

A NLP Method for Estimating the Alignment of University Curricula with Essential Job Skills

1st Daria Hypolite*

Department of Computer Science
The University of the West Indies
St. Augustine, Trinidad and Tobago
dariah1707@gmail.com

*Corresponding author

2nd Denelle Mohammed

Department of Computer Science
The University of the West Indies
St. Augustine, Trinidad and Tobago
denellemohammed164@gmail.com

3rd Patrick Hosein

Department of Computer Science
The University of the West Indies
St. Augustine, Trinidad and Tobago
patrick.hosein@uwi.edu

Abstract—Computing curricula must be systematically and continuously revised to respond to rapidly evolving industry requirements and an increasingly AI-mediated labor market. This study applies a dual-method Natural Language Processing (NLP) framework to evaluate the alignment between the program content of the Computer Science department of a specific university and the contemporary job market pertinent to its graduates. Job descriptions were collected from LinkedIn, Reed UK, Workopolis, and CaribbeanJobs using Selenium and BeautifulSoup in parallel with course syllabi from the Computer Science and Information Technology degree programs. SkillNER is employed for domain-specific skill extraction, and Latent Dirichlet Allocation (LDA) is used for topic modeling. A token-level closeness metric is defined to quantify surface vocabulary alignment, producing an overall weighted LDA-based alignment score of $\rho_w = 0.0412$. As a complementary measure, semantic embedding alignment based on the all-MiniLM-L6-v2 sentence-transformer model captures conceptual similarity in the presence of lexical divergence, yielding substantially higher alignment scores ranging 0.28 to 0.67. The gap between the two measures indicates that vocabulary register divergence, as opposed to a complete absence of relevant preparation, is the primary driver of weak token-level alignment. The findings suggest that the curricula preserve robust theoretical foundations but exhibit a deficit in explicit, industry-oriented tooling vocabulary.

Index Terms—curriculum alignment, job market skills, natural language processing, LDA topic modelling, SkillNER, semantic embedding, sentence transformers, computing education

I. INTRODUCTION

The rapid evolution of computing technologies has significantly changed the skills needed in the modern workforce, especially in Computer Science (CS), Information Technology (IT) and Artificial Intelligence (AI). As industries adopt emerging technologies, concerns have grown about whether university curricula equip graduates with the competencies employers require [4]. Studies report a widening gap between academic preparation and industry expectations, as job roles increasingly prioritize practical, current technical skills [7].

Computing curricula have traditionally emphasized theoretical foundations over practical, industry-focused skills. While this ensures academic rigor, it can exacerbate misalignment between graduate capabilities and employer needs. In the Caribbean, this issue is compounded by scarce local job market data and high graduate mobility to international markets.

Recent work has shown that NLP techniques such as skill extraction and topic modeling can analyze relationships between curricula and job requirements [1]. However, token-level methods like LDA-based closeness metrics are sensitive to vocabulary differences between academic and industry texts, and may underestimate true conceptual alignment. Semantic embedding techniques mitigate this by mapping text into a shared vector space in which conceptually related terms are close regardless of wording [16], [17].

This study makes four contributions. First, it proposes a dual-method evaluation framework combining LDA-based token-level alignment with semantic embedding alignment, explicitly quantifying vocabulary mismatch as the gap between the two measures. Second, it applies this framework comparatively across two computing disciplines—Computer Science and Information Technology—at the Department of Computing and Information Technology (DCIT) at the University of the West Indies. Third, it introduces regional stratification by separating Caribbean and international (US, UK, Canada) job markets, addressing a gap in prior work that treats job markets as homogeneous. Fourth, AI job descriptions are included to assess alignment with emerging industry demands.

II. LITERATURE REVIEW

Aligning computing syllabi with industry demands has emerged as a significant challenge in tertiary education as technology evolution accelerates and graduate employability becomes tied more to practical skills [4]. This review examines two themes: curriculum–industry alignment methodologies and academic versus vocational tension.

Ramnarine et al. [1] established an NLP-based framework combining web scraping, SkillNER skill extraction and LDA topic modeling to assess electrical and computer engineering programs. Their closeness metric quantifies token-level alignment between curricula and job requirements, providing a replicable methodology we adapt for CS, IT and AI contexts. Gurcan and Cagiltay [5] similarly employed LDA for analyzing big data engineering competencies. However, both studies treat geographic markets as homogeneous and do not account for the vocabulary register divergence between academic and industry text as a distinct methodological problem. Walker [6]

advanced skill extraction using SkillNER, achieving high accuracy in identifying specialized competencies, though soft skills remained under-represented.

