Enigma in Education
https://enigma.or.id/index.php/edu

e-ISSN: 3026-5827

ORIGINAL ARTICLE
Enigma in Education | Vol. 4, No. 2 (2026) | pp. 50–68

Strategic Orientation Profiles and Lexical Co-Occurrence Networks Across 51 Indonesian Higher Education Strategic Plans Covering 2025: An Exploratory Cluster Stability Analysis

Muhammad Hasan[1,*], Grace Olivia Silalahi[2]

[1]*Department of Social Sciences, Enigma Institute, Palembang, Indonesia*
[2]*Department of Management, CMHC Research Center, Palembang, Indonesia*

ARTICLE INFO

Article history
Received: August 19, 2026
Revised: September 20, 2026
Accepted: October 15, 2026
Available online: October 28, 2026
Published: November 1, 2026

Keywords
cluster stability; digital transformation; higher education governance; internationalization; strategic planning

Corresponding author
Muhammad Hasan
muhammad.hasan@enigma.or.id

ORCID
0009-0008-5296-1299

DOI
https://doi.org/10.61996/edu.v4i2.135

ABSTRACT

Background. Higher education strategic plans (Renstra) tie institutional visions to operational targets. Yet document length, format, and mission differ across institutions. That makes it hard to read lexical frequencies as discrete governance typologies.

Objective. To measure lexical attention to four strategic orientations (internationalization, research downstreaming, inclusivity, and digital transformation). It also tests whether Indonesian higher education institutions form stable, well-separated clusters.

Methods. This exploratory computational text-as-data analysis covered 51 official Renstra documents for 2025 (39 public [PTN] and 12 private [PTS] institutions in six regions) and yielded 13,275 non-overlapping 80-token windows (923,823 alphabetic words). A pre-specified 24-concept dictionary was combined with Normalized Pointwise Mutual Information (NPMI) co-occurrence networks, Ward clustering, 1,000 marginal feature-permutation null models, 1,000 subsampling runs, and 14 sensitivity analyses.

Results. Equal-weight institutional mean positive window proportions were 28.0% (95% bootstrap CI: 25.4–30.8%) for internationalization, 12.1% (10.8–13.5%) for digital transformation, 11.5% (10.0–13.1%) for research downstreaming, and 6.8% (5.8–8.0%) for inclusivity. The co-occurrence network had 24 nodes, 138 edges, and four communities. The best four-cluster partition (n = 14, 13, 14, and 10; silhouette = 0.166) did not beat the marginal permutation null (median = 0.201; empirical p = 0.961). Subsampling gave a median Adjusted Rand Index (ARI) of 0.389, selected k = 4 in 50.8% of runs, and left 58.6% of institution pairs ambiguous. Sensitivity ARIs ranged from 0.091 to 0.699.

Conclusion. Across 51 strategic plans, lexical attention to the four orientations varies considerably. The data do not justify discrete institutional typologies. Continuous profile reporting is preferable.

Copyright: © 2026 The Author(s). Published by Enigma Institute on behalf of Enigma in Education. The authors retain copyright of their work.

Access license: This is an open access article distributed under the terms of the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and any derivative work is distributed under the same license.

1. Introduction

1.1 Strategic Planning Documents as Instruments of Higher Education Governance

Institutional strategic plans, formally codified across Indonesian universities, institutes, and polytechnics as Rencana Strategis (Renstra), are the main governance link between academic visions and multi-year operational objectives, and they also connect to key performance indicators (Indikator Kinerja Utama), resource allocation priorities, and internal quality assurance systems.[1–4] Such plans do two jobs. They coordinate work inside the institution. They also signal legitimacy to the outside by showing how it answers ministerial regulatory mandates, accreditation standards, labor market change, and global competitiveness benchmarks.[5–9] Renstra documents are promulgated publicly by university senates, boards of trustees (Majelis Wali Amanat), or foundation boards, which makes them a rich empirical archive for comparative educational policy analysis.[2,3,10,11] Turning planning prose into valid governance indicators is hard, though. Plans differ in structure, page length, planning horizon, administrative template, and institutional mission, and they range from comprehensive research-intensive universities and state Islamic universities to vocational polytechnics and regional private colleges.[4,12–15] Over-interpretation is the main danger. Raw keyword frequencies or heuristic cluster assignments are not direct measures of policy execution, budgetary commitment, or governance quality.[2,12,16]

The text-as-data approach in computational social science and educational policy analytics treats automated content analysis as a measurement instrument. It does not replace close reading or checks against institutional context.[16,17] Applied to higher education strategic plans, lexical detection captures textual salience, and that salience mixes aspirational goal statements, summaries of past achievement, SWOT (strengths, weaknesses, opportunities, and threats) statements of deficit, and mandatory regulatory references.[2,16] Several safeguards are therefore needed. Documents must be screened for eligibility, denominators normalized, statutory reference text filtered, and stability checked by resampling. Without them, computational classifications of strategy documents can easily confuse

document length or template compliance with real institutional differences.[18–20]

Neo-institutional organization theory places university strategic plans in a dense field of coercive, normative, and mimetic pressures, which encourage thematic convergence and local operational adaptation at once.[2,21–24] In Indonesia, coercive isomorphism works through statutory regulation, ministerial strategic targets, and the national Key Performance Indicators (Indikator Kinerja Utama or IKU Perguruan Tinggi), which are tied to performance-based block grants and to accreditation instruments run by the National Accreditation Board for Higher Education (BAN-PT).[1,9,25] Normative isomorphism has another source. International ranking methodologies (such as QS World University Rankings and Times Higher Education Impact Rankings) and professional academic associations value globally indexed publications, patent portfolios, digital campus ecosystems, and inclusive student support services.[9,10,26–31] Mimetic isomorphism is different. Emerging public and private universities copy the five-year planning templates of established autonomous legal-entity state universities (Perguruan Tinggi Negeri Badan Hukum or PTN-BH).[24,25] Comparative studies of university strategic plans in the United States, Denmark, South Korea, and the Philippines show the result. Shared regulatory and competitive environments push institutions toward a common vocabulary of modernization that covers globalization, digital transformation, industry commercialization, and social equity, while the textual space given to each priority varies continuously.[2,11,12,24]

1.2 Four Core Modernization Agendas in Indonesian Higher Education Covering 2025

Four modernization agendas dominate the Indonesian higher education landscape for the 2025 planning horizon: internationalization, research downstreaming (hilirisasi riset), inclusivity, and digital transformation. They feature in national development frameworks, including the 2025–2045 National Long-Term Development Plan, and in institutional governance debates.[1,9,26–32] Internationalization covers global academic orientation, foreign institutional partnerships, international student and faculty mobility, recruitment of inbound foreign students, collaborative dual or joint degree programs, and internationally indexed scholarly publications.[9,10,29–31] Research downstreaming is a flagship national priority. It links academic discovery to industrial value creation through technology transfer, commercialization, intellectual property protection (patents and copyrights), business incubation and academic startups, and research–industry collaboration.[8,13,33]

Inclusivity reflects the constitutional and ethical mandate to widen equitable access to higher education across Indonesia's diverse socio-economic and geographic strata. It includes scholarship and tuition support (such as Kartu Indonesia Pintar Kuliah), affirmative admission pathways for vulnerable and frontier/outermost (3T) communities, disability accommodation and accessible campus services, gender equality and non-discrimination, and explicit inclusive educational framing.[34–40] Digital transformation is the modernization of academic and administrative operations. Its elements are enterprise information systems, campus information and communication technology (ICT) infrastructure, digitalization workflows, blended and online learning ecosystems, institutional data governance, and cybersecurity or information security stewardship.[26–28,41,42] Understanding how institutions balance these four agendas takes two kinds of evidence. One is lexical prevalence for each agenda. The other is concept-level co-occurrence structure within planning texts.[17,43,44]

1.3 Comparative Institutional Scope, Research Questions, and Analytical Objectives

A basic question in comparative institutional research is whether multidimensional profiles of strategic attention gather into discrete, reproducible typologies or form a continuous, overlapping spectrum that shifts when the comparator sample or the text-processing rules change.[18–20,45] Answering it needs a diverse corpus. That means going beyond flagship public universities on Java to cover several governance sectors, institutional forms, and island regions. The 51 institutions examined here comprise 39 public higher education institutions (Perguruan Tinggi Negeri or PTN) and 12 private institutions (Perguruan Tinggi Swasta or PTS), made up of 45 universities, five polytechnics, and one technological institute in six regional groupings. They allow a rigorous empirical test of whether lexical strategic orientations sort institutions into stable clusters or vary along continuous dimensions under standardized dictionary and segmentation rules.[16,18,19,43]

This exploratory computational documentary study asks three questions. (1) What are the cross-institutional lexical prevalence and the concept co-occurrence network structure of internationalization, research downstreaming, inclusivity, and digital transformation across 51 Indonesian higher education Renstra documents covering 2025? (2) Do combinations of these four orientations form discrete, well-separated, resampling-stable clusters of institutions when tested against marginal feature-permutation null models and 80% institutional subsampling? (3) How far do cluster memberships and thematic profiles depend on sample composition (the full 51-institution corpus, universities only, PTN only, and non-Java institutions), document length, statutory reference text filtering, dictionary specificity, and segmentation window size? A transparent computational workflow answers them. In doing so, the study sets an empirical benchmark for responsible text-as-data governance analysis in higher education.

2. Methods

2.1 Comparative Documentary Design, Purposive Acquisition Protocol, and Eligibility Criteria

This study used a comparative computational documentary design built on purposive web acquisition of publicly accessible higher education strategic plans.[2,16,41] Search queries combined institutional identifiers, institutional form (university, institute, polytechnic, state Islamic university), and planning terms (Renstra, Rencana Strategis), and targeted multi-year planning horizons that include 2025. The search was designed to reach beyond Java-based public universities. It also covered private universities (PTS), state Islamic higher education institutions (PTKIN/UIN), polytechnics, and technological institutes on Java and in the outer islands. Document availability depends on institutional web transparency and on how well public repositories are maintained. The corpus is therefore a purposive comparative sample, not a probabilistic national census or a systematic review of all Indonesian higher education institutions.[2,16]

Five inclusion criteria were applied to keep the documents comparable: (i) the document is an institution-level strategic plan (Renstra tingkat institusi); (ii) its stated multi-year planning period explicitly covers 2025; (iii) the file comes from an official institutional domain or from an official public repository linked on the institution's portal; (iv) institutional identity and planning period can be verified from the document contents; and (v) the main substantive chapters contain enough extractable text for window segmentation. Faculty-level, study-program-level, and research-unit plans were excluded, as were annual performance accountability reports (LAKIP), brief web summaries, and standalone business strategic plans (Rencana Strategis Bisnis), because they are not equivalent documents. Files on official portals or linked public repositories were retrieved without authentication. Password-protected or restricted intranet files were never bypassed. Legal adoption status (such as senate or board ratification) was recorded when the document stated it, but online availability alone was not taken as verification.

Across 83 documented retrieval records covering 78 unique official URLs, 55 unique strategic planning documents were acquired across 54 distinct institutions (including two versions for Universitas Negeri Malang, which were consolidated to one authoritative version per institution). Among the 54 candidate institutions, Universitas Lambung Mangkurat (ULM) was excluded because content verification confirmed a 2020–2024 revision period ending prior to 2025. Two additional institutions were separated from the primary corpus and reserved exclusively for sensitivity analysis due to internal multi-period phrasing or unconfirmed final version status: Universitas Mulawarman (UNMUL, predominantly covering 2025–2029 but containing internal 2024–2029 phrasing) and UIN Raden Fatah Palembang (whose introductory sections state 2020–2024 whereas the closing chapter states 2025–2029). This multistage screening yielded the primary analytical corpus of 51 institutions.

Analysis of official plans from 54 candidate institutions showed recurring structural and temporal complications that comparative higher education research has to handle explicitly.[2,3,16] Indonesian institutions ratify five-year Renstra cycles on their own rectorate or senate calendars, not on one synchronized national date. Active plans covering 2025 therefore span staggered multi-year windows, including 2021–2025, 2022–2026, 2022–2027, 2023–2026, 2023–2027, 2023–2028, 2024–2027, 2024–2028, 2024–2029, and 2025–2029, together with mid-cycle regulatory amendments (such as in Universitas Riau, Universitas Negeri Malang, and Universitas Sam Ratulangi). Document design also differs across governance sectors. Some institutions publish compact narrative chapters on strategic goals and key performance indicators. Others embed target matrices that run to hundreds of pages, two-page horizontal spread layouts (such as Universitas Hasanuddin), or extensive technical glossaries (such as Politeknik Negeri Jakarta). Uniform eligibility verification, single-version deduplication, and delimitation of substantive chapters help ensure that analytical windows capture comparable planning discourse before any statistical or clustering model is estimated.[2,16] Institutional names are reported strictly to document public strategic planning profiles and are not used for commercial ranking, accreditation adjudication, or funding allocation.

2.2 Text Extraction, Substantive Chapter Delimitation, and Window Segmentation

Text was extracted with coordinate-aware bounding-box layout parsing, which preserves multi-column reading order and block structure (including two-page horizontal spread layouts such as Universitas Hasanuddin). Before tokenization, the boundaries of the main substantive strategic planning chapters were audited and delimited manually for each of the 51 documents. Front covers, tables of contents, standalone ratification decrees, external glossaries (such as in Politeknik Negeri Jakarta), and post-closing appendices were left out of the analytical text range. Target matrices placed inside main substantive chapters (such as in Universitas Syiah Kuala and Universitas Muhammadiyah Makassar) were kept as part of the planning text, although tabular formatting can affect token proximity. Where cover pages were raster images (such as Universitas Indraprasta PGRI), visual inspection confirmed the 2025–2029 planning horizon, while lexical extraction used only the digital main-chapter text.

Standardized page-cleaning and windowing parameters were applied uniformly across all 51 documents. Lines falling within the top 4.5% and bottom 5.0% of each page's vertical bounding coordinates were stripped to remove running headers and footers, alongside short repetitive header strings, dotted table-of-contents lines, and numeric-dominant table rows. Cleaned page text was segmented into non-overlapping, within-page windows of 80 space-delimited tokens; windows never crossed page boundaries, and the trailing window on any page was retained only if it contained at least 20 alphabetic tokens. Because 80-token windows represent standardized analytical blocks rather than natural syntactic paragraphs, sensitivity models explicitly tested 40-token windows, 120-token windows, and approximate sentence units.

