

Auditing Strategic-Renewal Disclosure in SEC 10-K Reports: A Traceable Text-Analytic Framework

Zixuan Yin*

Beijing University of Financial Technology, Beijing, China

**Corresponding author: Zixuan Yin.*

Abstract

An SEC 10-K often places discussion of new markets, capabilities, and business scope beside routine growth and accounting language. We constructed a traceable audit for the strategic-renewal signal in those disclosures. It combines a theory-guided vocabulary, a sentence-level context screen, and exact-passage repetition adjustment, while retaining the source passages used in every filing score. The corpus comprises 1,663 filings from 1,530 companies in the 2018–2020 source test partitions of EDGAR-CORPUS. We drew 320 passages for a company-disjoint audit, including 120 length-matched renewal–growth pairs, and had a blinded language model code them. Context screening increased precision from 0.2333 to 0.2714 but reduced recall from 0.5600 to 0.3800. The resulting F1 was 0.3167, below the dictionary trigger’s 0.3294. Renewal-cue passages exceeded growth-only passages by 0.0750 in concrete-initiative proportion, yet the 95% bootstrap interval, [-0.0250, 0.1833], included zero. The evidence points to a rule-design trade-off, not classifier superiority or a verified renewal measure. Component and parameter checks identify the filing scores most affected by processing choices. Pending independent expert coding and verified organizational events, the scores are exploratory disclosure signals.

Keywords

content analysis, corporate annual reports, SEC 10-K, strategic renewal disclosure, text analysis

1. Introduction

An SEC 10-K provides a longitudinal record of how a firm describes its markets, capabilities, and business scope. Those sections also carry routine growth statements, aspirations, and accounting discussion. A sentence that mentions change may describe an implemented initiative, a plan, or ordinary operations; the wording alone does not decide among them. Strategic renewal concerns changes in organizational attributes and in the processes that produce them [1, 2]. The measurement task is therefore to distinguish disclosure about those changes from language that only shares their vocabulary.

A computer-aided text analysis pipeline can make long filings comparable, but its credibility depends on how the construct is defined and checked in the application at hand [3, 4]. Financial dictionaries illustrate the problem. A word can take on a specialized meaning in business reporting [5]; “acquisition” may refer to an

accounting balance or to a strategic initiative. “Expansion” can describe either organizational change or routine growth. Annual-report content analysis has been used to study corporate strategy for decades, yet the inference from disclosed language to organizational change remains delicate [6, 7].

We define the target as strategic-renewal disclosure: a passage that describes a concrete initiative or change related to the renewal dimensions. It is narrower than realized renewal and is not treated as a validated outcome measure. The audit asks:

1. Does contextual screening improve identification of concrete strategic-renewal disclosures over a dictionary trigger?
2. Do renewal-cue passages distinguish concrete initiatives from generic growth language?
3. How sensitive are filing-level disclosure scores to alternative screening and repetition-adjustment choices?

The analysis combines a theory-guided vocabulary, sentence context, repetition adjustment, and a paired audit of SEC 10-K reports. It contributes an inspectable procedure and a boundary condition for interpretation. Context cues raise precision but cost recall; a positive paired contrast alone is insufficient evidence of specificity; and ranking stability can coexist with shared errors. Because the coding is model-assisted, the paper keeps the source passages and comparison rules visible and reserves claims about organizational change for future validation.

2. Related Work

We treat strategic renewal as change in strategic intent, capabilities, or organizational trajectories. Earlier research separates renewal from nearby forms of strategic change and identifies the field’s theoretical tensions [1, 2]. Work on cognitive change and strategic action further suggests that plans alone cannot establish implementation [7]. Disclosed annual-report text has long served as evidence for studying corporate strategy [6]. A recent manufacturing study uses exploration–exploitation terms in CEO letters and management discussion while keeping behavioral manifestations separate [8]. We follow that distinction. The practical question is whether a sentence records a concrete initiative when the same vocabulary can also describe ordinary growth, accounting for an acquisition, or an aspiration.

Computer-aided content analysis makes the link between a theoretical construct and a text indicator explicit. Prior methodological work sets out construct-validation procedures and identifies sources of measurement error [3, 9]. A related approach combines corpus statistics with close reading [10], while later work stresses that organizational narratives represent only selected aspects of behavior [11]. These contributions motivate source retention and failure-case reporting. They also caution that an algorithm can be stable while the construct remains under-specified.