Sentence-transformer models, introduced by Reimers and Gurevych [16], enable efficient dense semantic similarity computation and have been applied to skill-matching tasks [17]. Januzaj et al. [18] demonstrated TF-IDF cosine similarity for curriculum to job alignment, finding it more robust than keyword matching against vocabulary variation. These embedding-based approaches represent a methodological advance over strict token-matching methods, particularly for cross-register comparisons between academic syllabi and industry jobs.

Lastly, there is a persistent tension between theoretical foundations and practical training in computing education. Gope and Gope [2] identified significant disconnects between student preparation and industrial requirements. Leandro et al. [3] documented how traditional programs struggle to integrate emerging technologies into established curricula. Radermacher and Walia [7] document industry frustration with graduates lacking practical competencies including version control, testing frameworks and agile methodologies.

III. METHODOLOGY

A. Data Collection

The data collection process compiled two primary datasets: job descriptions and course outlines. Job descriptions were sourced from four major online platforms. LinkedIn, Reed UK and Workopolis were selected to represent international markets across the US, UK and Canada, while CaribbeanJobs was used to capture local Caribbean employment demand.

Job postings were collected during the February 2026 period using a two-stage scraping process. In the first stage, Selenium was used to navigate dynamic pages and extract job posting URLs. In the second stage, BeautifulSoup was used to parse the HTML content of each posting and extract relevant sections including responsibilities, requirements and skills. Job postings were filtered using standardized, role-specific search terms across three categories: Computer Science roles (e.g., Software Engineer, Backend Developer), Information Technology roles (e.g., IT Support Specialist, Cybersecurity Analyst) and AI and ML roles (e.g., Data Scientist, Machine Learning Engineer).

Course outlines were obtained in PDF format directly from the department. Each course outline was assigned to one or both thematic areas, Computer Science and Information Technology, based on its program classification. An additional dataset of unrelated job descriptions was collected to serve as a lower bound baseline in the closeness metric calculation. All data was collected from publicly available sources. No personal or sensitive information was retained. Identifiable details were removed from job descriptions.

B. Data Preprocessing

The preprocessing pipeline standardizes raw text and reduces noise prior to analysis [12]. SpaCy [15] was selected

for its tokenization capability and processing speed. The pipeline applied the following steps in sequence to both job descriptions and course outlines.

First, non-alphabetic character removal filtered out punctuation, numbers and special characters [14]. Text was then converted to lowercase to ensure consistent token recognition [12]. Tokenization decomposed each document into individual word tokens [13]. Stopword removal eliminated common words such as *the*, *and* and *is*, which carry negligible semantic value for topic inference [12]. Following the approach of Gurcan and Cagiltay [5], stemming and lemmatisation were deliberately excluded from the pipeline to preserve domain-specific terminology such as *machine learning*, *object-oriented programming* and *cloud computing*.

C. Skill Extraction

Technical and soft skills were extracted from all datasets using SkillNER [9]. SkillNER employs a multi-stage pipeline including pattern-based and database-driven matching, POS-based filtering and n-gram detection to identify normalized skill entities.

Unlike the original study by Ramnarine et al. [1], which mapped documents to five ECE thematic areas, this study organizes extracted skills into two program-aligned categories, Computer Science and Information Technology, with AI job skills treated as a separate corpus for regional comparison analysis. Skills were extracted separately for Caribbean and international job postings to enable regional stratification.

D. Coherence Score

An essential step in LDA topic modeling is determining the optimal number of topics k [10]. To guide this selection, topic coherence was used as a diagnostic metric. The C_v coherence measure [11] combines several desirable properties of earlier measures, making it suitable for moderately sized corpora. Models were trained for values of k ranging from 1 to 20 and coherence scores were computed. Coherence evaluation was performed separately on the CS and IT course skill corpora.