2.3 Pre-Specified 24-Concept Lexical Dictionary and Theme Prevalence Measurement

Lexical measurement relied on a pre-specified operational dictionary of 24 concepts organized symmetrically into four strategic themes (six concepts per theme).[16,43] Internationalization comprised: (1) international orientation (orientasi internasional), (2) foreign partnerships (kemitraan luar negeri), (3) international mobility (mobilitas internasional), (4) foreign students (mahasiswa asing), (5) collaborative degrees (gelar kolaboratif), and (6) international publication (publikasi internasional). Research downstreaming comprised: (7) downstreaming (hilirisasi), (8) commercialization (komersialisasi), (9) technology transfer (transfer teknologi), (10) intellectual property (kekayaan intelektual), (11) incubation and startups (inkubasi dan usaha rintisan), and (12) research and industry linkage (riset dan industri). Inclusivity comprised: (13) inclusive orientation (orientasi inklusif), (14) disability (disabilitas), (15) equality and non-discrimination (kesetaraan dan nondiskriminasi), (16) affirmation and vulnerable groups (afirmasi dan kelompok rentan), (17) educational cost support (dukungan biaya pendidikan), and (18) equitable educational access (pemerataan akses pendidikan). Digital transformation comprised: (19) digitalization (digitalisasi), (20) information systems (sistem informasi), (21) online learning (pembelajaran daring), (22) data governance (tata kelola data), (23) information security (keamanan informasi), and (24) ICT infrastructure (infrastruktur TIK).

Dictionary matching was case-insensitive. Within each 80-token window, each of the 24 concepts was coded as a binary indicator (0 or 1), and a strategic theme was scored positive (1) if at least one of its six constituent concepts appeared in that window. Because a single window can mention multiple concepts across one or more themes, theme indicators are non-mutually exclusive (overlapping). Windows detected by heuristic rules as containing statutory or regulatory reference text were excluded in the primary analysis (leaving 13,275 analytical windows from 13,761 total extracted windows) and re-included in sensitivity testing. For each institution $i \in \{1, ..., 51\}$ and theme $t \in \{1, ..., 4\}$, the institutional theme proportion $P_{i,t}$ was computed as the number of positive windows divided by the institution's total analytical windows, multiplied by 100. Across the corpus, the primary summary estimator was the equal-weight institutional mean $\bar{P}_t = (1/51) \sum_{i=1}^{51} P_{i,t}$, accompanied by pooled window proportions (denominator = 13,275) and 95% percentile confidence intervals (CIs) derived from 5,000 non-parametric bootstrap resamples of the 51 institutions. Because lexical pattern matching captures aspirational targets, historical achievements, and diagnostic deficit statements alike without semantic disambiguation, automated lexical proportions measure documentary topic salience rather than verified policy implementation.

2.4 Lexical Co-Occurrence Network Construction and Centrality Metrics

An undirected weighted lexical co-occurrence network was built to map how strategic concepts co-occur in planning prose. The 24 operational concepts served as candidate nodes, and within-window co-occurrence defined candidate edges.[17,43,44] To avoid spurious associations from rare terms, a concept node had to appear in at least 10 analytical windows across at least three institutions. A candidate edge between concepts u and v was kept only if both co-occurred in at least three analytical windows across at least three institutions and had a strictly positive Normalized Pointwise Mutual Information (NPMI > 0).[43] Let p(u), p(v), and p(u,v) be the empirical window probabilities across the 13,275 analytical windows. Pointwise Mutual

Information is PMI(u,v) = ln[p(u,v) / (p(u)p(v))], and Bouma's normalization rescales it to [-1, +1] as NPMI(u,v) = PMI(u,v) / [−ln p(u,v)].[43]

Alongside NPMI, pairwise Jaccard similarity was computed as the ratio of co-occurring windows to the union of windows containing either concept. Node position within the 24-node, 138-edge graph was characterized by degree (number of incident supported edges), node strength (sum of incident NPMI edge weights), and shortest-path betweenness centrality, where edge distance was defined as d(u,v) = 1 / NPMI(u,v). Concept communities were identified using Louvain modularity optimization. Network links represent lexical proximity within 80-token windows and do not imply causal policy integration or inter-institutional collaboration.

2.5 Ward Hierarchical Clustering, Resampling Stability, Null Permutations, and Sensitivity Suite

To test whether Indonesian higher education institutions fall into discrete strategic orientation types, the four institutional theme proportions $P_{i,t}$ were transformed as $\log(1 + P_{i,t})$ (log1p) and z-score standardized across institutions. Agglomerative hierarchical clustering with Ward's minimum-variance linkage followed.[18,19,45] The primary analysis used a pre-specified admissibility rule. Candidate cluster counts $k \in \{2, 3, 4\}$ were evaluated, every cluster had to contain at least three institutions (min size ≥ 3), and the partition with the highest mean silhouette coefficient was selected.[18,45] As an extended diagnostic, candidate counts $k \in \{2, ..., 8\}$ with at least five institutions per cluster (min size ≥ 5) were also evaluated.

Internal cluster stability and over-claiming safeguards were implemented through four complementary protocols.[19,20] First, against a marginal feature-permutation null model, each of the four theme features was permuted independently across the 51 institutions 1,000 times—preserving exact univariate marginal distributions while destroying cross-theme covariance—and the optimal $k \in \{2, 3, 4\}$ (with min size ≥ 3) was re-selected on each null draw, computing the empirical p-value with the standard $(r + 1)/(B + 1)$ plus-one correction. Second, leave-one-institution-out (LOIO) jackknife analysis removed each institution in turn (n = 50), re-standardized the features, re-estimated the four-cluster Ward model, and computed the Adjusted Rand Index (ARI) against the full-sample partition on the 50 shared institutions. Third, 1,000 subsampling replications drew 80% of institutions without replacement ($n = \lceil 0.80 \times 51 \rceil = 41$), re-standardized and re-clustered the subsample, and recorded: (a) the ARI against the primary partition at k = 4, (b) the maximum Jaccard cluster recovery similarity for each primary cluster, (c) the frequency with which $k \in \{2, 3, 4\}$ was re-selected, and (d) the pairwise co-clustering consensus probability across all C(51, 2) = 1,275 institutional pairs, flagging the proportion of ambiguous pairs with consensus probability strictly between 0.1 and 0.9.[19]

Fourth, 14 deterministic sensitivity models (with k held fixed at 4) audited partition robustness across: (S1) retaining statutory and regulatory reference text (13,761 windows); (S2) a narrower dictionary removing generic 'internasional'; (S3) term frequency per 10,000 alphabetic words; (S4) excluding the longest document, Unismuh Makassar (n = 50; 11,976 windows); (S5) excluding the amended UNRI document (n = 50); (S6) adding the candidate UNMUL document (n = 52); (S7) adding the candidate UIN Raden Fatah document (n = 52); (S8) universities only (n = 45); (S9) public PTN only (n = 39); (S10) non-Java institutions only (n = 26); (S11) excluding the amended UNRI, UM, and UNSRAT documents simultaneously (n = 48); (S12) approximate sentence units (38,336 units); (S13) 40-token windows (24,419 windows); and (S14) 120-token windows (9,470 windows). In addition, exploratory Mann–Whitney U subgroup tests (PTN vs. PTS; Java vs. Non-Java) and Spearman rank correlations (ρ) with document length were conducted to examine structural correlates of thematic salience across the 51 institutions; because 12 exploratory tests were run, Benjamini–Hochberg adjusted q values are reported with the results. All statistical, network, and clustering computations were executed in Python 3.12.14 (pandas 2.2.3, NumPy 2.5.2, SciPy 1.18.0, scikit-learn 1.9.0, NetworkX 3.6.1).

The mathematical estimators governing theme prevalence, network centrality, cluster separation, null permutation testing, and resampling stability were formalized as follows. Let W_i denote the number of post-filtered 80-token analytical windows in institution $i \in \{1, ..., 51\}$, with total corpus windows $W_{total} = \sum_{i=1}^{51} W_i = 13{,}275$, and let $Y_{i,t} \in \{0, ..., W_i\}$ denote the count of windows in institution i where strategic theme $t \in \{1, 2, 3, 4\}$ is positive. The institutional theme percentage is $P_{i,t} = 100 \times (Y_{i,t} / W_i)$. The primary equal-weight institutional mean estimator is $\bar{P}_t = (1/51) \sum_{i=1}^{51} P_{i,t}$, which assigns equal analytical weight (1/51) to each institution regardless of document length W_i, whereas the pooled window estimator $\tilde{P}_t = 100 \times (\sum_{i=1}^{51} Y_{i,t}) / W_{total}$ weights institutions proportionally to W_i. Prior to Ward clustering, right-skewness in $P_{i,t}$ was stabilized via the natural logarithmic transformation $X_{i,t} = \ln(1 + P_{i,t})$, followed by column-wise z-score standardization $Z_{i,t} = (X_{i,t} - \mu_t) / \sigma_t$. In the marginal feature-permutation null model (B = 1,000 replications), each column t of the standardized matrix Z was permuted independently via a random bijection $\pi_t(b)$ across the 51 institutions, preserving exact marginal distributions while setting population cross-theme correlations to zero; on each permuted matrix Z*(b), Ward clustering was executed across $k \in \{2, 3, 4\}$ (subject to $\min_c |C_c| \geq 3$) to obtain the maximum admissible silhouette S*(b), and the empirical p-value was computed with the plus-one correction as $p_{emp} = [1 + \sum_{b=1}^{1,000} I(S^*(b) \geq S_{obs})] / (1{,}000 + 1)$. Across the 1,000 subsampling runs drawing n = 41 institutions without replacement, pairwise co-clustering consensus for any pair (i, j) was defined strictly over the subset of replications $B_{i,j}$ in which both institutions were jointly sampled: $C_{i,j} = [\sum_{b \in B(i,j)} I(c_i(b) = c_j(b))] / |B_{i,j}|$, classifying pairs with $0.1 < C_{i,j} < 0.9$ as ambiguous boundary pairs.[18,19,45]

3. Results

3.1 Corpus Acquisition Attrition, Institutional Stratification, and Text Window Distribution

Across 83 documented retrieval records spanning 78 unique URLs, 55 unique strategic planning documents were obtained for 54 higher education institutions. Following single-version deduplication (selecting one authoritative institutional Renstra for Universitas Negeri Malang), 54 candidate documents underwent eligibility verification: one institution (Universitas Lambung Mangkurat) was excluded because its revised Renstra covered 2020–2024, and two institutions (Universitas Mulawarman and UIN Raden Fatah Palembang) were assigned exclusively to sensitivity analysis due to internal multi-period phrasing, establishing a primary analytical corpus of 51 higher education institutions (94.4% of candidate institutions; Table 1 and Figure 1A).

Within the 51-institution primary corpus, public higher education institutions (PTN) accounted for 39/51 (76.5%) and private institutions (PTS) accounted for 12/51 (23.5%; Table 1 and Figure 1B). By institutional form, the corpus comprised 45 universities (88.2%), five polytechnics (9.8%; PNJ, POLBAN, Poltekkes Banjarmasin, Poltekpar Bali, and POLTERA), and one technological institute (2.0%; Institut Teknologi Kalimantan [ITK]). Geographically, main campuses were distributed across six regional groupings: Java represented nearly half of the corpus at 25/51 institutions (49.0%), followed by Sumatra with 10/51 (19.6%), Sulawesi with 7/51 (13.7%), Bali and Nusa Tenggara with 5/51 (9.8%), Kalimantan with 2/51 (3.9%), and Maluku with 2/51 (3.9%), while Papua was unrepresented (0/51; 0.0%; Table 1 and Figure 1C).

Substantive chapter extraction and statutory reference text filtering across the 51 Renstra documents yielded 13,275 non-

overlapping 80-token analytical windows containing 923,823 alphabetic words (from 13,761 pre-filtered windows; Table 1). Across the 51 institutions, analytical document length averaged 260.3 ± 194.0 (standard deviation, SD) windows per institution, with a median of 225 windows (interquartile range [IQR]: 140–312 windows) and a wide empirical span ranging from 54 windows (Universitas Islam Ibrahimy Banyuwangi [UNIIB]) and 66 windows (Universitas Bung Karno [UBK]) to 711 windows (UIN Alauddin Makassar) and 1,299 windows (Universitas Muhammadiyah Makassar). Public institutions contributed 10,442 analytical windows across 39 PTN documents, while private institutions contributed 2,833 analytical windows across 12 PTS documents, reflecting substantial variation in narrative detail and operational matrix formatting across Indonesian higher education strategic plans.

Table 1. Multi-stage sample selection and institutional stratification of the 51 Indonesian higher education strategic plans (Renstra) covering the 2025 planning horizon.

Analytical Panel	Screening Stage / Institutional Stratum	Count (n or N)	Proportion (%)	Analytical Scope and Institutional Classification
Panel A: Sample Selection	Initial document retrieval records screened	83	—	Purposive search across 78 unique official institutional web portals
Panel A: Sample Selection	Unique strategic planning documents identified	55	100.0% of files	Representing 54 higher education institutions prior to single-document consolidation
Panel A: Sample Selection	Candidate institutional documents evaluated for eligibility	54	100.0% of candidates	One authoritative institutional-level strategic plan retained per institution
Panel A: Sample Selection	Excluded due to planning horizon ending prior to 2025	−1	−1.9%	Strategic plan covering the 2020–2024 period outside the 2025 window
Panel A: Sample Selection	Reserved exclusively for sensitivity analysis	−2	−3.7%	Documents exhibiting internal multi-period phrasing across sections
Panel A: Sample Selection	Primary analytical corpus (eligible Renstra covering 2025)	51	94.4% of candidates	13,275 non-overlapping 80-token windows and 923,823 alphabetic words
Panel B: Governance Status	Public higher education institutions (Perguruan Tinggi Negeri / PTN)	39	76.5%	State universities, state Islamic universities (UIN), state polytechnics, and state institute
Panel B: Governance Status	Private higher education institutions (Perguruan Tinggi Swasta / PTS)	12	23.5%	Private universities governed by educational foundations across Indonesia
Panel C: Institutional Form	Universities (Universitas)	45	88.2%	Comprehensive academic institutions offering multi-faculty degree programs
Panel C: Institutional Form	Polytechnics (Politeknik)	5	9.8%	Applied vocational higher education institutions (PNJ, POLBAN, Poltekkes Banjarmasin, Poltekpar Bali, POLTERA)
Panel C: Institutional Form	Technological Institute (Institut)	1	2.0%	Specialized science and engineering institute (Institut Teknologi Kalimantan)
Panel D: Regional Group	Java (Jawa)	25	49.0%	Main campus located in Banten, DKI Jakarta, West Java, Central Java, DI Yogyakarta, or East Java
Panel D: Regional Group	Sumatra (Sumatera)	10	19.6%	Main campus located in Aceh, North Sumatra, West Sumatra, Riau, Riau Islands, or Lampung
Panel D: Regional Group	Sulawesi	7	13.7%	Main campus located in South Sulawesi, Central Sulawesi, North Sulawesi, or Southeast Sulawesi
Panel D: Regional Group	Bali and Nusa Tenggara	5	9.8%	Main campus located in Bali, West Nusa Tenggara, or East Nusa Tenggara
Panel D: Regional Group	Kalimantan	2	3.9%	Main campus located in East Kalimantan and South Kalimantan
Panel D: Regional Group	Maluku	2	3.9%	Main campus located in Maluku and North Maluku (Papua: 0; 0.0%)

Note: Proportions in Panels B, C, and D sum to 100.0% across the 51 institutions in the primary analytical corpus. PTN = public higher education institutions; PTS = private higher education institutions.