Financial-text research supplies the domain context for our design. Evidence from 10-K analysis shows why general-purpose word lists misread financial language; subsequent work develops application-specific lexicons for that setting [5, 12]. Topic analysis has also been used to trace changes in 10-K disclosure content [13], while EDGAR-CORPUS provides reusable, sectioned filings [14]. The outcome here is narrower than forecasting or complexity: the question is whether a selected passage describes a concrete strategic initiative.

Representation and evaluation address different questions [15]. We retain the model responses and coding path so that repeatability within one model can be read separately from independent validation. The paired growth comparison probes one ambiguity in the reports, while the sensitivity analysis shows how alternative processing choices alter the result. Together, these checks leave a measurement starting point that can be inspected and extended.

Financial outcome terms can still select routine operations, and an initiative statement need not imply an economic effect. Sentence comparisons thus evaluate language selection, whereas document sensitivity evaluates what processing choices do to filing rankings. The manuscript reports those evaluations separately.

3. Methodology

For each filing, the pipeline records the passages that contribute to a disclosure score. Figure 1 summarizes corpus preparation, sentence screening, document aggregation, and the audit. Labels are applied after scoring, so they never feedback into a filing value.

3.1 Corpus and Operational Definition

The corpus is the 2018–2020 source test partitions of EDGAR-CORPUS, a public collection of extracted SEC filings [14]. We pinned and hashed the downloaded revision. Direct SEC body retrieval returned an access restriction; consequently, the analysis uses the corpus authors' extracted text rather than a new HTML extraction. A record enters the sample only when both Item 1, Business, and Item 7, Management's Discussion and Analysis, contain at least 300 words. This rule yields 1,663 filings from 1,530 companies out of 2,042 records. Corpus year denotes report year; exact filing dates were not independently verified. The cohorts form one corpus rather than a balanced longitudinal panel.

We organize the vocabulary into three dimensions: market and product change; capabilities and organization; and resource and business scope. The scheme is an operationalization rather than a validated scale. The first dimension covers new markets, products, business models, and diversification. The second covers transformation, restructuring, reorganization, reconfiguration, new technologies, and modernization. The third covers acquisitions, divestment, reallocation, spin-offs, and business exits. Frozen regular expressions capture inflected forms. A sentence can belong to more than one dimension, so dimension counts need not add to the aggregate.

3.2 Sentence Screening and Document Scores

We retain tokens with alphabetic characters, internal hyphens, or apostrophes. A punctuation-based splitter keeps sentences between 15 and 120 words and records the company identifier, report year, section, character offsets, and normalized-text hash. The dictionary trigger fires on any renewal expression. The context rule is stricter: the sentence must contain a focal-organization cue and an action cue within 12 tokens of a renewal match. Organization cues are we, our, company, business, firm, and group; action cues include launching, developing, investing, acquiring, reorganizing, implementing, planning, and their inflections. A match is discarded when negation appears in the preceding four tokens, although another qualifying match may still admit the sentence.

Within each filing, repeated normalized passages are collapsed. Passages appearing in at least 1% of eligible filings (17 distinct documents) are also removed. This exact-match adjustment targets repetition rather than semantic similarity. For filing i , let S_i denote the accepted sentence count after both checks and let W_i denote the total word count across the two retained sections. The disclosure score is

$$\text{Disclosure score} = \frac{S_i}{W_i} \times 1000 \quad (1)$$

Equation 1 reports accepted sentences per 1,000 section words. The denominator includes all section words, including text outside the candidate sentence-length range. Dimension scores use the same denominator. The resulting quantity is a rule-selected disclosure density, not an estimate of realized organizational renewal.

3.3 Reserved Audit and Comparisons

The audit deliberately enriches the sample for potentially confusing language. In each selected filing, one renewal-cue sentence is paired with the closest-length growth-only sentence available in that filing. Growth-only sentences contain growth, increase, improvement, expansion, opportunity, success, profitability, or related forms but no renewal cue. Pair members can come from different sections. Background sentences contain neither vocabulary. The pilot contains 24 pairs and 16 background passages; the reserved audit contains 120 pairs and 80 background passages. Companies are disjoint across splits and sampling units. The seed is 20260905, and rules, baselines, sampling, and coding instructions were frozen before labels were requested.