E. Closeness Metric

The closeness metric quantifies the alignment between the DCIT curriculum and computing job market requirements. Following the approach of Chamansingh and Hosein [8] as adapted by Ramnarine et al. [1], LDA is applied to generate topic clusters that serve as base topics T . In this study, base topics T are derived from the course syllabus corpus rather than the job description corpus. The LDA model is trained on the course skill corpus using the optimal k values determined from the coherence evaluation. The resulting topics T capture the dominant skill themes present in the DCIT curriculum. Three datasets are then compared against T :

- **Course Syllabi (CS):** the curriculum itself, serving as the upper bound C_{CS}
- **Job Descriptions (JD):** computing job postings, split into Caribbean ($C_{JD_{local}}$) and International ($C_{JD_{intl}}$)

- **Random Unrelated (RU)**: unrelated job descriptions serving as the lower bound C_{RU}

For each document, the skill tokens are converted to a bag-of-words representation. For each topic in T , the probability of each topic word is multiplied by its frequency in the document and summed. The topic with the highest summed probability is selected and the closeness score for a dataset is computed as the mean of the maximum topic scores across all documents. The alignment metric, ρ , is then computed as:

$$\rho \equiv \frac{C_{JD} - C_{RU}}{C_{CS} - C_{RU}} \quad (1)$$

where C_{CS} represents the upper bound, C_{RU} the lower bound and C_{JD} the closeness of job descriptions to the curriculum-derived topics. A ρ value approaching 1 indicates strong alignment between job market demands and the curriculum. A weighted overall alignment score ρ_w is computed across both thematic areas to account for differences in dataset size:

$$\rho_w = \frac{\sum(\rho_{thematic} \times N_{thematic})}{N_{total}} \quad (2)$$

where $\rho_{thematic}$ is the alignment score for a thematic area, $N_{thematic}$ is the number of job descriptions in that area and N_{total} is the total number of job descriptions.

F. Semantic Embedding Alignment

While the LDA closeness metric captures token-level overlap, it is sensitive to the vocabulary divergence between academic syllabi and industry job postings. To address this, a semantic embedding approach is employed as a complementary alignment measure.

Each document's extracted skill tokens are encoded into a 384-dimensional dense vector using the all-MiniLM-L6-v2 sentence-transformer model [16], which maps semantically related terms to geometrically proximate points in the embedding space regardless of surface vocabulary differences. The mean pooled embedding of each document's skill token set is computed to produce a single document-level vector. For each job description or course document, the cosine similarity to the curriculum corpus centroid is computed:

$$\text{sim}(d, CS) = \frac{\mathbf{e}_d \cdot \mathbf{e}_{CS}}{\|\mathbf{e}_d\| \|\mathbf{e}_{CS}\|} \quad (3)$$

where $\mathbf{e}_d$ is the document embedding and $\mathbf{e}_{CS}$ is the mean curriculum embedding. The closeness values C_{CS} , C_{JD} and C_{RU} are then defined as the mean cosine similarity of each dataset to the curriculum centroid, and the same normalized ρ formula (Eq. 1) is applied to these cosine similarity values.

IV. RESULTS

A. Dataset Distribution

Table I summarizes the document distribution across datasets.

TABLE I
DATASET DISTRIBUTION

Dataset	Courses	Local JDs	Intl. JDs
Computer Science	37	13	1,313
Information Technology	31	13	1,351
AI / Data Science	–	–	1,092
Unrelated (baseline)	–	–	1,627

B. Coherence Measure

Both programs achieved an optimal topic count of $k = 18$, with C_v coherence scores of 0.78 (CS) and 0.86 (IT). Tables II and III present the 18 thematic areas for each program, labeled by top contributing skill terms.

TABLE II
CS PROGRAM TOPIC THEMES ($k = 18$, COHERENCE: 0.7844)

T	Theme	T	Theme
0	HCI & Social Computing	9	Software Architecture
1	Databases & Info Systems	10	AI & Machine Learning
2	Software Project Mgmt.	11	OOP & Language Design
3	E-Commerce & Business	12	Networking
4	Cloud & Distributed Sys.	13	Computer Architecture
5	Digital Rights & Society	14	Programming Fundamentals
6	Software Requirements	15	Discrete Mathematics
7	Web Development	16	Operating Systems
8	Cybersecurity	17	Formal Methods & Automata

TABLE III
IT PROGRAMME TOPIC THEMES ($k = 18$, COHERENCE: 0.8573)

T	Theme	T	Theme
0	Simulation & Modelling	9	Data Analysis & Scalability
1	Game Dev. & HCI	10	OOP & Debugging
2	Web Applications & Security	11	Discrete Maths & Security
3	Social Computing & Docs.	12	Social Media & Decision Mgmt.
4	Networking & Infrastructure	13	Software Dev. & Documentation
5	Software Engineering & Test.	14	Algorithms & Data Structures
6	Project Mgmt. & Requirements	15	Computer Architecture
7	Systems & Data Storage	16	Programming Fundamentals
8	Database Systems & Networks	17	OOP Design & UI

C. LDA Closeness and Alignment Scores

Table IV presents the raw LDA closeness values.

