresponding email: tristanjay.calaguas@eac.edu.ph; gio.friginal@eac.edu.ph

## ABSTRACT

Generating a concise central theme from a heterogeneous set of research titles is a recurring but time-sensitive task in research conferences. This study evaluated whether latent Dirichlet allocation (LDA) can serve as an assistive approach for generating three-word central themes from mixed-domain research titles. Specifically, it examined the coherence of the generated themes, described process-efficiency estimates across increasing document volumes, and determined respondents' evaluation of the highest-scoring theme. A quantitative descriptive-evaluative design was used. Twenty publicly available English-language titles were compiled, preprocessed through lowercasing, tokenization, stop-word removal, and trigram construction, and then analyzed through an LDA workflow that produced five candidate themes. Coherence scores were computed for the generated themes, while process efficiency was described using estimated manual turnaround-time comparisons. The theme with the highest coherence score was further evaluated by 10 respondents using a structured 5-point Likert scale questionnaire administered online. The model generated five candidate themes with coherence scores ranging from 0.46 to 0.55 and an overall mean coherence score of 0.502, indicating generally coherent topic formation within the study rubric. Process-efficiency estimates showed increasing human turnaround time as document load expanded. The highest-scoring theme, "health digital telemedicine," also received an overall mean rating of 4.45, interpreted as Strongly Agree, while the six-item instrument showed excellent internal consistency (Cronbach's alpha = 0.93). The findings indicate that LDA can function as a preliminary decision-support tool for deriving concise conference themes from small title corpora, although broader validation with larger and more homogeneous datasets remains necessary.

---

## Article Info

### *Article history:*

Received: 30 March 2026

Revised: 27 April 2026

Accepted: 30 April 2026

---

*Keywords:* Medical imaging, parental knowledge, radiation exposure, radiation safety, and pediatric imaging

---

## INTRODUCTION

Research conferences and scholarly events often require facilitators to synthesize multiple submitted titles into a concise central theme. This task is commonly done manually even though topic-modeling methods already provide a systematic way

to identify latent semantic patterns in text collections. Latent Dirichlet allocation is one of the foundational probabilistic topic-modeling approaches because it represents documents as mixtures of topics and topics as distributions over words (Blei et al., 2003; Blei, 2012). Early applications likewise showed that topic models can reveal interpretable thematic structures in scientific corpora (Griffiths & Steyvers, 2004). Because model outputs must still be interpretable to end users, topic coherence measures are often used to judge whether the top words of a generated topic are meaningfully related (Röder et al., 2015).

Applying LDA to short text remains challenging because titles contain limited word co-occurrence information and sparse lexical cues. Studies on short-text topic modeling have shown that conventional topic models may become less stable when documents are brief and semantically heterogeneous (Albalawi et al., 2020; Qiang et al., 2022). Methodological guidance for applied LDA studies also emphasizes transparent preprocessing, parameter selection, and interpretation because these choices affect validity and reliability (Maier et al., 2018). This limitation is especially relevant in conference title aggregation, where organizers often need a quick, defensible, and understandable umbrella theme from a small and mixed set of titles. The present study addresses this practical gap by applying LDA to a narrowly defined operational problem rather than presenting it as a new algorithmic innovation.

Specifically, the study aimed to: (1) determine the coherence scores of the three-word themes generated from twenty research titles; (2) describe process-efficiency estimates across increasing document volumes using manual turnaround-time comparisons; and (3) evaluate the highest-scoring LDA-generated theme using a

respondent survey focused on relatedness, clarity, comprehensibility, task fulfillment, and overall satisfaction. The novelty of the study lies not in proposing a new topic-modeling algorithm, but in positioning LDA as a preliminary decision-support framework for conference-theme generation from mixed-domain titles in an institutional setting.

## **METHODS**

### **Study Design**

This study used a quantitative descriptive-evaluative design. The workflow examined how an LDA-based topic-modeling procedure generated candidate three-word themes from a small corpus of research titles, and it subsequently evaluated the highest-scoring theme through an online respondent survey. Because the study focused on the descriptive assessment of algorithmic output and respondent perceptions rather than on intervention or controlled experimentation, it is presented as an applied text-analytics evaluation. The modeling stage used a fixed exploratory configuration that generated five candidate topics ( $k = 5$ ), which was considered a manageable number of alternative theme options for comparison within a 20-title corpus. Hyperparameter optimization was not a primary aim of the study; instead, the analysis prioritized a transparent, practical workflow consistent with methodological guidance for applied LDA research.

### **Ethics Statement**

The title-analysis component used publicly available research and seminar titles only. For the respondent-validation component, participation in the online survey was voluntary, names were optional, and

no sensitive personal data were collected. Respondents were informed of the purpose of the questionnaire and submitted their answers electronically. No formal ethics approval protocol number was available at the time of manuscript preparation.