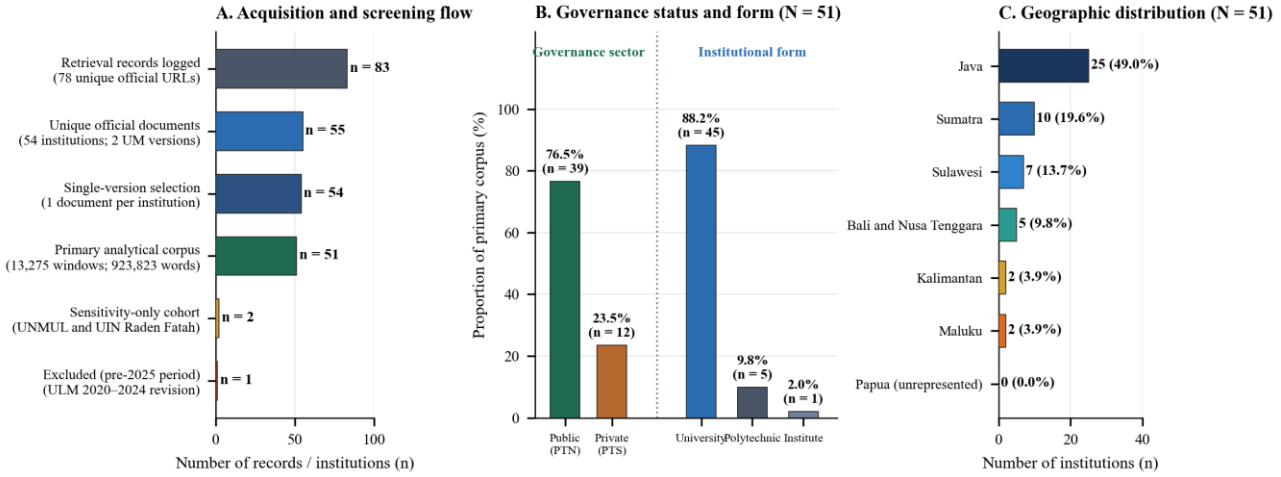


Figure 1. Multi-stage document acquisition flow and institutional stratification of the 51 Indonesian higher education strategic plans (Renstra) covering 2025: (A) document acquisition and screening flow from 83 retrieval records across 78 unique official URLs through 55 unique documents (across 54 institutions, consolidating two versions for Universitas Negeri Malang), 54 single-version institutional candidates, and the primary analytical corpus of 51 institutions (13,275 windows; 923,823 alphabetic words), alongside two sensitivity-only documents (UNMUL and UIN Raden Fatah) and one excluded pre-2025 document (ULM 2020–2024 revision); (B) governance status (39 public [PTN, 76.5%] and 12 private [PTS, 23.5%]) and institutional form (45 universities [88.2%], five polytechnics [9.8%], and one technological institute [2.0%]); and (C) geographic distribution across six regional groupings (Java 25 [49.0%], Sumatra 10 [19.6%], Sulawesi 7 [13.7%], Bali and Nusa Tenggara 5 [9.8%], Kalimantan 2 [3.9%], Maluku 2 [3.9%], and Papua 0 [0.0%]).

3.2 Cross-Institutional Theme Prevalence and Concept Dictionary Support

Across the 51 strategic plans, three orientations—internationalization, digital transformation, and research downstreaming—were detected in all 51/51 institutions (100.0%), whereas inclusivity was detected in 50/51 institutions (98.0%; unflagged only in Universitas Muhammadiyah Riau [UMRI], which does not imply the absence of campus inclusion policies outside dictionary terms; Table 2 and Table 3). Cumulative positive window counts across the 13,275 analytical windows were 3,522 windows for internationalization (pooled proportion: 26.5%), 1,615 windows for digital transformation (pooled: 12.2%), 1,517 windows for research downstreaming (pooled: 11.4%), and 990 windows for inclusivity (pooled: 7.5%; Table 2 and Figure 2A).

When weighted equally across all 51 institutions, internationalization exhibited the highest mean positive window proportion at 28.0% ± 9.9% (95% institutional bootstrap CI: 25.4%–30.8%; median: 26.1% [IQR: 21.8%–32.1%; min–max: 9.9%–55.3%]), followed by digital transformation at 12.1% ± 4.9% (95% CI: 10.8%–13.5%; median: 11.1% [IQR: 8.8%–14.2%; min–max: 4.2%–27.8%]), research downstreaming at 11.5% ± 5.7% (95% CI: 10.0%–13.1%; median: 10.7% [IQR: 6.9%–15.9%; min–max: 2.3%–24.6%]), and inclusivity at 6.8% ± 4.1% (95% CI: 5.8%–8.0%; median: 5.4% [IQR: 3.5%–8.9%; min–max: 0.0%–18.1%; Table 2 and Figure 2A]). Shapiro–Wilk normality tests indicated significant departures from normality across all four themes ($W = 0.942$, $p = 0.0155$ for internationalization; $W = 0.928$, $p = 0.0044$ for digital transformation; $W = 0.954$, $p = 0.0452$ for research downstreaming; and $W = 0.933$, $p = 0.0064$ for inclusivity), justifying the joint reporting of parametric means, non-parametric bootstrap intervals, and medians with IQRs.

Concept-level decomposition across the 24 operational dictionary concepts demonstrated why broad thematic prevalence must be interpreted alongside concept specificity (Table 2 and Figure 2B). Within internationalization, general international orientation accounted for 3,203 windows (51/51 institutions) and foreign partnerships for 921 windows (51/51), followed by international publication (515 windows; 50/51), foreign students (232 windows; 45/51), international mobility (193 windows; 46/51), and collaborative degrees (73 windows; 22/51). When the generic term 'internasional' was excluded in sensitivity model S2, mean internationalization prevalence dropped by more than half from 28.0% to 13.1%. Within digital transformation, information systems (642 windows; 51/51) and ICT infrastructure (523 windows; 51/51) were universally present, followed by digitalization (347 windows; 37/51), online learning (303 windows; 47/51), data governance (133 windows; 38/51), and information security (64 windows; 18/51). Within research downstreaming, research and industry linkage (684 windows; 49/51) and intellectual property (470 windows; 51/51) led detection, followed by explicit downstreaming/hilirisasi (400 windows; 41/51), incubation and startups (270 windows; 39/51), commercialization (129 windows; 30/51), and technology transfer (61 windows; 23/51). Within inclusivity, educational cost support (370 windows; 48/51) and equitable educational access (291 windows; 42/51) were most prevalent, followed by inclusive orientation (274 windows; 32/51), affirmation and vulnerable groups (148 windows; 36/51), disability (122 windows; 30/51), and equality and non-discrimination (48 windows; 22/51; Table 2).

Institution-specific profiles varied substantially across the 51 higher education institutions (Table 3 and Figure 3A–C). Internationalization reached its highest window proportions in Universitas Trunojoyo Madura (55.3%) and Universitas Negeri Malang (55.3%), followed by Universitas PGRI Semarang (47.2%), Universitas Andalas (43.4%), Universitas Negeri Semarang (42.2%), Universitas Negeri Jakarta (41.6%), and Politeknik Negeri Jakarta (40.9%), whereas Politeknik Negeri Bandung (9.9%), Politeknik Negeri Madura (14.7%), Institut Teknologi Kalimantan (14.9%), and Universitas Muhammadiyah Makassar (14.9%) recorded the lowest proportions. Research downstreaming peaked in Universitas Trunojoyo Madura (24.6%), Universitas PGRI Semarang (23.1%), Politeknik Negeri Jakarta (22.0%), Universitas Negeri Malang (22.0%), and Politeknik Negeri Madura (21.7%). Inclusivity was highest in Universitas Terbuka (18.1%), Universitas Gadjah Mada (15.4%), UIN Alauddin Makassar (15.2%), and UIN Sultan Maulana Hasanuddin Banten (14.5%). Digital transformation peaked in UIN Alauddin Makassar (27.8%), Universitas Terbuka (25.8%), Universitas Muhammadiyah Surakarta (20.3%), Universitas Islam Ibrahimy Banyuwangi (18.5%), and Universitas Wiraraja (18.5%; Table 3 and Figure 3).

Table 2. Cross-institutional prevalence of the four strategic orientation themes and dictionary support across the 24 operational lexical concepts (N = 51 institutions; 13,275 analytical text windows).

Strategic Theme / Operational Concept	Positive Windows (n)	Pooled Window %	Institution Detection (n/51, %)	Equal-Weight Mean % ± SD (95% Bootstrap CI)	Median % [IQR; Min–Max]
THEME 1: INTERNATIONALIZATION (Overall Theme)	3,522	26.5%	51/51 (100.0%)	28.0% ± 9.9% (25.4–30.8%)	26.1% [21.8–32.1%; 9.9–55.3%]
1.1. International orientation (Orientasi internasional)	3,203	24.1%	51/51 (100.0%)	—	—
1.2. Foreign partnerships (Kemitraan luar negeri)	921	6.9%	51/51 (100.0%)	—	—
1.3. International publication (Publikasi internasional)	515	3.9%	50/51 (98.0%)	—	—
1.4. Foreign students (Mahasiswa asing)	232	1.7%	45/51 (88.2%)	—	—
1.5. International mobility (Mobilitas internasional)	193	1.5%	46/51 (90.2%)	—	—
1.6. Collaborative degrees (Gelar kolaboratif)	73	0.5%	22/51 (43.1%)	—	—
THEME 2: DIGITAL TRANSFORMATION (Overall Theme)	1,615	12.2%	51/51 (100.0%)	12.1% ± 4.9% (10.8–13.5%)	11.1% [8.8–14.2%; 4.2–27.8%]
2.1. Information systems (Sistem informasi)	642	4.8%	51/51 (100.0%)	—	—
2.2. ICT infrastructure (Infrastruktur TIK)	523	3.9%	51/51 (100.0%)	—	—
2.3. Digitalization (Digitalisasi)	347	2.6%	37/51 (72.5%)	—	—
2.4. Online learning (Pembelajaran daring)	303	2.3%	47/51 (92.2%)	—	—
2.5. Data governance (Tata kelola data)	133	1.0%	38/51 (74.5%)	—	—
2.6. Information security (Keamanan informasi)	64	0.5%	18/51 (35.3%)	—	—
THEME 3: RESEARCH DOWNSTREAMING (Overall Theme)	1,517	11.4%	51/51 (100.0%)	11.5% ± 5.7% (10.0–13.1%)	10.7% [6.9–15.9%; 2.3–24.6%]
3.1. Research and industry linkage (Riset dan industri)	684	5.2%	49/51 (96.1%)	—	—
3.2. Intellectual property (Kekayaan intelektual)	470	3.5%	51/51 (100.0%)	—	—
3.3. Downstreaming (Hilirisasi)	400	3.0%	41/51 (80.4%)	—	—
3.4. Incubation and startups (Inkubasi dan usaha rintisan)	270	2.0%	39/51 (76.5%)	—	—
3.5. Commercialization (Komersialisasi)	129	1.0%	30/51 (58.8%)	—	—
3.6. Technology transfer (Transfer teknologi)	61	0.5%	23/51 (45.1%)	—	—
THEME 4: INCLUSIVITY (Overall Theme)	990	7.5%	50/51 (98.0%)	6.8% ± 4.1% (5.8–8.0%)	5.4% [3.5–8.9%; 0.0–18.1%]
4.1. Educational cost support (Dukungan biaya pendidikan)	370	2.8%	48/51 (94.1%)	—	—
4.2. Equitable educational access (Pemerataan akses pendidikan)	291	2.2%	42/51 (82.4%)	—	—
4.3. Inclusive orientation (Orientasi inklusif)	274	2.1%	32/51 (62.7%)	—	—
4.4. Affirmation & vulnerable groups (Afirmasi & kelompok rentan)	148	1.1%	36/51 (70.6%)	—	—
4.5. Disability (Disabilitas)	122	0.9%	30/51 (58.8%)	—	—
4.6. Equality and non-discrimination (Kesetaraan & nondiskriminasi)	48	0.4%	22/51 (43.1%)	—	—

Note: A theme is scored positive in an 80-token window if at least one of its six constituent concepts is detected; themes overlap across windows, and constituent concept window counts sum to more than the theme total because multiple concepts co-occur within the same window. Equal-weight institutional means weight each of the 51 institutions equally; 95% confidence intervals (CIs) are derived from 5,000 non-parametric institutional bootstrap replications conditional on the purposive sample. Shapiro–Wilk normality tests across the 51 institutions: Internationalization W = 0.942 (p = 0.0155), Digital Transformation W = 0.928 (p = 0.0044), Research Downstreaming W = 0.954 (p = 0.0452), and Inclusivity W = 0.933 (p = 0.0064).

