# Semantic Trace Topics: Text-Based Encodings for Interpretable Multi-Faceted Process Exploration

G. Andrienko[†1] and N. Andrienko[1] and M. Resinas[2] and S. van den Elzen[3] and B. Weber[4]

[1] Fraunhofer Institute IAIS, Germany, and City St George's University of London, UK
[2] SCORE Lab, Universidad de Sevilla, Spain
[3] Eindhoven University of Technology, The Netherlands
[4] University os St. Gallen, Switzerland

**Abstract**

*Trace encodings are central to process mining, as they determine what aspects of an event log can be explored and how results can be interpreted. Existing approaches to multi-faceted exploration often rely on high-dimensional boolean feature vectors over events, transitions, and attributes, which suffer from the curse of dimensionality and hinder semantic interpretation. We propose to represent event sequences as text documents and apply non-negative matrix factorization (NMF) topic modeling to obtain low-dimensional, interpretable trace features. A flexible, domain-specific vocabulary captures events, n-grams, and discretized event attributes, yielding topics that describe recurrent behavioral patterns as weighted sets of human-readable terms. Each trace is embedded as a topic-weight vector that supports classification, clustering, and dimensionality reduction. On a real-world traffic fines log, topic-based representations reveal meaningful process variants and their relations to outcomes such as success, duration, and cost.*

**CCS Concepts**

• ***Human-centered computing*** → *Visual Analytics;* • ***Applied computing*** → *Business process monitoring;*

## 1. Introduction and Motivation

Process exploration is a core activity in process mining [Sv08, vdA16]. Practitioners seek to familiarize themselves with the data, gain an initial understanding of the process, formulate or refine questions, and uncover unexpected patterns or problems. This exploratory phase is iterative and strongly driven by hypothesis generation and sense-making [MCSW24]. In practice, exploration is frequently centered around process maps such as Directly-Follows Graphs (DFGs), where nodes represent activities and edges encode temporal succession. These visualizations privilege control-flow as the primary facet [vJM*25]: other facets are only overlaid, which constrains the questions that can be answered and makes it hard to switch perspectives.

Previous work on multi-faceted process exploration has addressed part of this limitation by introducing coordinated projections and multiple encodings of event sequences (traces) [vdEAK*25]. Traces are mapped to high-dimensional feature spaces (e.g., control-flow features or outcome-related attributes), reduced to two dimensions using non-linear dimensionality reduction (DR) [EMK*21, VPvM25], and visualized in coordinated views [Rob07] with brushing-and-linking. This enables analysts to jointly explore, for instance, clusters of similar control-flow behavior and their corresponding distributions of outcomes or other attributes. However, these approaches typically rely on generic trace encodings in the form of high-dimensional boolean vectors that indicate the presence of activities, transitions, or categorical event-level attributes at all positions in the trace. This design has several drawbacks:

- **No latent structure.** Binary indicators treat each feature independently, missing co-occurrence patterns that characterize coherent process behaviors.
- **Curse of dimensionality.** As the number of possible activities, transitions, attribute values, and their combinations grows, the resulting feature spaces become extremely sparse and high-dimensional, complicating similarity computation [AHK01].
- **Limited interpretability.** Clusters in projection plots are defined by complex combinations of binary features that are difficult to explain in terms of meaningful process concepts.

We propose an alternative trace representation strategy that addresses these limitations by treating event sequences as text

† This work was done with support of Lamarr Institute for Machine Learning and Artificial Intelligence. This paper is a result of the discussions that took place at the Dagstuhl Seminar 25152, "Multi-Faceted Visual Process Mining and Analytics".

and applying topic modeling. We construct a domain-specific vocabulary—in its simplest form activities and *n*-grams, extensible to attribute-augmented and temporal tokens—and apply non-negative matrix factorization (NMF) [LNC*17] to discover topics: latent, recurring combinations of terms that define coherent behavioral patterns. Each trace is then compactly represented as a vector of topic weights. This has three key advantages:

- **Latent pattern discovery.** Even with a simple vocabulary, topic modeling uncovers co-occurrence structure, grouping related terms into interpretable behavioral archetypes not visible in individual features.
- **Compact, interpretable representations.** Topic vectors are low-dimensional and directly interpretable because each dimension corresponds to an understandable topic rather than to a single sparse feature.
- **Better support for exploration.** Topic-weighted trace embeddings can be used for classification, clustering, and DR in a smaller, more meaningful space, improving both visual overview and explainability.