TABLE IV
LDA CLOSENESS AND ALIGNMENT RESULTS

Dataset	CS	IT
Courses (upper bound)	0.2540	0.3518
Local (Caribbean) JDs	0.0189	0.0224
International JDs	0.0244	0.0265
AI JDs (CS only)	0.0233	–
Unrelated (lower bound)	0.0127	0.0148

LDA alignment scores ρ (Eq. 1) are presented in Table VI. All LDA scores indicate weak alignment. The overall weighted alignment score is $\rho_w = 0.0412$.

D. Semantic Embedding Alignment

Table V presents embedding-based cosine similarity values and the corresponding ρ scores.

TABLE V
SEMANTIC EMBEDDING ALIGNMENT RESULTS

Segment	C_{CS}	C_{JD}	C_{RU}	ρ
CS Caribbean	0.4243	0.2808	0.2175	0.31
CS International	0.4243	0.3330	0.2175	0.56
CS AI Jobs	0.4243	0.3566	0.2175	0.67
IT Caribbean	0.4582	0.2943	0.2316	0.28
IT International	0.4582	0.3570	0.2316	0.55

E. Method Comparison

Table VI presents side-by-side LDA and embedding ρ values across all segments.

TABLE VI
LDA VS. EMBEDDING ALIGNMENT SCORE COMPARISON

Segment	LDA ρ	Embedding ρ
CS Caribbean	0.0256	0.3061 (30.6%)
CS International	0.0483	0.5585 (55.8%)
CS AI Jobs	0.0436	0.6725 (67.2%)
IT Caribbean	0.0225	0.2767 (27.7%)
IT International	0.0346	0.5535 (55.3%)

V. DISCUSSION

A. Vocabulary Mismatch and Conceptual Alignment

The central finding of this study is that the 11–16 $\times$ gap between LDA and embedding scores across all segments quantifies vocabulary register divergence as the dominant suppressor of apparent alignment. The LDA $\rho_w = 0.0412$ represents a lower bound that measures only token-level vocabulary overlap: course outlines use multi-word academic terminology such as *object-oriented programming*, *distributed systems* and *finite state machine*, while job postings use concise industry shorthand such as *docker*, *agile*, *react* and *sql*. This divergence suppresses LDA scores regardless of conceptual overlap, and is itself a meaningful finding—it indicates the extent to which graduates must translate academic knowledge into industry-relevant language during recruitment.

The embedding scores of 0.28–0.67 capture conceptual proximity between curriculum content and job market vocabulary, and should be interpreted as an upper estimate of conceptual alignment rather than a direct measure of practical proficiency. High embedding similarity between a course’s

distributed systems content and a job posting requiring *Kubernetes* experience indicates conceptual proximity, not that the course develops hands-on Kubernetes competency. Validating these text-based findings against employer surveys, graduate outcomes or competency assessments is an important direction for future work. Together, the two measures provide a more complete diagnostic than either alone: LDA quantifies the vocabulary gap students must bridge, while embedding estimates the conceptual overlap that this gap obscures. This reversal is most pronounced for AI job descriptions, where the CS curriculum’s strong conceptual AI coverage ($\rho = 0.6725$ embedding) contrasts with its weakest LDA score ($\rho = 0.0436$), indicating that graduates may have strong conceptual AI preparation without exposure to the tooling vocabulary AI employers expect.

B. CS vs IT Alignment

The Computer Science program demonstrated slightly stronger semantic alignment than Information Technology across most segments, possibly due to CS disciplines use more globally standardized terminology that transfers more directly between academia and industry, whereas IT roles frequently emphasize vendor-specific technologies not explicitly represented in course outlines. Both programs nonetheless demonstrated moderate semantic alignment overall. It should be noted that the CS, IT and AI categories aggregate diverse roles (e.g., software engineering and cybersecurity within CS; network administration and IT support within IT), and the programme-level scores reported here represent an average across this heterogeneity.