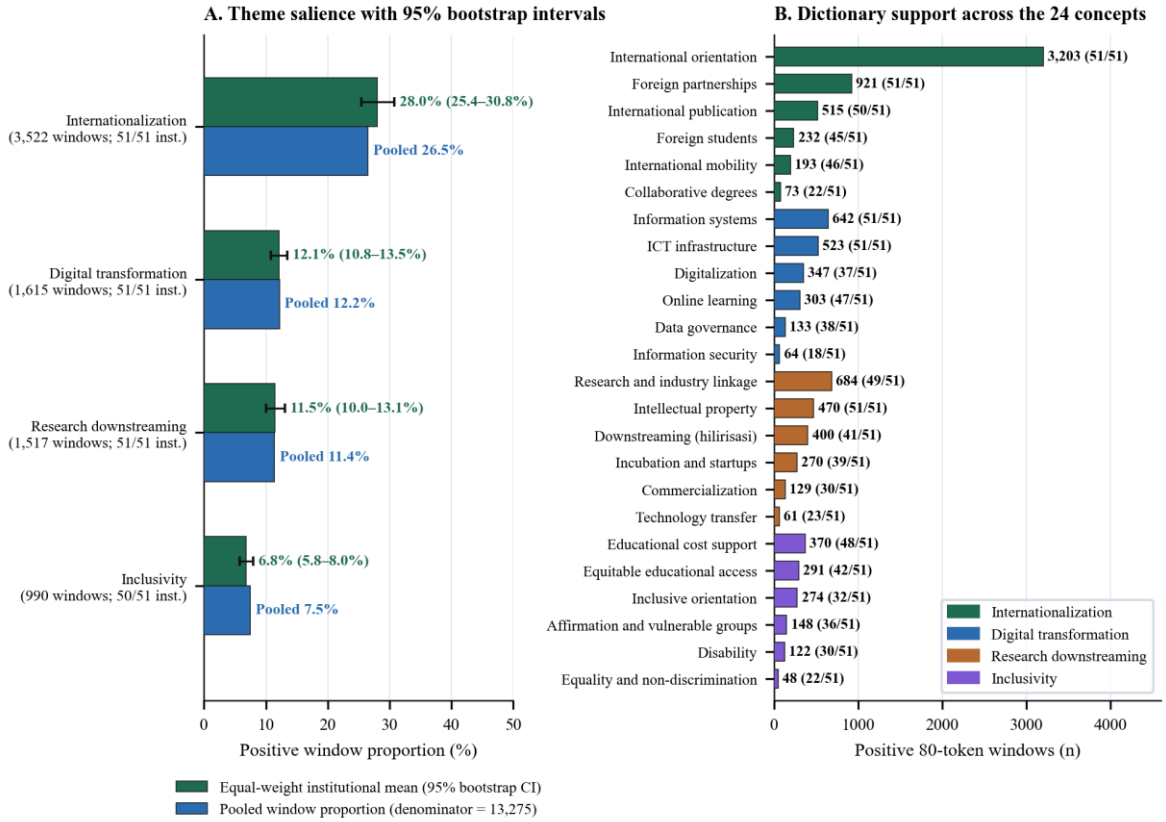


Figure 2. Cross-institutional prevalence of the four strategic orientations and operational dictionary support across the 24 pre-specified concepts (N = 51 institutions; 13,275 analytical 80-token windows): (A) equal-weight institutional mean positive window proportions with 95% non-parametric bootstrap confidence intervals (5,000 institutional resamples) compared against pooled window proportions (Internationalization: 28.0% [95% CI: 25.4–30.8%], pooled 26.5% [3,522 windows]; Digital Transformation: 12.1% [10.8–13.5%], pooled 12.2% [1,615 windows]; Research Downstreaming: 11.5% [10.0–13.1%], pooled 11.4% [1,517 windows]; Inclusivity: 6.8% [5.8–8.0%], pooled 7.5% [990 windows]), and (B) positive window counts and institutional detection frequencies (n/51) across all 24 operational concepts.

Table 3. Institutional characteristics, planning horizon, analytical window counts, lexical theme proportions, and exploratory cluster assignments across the 51 Indonesian higher education institutions.

No.	Institution Code and Name (Governance, Form, Region)	Planning Horizon	Windows (n)	Intl (%)	Downstr (%)	Incl (%)	Digital (%)	Cluster
1	Hang Tuah · Universitas Hang Tuah (PTS, University, Java)	2021–2025	106	27.4%	10.4%	2.8%	11.3%	4
2	ITK · Institut Teknologi Kalimantan (PTN, Institute, Kalimantan)	2021–2025	281	14.9%	12.8%	5.3%	17.1%	4
3	PNJ · Politeknik Negeri Jakarta (PTN, Polytechnic, Java)	2025–2029	186	40.9%	22.0%	4.8%	17.2%	2
4	POLBAN · Politeknik Negeri Bandung (PTN, Polytechnic, Java)	2025–2029	203	9.9%	16.3%	3.4%	11.8%	4
5	Poltekkes Banjarmasin · Poltekkes Kemenkes Banjarmasin (PTN, Polytechnic, Kalimantan)	2025–2029	118	21.2%	6.8%	3.4%	5.1%	3
6	Poltekpar Bali · Politeknik Pariwisata Bali (PTN, Polytechnic, Bali–NT)	2025–2029	266	29.3%	9.4%	5.3%	7.5%	3
7	POLTERA · Politeknik Negeri Madura (PTN, Polytechnic, Java)	2025–2029	300	14.7%	21.7%	9.3%	11.0%	2
8	TRUNOJOYO · Universitas Trunojoyo Madura (PTN, University, Java)	2025–2029	179	55.3%	24.6%	3.4%	11.2%	2
9	UBK · Universitas Bung Karno (PTS, University, Java)	2024–2029	66	22.7%	6.1%	3.0%	12.1%	1
10	UGM · Universitas Gadjah Mada (PTN, University, Java)	2022–2027	156	29.5%	7.1%	15.4%	12.8%	1
11	UHO · Universitas Halu Oleo (PTN, University, Sulawesi)	2025–2029	350	34.0%	17.4%	11.4%	11.7%	2
12	UI · Universitas Indonesia (PTN, University, Java)	2025–2029	478	20.7%	12.1%	13.6%	9.2%	2
13	UIN Alauddin · UIN Alauddin Makassar (PTN, University, Sulawesi)	2025–2029	711	27.0%	16.6%	15.2%	27.8%	1
14	UIN Banten · UIN Sultan Maulana Hasanuddin Banten (PTN, University, Java)	2025–2029	276	22.5%	5.8%	14.5%	10.1%	1

No.	Institution Code and Name (Governance, Form, Region)	Planning Horizon	Windows (n)	Intl (%)	Downstr (%)	Incl (%)	Digital (%)	Cluster
15	UIN Saizu · UIN Prof. K.H. Saifuddin Zuhri (PTN, University, Java)	2025–2029	162	28.4%	6.2%	8.6%	9.9%	1
16	UIN Syahada · UIN Syekh Ali Hasan Ahmad Addary (PTN, University, Sumatra)	2025–2029	259	16.2%	3.9%	3.9%	7.7%	3
17	UM · Universitas Negeri Malang (PTN, University, Java)	2025–2029	150	55.3%	22.0%	10.7%	17.3%	2
18	UMRAH · Universitas Maritim Raja Ali Haji (PTN, University, Sumatra)	2025–2029	282	22.7%	8.5%	5.3%	9.6%	3
19	UMRI · Universitas Muhammadiyah Riau (PTS, University, Sumatra)	2022–2026	127	26.0%	6.3%	0.0%	13.4%	4
20	UMS · Universitas Muhammadiyah Surakarta (PTS, University, Java)	2025–2029	350	36.9%	11.4%	6.3%	20.3%	4
21	UM Surabaya · Universitas Muhammadiyah Surabaya (PTS, University, Java)	2021–2025	203	31.0%	6.4%	5.4%	8.4%	3
22	UNAND · Universitas Andalas (PTN, University, Sumatra)	2025–2029	318	43.4%	19.8%	8.2%	16.0%	2
23	UNDANA · Universitas Nusa Cendana (PTN, University, Bali–NT)	2021–2025	357	22.4%	5.9%	6.2%	17.9%	1
24	Undip · Universitas Diponegoro (PTN, University, Java)	2025–2029	317	32.5%	16.7%	4.7%	13.6%	4
25	Unhas · Universitas Hasanuddin (PTN, University, Sulawesi)	2025–2029	230	38.3%	13.0%	7.8%	7.0%	3
26	UNIIB · Universitas Islam Ibrahimy Banyuwangi (PTS, University, Java)	2023–2026	54	16.7%	11.1%	1.9%	18.5%	4
27	Unila · Universitas Lampung (PTN, University, Sumatra)	2025–2029	315	28.6%	9.2%	5.4%	10.2%	3
28	UNIMAL · Universitas Malikussaleh (PTN, University, Sumatra)	2025–2029	151	31.8%	16.6%	4.6%	10.6%	4
29	Unimor · Universitas Timor (PTN, University, Bali–NT)	2024–2028	110	25.5%	7.3%	10.0%	10.0%	1
30	UNINDRA · Universitas Indraprasta PGRI (PTS, University, Java)	2025–2029	120	17.5%	4.2%	5.0%	14.2%	1
31	Unismuh Makassar · Univ. Muhammadiyah Makassar (PTS, University, Sulawesi)	2021–2025	1,299	14.9%	3.9%	3.4%	7.4%	3
32	UNJ · Universitas Negeri Jakarta (PTN, University, Java)	2025–2029	298	41.6%	9.1%	4.0%	8.7%	3
33	Unkhair · Universitas Khairun (PTN, University, Maluku)	2025–2029	262	29.0%	13.7%	11.1%	8.0%	2
34	UNM · Universitas Negeri Makassar (PTN, University, Sulawesi)	2024–2028	113	27.4%	11.5%	1.8%	16.8%	4
35	UNNES · Universitas Negeri Semarang (PTN, University, Java)	2023–2028	192	42.2%	9.4%	1.6%	6.2%	3
36	UNPATTI · Universitas Pattimura (PTN, University, Maluku)	2025–2029	455	21.5%	15.6%	9.2%	9.0%	2
37	UNRAM · Universitas Mataram (PTN, University, Bali–NT)	2025–2029	442	21.7%	8.1%	11.1%	17.6%	1
38	Unri · Universitas Riau (PTN, University, Sumatra)	2025–2029	225	19.6%	10.7%	8.0%	8.9%	2
39	UNSIL · Universitas Siliwangi (PTN, University, Java)	2022–2026	126	19.8%	7.1%	3.2%	14.3%	1
40	UNSRAT · Universitas Sam Ratulangi (PTN, University, Sulawesi)	2025–2029	310	26.1%	17.7%	13.2%	9.7%	2
41	Unsurya · Univ. Dirgantara Marsekal Suryadarma (PTS, University, Java)	2023–2027	168	24.4%	13.7%	4.8%	9.5%	3
42	UNTAD · Universitas Tadulako (PTN, University, Sulawesi)	2025–2029	393	30.3%	8.4%	8.1%	14.0%	1
43	UNUD · Universitas Udayana (PTN, University, Bali–NT)	2025–2029	191	34.0%	15.2%	7.3%	11.5%	2
44	UNY · Universitas Negeri Yogyakarta (PTN, University, Java)	2023–2026	194	29.4%	14.4%	3.1%	12.9%	4
45	UPGRIS · Universitas PGRI Semarang (PTS, University, Java)	2025–2029	108	47.2%	23.1%	8.3%	11.1%	2
46	USK · Universitas Syiah Kuala (PTN, University, Sumatra)	2025–2029	426	24.4%	11.3%	5.9%	4.2%	3

No.	Institution Code and Name (Governance, Form, Region)	Planning Horizon	Windows (n)	Intl (%)	Downstr (%)	Incl (%)	Digital (%)	Cluster
47	USTI · Univ. Sains dan Teknologi Indonesia (PTS, University, Sumatra)	2024–2028	124	22.6%	6.5%	3.2%	6.5%	3
48	UT · Universitas Terbuka (PTN, University, Java)	2025–2029	248	21.8%	7.3%	18.1%	25.8%	1
49	UTU · Universitas Teuku Umar (PTN, University, Sumatra)	2025–2029	283	23.0%	17.7%	3.5%	5.3%	3
50	WALISONGO · UIN Walisongo Semarang (PTN, University, Java)	2025–2029	129	38.0%	2.3%	8.5%	8.5%	1
51	Wiraraja · Universitas Wiraraja (PTS, University, Java)	2024–2027	108	25.9%	2.8%	5.6%	18.5%	1

Note: Intl = Internationalization; Downstr = Research Downstreaming (Hilirisasi riset); Incl = Inclusivity (Inklusivitas); Digital = Digital Transformation (Transformasi digital). Cluster numbers (1 to 4) are arbitrary identifiers from exploratory Ward clustering on log1p-transformed, standardized theme proportions and do not represent institutional rankings or definitive governance typologies. PTN, PTS, and institution codes as in Table 1; Cluster = exploratory Ward cluster assignment.



Figure 3. Complete institutional lexical attention profile matrix across all 51 Indonesian higher education strategic plans on a common percentage scale (0% to 60% positive 80-token windows) with exploratory Ward cluster assignments (C1 to C4): (A) institutions 1 to 17 (Hang Tuah through UM), (B) institutions 18 to 34 (UMRAH through UNM), and (C) institutions 35 to 51 (UNNES through Wiraraja). Percentages quantify documentary lexical salience rather than institutional governance quality or realized program performance.

3.3 Lexical Co-Occurrence Network Topology and Concept Centrality

All 24 dictionary concepts satisfied the node inclusion threshold (≥ 10 windows and ≥ 3 institutions), generating an undirected lexical co-occurrence graph with 24 concept nodes, 138 supported edges (≥ 3 co-occurring windows, ≥ 3 institutions, and NPMI > 0), and 4 Louvain modularity communities (Table 4 and Figure 4A). Across the 13,275 analytical windows, the density of supported edges reflects frequent within-window co-location of related strategic terms across institutional planning documents rather than verified operational integration.

As detailed in Table 4 (Panel A) and illustrated in the 12-node high-strength subgraph (Figure 4A), the eight concepts with the highest node strength (sum of incident NPMI weights) spanned three strategic themes: downstreaming/hilirisasi ranked first (degree = 17, strength = 3.094, betweenness = 0.265), followed by affirmation and vulnerable groups (degree = 14, strength = 2.713, betweenness = 0.166), intellectual property (degree = 9, strength = 2.378, betweenness = 0.107), foreign partnerships (degree = 15, strength = 2.374, betweenness = 0.063), disability (degree = 10, strength = 2.361, betweenness = 0.178), research and industry linkage (degree = 17, strength = 2.243, betweenness = 0.032), commercialization (degree = 8, strength = 2.170, betweenness = 0.000), and foreign students (degree = 15, strength = 2.158, betweenness = 0.087). Downstreaming (0.265), disability (0.178), and affirmation and vulnerable groups (0.166) exhibited the highest shortest-path betweenness centralities, acting as lexical bridges connecting distinct thematic neighborhoods.

Across the 138 supported edges, the eight concept pairs with the highest NPMI values highlighted strong within-theme and cross-theme lexical coupling (Table 4 Panel B and Figure 4B). The strongest normalized association linked data governance and information security within the digital transformation theme (20 co-occurring windows across 5 institutions; Jaccard = 0.113; NPMI = 0.529), followed closely by affirmation and vulnerable groups paired with educational cost support (67 windows across 28 institutions; Jaccard = 0.149; NPMI = 0.527), and disability paired with affirmation and vulnerable groups (29 windows across 14 institutions; Jaccard = 0.120; NPMI =

0.499). Within research downstreaming, explicit downstreaming paired strongly with commercialization (51 windows across 12 institutions; Jaccard = 0.107; NPMI = 0.463), and commercialization paired with technology transfer (13 windows across 9 institutions; Jaccard = 0.073; NPMI = 0.446). Within internationalization, general international orientation and foreign partnerships co-occurred in 791 windows across 50 institutions (Jaccard = 0.237; NPMI = 0.450). Affirmation and vulnerable groups also linked strongly with equitable educational access (40 windows across 13 institutions; Jaccard = 0.100; NPMI = 0.433), while a prominent cross-theme bridge connected international publication and intellectual property across 130 windows in 44 institutions (Jaccard = 0.152; NPMI = 0.425; Table 4).