For the reserved labels, GPT-4o-mini-2024-07-18 assigns each shuffled passage one of four categories: concrete initiative, intention, generic discussion, or unrelated. Temperature is zero, batches contain 12 passages, and neither method scores nor vocabularies are shown. We call a passage concrete when it names a specific

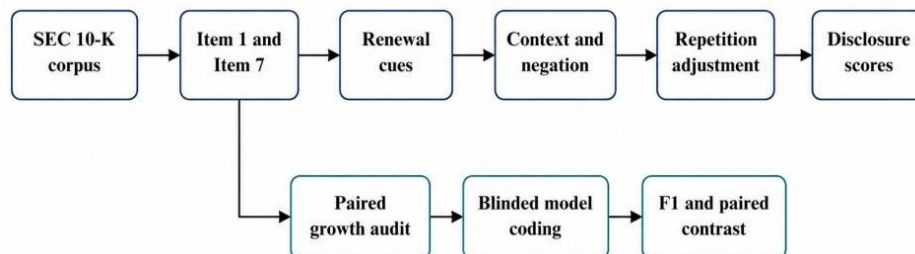
focal-firm change or definite strategic initiative; implementation itself need not be verified. Acquisition accounting without an initiative is generic, whereas an explicit acquisition plan may qualify. Only the concrete category is positive. We retain the raw requests and responses, and run a second pass on 48 reserved passages with the same model. This produces a model-assisted reference, not independent human validation.

The comparison includes the context rule, a raw dictionary, a length-normalized dictionary, TF-IDF similarity, a generic-growth rule, and an all-negative rule. The two dictionary variants use the same binary sentence trigger; normalization therefore changes document scores, not sentence-level classification. For TF-IDF, we remove English stop words, use sublinear term frequency and at most 5,000 features, and apply a fixed renewal-description query. The vocabulary is fitted on cue-bearing corpus sentences and the query without audit labels; cosine similarity of at least 0.10 marks a positive. Precision, recall, and F1 use the common reference, with no threshold tuned on the reserved audit.

The paired contrast subtracts the concrete-initiative proportion among growth-only passages from the proportion among renewal-cue passages. Percentile intervals use 2,000 bootstrap resamples of complete pairs; F1 intervals resample companies, keeping their passages together. Context-window sensitivity uses 6, 12, and 24 tokens, and repetition thresholds use 0.5%, 1%, and 2%. Separate removals test context, negation, and duplicate adjustment. Five bootstrap streams assess interval simulation stability without adding labels, firms, trained models, or independent validation studies. All metric-producing computations run on the remote AutoDL CPU backend, with environment versions and commands retained. Every one of the 384 sampled passage offsets is checked against the stored source sections.

The accompanying Supplementary Reproducibility Note lists the frozen vocabularies, regular expressions, negation set, matching order, sentence-boundary rule, and executable extraction logic. Splitting precedes whitespace normalization; offsets index original section strings. Duplicate hashes use SHA-256 on lowercased, whitespace-normalized UTF-8 passages. The note also gives the verbatim TF-IDF query and scikit-learn 1.7.2 parameters, including word unigrams, lowercasing, token pattern, document-frequency limits, and Euclidean normalization. The source manifest links filings to company identifiers, report years, download sources, and section hashes. These materials accompany the research files; no public deposit is claimed.

Figure 1: Disclosure measurement and audit workflow. Sentence labels evaluate screening decisions independently of document-score aggregation. Source: Authors' implemented framework



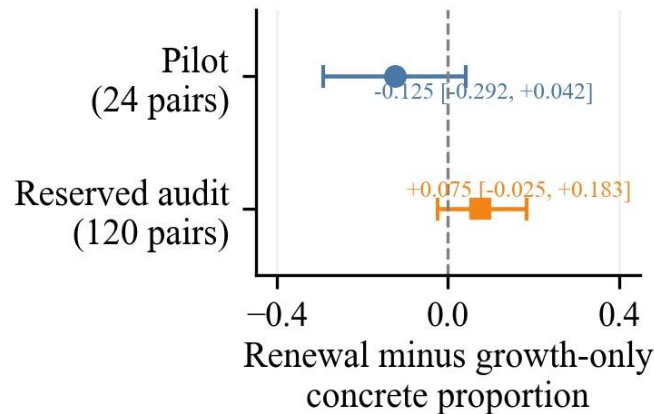
4. Results

The reserved-passage test addresses RQ1. Context screening is more selective than the dictionary trigger, but its F1 is lower. Table 1 gives precision, recall, and F1 for the same 320 passages. Screening raises precision while sacrificing recall; the two dictionary rows share a binary trigger, so normalization changes filing scores without changing sentence classification. The reference contains 50 concrete and 270 nonconcrete passages. The all-negative rule provides the zero-recall floor. The context rule finds 19 of the 50 positives, compared with 28 for the dictionary, and the added selectivity does not recover the missed initiatives in F1.