We outline a human-centered trust-preserving workflow [AAAW22, vdEAA*23] that integrates the proposed encoding and topic modeling approach into the larger paradigm of multi-faceted process exploration [ZSW21] and knowledge-assisted process mining [SAD*26] and illustrate it using the Road Traffic Fines event log [dLM15].

## 2. Background and Related Work

### 2.1. Trace encodings in process mining

Trace encoding—the transformation of event sequences into feature representations—is a prerequisite for many process mining tasks, including clustering, anomaly detection, classification, and predictive monitoring [ZRSH20, TOBC23, RAS25]. Existing approaches can be grouped into three broad categories.

**Control-flow encodings.** These capture the order and structure of activities, often using *n*-grams, bags of events, or transition abstractions [BvdA09]. For example, traces may be represented as sets of activities, sets of directly-follows pairs, or more sophisticated abstractions such as patterns of concurrency and choice.

**Data-aware encodings.** These augment control-flow information with case- and event-level attributes, such as monetary amounts, durations, or resource identifiers [JRSW19]. Attributes may be included directly, discretized into bins, or aggregated over the trace. Time-related information, such as inter-event times or throughput times, is often represented as additional features.

**Embedding-based encodings.** These leverage machine learning models (e.g., LSTMs, Transformers, autoencoders) [DKvBDW18] to learn dense, fixed-length vector representations from raw sequences [CFDS24]. Such methods often behave as black boxes and require substantial data, parameter tuning, and pre-processing.

Surveys and benchmarks of trace encodings focus mainly on performance in downstream tasks such as classification or clustering [BvdA09, JRSW19]. Interpretability is discussed far less, especially in combination with interactive visualization.

Our work is positioned between these worlds. Like simple control-flow and data-aware encodings, we build on discrete, human-readable elements (activities, transitions, attributes) and maintain interpretability. However, instead of treating each element as an independent feature, we apply topic modeling to discover higher-level behavioral patterns and use them as a compact basis for exploration.

### 2.2. Topic modeling and visual analytics for processes

Topic modeling has been widely used in text analysis to uncover latent themes in document collections [VK20]. In the context of processes, topic modeling can be applied after transforming traces into textual documents, for instance by mapping activities or transitions to tokens. Topics then correspond to frequently co-occurring subsets of activities or subsequences and can be used to define process variants or behavioral patterns.

Building on our previous use of topic modeling with DR and coordinated views [vdEAK*25] and established VA principles [EMK*21], we now focus on richer, semantics-preserving trace encodings beyond simple event/transition sets. We retain these interaction principles as the foundation of our framework, but extend them with semantics-preserving, topic-based encodings and explicit topic-centered views that bridge semantic modeling and process behavior. Topic–term weights are mapped onto the event/2-gram graph to reveal each topic's semantic signature, and cases grouped by dominant topic are summarized in DFG-like graphs with frequency-encoded nodes and edges and compact outcome distributions.

## 3. Method: Text-Based Encodings and NMF Topic Modeling

Our analytical workflow comprises five main stages: (1) data and case preparation, (2) sequence encoding as text, (3) topic modeling with an NMF ensemble, (4) topic interpretation and trace representation, and (5) pattern analysis and validation (see Fig. 1).