### **Setting**

The title corpus was assembled from publicly accessible online sources, while respondent validation was conducted online through Google Forms. Data collection and online validation were completed in March 2026. The study context reflects a practical conference or research-management environment in which facilitators need to derive a central theme from a limited number of mixed-domain titles.

### **Materials**

The materials consisted of twenty English-language research or seminar titles covering diverse themes such as healthcare, finance, sustainability, cybersecurity, education, transportation, agriculture, and computing. For Objective 3, ten respondents were purposively included to evaluate the highest-scoring generated theme. The respondent group represented academic and technology-related roles, including faculty members, students, a programmer, and a laboratory custodian, reflecting the types of end users who may encounter thematic grouping tasks in institutional settings.

### **Data Collection**

Relevant titles were identified from publicly accessible online sources and manually compiled into a text file for analysis. Duplicate entries were avoided. After the highest-coherence theme was identified, a structured online

questionnaire was administered to 10 respondents. The questionnaire asked respondents to rate the generated theme against six indicators using a 5-point Likert scale and also provided an optional comments section.

### **Data Processing and Analysis**

Following common text-preprocessing practices for natural language processing and topic modeling (Bird et al., 2009; Uysal & Gunal, 2014), the titles were converted to lowercase, tokenized into word units, and cleaned through stop-word removal. Since the study aimed to generate concise three-word themes, trigram construction was applied to preserve meaningful lexical groupings prior to topic modeling. The processed corpus was then analyzed through LDA to generate five candidate topics, each expressed as a three-word theme. The use of five candidate topics ( $k = 5$ ) served as a pragmatic exploratory setting intended to produce a small, comparable set of possible umbrella themes from the limited title corpus. The study did not perform alpha/beta tuning or compare multiple  $k$  values, because it was designed as a preliminary applied evaluation rather than a hyperparameter-optimization study. Topic relatedness was evaluated using coherence scores, and the arithmetic mean coherence score was used to summarize overall topic quality. Process efficiency was described using estimated, scenario-based manual turnaround-time comparisons across document volumes of 20, 50, 100, and 150 titles. The theme with the highest coherence score was selected for respondent evaluation.

### Statistical methods

Descriptive statistics were used throughout the study. Individual coherence scores and their arithmetic mean were reported to describe topical relatedness, while total manual turnaround time and adjusted human turnaround time were presented across four document loads to describe process-efficiency trends. For the respondent survey, mean, standard deviation, and overall mean were computed for each indicator. The following verbal interpretation was applied: 4.21–5.00, Strongly Agree; 3.41–4.20, Agree; 2.61–3.40, Neutral; 1.81–2.60, Disagree; and 1.00–1.80, Strongly Disagree. Cronbach’s alpha was also computed to assess the internal

consistency of the six-item questionnaire. Because topic coherence values may vary across corpora, preprocessing choices, and implementations, the terms good topic and moderate were used only as internal descriptive labels for comparing outputs within this dataset rather than as universal threshold categories.

### RESULTS

The corpus consisted of twenty publicly available titles drawn from multiple domains. After preprocessing and topic modeling, the LDA workflow generated five candidate three-word themes.

Table 1. Generated three-word themes and coherence scores.

Generated Theme	Coherence Score	Interpretation	Rank
health digital telemedicine	0.55	Good Topic	1
energy renewable smart	0.52	Good Topic	2
learning deep image	0.50	Good Topic	3
data analytics decision	0.48	Moderate	4
security cybersecurity privacy	0.46	Moderate	5
Mean	0.502	Good Topic	—

Table 1 presents the resulting themes and their corresponding coherence scores. For descriptive comparison within this study, themes with coherence scores of 0.50 and above were labeled good topics, whereas those below 0.50 but within the observed range were labeled moderate. These labels were used only as internal

comparative descriptors for this corpus and should not be interpreted as universal topic-quality thresholds. The highest coherence score was obtained by the theme “health digital telemedicine” (0.55), followed by “energy renewable smart” (0.52). The overall mean coherence score was 0.502.

Table 2. Estimated manual and adjusted turnaround time across document volumes

Documents (D)	Reading Prep (min)	Grouping (min)	Summary (min)	Total Manual TAT	Adjusted Human TAT
20	10	80	10	100	108
50	25	200	10	235	254
100	50	400	10	460	497
150	75	600	10	685	740

Table 2 presents the study's estimated manual turnaround-time scenarios across increasing document volumes. The values indicate that total manual turnaround time would increase as document load rose from 20 to 150 titles. The same upward pattern was observed for adjusted human turnaround

time, suggesting that workload expansion would proportionally increase the time required to organize and summarize larger sets of titles without computational support. These values should be interpreted as scenario-based estimates rather than empirically timed observations.