Table 1: Model-reference classification on 320 reserved passages. Source: Authors' remote audit

Method	Precision	Recall	F1
Dictionary count	0.2333	0.5600	0.3294
Normalized dictionary	0.2333	0.5600	0.3294
TF-IDF similarity	0.1538	0.0400	0.0635
Generic growth	0.1678	0.5000	0.2513
All-negative	0.0000	0.0000	0.0000
Context rule (ours)	0.2714	0.3800	0.3167

Figure 2: Paired contrasts with 95% bootstrap intervals; both include zero. Source: Authors' remote audit



The 120 reserved pairs used for RQ2 yield concrete-initiative proportions of 0.2333 for renewal-cue passages and 0.1583 for growth-only passages. The difference is 0.0750, with a 95% bootstrap interval of $[-0.0250, 0.1833]$; the interval crosses zero, so the paired audit does not establish specificity. The pilot contrast also remains inconclusive at pilot estimate (interval $[-0.292, 0.042]$). Both estimates appear in Figure 2. Their opposite signs make it safer to retain the uncertainty than to interpret the reserved estimate by itself.

For RQ3, removing context raises F1 to 0.3294, while removing negation leaves it at 0.3167. Duplicate adjustment changes the outcome space: removing it changes scores for 467 filings but retains a Spearman rank correlation of 0.9960 with the full score. Across prespecified settings and removals, document-score rank correlations range from 0.9060 to approximately 1, with the largest departure after context removal. These correlations describe numerical stability, not construct validity.

Section choice has a visible effect: independently normalized Item 1 and Item 7 scores correlate at 0.4877 by Spearman's rho across eligible filings. In the 48-passage repeat pass, category agreement is 0.7917, category kappa is 0.7043, and binary agreement is 0.8958. These figures describe repeatability for one coding model. Five bootstrap streams serve as a simulation check; they add no labels, firms, trained models, or independent validation studies.

The first four passages in the frozen random audit order illustrate individual decisions without selection by classifier outcome. They preserve the model's interpretations rather than adjudicating which decisions are correct. A 2019 Business passage discusses extending pulse oximetry to other patient-care settings. The coding model labels it concrete, while the renewal dictionary and context rule reject it and the growth trigger accepts it. A separate passage defining average restaurant sales receives a generic label and negative decisions from all comparators. Another lists capital expenditures and receives a generic label despite a positive growth trigger. The remaining passage concerns a reclamation-plan requirement and receives an unrelated label. These cases show how vocabulary membership and passage interpretation can diverge.

5. Discussion and Limitations

5.1 Main Findings and Implications

The audit yields three answers. Contextual screening changes which sentences are selected, but F1 falls even as precision moves from 0.2333 to 0.2714 and recall moves from 0.5600 to 0.3800. The reserved paired contrast is positive yet uncertain, so renewal vocabulary alone does not consistently separate concrete initiatives from generic growth language. Filing rankings are numerically stable under most parameter settings, with the largest change occurring when context is removed. Such stability speaks to repeatability of the procedure; it does not show that the intended construct has been measured.

This distinction gives the paper a practical methodological contribution. Transparency, robustness, and construct validity are separate properties of a text-based strategic measure. Source-linked scores let a researcher inspect how vocabulary, context, and repetition choices produce a filing value. The component and parameter checks identify which choices deserve attention in later validation. Used with an explicit disclosure

label, the protocol offers a reproducible starting point for strategic-management research and a way to report uncertainty alongside aggregate results.

The error cases point to specific refinements. Organization and action cues sometimes admit routine corporate language, while a valid initiative may lack the cue pattern used here. Acquisition accounting and definite plans remain boundary cases for the coding reference. Shared boilerplate can keep ranking errors in place when different rules select the same statements. These are explanations to test in future work, not causal interpretations of the current metrics.

5.2 Limitations and Future Validation

The audit deliberately stress-tests confusing language. Its precision, recall, and F1 therefore describe the selected passages, not population accuracy. Length matching controls one observable feature within a filing but leaves topic, section, and rhetorical purpose uncontrolled. A growth-only passage may contain a genuine initiative without a renewal cue, so it is a diagnostic comparison rather than a guaranteed negative. The design is informative about ambiguity, provided that its interpretation stays local.

The coding reference is model-assisted. Blinding prevents the model from seeing method scores or vocabularies, and the repeat pass indicates how consistently that model applies the categories. It does not replace independent expert judgment. A concrete label can also describe a planned initiative whose implementation remains unverified. The next practical step is double coding of a reserved sample, adjudication of disagreements, and reporting human–human agreement before stronger validity claims are made.

The corpus covers one pre-extracted collection, two sections, and three report-year cohorts. That scope is sufficient for a controlled audit, but it does not establish cross-corpus generalization or historical industry comparability. Exact repetition removal may discard substantive statements shared across reports while retaining paraphrased boilerplate, and the length filter and denominator operate on different text sets. These choices are fully specified so that a second SEC-derived corpus can test their portability.