Table 4. Lexical co-occurrence network topology across the 24 concepts and 138 supported edges (4 Louvain communities): top eight concepts by node strength and top eight concept pairs by Normalized Pointwise Mutual Information (NPMI).

Panel	Rank. Concept / Konsep (theme)	Degree	Node strength (sum of NPMI)	Betweenness	NPMI
Panel A: Top 8 Concepts by Node Strength	1. Downstreaming / Hilirisasi (Research Downstreaming)	17	3.094	0.265	—
Panel A: Top 8 Concepts by Node Strength	2. Affirmation and vulnerable groups / Afirmasi dan kelompok rentan (Inclusivity)	14	2.713	0.166	—
Panel A: Top 8 Concepts by Node Strength	3. Intellectual property / Kekayaan intelektual (Research Downstreaming)	9	2.378	0.107	—
Panel A: Top 8 Concepts by Node Strength	4. Foreign partnerships / Kemitraan luar negeri (Internationalization)	15	2.374	0.063	—
Panel A: Top 8 Concepts by Node Strength	5. Disability / Disabilitas (Inclusivity)	10	2.361	0.178	—
Panel A: Top 8 Concepts by Node Strength	6. Research and industry linkage / Riset dan industri (Research Downstreaming)	17	2.243	0.032	—
Panel A: Top 8 Concepts by Node Strength	7. Commercialization / Komersialisasi (Research Downstreaming)	8	2.170	0.000	—
Panel A: Top 8 Concepts by Node Strength	8. Foreign students / Mahasiswa asing (Internationalization)	15	2.158	0.087	—
Panel	**Rank. Concept pair (themes)**	**Co-occurring windows**	**Institutions (n)**	**Jaccard**	**NPMI**
Panel B: Top 8 Edges by NPMI	1. Data governance — Information security (Digital — Digital)	20 windows	5 inst.	0.113	0.529
Panel B: Top 8 Edges by NPMI	2. Affirmation & vulnerable groups — Educational cost support (Incl. — Incl.)	67 windows	28 inst.	0.149	0.527
Panel B: Top 8 Edges by NPMI	3. Disability — Affirmation & vulnerable groups (Incl. — Incl.)	29 windows	14 inst.	0.120	0.499
Panel B: Top 8 Edges by NPMI	4. Downstreaming — Commercialization (Downstr. — Downstr.)	51 windows	12 inst.	0.107	0.463
Panel B: Top 8 Edges by NPMI	5. International orientation — Foreign partnerships (Intl. — Intl.)	791 windows	50 inst.	0.237	0.450
Panel B: Top 8 Edges by NPMI	6. Commercialization — Technology transfer (Downstr. — Downstr.)	13 windows	9 inst.	0.073	0.446
Panel B: Top 8 Edges by NPMI	7. Affirmation & vulnerable groups — Equitable educational access (Incl. — Incl.)	40 windows	13 inst.	0.100	0.433
Panel B: Top 8 Edges by NPMI	8. International publication — Intellectual property (Intl. — Downstr.)	130 windows	44 inst.	0.152	0.425

Note: In the 24-node, 138-edge co-occurrence graph, candidate nodes required support in ≥10 windows and ≥3 institutions; candidate edges required co-occurrence in ≥3 windows across ≥3 institutions and positive NPMI (>0). Node strength is the sum of incident NPMI edge weights; shortest-path distances for betweenness centrality are defined as 1/NPMI. NPMI = Normalized Pointwise Mutual Information; Jaccard = ratio of co-occurring windows to the union of windows containing either concept.

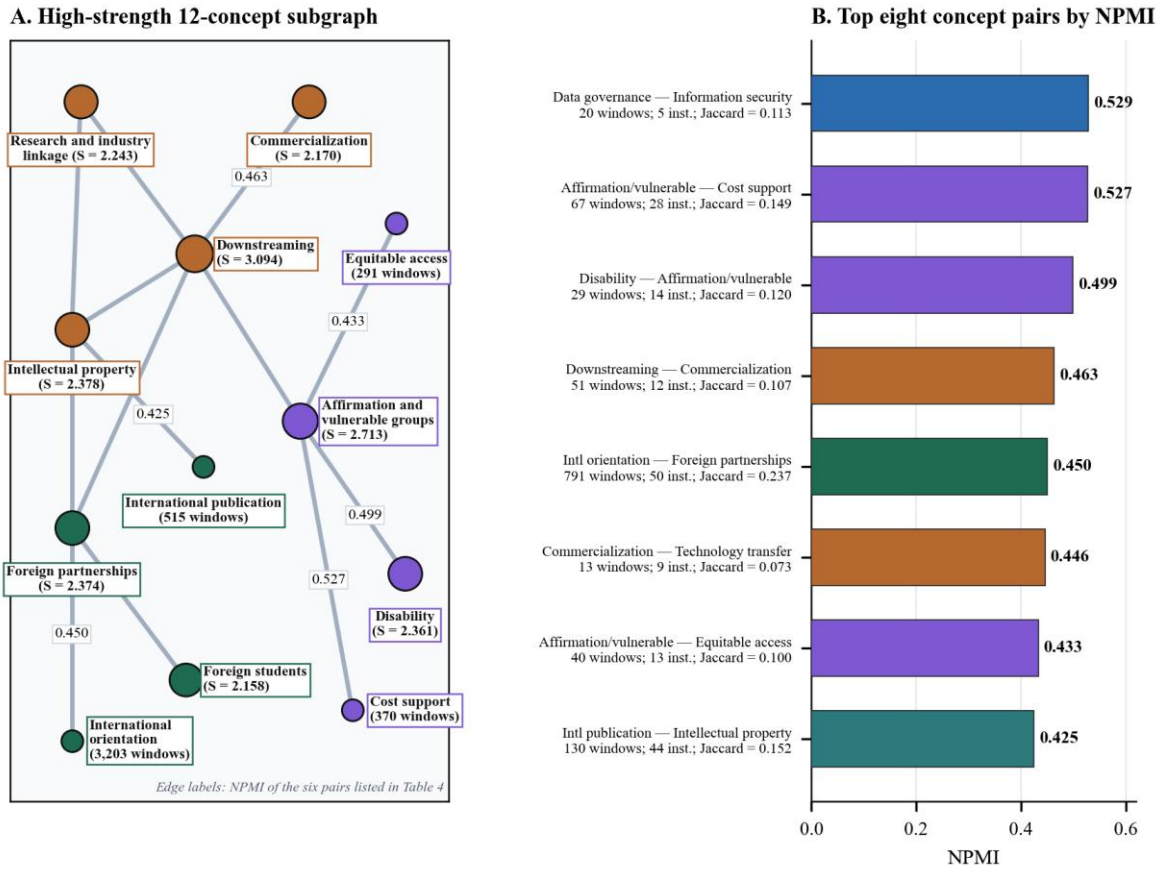


Figure 4. Lexical co-occurrence network topology across the 24 operational concepts and 138 supported edges (four Louvain communities): (A) high-strength 12-concept subgraph illustrating within-theme and cross-theme co-occurrence bridges (node size proportional to NPMI strength; led by Downstreaming [strength = 3.094, betweenness = 0.265], Affirmation and Vulnerable Groups [2.713, 0.166], Intellectual Property [2.378, 0.107], Foreign Partnerships [2.374, 0.063], and Disability [2.361, 0.178]), and (B) top eight concept pairs ranked by Normalized Pointwise Mutual Information (led by Data Governance—Information Security [NPMI = 0.529], Affirmation/Vulnerable Groups—Educational Cost Support [0.527], and Disability—Affirmation/Vulnerable Groups [0.499]).

3.4 Exploratory Clustering, Null-Permutation Benchmarking, and Multi-Scenario Sensitivity

In the primary Ward hierarchical clustering analysis across $k \in \{2, 3, 4\}$ (min size ≥ 3), all three candidate partitions satisfied the minimum cluster size requirement: k = 2 yielded a silhouette coefficient of 0.142 (0.1420; minimum cluster size = 14), k = 3 yielded 0.140 (0.1405; minimum size = 13) and k = 4 achieved the highest silhouette score of 0.166 (0.1663; cluster sizes of 14, 13, 14, and 10 institutions; Table 5 Panel A and Figure 5A). In the extended diagnostic range $k \in \{2, ..., 8\}$ requiring min size ≥ 5, k = 5 remained admissible (minimum size = 5) but yielded a slightly lower silhouette (0.1657 vs. 0.1663), whereas k = 6 (silhouette = 0.1819), k = 7 (0.1949), and k = 8 (0.2151) all failed the admissibility criterion by fragmenting into tiny two-institution clusters (min size = 2; Table 5 Panel B and Figure 5A). When benchmarked against 1,000 independent marginal feature permutations, the observed four-cluster silhouette of 0.166 fell below the null median silhouette of 0.201, yielding an empirical p-value of 0.961 (Table 5 and Table 6). Thus, the observed cluster separation does not exceed what is expected by chance when the four strategic themes vary independently with their observed marginal distributions.

Descriptive theme averages across the four exploratory clusters summarize how the partition divides the continuous four-dimensional space without justifying reified typological labels (Table 5 Panel C and Figure 5B). Cluster 1 (n = 14; 27.5%) exhibited relatively higher digital transformation (15.3%) and inclusivity (9.5%) alongside moderate internationalization (25.2%) and lower research downstreaming (6.8%). Cluster 2 (n = 13; 25.5%) displayed the highest mean internationalization (34.0%) and research downstreaming (18.1%), paired with 9.1% inclusivity and 11.7% digital transformation. Cluster 3 (n = 14; 27.5%) showed moderate internationalization (27.2%) and downstreaming (9.2%) but lower digital transformation (7.4%) and inclusivity (4.5%). Cluster 4 (n = 10; 19.6%) combined moderate internationalization (25.3%), higher digital transformation (14.6%), and above-average downstreaming (12.7%) with the lowest inclusivity (3.4%).

Internal resampling diagnostics showed marked partition instability (Table 5 Panel C, Table 6 Panel A, and Figure 5B). Across 1,000 subsampling runs drawing 80% of institutions (n = 41), the median ARI against the full-sample partition was only 0.389 (2.5th–97.5th percentile interval: 0.110 to 0.763). Cluster-wise mean Jaccard recovery across the 1,000 subsamples was 0.625 (0.235–1.000) for Cluster 1, 0.445 (0.273–0.819) for Cluster 2, 0.594 (0.286–1.000) for Cluster 3, and 0.401 (0.188–0.889) for Cluster 4. When the optimal cluster count $k \in \{2, 3, 4\}$ was re-selected within each of the 1,000 subsamples, k = 4 was chosen in only 508 runs (50.8%), whereas k = 2 was chosen in 351 runs (35.1%) and k = 3 in 141 runs (14.1%). Leave-one-institution-out jackknife analysis (n = 50) produced ARIs ranging from 0.237 to 1.000, demonstrating that removing a single institution can alter global cluster assignments. Furthermore, across all 1,275 institutional pairs, 58.6% exhibited pairwise co-clustering consensus probabilities strictly between 0.1 and 0.9, indicating widespread co-membership ambiguity.

Across the 14 deterministic sensitivity specifications with k held constant at 4, ARI values ranged widely from 0.091 to 0.699 (Table 6 Panel B and Figure 5C). The highest agreement occurred when excluding the 1,299-window Unismuh Makassar document (ARI = 0.699; n = 50; 11,976 windows) or adding either candidate document one at a time (ARI = 0.697 for UNMUL [n = 52; 13,558 windows] and ARI = 0.697 for UIN Raden Fatah [n = 52; 13,593 windows]), followed by restricting to the 39 public PTN institutions (ARI = 0.634; 10,442 windows) or simultaneously excluding the three amended documents from UNRI, UM, and UNSRAT (ARI = 0.629; n = 48; 12,590 windows). Moderate sensitivity appeared when excluding the amended

UNRI document alone (ARI = 0.516; n = 50), applying a narrower dictionary without generic 'internasional' (ARI = 0.497; n = 51), halving the window size to 40 tokens (ARI = 0.451; 24,419 windows), or retaining statutory reference text (ARI = 0.352; 13,761 windows). In contrast, cluster membership reorganized sharply when using approximate sentence units (ARI = 0.290; 38,336 sentences), restricting to the 45 universities by removing the five polytechnics and one institute (ARI = 0.253; 11,921 windows), measuring term frequency per 10,000 alphabetic words (ARI = 0.183), restricting to the 26 non-Java institutions (ARI = 0.182; 8,398 windows), or expanding window length to 120 tokens (ARI = 0.091; 9,470 windows). Finally, exploratory subgroup and correlation analyses (Table 7 and Figure 6A–B) revealed that document window count correlated positively with inclusivity proportion (Spearman ρ = 0.426, nominal p = 0.0019), and that PTN institutions (n = 39) exhibited nominally higher lexical proportions than PTS institutions (n = 12) in inclusivity (7.6% ± 4.2% vs. 4.1% ± 2.2%; Mann–Whitney U = 351.0, p = 0.0097) and research downstreaming (12.3% ± 5.6% vs. 8.8% ± 5.6%; U = 329.0, p = 0.0358), whereas Java (n = 25) and non-Java (n = 26) institutions showed no statistically significant differences across any of the four themes (all p ≥ 0.0750; Table 7). After Benjamini–Hochberg adjustment across the 12 exploratory tests, q = 0.023 for the length–inclusivity correlation, 0.058 for the PTN versus PTS inclusivity difference, and 0.143 for the PTN versus PTS research downstreaming difference.

Examining the structural mechanics across the 14 sensitivity models (Table 6 Panel B and Figure 5C) clarifies why seemingly minor analytical choices induce substantial cluster reassignments (ARI spanning 0.091 to 0.699). First, text segmentation granularity directly alters baseline concept probabilities: shrinking the window from 80 tokens (13,275 windows) to 40 tokens (24,419 windows; ARI = 0.451) or approximate sentence units (38,336 units; ARI = 0.290) reduces the probability that multiple diffuse concepts fall within the same unit, whereas expanding the window to 120 tokens (9,470 windows; ARI = 0.091) inflates positive window proportions in dense documents and blurs thematic boundaries. Second, replacing binary window prevalence with term frequency per 10,000 alphabetic words (ARI = 0.183) shifts leverage toward documents that repeat key terms multiple times within concentrated tables or paragraphs rather than distributing attention across sections. Third, altering sample composition—such as removing the five polytechnics and one institute to analyze the 45 universities alone (ARI = 0.253) or restricting to the 26 outer-island institutions outside Java (ARI = 0.182)—modifies the sample mean μ_t and standard deviation σ_t during z-score re-standardization, causing institutions situated near the center of the continuous four-dimensional distribution to cross Ward decision boundaries.