C. Caribbean vs International Markets

A notable finding was the stronger semantic alignment observed for international job postings compared to Caribbean postings in both programs (CS: 0.56 vs 0.31; IT: 0.55 vs 0.28), likely because international postings are more detailed and overlap more strongly with academic content. However, with only 13 Caribbean postings per thematic area compared to over 1,300 international postings, the Caribbean-specific results lack statistical reliability and should be treated as exploratory instead of definitive. Expanding Caribbean job posting data is a priority for future work.

D. Academic Rigor vs Vocational Preparation

The identified curriculum topics show that the DCIT programs maintain strong emphasis on theoretical foundations—*formal methods and automata*, *discrete mathematics* and *computer architecture* rarely appear in job postings but contribute to broader conceptual alignment per the embedding analysis. Conversely, contemporary technologies frequently referenced in job postings—including *Docker*, *React*, *Jira*, *Kubernetes*, *AWS* and *GitHub*—were minimally represented in curriculum vocabulary. The appropriate response is not to reduce theoretical content but to supplement it with explicit practical tooling.

E. Limitations

Several limitations must be acknowledged. The Caribbean data imbalance (13 vs. 1,300+ postings per thematic area) is the most significant, limiting the reliability of regional conclusions. The study relies entirely on text similarity as a proxy for alignment, without external validation from employer interviews, graduate outcomes, or competency assessments—text proximity measures what is described in course outlines, not what students actually learn or what employers observe. The broad CS/IT/AI categorization aggregates roles with different skill profiles, potentially masking within-category variation. SkillNER may under-represent soft skills [6] such as communication and teamwork, which are central to employability however difficult to extract automatically. Job postings were not stratified by experience or education level, and the pre-trained embedding model was not fine-tuned on Caribbean or academic computing corpora. Finally, while the dual-method design combining LDA with semantic embedding to explicitly quantify vocabulary mismatch is a novel contribution, the individual components draw on established NLP techniques.

VI. CONCLUSIONS AND FUTURE WORK

This study assessed how well the DCIT Computer Science and Information Technology undergraduate curricula at UWI align with computing job requirements in Caribbean and international markets using a dual-method NLP framework. The key contribution is the joint use of LDA-based token-level alignment and semantic embedding conceptual alignment to explicitly distinguish vocabulary register divergence from conceptual alignment.

The LDA results ($\rho_w = 0.0412$) reveal weak token-level vocabulary overlap, reflecting terminological divergence between academic and industry language rather than an absence of relevant content. The embedding results ($\rho = 0.28\text{--}0.67$) suggest moderate conceptual alignment, with CS outperforming IT and international postings outperforming Caribbean postings. The largest method-level gap was for AI job descriptions, where the CS curriculum shows strong conceptual coverage of AI foundations but minimal token-level overlap with industry AI terminology.

Both programs maintain strong theoretical foundations and would benefit from supplementary practical content—particularly contemporary tooling—without reducing academic rigor, and from closer alignment with ACM/IEEE CS2023 guidelines. Future work should: (1) substantially expand Caribbean job posting data; (2) stratify postings by experience and education level; (3) validate text-similarity findings against employer surveys and graduate outcomes; and (4) extend the analysis to finer-grained role categories to reduce within-category heterogeneity.