A useful validation sequence follows directly from the present evidence: independent expert coding with explicit distinctions among intention, initiative, and completed change; linkage to verified renewal events; and an external corpus test of the frozen procedure. Until those checks are completed, scores should be read as exploratory disclosure signals. That qualification preserves the practical value of the audit while keeping the inference about realized strategic renewal open for empirical testing.

6. Conclusion

This paper gives a traceable audit for strategic-renewal disclosure in SEC 10-K reports. It combines a theory-guided vocabulary, contextual screening, repetition adjustment, retained source passages, and explicit comparison rules. In the audit, context screening trades recall for precision; the renewal–growth contrast remains uncertain; and numerical stability does not resolve construct validity. The resulting scores are best used as rule-selected disclosure signals. Independent expert coding, verified organizational events, and tests on an external corpus are the steps needed to develop the protocol into a more defensible indicator of strategic change.

References

- [1] Agarwal, Rajshree, and Constance E. Helfat. 2009. “Strategic Renewal of Organizations.” *Organization Science* 20 (2): 281–93. <https://doi.org/10.1287/orsc.1090.0423>.
- [2] Schmitt, Achim, Sebastian Raisch, and Henk W. Volberda. 2018. “Strategic Renewal: Past Research, Theoretical Tensions and Future Challenges.” *International Journal of Management Reviews* 20 (1): 81–98. <https://doi.org/10.1111/ijmr.12117>.
- [3] Short, Jeremy C., J. Christian Broberg, Claudia C. Coglisier, and Keith H. Brigham. 2010. “Construct Validation Using Computer-Aided Text Analysis (CATA).” *Organizational Research Methods* 13 (2): 320–47. <https://doi.org/10.1177/1094428109335949>.

- [4] Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. <https://doi.org/10.1093/pan/mps028>.
- [5] Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1): 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.
- [6] Bowman, Edward H. 1984. "Content Analysis of Annual Reports for Corporate Strategy and Performance." *Academy of Management Journal* 27 (1): 61–73. <https://doi.org/10.2307/255958>.
- [7] Barr, Pamela S., J. L. Stimpert, and Anne Sigismund Huff. 1992. "Cognitive Change, Strategic Action, and Organizational Renewal." *Strategic Management Journal* 13 (S1): 15–36. <https://doi.org/10.1002/smj.4250131004>.
- [8] Maharani, Ida Ayu Kartika, Badri Munir Sukoco, Indrianawati Usman, and David Ahlstrom. 2025. "Balancing Exploration-Exploitation Renewal and Indonesian Firm Performance: Do CEO Power and TMT Diversity Matter?" *Humanities and Social Sciences Communications* 12 (1): 1872. <https://doi.org/10.1057/s41599-025-06139-2>.
- [9] McKenny, Aaron F., Herman Aguinis, Jeremy C. Short, and Aaron H. Anglin. 2018. "What Doesn't Get Measured Does Exist: Improving the Accuracy of Computer-Aided Text Analysis." *Journal of Management* 44 (7): 2909–33. <https://doi.org/10.1177/0149206316657594>.
- [10] Pollach, Irene. 2012. "Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis." *Organizational Research Methods* 15 (2): 263–87. <https://doi.org/10.1177/1094428111417451>.
- [11] Short, Jeremy C., Aaron F. McKenny, and Shane W. Reid. 2018. "More Than Words? Computer-Aided Text Analysis in Organizational Behavior and Psychology Research." *Annual Review of Organizational Psychology and Organizational Behavior* 5 (1): 415–35. <https://doi.org/10.1146/annurev-orgpsych-032117-104622>.
- [12] Loughran, Tim, and Bill McDonald. 2024. "Measuring Firm Complexity." *Journal of Financial and Quantitative Analysis* 59 (6): 2487–514. <https://doi.org/10.1017/s0022109023000716>.
- [13] Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence. 2017. "The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation." *Journal of Accounting and Economics* 64 (2-3): 221–45. <https://doi.org/10.1016/j.jacceco.2017.07.002>.
- [14] Loukas, Lefteris, Manos Fergadiotis, Ion Androutsopoulos, and Prodromos Malakasiotis. 2021. "EDGAR-CORPUS: Billions of Tokens Make the World Go Round." *Proceedings of the Third Workshop on Economics and Natural Language Processing*, 13–18. <https://doi.org/10.18653/v1/2021.econlp-1.2>.
- [15] Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57 (3): 535–74. <https://doi.org/10.1257/jel.20181020>.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