Table 5. Evaluation of candidate cluster counts (k = 2 to 8) under Ward hierarchical clustering and descriptive theme profiles with subsampling Jaccard stability across the primary four-cluster solution (N = 51).

Panel	Candidate k	Silhouette	Minimum cluster size	Admissibility	Diagnostic status
Panel A: Primary Range (k = 2–4; min ≥ 3)	k = 2 clusters	0.142 (0.1420)	Minimum cluster size = 14	Meets min ≥ 3 (Yes)	Meets diagnostic min ≥ 5 (Yes)
Panel A: Primary Range (k = 2–4; min ≥ 3)	k = 3 clusters	0.140 (0.1405)	Minimum cluster size = 13	Meets min ≥ 3 (Yes)	Meets diagnostic min ≥ 5 (Yes)
Panel A: Primary Range (k = 2–4; min ≥ 3)	k = 4 clusters (Selected Primary)	0.166 (0.1663)	Minimum cluster size = 10	Meets min ≥ 3 (Yes)	Meets diagnostic min ≥ 5 (Yes); Null median = 0.201 (p = 0.961)
Panel B: Extended Range (k = 5–8; min ≥ 5)	k = 5 clusters	0.1657	Minimum cluster size = 5	Meets min ≥ 5 (Yes)	Below k = 4 silhouette (0.1657 vs. 0.1663)
Panel B: Extended Range (k = 5–8; min ≥ 5)	k = 6 clusters	0.1819	Minimum cluster size = 2	Fails min ≥ 5 (No)	Inadmissible due to fragmentation (min size = 2)
Panel B: Extended Range (k = 5–8; min ≥ 5)	k = 7 clusters	0.1949	Minimum cluster size = 2	Fails min ≥ 5 (No)	Inadmissible due to fragmentation (min size = 2)
Panel B: Extended Range (k = 5–8; min ≥ 5)	k = 8 clusters	0.2151	Minimum cluster size = 2	Fails min ≥ 5 (No)	Inadmissible due to fragmentation (min size = 2)
Panel	**Cluster**	**Size, n (%)**	**Theme profile (Intl, Downstr, Incl, Digital %)**	**Mean Jaccard**	**2.5th–97.5th percentile**
Panel C: 4-Cluster Profiles & Stability	Cluster 1 (Higher Digital & Inclusivity)	n = 14 (27.5%)	Intl: 25.2%, Downstr: 6.8%, Incl: 9.5%, Digital: 15.3%	Mean Jaccard = 0.625	2.5th–97.5th percentile: 0.235 to 1.000
Panel C: 4-Cluster Profiles & Stability	Cluster 2 (Higher Intl & Downstreaming)	n = 13 (25.5%)	Intl: 34.0%, Downstr: 18.1%, Incl: 9.1%, Digital: 11.7%	Mean Jaccard = 0.445	2.5th–97.5th percentile: 0.273 to 0.819
Panel C: 4-Cluster Profiles & Stability	Cluster 3 (Moderate Intl, Lower Other)	n = 14 (27.5%)	Intl: 27.2%, Downstr: 9.2%, Incl: 4.5%, Digital: 7.4%	Mean Jaccard = 0.594	2.5th–97.5th percentile: 0.286 to 1.000
Panel C: 4-Cluster Profiles & Stability	Cluster 4 (Higher Digital & Downstr, Low Incl)	n = 10 (19.6%)	Intl: 25.3%, Downstr: 12.7%, Incl: 3.4%, Digital: 14.6%	Mean Jaccard = 0.401	2.5th–97.5th percentile: 0.188 to 0.889

Note: Theme differences across Clusters 1–4 describe a partition constructed from those same four features and do not constitute an independent hypothesis test or external validation of discrete institutional types. Jaccard = maximum Jaccard cluster recovery similarity across subsamples; k = number of clusters; Intl = internationalization; Downstr = research downstreaming; Incl = inclusivity; Digital = digital transformation.

Table 6. Internal cluster stability diagnostics, marginal feature-permutation null benchmarking, and 14 sensitivity analyses across measurement and sample specifications (k fixed at 4).

Panel	Resampling or null diagnostic	Institutions (n)	Text units (n)	Result	Interpretation
Panel A: Resampling & Null Diagnostics	80% institutional subsampling (1,000 draws of n = 41)	41 per draw	—	Median ARI = 0.389 (95% PI: 0.110–0.763)	Moderate-to-low partition recovery under 20% sample perturbation
Panel A: Resampling & Null Diagnostics	Leave-one-institution-out (LOIO) jackknife	50 per run	—	ARI range = 0.237 to 1.000	Removing a single institution can substantially alter cluster assignments
Panel A: Resampling & Null Diagnostics	Resampling re-selection of optimal k (1,000 draws)	41 per draw	—	k=2: 351 (35.1%); k=3: 141 (14.1%); k=4: 508 (50.8%)	Optimal cluster count selects k = 4 in only half of subsamples
Panel A: Resampling & Null Diagnostics	Pairwise co-clustering consensus ambiguity	51 (1,275 pairs)	—	58.6% of pairs in (0.1, 0.9)	Over half of institutional pairs lack consistent co-membership
Panel A: Resampling & Null Diagnostics	Marginal feature-permutation null (1,000 permutations)	51	13,275	Null median silhouette = 0.201; p = 0.961	Observed silhouette (0.166) does not exceed independent-feature null
Panel	**Sensitivity specification (k = 4)**	**Institutions (n)**	**Text units (n)**	**Adjusted Rand Index (ARI)**	**Interpretation**
Panel B: 14 Sensitivity Specifications	S1. Without regulatory text removal	51	13,761 windows	ARI = 0.352	Retaining statutory reference text shifts partition substantially
Panel B: 14 Sensitivity Specifications	S2. Narrower dictionary (excluding generic 'internasional')	51	13,275 windows	ARI = 0.497	Internationalization mean drops from 28.0% to 13.1%; moderate ARI
Panel B: 14 Sensitivity Specifications	S3. Term frequency per 10,000 alphabetic words	51	13,275 windows	ARI = 0.183	Switching from binary window proportion to rate alters grouping
Panel B: 14 Sensitivity Specifications	S4. Excluding longest document (Unismuh Makassar; 1,299 windows)	50	11,976 windows	ARI = 0.699	Highest agreement among sensitivity runs, yet still non-identical
Panel B: 14 Sensitivity Specifications	S5. Excluding amended UNRI document	50	13,050 windows	ARI = 0.516	Moderate sensitivity to single amended document removal
Panel B: 14 Sensitivity Specifications	S6. Including Universitas Mulawarman candidate document	52	13,558 windows	ARI = 0.697	Moderate-high consistency when adding single candidate document
Panel B: 14 Sensitivity Specifications	S7. Including UIN Raden Fatah Palembang candidate document	52	13,593 windows	ARI = 0.697	Moderate-high consistency when adding single candidate document
Panel B: 14 Sensitivity Specifications	S8. Universities only (excluding 5 polytechnics and 1 institute)	45	11,921 windows	ARI = 0.253	High sensitivity to institutional form and comparator context
Panel B: 14 Sensitivity Specifications	S9. Public institutions (PTN) only (excluding 12 PTS)	39	10,442 windows	ARI = 0.634	Moderate consistency within public governance sector
Panel B: 14 Sensitivity Specifications	S10. Non-Java institutions only (excluding 25 Java institutions)	26	8,398 windows	ARI = 0.182	Strong dependence on regional sample composition
Panel B: 14 Sensitivity Specifications	S11. Excluding amended UNRI, UM, and UNSRAT documents	48	12,590 windows	ARI = 0.629	Moderate stability after removing all three amended documents
Panel B: 14 Sensitivity Specifications	S12. Approximate sentence units instead of 80-token blocks	51	38,336 sentences	ARI = 0.290	Substantial sensitivity to segmentation unit definition
Panel B: 14 Sensitivity Specifications	S13. Shorter text window size (40 space-delimited tokens)	51	24,419 windows	ARI = 0.451	Moderate sensitivity to halving window length
Panel B: 14 Sensitivity Specifications	S14. Longer text window size (120 space-delimited tokens)	51	9,470 windows	ARI = 0.091	Lowest agreement; wide windows blur theme co-occurrence boundaries

Note: In Panel B, Adjusted Rand Index (ARI) values compare partitions across the shared institutions with the cluster count fixed at k = 4 to isolate measurement and sample sensitivity from k-selection variance. ARI = Adjusted Rand Index; LOIO = leave-one-institution-out; PI = percentile interval; UNRI = Universitas Riau; UM = Universitas Negeri Malang; UNSRAT = Universitas Sam Ratulangi; UNMUL = Universitas Mulawarman.

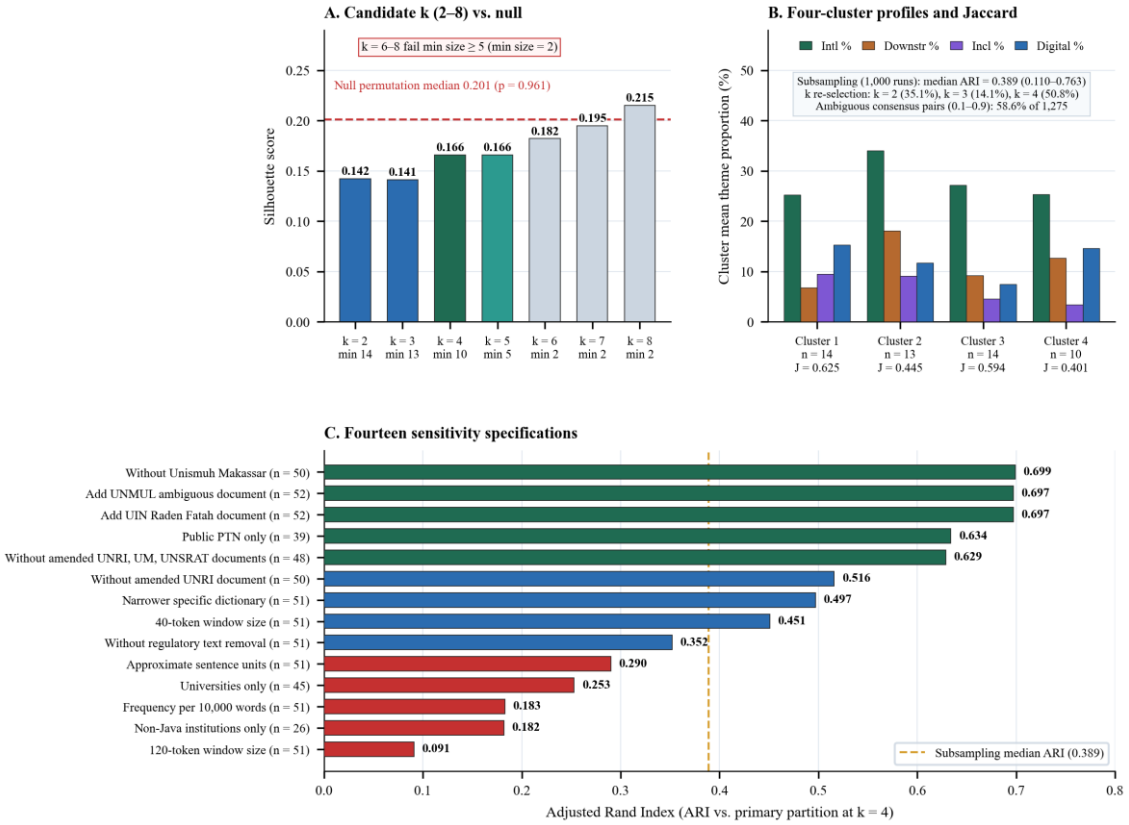


Figure 5. Exploratory Ward clustering evaluation, marginal feature-permutation null benchmarking, and 14 partition sensitivity analyses: (A) silhouette coefficients across candidate cluster counts k = 2 to 8 showing that the optimal primary partition at k = 4 (silhouette = 0.1663; cluster sizes 14, 13, 14, and 10) fails to exceed the 1,000-draw marginal feature-permutation null median (0.201; empirical p = 0.961) and that k = 6–8 fail the minimum five-member diagnostic criterion (minimum size = 2); (B) descriptive theme profiles across the four exploratory clusters alongside 1,000-draw 80% subsampling Jaccard recovery (0.625, 0.445, 0.594, and 0.401; overall median ARI = 0.389; k = 4 selected in 50.8% of runs; 58.6% ambiguous pairs); and (C) Adjusted Rand Index (ARI) across 14 deterministic sensitivity specifications at fixed k = 4 (spanning 0.091 for 120-token windows to 0.699 when excluding Unismuh Makassar).

Table 7. Exploratory subgroup comparisons across governance status (PTN vs. PTS) and geographic region (Java vs. Outside Java), alongside document-length rank correlations across the four strategic themes (N = 51).