Table 2. Estimated manual and adjusted turnaround time across document volumes

<b>Documents (D)</b>	<b>Reading Prep (min)</b>	<b>Grouping (min)</b>	<b>Summary (min)</b>	<b>Total Manual TAT</b>	<b>Adjusted Human TAT</b>
20	10	80	10	100	108
50	25	200	10	235	254
100	50	400	10	460	497
150	75	600	10	685	740

Table 3 presents the respondents' evaluation of the highest-scoring LDA-generated theme, "health digital telemedicine." All six indicators obtained high mean ratings, ranging from 4.30 to 4.60, with an overall mean of 4.45, interpreted as Strongly Agree. The highest mean was obtained by the indicator on ease of

understanding ( $M = 4.60$ ,  $SD = 0.52$ ), while the lowest mean was observed for whether the theme summarized the central idea of the titles ( $M = 4.30$ ,  $SD = 0.82$ ). The six-item questionnaire also showed excellent internal consistency (Cronbach's  $\alpha = 0.93$ ), indicating that the respondent ratings were sufficiently reliable for descriptive interpretation.

Table 3. Respondents' evaluation of the highest-scoring LDA-generated theme (n = 10)

<b>Indicator</b>	<b>Mean</b>	<b>SD</b>	<b>Interpretation</b>
The generated theme is related to the collection of research titles.	4.40	0.52	Strongly Agree
The generated theme is easy to understand.	4.60	0.52	Strongly Agree
The generated theme is clear and meaningful.	4.50	0.53	Strongly Agree
The generated theme summarizes the central idea of the given titles.	4.30	0.82	Strongly Agree
The generated theme fulfills the intended task of identifying a common topic from multiple titles.	4.40	0.52	Strongly Agree
Overall, I am satisfied with the generated theme produced by the algorithm.	4.50	0.53	Strongly Agree
<b>Overall Mean</b>	<b>4.45</b>	—	<b>Strongly Agree</b>

---

## DISCUSSION

The study showed that LDA was able to generate five candidate three-word central themes from a heterogeneous set of twenty titles, with coherence scores ranging from 0.46 to 0.55. Based on the study's internal comparative rubric, the average coherence score of 0.502 indicates generally coherent topic formation. The survey results further showed that respondents strongly agreed that the highest-scoring theme was understandable, meaningful, related to the given titles, and capable of fulfilling the intended thematic-grouping task.

These findings are consistent with the general role of LDA as a topic-discovery method for text corpora (Blei et al., 2003; Blei, 2012). Prior work has also shown that topic interpretation remains partly human-centered, which makes end-user validation important when generated labels are intended for practical use (Chang et al., 2009). In this study, coherence-based evaluation helped verify whether the top words of each generated topic were semantically related, consistent with the broader literature on topic-quality assessment (Newman et al., 2010; Mimno et al., 2011; Lau et al., 2014; Röder et al., 2015). However, coherence score alone should not be treated as a complete measure of topic quality. Alternative evaluation approaches, such as other coherence formulations, direct human benchmarking, or comparative labeling tasks, could provide a fuller assessment in future work. At the same time, the study should be interpreted as an applied evaluation rather than as evidence of a new algorithmic contribution. Prior work on short-text topic modeling has shown that sparse word co-occurrence can make topic inference more difficult when documents are very short, such as titles, headlines, or snippets (Albalawi et al., 2020; Qiang et al., 2022). In this context, the

moderate-to-good coherence scores and strong respondent ratings suggest that the approach is useful as a preliminary aid for conference-theme generation. This practical contribution is best understood as an institutional decision-support application of LDA rather than a claim of theoretical innovation.

The process-efficiency estimates also have practical implications. As the number of titles increased, estimated manual turnaround time rose substantially, indicating that a computational aid may be valuable when conference committees or academic offices must process larger batches of submissions. However, these turnaround-time values were estimated rather than empirically timed, so they should be interpreted as simulated workload comparisons rather than direct performance measurements. The generated theme should still be treated as a draft or assistive output rather than as a definitive final label. Human review remains important, particularly when the source titles are multidisciplinary or semantically uneven. The study has several limitations. First, the dataset was small and consisted of only twenty titles, which limits topic stability and generalizability. Second, the corpus combined titles from different disciplines, which may have reduced semantic concentration within topics. Third, the respondent validation involved only ten evaluators drawn from a limited institutional context, which introduces possible sampling bias and limited statistical power for broader inference. These features should therefore temper the interpretation of the findings. Future work should examine larger and more homogeneous datasets, compare LDA with online, dynamic, aggregation-based, and short-text-specific topic models (Hoffman et al., 2010; Diao & Jiang, 2013; Sasaki et al., 2014; Steinskog et al., 2017), and evaluate generated themes using broader

benchmarking and validation procedures (Wallach et al., 2009; Shi et al., 2019). Additional comparisons between title-only corpora and longer text representations may also clarify how document granularity influences coherence behavior (Syed & Spruit, 2017).