### 3.1. Sequence encoding as text

Given an event log segmented into cases (traces), we first transform each trace into a textual document. The central design choice is the construction of a *vocabulary* of terms that reflect the aspects of process behavior we deem relevant. We support the following types of terms, which can be freely combined:

- **Activity tokens.** Each activity is represented as a token, preserving the original order within the trace (e.g., `createFine`, `sendFine`).
- ***n*-gram tokens.** Local patterns of successive activities are encoded as *n*-grams (e.g., 2-grams (or transitions) `createFine→sendFine`, 3-grams `sendFine→insertNotification→addPenalty`), capturing frequent subsequences and local structure.
- **Lagged-follow tokens.** Pairs of activities are encoded as tokens when activity B follows activity A within a bounded gap, expressed either as a maximum number of intermediate events between the two or as a time window. For example, `A⇒B_max3steps` denotes that B occurs after A with at most

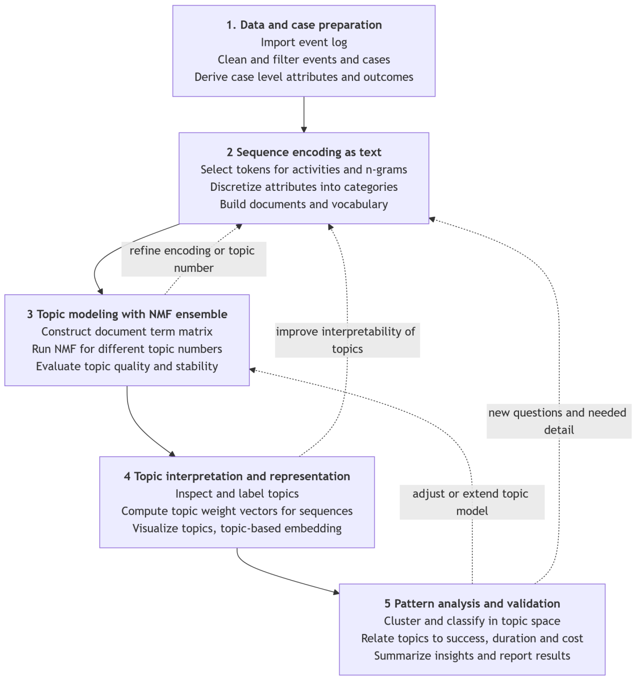


**Figure 1:** *Human-centered analysis workflow.*

three intermediate events, and `A⇒B_max2days` denotes that B occurs within two days after A. These tokens capture longer-range dependencies that are not visible in local *n*-grams.

- **Attribute-augmented tokens.** Event attributes such as duration, waiting time, or amount are discretized into categorical bins (e.g., `short`, `medium`, `long`; `low`, `high`) and attached to event names. For example, a payment event might yield tokens such as `Payment_highAmount` or `Payment_late`.
- ***n*-gram tokens with attributes.** *n*-grams can also be augmented with attributes of one or more constituent events, enabling patterns such as "long waiting time followed by a dismissal".
- **Constraint tokens.** Tokens derived from declarative process constraints (e.g., `Declare` templates [DCM22]) encode relations between activities within a trace such as precedence, response, or co-existence.
- **Composite vocabularies.** In practice, analysts often wish to consider multiple of these aspects simultaneously.

For each trace, we instantiate the vocabulary as follows: (1) extract the sequence of activities and their attributes; (2) generate tokens according to the chosen vocabulary definition; (3) concatenate tokens into a "document". From these documents we build a document–term matrix using standard text mining techniques such as term frequency (TF) or TF–IDF weighting.

Compared to previous boolean encodings, this approach preserves more semantics (order via *n*-grams, attribute context via augmented tokens) while remaining transparent.

### 3.2. Ensemble NMF topic modeling

We apply non-negative matrix factorization (NMF) [LNC*17] to the document–term matrix to obtain an interpretable, low-dimensional representation. NMF factorizes the non-negative matrix $V$ (documents $\times$ terms) into two non-negative matrices $W$ (documents $\times$ topics) and $H$ (topics $\times$ terms), such that $V \approx WH$. Each row of $H$ describes a topic as a sparse combination of weighted terms, and each row of $W$ provides the topic weights for a document (trace).