Panel	Variable	Sample scope	Estimate (mean ± SD; median)	Test and effect metric	Interpretation
Panel A: Governance Status (PTN vs. PTS)	Internationalization prevalence	39 PTN vs. 12 PTS	PTN: 28.6% ± 10.1% (med 27.4%) vs. PTS: 26.1% ± 9.1% (med 25.1%)	Mann–Whitney U = 267.0, p = 0.4704	Comparable lexical internationalization attention across public and private sectors
Panel A: Governance Status (PTN vs. PTS)	Research Downstreaming prevalence	39 PTN vs. 12 PTS	PTN: 12.3% ± 5.6% (med 11.5%) vs. PTS: 8.8% ± 5.6% (med 6.5%)	Mann–Whitney U = 329.0, p = 0.0358	Higher nominal downstreaming salience in PTN under national innovation mandates
Panel A: Governance Status (PTN vs. PTS)	Inclusivity prevalence	39 PTN vs. 12 PTS	PTN: 7.6% ± 4.2% (med 7.3%) vs. PTS: 4.1% ± 2.2% (med 4.1%)	Mann–Whitney U = 351.0, p = 0.0097	Higher nominal inclusivity salience in PTN reflecting statutory scholarship/affirmation reporting
Panel A: Governance Status (PTN vs. PTS)	Digital Transformation prevalence	39 PTN vs. 12 PTS	PTN: 11.9% ± 5.1% (med 10.6%) vs. PTS: 12.6% ± 4.6% (med 11.7%)	Mann–Whitney U = 203.0, p = 0.4982	Comparable enterprise digital systems and ICT infrastructure attention across sectors
Panel B: Geographic Region (Java vs. Non-Java)	Internationalization prevalence	25 Java vs. 26 Non-Java	Java: 30.1% ± 12.1% (med 28.4%) vs. Non-Java: 26.0% ± 6.8% (med 25.8%)	Mann–Whitney U = 380.5, p = 0.3000	Wide dispersion within Java; non-significant regional difference
Panel B: Geographic Region (Java vs. Non-Java)	Research Downstreaming prevalence	25 Java vs. 26 Non-Java	Java: 11.7% ± 6.7% (med 10.4%) vs. Non-Java: 11.3% ± 4.7% (med 11.0%)	Mann–Whitney U = 315.5, p = 0.8653	Near-identical downstreaming lexical salience across Java and outer-island regions
Panel B: Geographic Region (Java vs. Non-Java)	Inclusivity prevalence	25 Java vs. 26 Non-Java	Java: 6.8% ± 4.5% (med 5.0%) vs. Non-Java: 6.8% ± 3.7% (med 6.1%)	Mann–Whitney U = 296.0, p = 0.5911	Identical mean inclusivity prevalence (6.8%) across regional groupings
Panel B: Geographic Region (Java vs. Non-Java)	Digital Transformation prevalence	25 Java vs. 26 Non-Java	Java: 13.0% ± 4.5% (med 11.8%) vs. Non-Java: 11.2% ± 5.3% (med 9.8%)	Mann–Whitney U = 420.0, p = 0.0750	Modest nominal trend toward higher digital lexical share in Java
Panel	**Correlated variables**	**Sample scope**	**Range or description**	**Spearman ρ (nominal p)**	**Interpretation**

Panel	Variable	Sample scope	Estimate (mean ± SD; median)	Test and effect metric	Interpretation
Panel C: Document-Length Coupling	Window count vs. Inclusivity proportion	51 institutions	Window range: 54 (UNIIB) to 1,299 (Unismuh Makassar); mean 260.3 ± 194.0	Spearman ρ = 0.426, nominal p = 0.0019	Denominator normalization does not fully eliminate document-length dependence
Panel C: Document-Length Coupling	Window count vs. Internationalization, Downstreaming, Digital	51 institutions	Intl: ρ = −0.082 (p = 0.5684); Downstr: ρ = 0.201 (p = 0.1577); Digital: ρ = −0.074 (p = 0.6060)	Non-significant rank correlations	Length coupling is concentrated specifically in Inclusivity rather than all themes

Note: All subgroup Mann–Whitney U tests and Spearman rank correlations are exploratory and descriptive conditional on the purposive sample of 51 accessible Renstra documents; nominal p-values are reported without claiming national population inference. PTN = public institutions; PTS = private institutions; SD = standard deviation; ρ = Spearman rank correlation coefficient; U = Mann–Whitney U statistic.

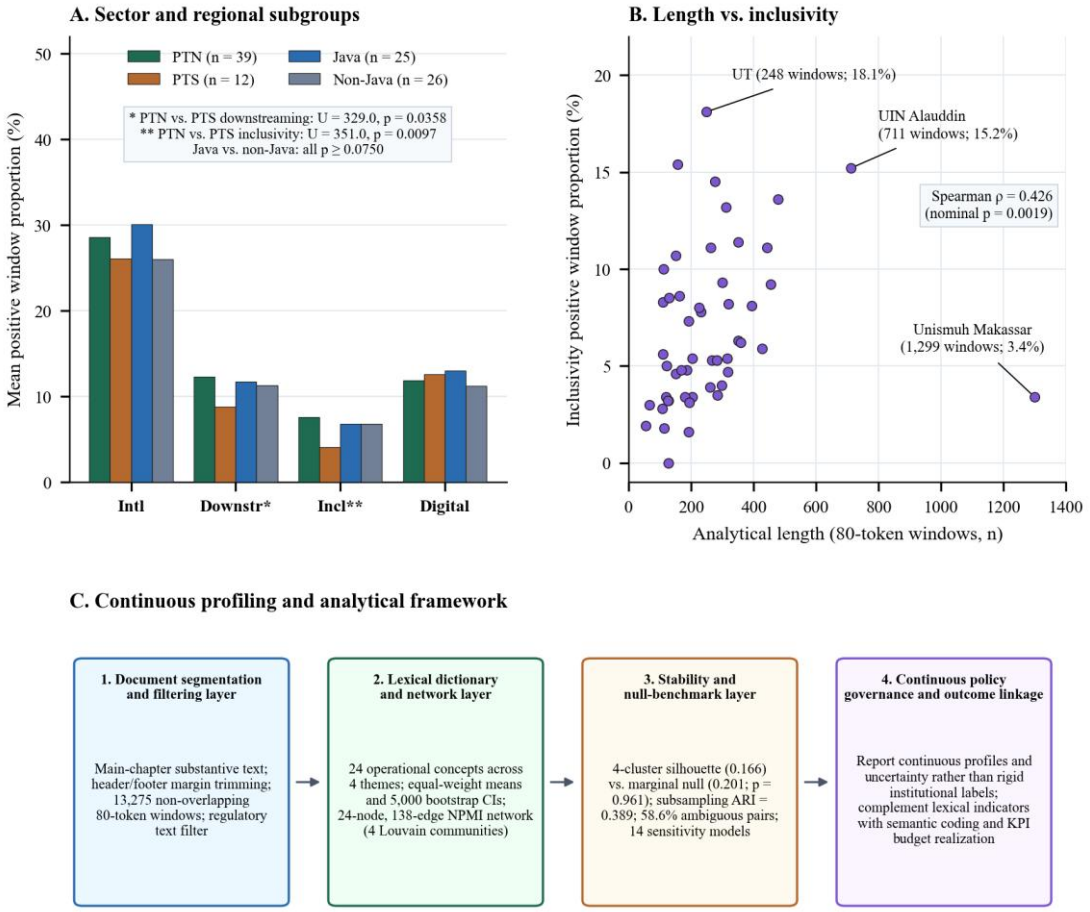


Figure 6. Exploratory subgroup comparisons, document-length correlation, and continuous text-as-data analytical architecture: (A) mean positive window proportions across public (PTN, n = 39) versus private (PTS, n = 12) institutions and Java (n = 25) versus non-Java (n = 26) institutions, showing nominally higher PTN attention in Inclusivity (7.6% vs. 4.1%; Mann–Whitney U = 351.0, p = 0.0097) and Research Downstreaming (12.3% vs. 8.8%; U = 329.0, p = 0.0358); (B) positive rank correlation between analytical document length (80-token windows) and Inclusivity proportion across the 51 institutions (Spearman ρ = 0.426, nominal p = 0.0019); and (C) four-layer continuous profiling and analytical framework linking text segmentation, lexical networks, null-model stability testing, and policy governance.

4. Discussion

4.1 Why Multi-Institutional Documentary Evidence Does Not Support Rigid Governance Typologies

Analyzing 51 diverse Indonesian higher education strategic plans (39 PTN and 12 PTS across six island regions) offers a clear methodological lesson for computational education policy research. Broad multi-institutional coverage strengthens the descriptive mapping of textual attention, and it also shows that strategic orientations vary along a continuous, overlapping spectrum and do not form discrete, stable institutional archetypes.[16,18–20,45] Reading the four-cluster Ward partition as validation of a four-type governance taxonomy would be unsupported by the data.[18,45] Two findings rule that reading out. First, a silhouette coefficient of 0.166 lies well below conventional thresholds for strong cluster separation. Second, and more decisively, it does not exceed the median silhouette of 0.201 that arises when the four theme features are permuted independently across institutions (p = 0.961; Table 5 and Figure 5A).

Resampling and sensitivity audits add a third line of evidence: cluster boundaries in this four-dimensional lexical space are highly permeable (Table 6 and Figure 5B–C). When 20% of institutions were withheld across 1,000 subsampling draws, median partition agreement fell to ARI = 0.389. The four-cluster solution was selected in barely half (50.8%) of iterations. In addition, 58.6% of all 1,275 institutional pairs had ambiguous co-clustering consensus probabilities between 0.1 and 0.9.[19] Changing the comparator frame has a similar effect. Restricting the sample to the 45 universities (ARI = 0.253) or to the 26 non-Java institutions (ARI = 0.182) reassigns many institutions. The pattern fits theories of institutional isomorphism and hybrid strategic planning. Indonesian universities and polytechnics draft their Renstra under shared national accreditation instruments (BAN-PT) and ministerial key performance indicators (IKU Perguruan Tinggi), so almost all of them articulate all four modernization themes at once (100.0% detection for internationalization, digital transformation, and downstreaming; 98.0% for inclusivity); they differ by continuous gradations of emphasis, not by sorting into mutually exclusive boxes.[21–24]

4.2 Interpreting Theme Salience, Concept Specificity, and Document-Length Coupling

The rank order of equal-weight institutional theme prevalence across the 51-institution corpus was clear. Internationalization led at 28.0%, followed by digital transformation (12.1%), research downstreaming (11.5%), and inclusivity (6.8%; Table 2 and Figure 2A). The concept-level and sensitivity analyses show why this order must not be read as a claim that Indonesian higher education institutions put four times more financial or administrative effort into internationalization than into inclusivity.[2,9,16,34–40] The word 'internasional' has a very broad lexical footprint in Indonesian academic planning prose. It captures global vision mottos, international accreditation targets, Scopus-indexed publication goals, inbound and outbound mobility, and global environmental scans (3,203 windows). Remove the generic occurrences and keep only specific operational mechanisms (such as collaborative degrees [73 windows] or foreign student mobility [193–232 windows]), and mean internationalization prevalence falls from 28.0% to 13.1% (ARI = 0.497; Table 2 and Table 6).

Inclusivity, in contrast, has the lowest lexical proportion (6.8%). It also correlates positively with analytical document length (Spearman ρ = 0.426, nominal p = 0.0019; Table 7 and Figure 6B). Short, condensed Renstra documents (such as UNIIB with 54 windows or UBK with 66 windows) concentrate on core vision statements and headline academic KPIs. Comprehensive plans that run to several hundred pages (such as UIN Alauddin Makassar with 711 windows or Universitas Terbuka with 248 windows) give extended narrative sections to student welfare, regional affirmation, tuition waivers, and disability services.[34–40] Many institutions may also handle student financial aid or regional access through operational tables or administrative terms that lexical roots do not fully capture. Normalizing by total windows does not automatically remove the link between document length and thematic breadth.

Public and private institutions differ too (PTN, n = 39; PTS, n = 12). Together with the positive correlation between document length and inclusivity prevalence (Spearman ρ = 0.426, nominal p = 0.0019; Table 7 and Figure 6A–B), these exploratory comparisons point to an interplay between statutory governance mandates and documentary genre.[34–40] PTN institutions are expected to report on national tuition support schemes (Kartu Indonesia Pintar Kuliah), regional affirmation admissions (ADik for Papua and 3T frontier regions), disability service units (Unit Layanan Disabilitas), and national research downstreaming roadmaps. That may raise their nominal lexical proportions in inclusivity (7.6% ± 4.2% vs. 4.1% ± 2.2%; p = 0.0097) and research downstreaming (12.3% ± 5.6% vs. 8.8% ± 5.6%; p = 0.0358). Private universities (PTS) often publish concise plans focused on institutional positioning, student enrollment, and digital learning infrastructure (12.6% ± 4.6% in digital transformation vs. 11.9% ± 5.1% in PTN; p = 0.4982), and they document scholarship schemes or community access in separate foundation decrees or operational manuals. Lower lexical inclusivity or downstreaming percentages in concise PTS Renstra documents therefore may reflect documentary scope and statutory reporting templates. The data do not measure institutional commitment.

4.3 Lexical Co-Occurrence Bridges and Institutional Form Context

The 24-node, 138-edge co-occurrence network (Table 4 and Figure 4) gives an auditable map of where agendas intersect within complex planning documents.[17,43,44] Several within-theme pairs have high NPMI: data governance with information security (0.529), affirmation/vulnerable groups with educational cost support (0.527) and disability (0.499), and downstreaming with commercialization (0.463). When institutions describe a specialized sub-agenda, then, they tend to place complementary policy instruments in the same 80-token window. The cross-theme bridge between international publication and intellectual property (130 windows in 44 institutions; NPMI = 0.425) is equally informative, as is the high betweenness of downstreaming (0.265) and disability (0.178). In Indonesian university Renstra matrices, research output sections often appear to bundle international journal targets with patent registration and industrial downstreaming indicators in the same performance paragraph.[8,13,33] Co-occurrence is still only textual proximity. Multi-column KPI tables or regulatory templates can produce it, so budget lines and program realization must be verified before any operational synergy is inferred.

Sensitivity model S8 restricted the corpus to the 45 universities by removing the five state polytechnics (PNJ, POLBAN, Poltekkes Banjarmasin, Poltekpar Bali, and POLTERA) and one technological institute (ITK). The ARI against the full 51-institution partition was low (0.253; Table 6). Vocational polytechnics and specialized technological institutes are expected to have mandates centered on applied industrial downstreaming (for example, PNJ at 22.0%, POLTERA at 21.7%, and POLBAN at 16.3%) and on technical systems, not on broad multi-faculty general education. Including them shifts the standardization mean and covariance structure of the corpus. Comparative policy audits therefore need to account for institutional mission and statutory form. Pooling heterogeneous institutions into unconditional rankings should be avoided.[1,3,4,12–14]

4.4 Methodological Limitations, Analytical Safeguards, and Governance Recommendations

This study has six explicit limitations, and they bound its scholarly and policy use (Figure 6C). First, the 51-institution corpus is a purposive sample conditioned on public web accessibility. PTN (76.5%) and Java-based institutions (49.0%) are over-represented relative to national institutional counts, and Papua (0.0%) is absent, so national prevalence cannot be inferred. Second, planning periods covering 2025 are not synchronized to one adoption date or implementation stage. Documents run from 2021–2025 to 2025–2029, and UNRI, UM, and UNSRAT include amended documents. Third, PDF layout extraction, multi-column spreads, embedded target matrices, and heuristic filtering of statutory text add unavoidable measurement noise, even with manual main-chapter delimitation.