## CONCLUSION

Latent Dirichlet allocation demonstrated potential as a preliminary decision-support approach for generating concise three-word central themes from research conference titles. In this study, the model produced five candidate themes with coherence scores from 0.46 to 0.55 and an overall mean coherence score of 0.502, indicating generally coherent topic generation within the study's internal comparative rubric. The scenario-based manual turnaround-time estimates further suggested that human processing demand increases as document volume expands. In addition, the highest-scoring generated theme received an overall respondent rating of 4.45, interpreted as Strongly Agree, indicating that the output was generally understandable, meaningful, and acceptable to users. Taken together, the findings suggest that LDA may assist conference facilitators and academic units in deriving umbrella themes from small title corpora, although broader validation using larger datasets, alternative evaluation metrics, and more diverse evaluators remains necessary before wider institutional adoption.

## Author's Contributions

Tristan Jay P. Calaguas and Frances Frangelico S. Friginal contributed to the development and design of the study, acquisition, analysis, and interpretation of data, drafting and revising the manuscript. Erlinda B. Gonzalez contributed to

manuscript critique, literature and reference support, and scholarly review of the study. All authors reviewed and approved the final manuscript.

## Ethics Approval Protocol Number

No formal ethics approval protocol number was available at the time of manuscript preparation. The online survey involved voluntary participation, optional identification, and no collection of sensitive personal data.

## Funding Source(s)

None.

## Conflict of Interest

The authors declare that no potential conflict of interest relevant to this article was reported.

## Declaration Concerning Generative AI

The authors did not employ generative artificial intelligence in the analysis of data or in deriving the study findings. AI-assisted tools were used only to improve language clarity and manuscript organization, and no AI system is listed as an author.

## Data Availability and Additional Supporting Documents

The title corpus, anonymized survey summaries, and supporting analytical outputs used in this study are available from the corresponding author upon reasonable request.

## Acknowledgements

None.

## REFERENCES

- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3, Article 42. <https://doi.org/10.3389/frai.2020.00042>

- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media. <http://www.nltk.org/>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems* (Vol. 22, pp. 288–296). Curran Associates, Inc.
- Diao, Q., & Jiang, J. (2013). A unified model for topics, events and users on Twitter. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1869–1879). Association for Computational Linguistics. <https://aclanthology.org/D13-1192/>
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Hoffman, M. D., Blei, D. M., & Bach, F. R. (2010). Online learning for latent Dirichlet allocation. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems* (Vol. 23, pp. 856–864). Curran Associates, Inc.
- Lau, J. H., Newman, D., & Baldwin, T. (2014). Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 530–539). Association for Computational Linguistics. <https://doi.org/10.3115/v1/E14-1056>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Association for Computational Linguistics. <https://aclanthology.org/D11-1024/>
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 100–108). Association for Computational Linguistics. <https://aclanthology.org/N10-1012/>

- Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X. (2022). Short text topic modeling techniques, applications, and performance: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(3), 1427–1445. <https://doi.org/10.1109/TKDE.2020.2992485>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399–408). <https://doi.org/10.1145/2684822.2685324>
- Sasaki, K., Yoshikawa, T., & Furuhashi, T. (2014). Online topic model for Twitter considering dynamics of user interests and topic trends. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1977–1985). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1212>
- Shi, H., Gerlach, M., Diersen, I., Downey, D., & Amaral, L. A. N. (2019). A new evaluation framework for topic modeling algorithms based on synthetic corpora. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (pp. 816–826). PMLR. <https://proceedings.mlr.press/v89/shi19a.html>
- Steinskog, A., Therkelsen, J., & Gambäck, B. (2017). Twitter topic modeling by tweet aggregation. In *Proceedings of the 21st Nordic Conference on Computational Linguistics* (pp. 77–86). Association for Computational Linguistics. <https://aclanthology.org/W17-0210/>
- Syed, S., & Spruit, M. (2017). Full text or abstract? Examining topic coherence scores using latent Dirichlet allocation. In *2017 IEEE International Conference on Data Science and Advanced Analytics* (pp. 165–174). IEEE. <https://doi.org/10.1109/DSAA.2017.61>
- Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, 50(1), 104–112. <https://doi.org/10.1016/j.ipm.2013.08.006>
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 1105–1112). ACM. <https://doi.org/10.1145/1553374.1553515>