A central modeling decision is the choice of the number of topics $K$. To avoid ad hoc choices and increase robustness, we adopt an ensemble strategy [CAA*20]: we define a plausible range for $K$, run multiple NMF initializations for each candidate value, evaluate coherence, sparsity, stability, and DR-based separation [AAH24, MAAS25], and then select the model that best balances these criteria. This procedure reduces the risk of over- or under-segmenting the behavioral space and aligns the topic structure with the needs of exploratory analysis.

### 3.3. Topic interpretation and trace representation

Once an NMF model has been selected, we interpret the topics and derive trace-level representations.

**Topic interpretation.** For each topic, we inspect the highest-weight terms in $H$. Because terms are meaningful tokens (activities, *n*-grams, attribute-augmented activities), domain experts can often readily label topics as, for example, "standard payment flow", "appeal followed by dismissal", or "escalation to credit collection with high costs". Visualizations such as weighted bar charts or term–term graphs can support this interpretation.

**Trace embeddings.** Each trace is represented by its row in $W$, i.e., a vector of topic weights. This embedding is compact (dimensionality = number of topics) and has an interpretable basis: each dimension corresponds to a known behavioral pattern.

**Derived labeling and grouping.** Based on the topic weight vectors, we can (i) assign traces to *dominant topics* (e.g., the topic with the highest weight), (ii) cluster traces in topic space, and (iii) use DR (e.g., UMAP) on the topic vectors to utilize 2D scatterplots. Compared to working directly in the original high-dimensional boolean space, these operations are more stable (due to lower dimensionality) and easier to interpret (due to topic semantics).

## 4. Use Case: Road Traffic Fines

To illustrate the approach, we apply it to the Road Traffic Fines event log [dLM15], previously used in studies of multi-faceted process exploration, including our earlier study [vdEAK*25]. The log records the handling of traffic fines by an Italian police force and contains over 150,000 traces consisting of about 560,000 events of 11 distinct activities, along with attributes such as fine amount, expenses, payment amount, and dismissal codes.

### 4.1. Data preparation and encoding

We first transform the event log into a case log and enrich it with case-level outcomes, distinguishing, for example, fully paid, dismissed, credit collected, and unresolved cases [VZSW23]. Additional derived attributes include the outstanding balance and indicators for whether the fine was appealed.

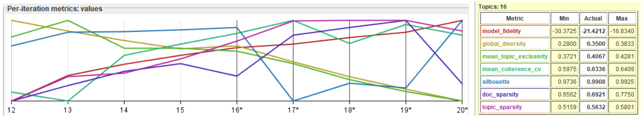


**Figure 2:** *Metrics of topics.*

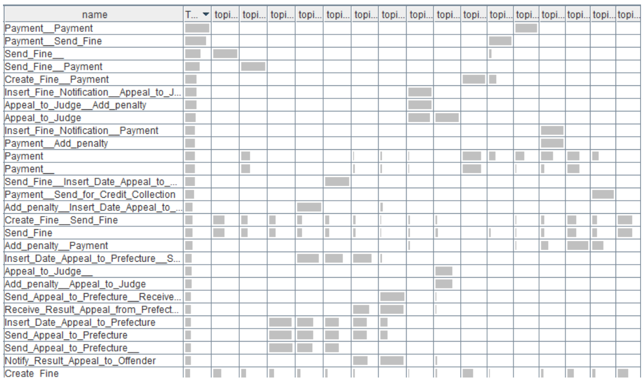


**Figure 3:** *Most important terms and their weights for 16 topics.*

Although the framework supports richer vocabularies, here we restrict ourselves to activity and 2-gram tokens to keep the vocabulary compact and visualizations readable; richer encodings are left for future work.

For each case, we generate a "document" consisting of its activity tokens and 2-gram tokens and construct a TF–IDF-weighted document–term matrix. We apply the ensemble NMF procedure from Section 3.2 for a range of K, summarizing reconstruction quality, sparsity, stability, and DR-based separation (Fig. 2, left). We selected $K = 16$ as a good compromise between stability, separation, and interpretability: the indicators show no clear improvement for larger $K$, and 2D projections exhibit well-separated topic regions with limited overlap. The right part of Fig. 2 summarizes the indicators for the chosen solution with 16 topics.