REFERENCES

- [1] V. Ramnarine, N. Chamansingh, and P. Hosein, “Using Natural Language Processing to Correlate University Curricula with Required Job Skills,” *Proc. Int. Conf. Artif. Intell., Comput., Data Sci. Appl. (ACDSA)*, 2025. [Online]. Available: <https://lab.tt/wp-content/uploads/2025/08/acdsa2025-ramnarine.pdf>
- [2] D. Gope and A. Gope, “Students and academicians views on the engineering curriculum and industrial skills requirement for a successful job career,” *Open Education Studies*, vol. 4, no. 1, pp. 173–186, Jan. 2022, doi: <https://doi.org/10.1515/edu-2022-0011>.
- [3] G. Leandro, G. Fernandes, and C. Ferreira, “Complementary Training Program in Electrical Engineering and Computer Engineering Undergraduate Courses,” *IEEE Trans. Educ.*, vol. 65, no. 4, pp. 670–683, Nov. 2022, doi: <https://doi.org/10.1109/te.2022.3161866>.
- [4] N. R. Aljohani, A. Aslam, A. O. Khadidos, and S.-U. Hassan, “Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature,” *J. Innov. Knowl.*, vol. 7, no. 3, p. 100190, Jul. 2022, doi: <https://doi.org/10.1016/j.jik.2022.100190>.
- [5] F. Gurcan and N. E. Cagiltay, “Big Data Software Engineering: Analysis of Knowledge Domains and Skill Sets Using LDA-Based Topic Modeling,” *IEEE Access*, vol. 7, pp. 82541–82552, 2019, doi: <https://doi.org/10.1109/access.2019.2924075>.
- [6] R. Walker, “Mapping curricula to skills and occupations using course descriptions,” in *Proc. IEEE World Eng. Educ. Conf. (EDUNINE)*, Mar. 2024, doi: <https://doi.org/10.1109/edunine60625.2024.10500452>.
- [7] A. Radermacher and G. Walia, “Gaps between industry expectations and the abilities of graduates,” in *Proc. 44th ACM Tech. Symp. Comput. Sci. Educ. (SIGCSE)*, pp. 525–530, 2013, doi: <https://doi.org/10.1145/2445196.2445351>.
- [8] N. Chamansingh and P. Hosein, “Using topic modelling to correlate a research institution’s outputs with its goals,” in *Adv. Intell. Syst. Comput.*, pp. 147–156, Jan. 2020, doi: https://doi.org/10.1007/978-3-030-39442-4_13.
- [9] S. Fareri, N. Melluso, F. Chiarello, and G. Fantoni, “SkillNER: Mining and mapping soft skills from any text,” *Expert Syst. Appl.*, vol. 184, p. 115544, Dec. 2021, doi: <https://doi.org/10.1016/j.eswa.2021.115544>.
- [10] D. Maier et al., “Applying LDA topic modeling in communication research: Toward a valid and reliable methodology,” *Commun. Methods Meas.*, vol. 12, pp. 93–118, Feb. 2018, doi: <https://doi.org/10.1080/19312458.2018.1430754>.
- [11] S. Sahoo, J. Mati, and K. Tewari, “Multivariate Gaussian Topic Modelling: A novel approach to discover topics with greater semantic coherence,” Mar. 2025.
- [12] L. M. Chandrapati and C. K. Rao, “Descriptive answers evaluation using natural language processing approaches,” *IEEE Access*, vol. 12, pp. 1–1, Jan. 2024, doi: <https://doi.org/10.1109/access.2024.3417706>.
- [13] C. P. Chai, “Comparison of text preprocessing methods,” *Nat. Lang. Eng.*, vol. 29, pp. 1–45, Jun. 2022, doi: <https://doi.org/10.1017/s1351324922000213>.
- [14] G. C. Banks et al., “A review of best practice recommendations for text analysis in R (and a user-friendly app),” *J. Bus. Psychol.*, vol. 33, pp. 445–459, Jan. 2018, doi: <https://doi.org/10.1007/s10869-017-9528-3>.
- [15] A. Omran and C. Treude, “Choosing an NLP library for analyzing software documentation: A systematic literature review and a series of experiments,” in *Proc. IEEE/ACM 14th Int. Conf. Mining Softw. Repositories (MSR)*, 2017, pp. 187–197.
- [16] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in *Proc. 2019 Conf. Empirical Methods Nat. Lang. Process. (EMNLP)*, Nov. 2019, doi: <https://doi.org/10.18653/v1/D19-1410>.
- [17] Y. Shi, Y. Yao, H. Xu, A. Wu, X. Li, and D. Shen, “SALIEN: An attentive learning-based approach for skill extraction from job postings,” in *Proc. 2020 IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2020, doi: <https://doi.org/10.1109/ICDMW51313.2020.00071>.
- [18] Y. Januzaj, A. Luma, and B. Raufi, “Alignment of study programs with job market requirements using text mining techniques,” *Int. J. Interact. Mob. Technol. (ijIM)*, vol. 18, no. 3, 2024, doi: <https://doi.org/10.3991/ijim.v18i03.46015>.