Fourth, the 24-concept dictionary relies on automated lexical pattern matching and has not yet been paired with multi-rater human coding of pragmatic discourse stance. Positive window proportions therefore combine aspirational targets, historical achievements, and diagnostic deficit statements. Fifth, four thematic proportions are a low-dimensional projection that leaves out curriculum pedagogy, community service (pengabdian kepada masyarakat), human resource governance, and financial sustainability. Sixth, all bootstrap intervals, ARI metrics, Jaccard similarities, and subgroup Mann–Whitney and Spearman tests are internal, conditional diagnostics, not causal or external performance validations. Lexical profiles of this kind should not be used to sanction institutions, allocate ministerial block grants, build league tables, or allege policy mimicry. Higher education leaders and ministry analysts can use continuous thematic profiles, uncertainty intervals, and concept co-occurrence maps in a different way. As a transparent screening tool, they guide qualitative reading and checks on how strategic prose, measurable KPIs, and realized institutional budgets align.

Confirmatory policy evaluation needs a formal three-stage validation protocol.[16,17] Stage 1 is multi-rater semantic annotation. At least two independent domain experts, blinded to institutional identity and automated dictionary flags, would annotate a stratified random sample of 80-token windows on two orthogonal dimensions. The first is concept presence or absence for each of the 24 dictionary concepts, from which false-positive precision, false-negative recall, macro-F1, Cohen's κ, and Krippendorff's α are computed. The second is pragmatic

discourse stance, which classifies each matched window as (i) an aspirational target or policy commitment, (ii) a historical achievement or baseline description, (iii) a SWOT deficit, weakness, or constraint statement, or (iv) a tabular or regulatory template artifact. Stage 2 is sampling frame expansion. Purposive web collection should be complemented by direct inter-library and ministerial repository requests, so that institutions in Papua and under-represented private colleges (STIE, STIKES, and regional academies) are included. Stage 3 is implementation linkage. Continuous lexical profiles should be linked longitudinally to realized budget allocations (RKAKL/RKAT), accredited study program outcomes, and verified KPI achievements. Governance scholars could then test where strategic planning rhetoric turns into measurable institutional change.[2,3,17,25,45]

5. Conclusion

5.1 Summary of Findings and Future Validation Roadmap

Across 51 Indonesian higher education strategic plans (Renstra) covering 2025 (13,275 analytical 80-token windows; 923,823 alphabetic words), lexical attention is highest for internationalization (28.0%), followed by digital transformation (12.1%), research downstreaming (11.5%), and inclusivity (6.8%). The corpus shows structured concept co-occurrence across 24 concepts and 138 supported network edges. The clustering result is different. A Ward clustering into four groups (silhouette = 0.166) does not exceed marginal feature-permutation null expectations (median = 0.201; empirical p = 0.961), and it is highly sensitive to 80% institutional subsampling (median ARI = 0.389; k = 4 selected in 50.8% of runs; 58.6% ambiguous consensus pairs) and to 14 methodological perturbations (ARI = 0.091–0.699). Comparative higher education governance research should therefore report continuous multidimensional attention profiles with resampling uncertainty. Premature typological labels should be avoided. Future confirmatory work should add multi-rater human semantic annotation to separate aspirational goals from diagnostic deficits, widen regional representation to Eastern Indonesia including Papua, and link textual strategic orientations to realized institutional budgets and audited performance outcomes.

Declaration

Artificial Intelligence Use Statement

The authors did not use generative artificial intelligence tools to conceive the study, analyze the data, draft the text, or edit the manuscript. They reviewed the final text and take full responsibility for its accuracy and integrity.

Author Contribution Statement

Conceptualization, M.H.; methodology, M.H.; software, M.H.; validation, M.H. and G.O.S.; formal analysis, M.H.; investigation, M.H.; data curation, M.H.; writing – original draft, M.H.; writing – review and editing, M.H. and G.O.S.; visualization, M.H.; supervision, G.O.S.; project administration, M.H. All authors read and approved the final manuscript.

Funding Statement

The study was carried out without financial support from any public, commercial, or not-for-profit funding body.

Conflict of Interest / Competing Interests

The authors declare that they have no competing financial or non-financial interests related to this work.

Ethics Approval and Consent to Participate

Not applicable. This comparative documentary study analyzed only publicly accessible institutional strategic planning documents (Rencana Strategis/Renstra) published on official higher education web portals and public repositories. No human participants, student records, clinical samples, or private personal data were collected or analyzed, and no access controls were bypassed.

Consent for Publication

Not applicable. No identifiable individual images, private correspondence, or personal clinical records are included in this manuscript.

Data Availability Statement

All 51 strategic plans in the primary corpus, together with the two sensitivity-only documents and the one excluded pre-2025 document, were retrieved from publicly accessible institutional portals. Institutional identifiers, planning periods, analytical window counts, and thematic profiles are documented in Table 3. The analytical code, institutional summary datasets, and co-occurrence network files (Python 3.12.14, pandas 2.2.3, NumPy 2.5.2, SciPy 1.18.0, scikit-learn 1.9.0, NetworkX 3.6.1) can be obtained from the corresponding author on reasonable request.

Acknowledgments

The authors acknowledge the public information officers (PPID) and quality assurance units (LPM/BPM/LPMPP) of the Indonesian universities, polytechnics, and technological institutes whose publicly accessible strategic planning documents made this comparative documentary investigation possible.

References

1. Gaus N, Hall D. Performance indicators in Indonesian universities: the perception of academics. High Educ Q. 2016;70(2):127-144. doi:10.1111/hequ.12085
2. Graves BM, Erickson FJ. Structure and content of strategic plans of public higher education institutions in the United States. J High Educ Policy Manag. 2026;48(1):1-16. doi:10.1080/1360080x.2026.2626617
3. Aljuwaiber A. Implementation of a strategic plan within a university college: a practical framework. Perspect Policy Pract High Educ. 2023;28(4):173-182. doi:10.1080/13603108.2023.2255851
4. Morales Muñoz K, Alfaro Arriagada N, Pino Sandoval Y, et al. Strategic plan development: a methodological proposal for public higher education institutions. Sapiens High Educ. 2025;2(6):1-21. doi:10.71068/xqrdns61
5. Anggara S. Exploring the effectiveness of Merdeka Belajar Kampus Merdeka policy in Indonesian higher education institutions: an in-depth case study analysis. Al-Ishlah J Pendidik. 2023;15(2):1563-1570. doi:10.35445/alishlah.v15i2.3885
6. Pratiwi M, Kusumah LH. Enhancing the accreditation of Indonesian private universities through the integration of EduQual and accreditation standards of the BAN-PT. REiD (Res Eval Educ). 2024;10(2):227-243. doi:10.21831/reid.v10i2.76406
7. Yusriadi Y, Farida U, Misnawati M. Merdeka Belajar Kampus Merdeka policies in Indonesia higher education institutions: new public policy reform. Al-Tanzim J Manaj Pendidik Islam. 2024;8(1):344-360. doi:10.33650/al-tanzim.v8i1.7173
8. Hidayatullah N. Analyzing the synergy efforts between the education and industry sectors. Enigma Educ. 2023;1(2):50-54. doi:10.61996/edu.v1i2.29
9. Brilyanti FA. Strategy for internationalizing Indonesian higher education institutions to enhance global competitiveness. J Indones Sustain Dev Plan. 2025;6(2):281-296. doi:10.46456/jisdep.v6i2.693
10. Bulut-Sahin B, Emil S, Okur S, et al. Strategic management of internationalization in higher education institutions: the lens of international office professionals. Tert Educ Manag. 2023;29(4):429-445. doi:10.1007/s11233-023-09121-2
11. Armas KL, Jugo RR. Strategic development planning for global integration: a comprehensive framework for state universities in the Philippines. Int Rev Manag Mark. 2024;14(6):167-177. doi:10.32479/irmm.17220
12. Ralund S, Weiss F. The university sector has become more application-oriented: evidence from course descriptions using text analysis for the case of higher education in Denmark. High Educ Q. 2025;79(3):e70044. doi:10.1111/hequ.70044

13. Djoundourian S, Shahin W. Academia–business cooperation: a strategic plan for an innovative executive education program. Ind High Educ. 2022;36(6):835-845. doi:10.1177/09504222221083852
14. Sharma K, Arya V, Mathur HP. New higher education policy and strategic plan: commensurate India's higher education in global perspective. FIIB Bus Rev. 2023;12(3):23197145221125351. doi:10.1177/23197145221125351
15. Phan Q, Sperling J. Legislating a strategic plan: anti-2SLGBTQIA+ discourse and the political agenda reshaping higher education in Oklahoma. Educ Sci. 2025;15(7):851. doi:10.3390/educsci15070851
16. Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Polit Anal. 2013;21(3):267-297. doi:10.1093/pan/mps028
17. Malandrino A. Comparing qualitative and quantitative text analysis methods in combination with document-based social network analysis to understand policy networks. Qual Quant. 2023;58(3):2543-2570. doi:10.1007/s11135-023-01753-1
18. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65. doi:10.1016/0377-0427(87)90125-7
19. Hennig C. Cluster-wise assessment of cluster stability. Comput Stat Data Anal. 2007;52(1):258-271. doi:10.1016/j.csda.2006.11.025
20. Aktaş D, Lokman B, İnkaya T, et al. Cluster ensemble selection and consensus clustering: a multi-objective optimization approach. Eur J Oper Res. 2024;314(3):1065-1077. doi:10.1016/j.ejor.2023.10.029
21. Choi E, Kim YJ. Confusion in the marketplace: a study of institutional isomorphism and organisational identity in South Korea. Policy Rev High Educ. 2024;9(1):108-140. doi:10.1080/23322969.2024.2439463
22. Guil Gorostidi SdC, Rubio-Arostegui JA. Quality management in higher education from the perspective of institutional isomorphism: a scoping review. Front Educ. 2026;10:1720224. doi:10.3389/feduc.2025.1720224
23. Zdonek I, Zdonek D, Król K, et al. Sustainability-oriented higher education activities: insights from institutional isomorphism perspective. Sustainability. 2025;17(24):11034. doi:10.3390/su172411034
24. Li M. Institutional isomorphism and differentiation in a state-led glocal reform: a textual analysis of 27 implementation plans under South Korea's Glocal University 30. High Educ Policy. 2026;39(1):1-24. doi:10.1057/s41307-026-00492-w
25. Endawansa A, Juwono V. Analysis of the principles of good university governance at the University of Indonesia as a State University with Legal Entity (PTN-BH) in 2014–2019. J Borneo Adm. 2024;20(1):57-70. doi:10.24258/jba.v20i1.1295
26. Priyanto P, Suhandi NPM. Indonesian higher education institutions competitiveness and digital transformation initiatives. J Pendidik Indones. 2022;11(1):86-95. doi:10.23887/jpi-undiksha.v11i1.34263
27. Basalamah J. Technology readiness as a bridge between digital transformation and curriculum reform: evidence from Indonesia. J Appl Res High Educ. 2026;18(1):1-17. doi:10.1108/jarhe-06-2025-0471
28. Akbari TT, Pratomo RR. Higher education digital transformation implementation in Indonesia during the COVID-19 pandemic. J Kajian Komun. 2022;10(1):52-65. doi:10.24198/jkk.v10i1.38052
29. Sibawaihi S, Fernandes V. Globalizing higher education through internationalization and multiculturalism: the case of Indonesia. High Educ Q. 2022;77(2):232-245. doi:10.1111/hequ.12391
30. Simbolon NE. English medium instruction (EMI) practice: higher education internationalization in Indonesia. Englisia J Lang Educ Humanit. 2021;8(2):72-83. doi:10.22373/ej.v8i2.8961
31. El Junusi R, Agriyanto R, Khusnul Mubarok F, et al. Nexus knowledge-oriented leadership, innovation capability, and internationalization of higher education in Indonesia. J Appl Res High Educ. 2024;17(5):1612-1625. doi:10.1108/jarhe-02-2024-0092
32. Hidayadi T, Zuhdi ML, Yasin HR. The strategic economic and soft-power value of Indonesia–China educational cooperation. Policy Rev High Educ. 2026;10(1):1-24. doi:10.1080/23322969.2026.2705837
33. Ravi R, Janodia MD. Factors affecting technology transfer and commercialization of university research in India: a cross-sectional study. J Knowl Econ. 2021;13(1):787-803. doi:10.1007/s13132-021-00747-4
34. Afiefa NK, Ahmad M. Kartu Indonesia Pintar (KIP) scholarship: target reached or missed? Edunesia J Ilm Pendidik. 2023;4(3):1316-1331. doi:10.51276/edu.v4i3.505
35. Nguyen CK, Nguyen A. Disability support in higher education in Australia: a metasynthesis. Int J Incl Educ. 2025;29(12):1-22. doi:10.1080/13603116.2025.2567427
36. Spaan N, Van Trigt P, Schippers A. Integration of a disability lens as prerequisite for inclusive higher education. Eur J Incl Educ. 2024;3(1):1-24. doi:10.7146/ejie.v3i1.137271
37. Wanti M, Wesselink R, Biemans H, et al. Further insight into social factors affecting access to and equity for low-SES students in higher education. Eur J Incl Educ. 2024;3(1):61-82. doi:10.7146/ejie.v3i1.140295
38. Dewi EP, Razak A, Tetteng B, et al. Factors supporting the academic success of students with disabilities in inclusive universities. IJDS Indones J Disabil Stud. 2025;12(2):203-219. doi:10.21776/ub.ijds.2025.12.2.1
39. Riswari F, Puspitasari FH, Yuniarti N, et al. The management gaps toward inclusive education implementation at higher education in Indonesia. Indones J Disabil Stud. 2022;9(1):153-162. doi:10.21776/ub.ijds.2022.009.01.12
40. Rivai AB. Kebijakan afirmasi pendidikan tinggi untuk Papua. CosmoGov. 2017;1(2):266. doi:10.24198/cosmogov.v1i2.11838
41. Si L, Liu L, He Y. Scientific data management policy in China: a quantitative content analysis based on policy text. Aslib J Inf Manag. 2023;76(2):269-292. doi:10.1108/ajim-05-2022-0257
42. Fernández-Concha R, Lazarte-Aguirre A, Álvarez-Salazar J. Friends or foes: decoding OPM partnerships in higher education through text mining. High Educ Policy. 2025;38(3):1-22. doi:10.1057/s41307-025-00420-4
43. Bouma G. Normalized (pointwise) mutual information in collocation extraction. Proc GSCL. 2009:31-40. Available from: https://svn.spraakdata.gu.se/repos/gerlof/pub/www/Docs/npmi-pfd.pdf
44. Pan G. Lexical co-occurrence network and semantic relation mining based on English corpus. Appl Math Nonlinear Sci. 2025;10(1):20250209. doi:10.2478/amns-2025-0209
45. Sträng H, Dinh T. An upper bound on the silhouette evaluation metric for clustering. Pattern Recognit. 2026;178:113402. doi:10.1016/j.patcog.2026.113402