### 4.2. Topics and exploratory analysis

We next analyze the semantics of the 16 topics and their relation to cases and outcomes. The factor matrix $H$ (topics $\times$ terms) provides a weight for each term in each topic. To support interpretation, we visualize these weights as a term–topic matrix (Fig. 3). The figure has 16 columns corresponding to the topics; rows correspond to terms and are ordered by the maximum weight that a term attains in any topic. This ordering emphasizes terms that strongly characterize at least one topic. Complementarily, we also project topic-term weights onto graphs (Fig. 4, left), where nodes represent activities and edges represent 2-grams. Node sizes and edge widths encode term weights per topic. Taken together, the matrix and the weighted graphs show which semantic features—individual activities and short subsequences—define each topic.

The topics capture recurring behavioral patterns such as standard execution with timely payment before penalty, escalation paths with credit collection and high outstanding balances, appeal trajectories leading to dismissal, and rarer deviant sequences with unusual combinations of activities.

Each case is represented by its topic-weight vector (row of $W$). We assign each case to its *dominant topic* (highest weight), partitioning the 150,000 cases into 16 classes. Figure 5 shows case counts and outcome proportions (fully paid vs. dismissed vs. credit collected) per dominant topic. The inset displays a UMAP [MHSG18] embedding of all cases in topic space, colored by outcome. Clearly separated regions correspond to major topic classes; mixed regions—where cases have balanced topic mixtures—may indicate borderline behaviors or alternative paths leading to similar results.

To relate topics back to familiar control-flow views, we aggregate the cases per dominant topic and derive topic-specific graphs (Fig. 4, right). Node size now encodes how often an activity occurs in the cases assigned to a topic, and edge width encodes the frequency of the corresponding 2-gram in those cases. Small insets next to each graph show again the number of cases and the distribution of outcomes for the respective topic. These graphs highlight the most frequent activities and transitions within each topic class and make it easy to compare dominant paths across topics.

Interestingly, the term-weight graphs (Fig. 4, left) and aggregated case graphs (Fig. 4, right) differ systematically: frequent activities may receive low NMF weights because they also appear in other topics and are not discriminative, while rare but distinctive edges act as "signature" features. Thus the weight-based view emphasizes what makes a topic distinct, while the aggregated-case view reflects what is typical for cases dominated by that topic.

When analysts identify an interesting topic or region in the UMAP view, they can move back and forth between these representations: from the case embedding to dominant topics, from topics to their defining terms, and from terms to topic-specific process graphs and outcomes. This tight coupling of interpretable trace embeddings and established process mining views supports both explanation and hypothesis generation.

### 4.3. Findings

Applied to a real-world traffic fines log, the learned topics can be grouped into five higher-level distinct and semantically coherent behavioral regimes based on their dominant activities and outcome distributions. We identify a Direct Compliance regime, characterized by linear traces like `Create Fine` $\rightarrow$ `Payment` (Topics 1, 5) with minimal branching and a dominant `paid_full` outcome. In contrast, an Escalated Enforcement regime exhibits additional penalty and credit-collection events and is strongly associated with `credit_collection` and `unresolved` outcomes (Topics 0, 8), reflecting structural divergence from voluntary payment toward institutional recovery. A separate Appeal / Contest regime is marked by appeal-related events and shows an increased proportion of dismissed cases, distinguishing it structurally and outcome-wise from enforcement-driven variants (Topics 4, 9, 11-15). We further observe Delayed or Partial Resolution topics, in which payment transitions occur only after additional procedural steps (Topics 3, 6, 7, 10). These topics exhibit more heterogeneous outcomes, including both `paid_full` and `unresolved` cases, indicating late or incomplete compliance rather than full escalation. Finally, we identify an Early Incomplete regime, consisting of short traces (`Create Fine` $\rightarrow$ `Send Fine`) with exclusively

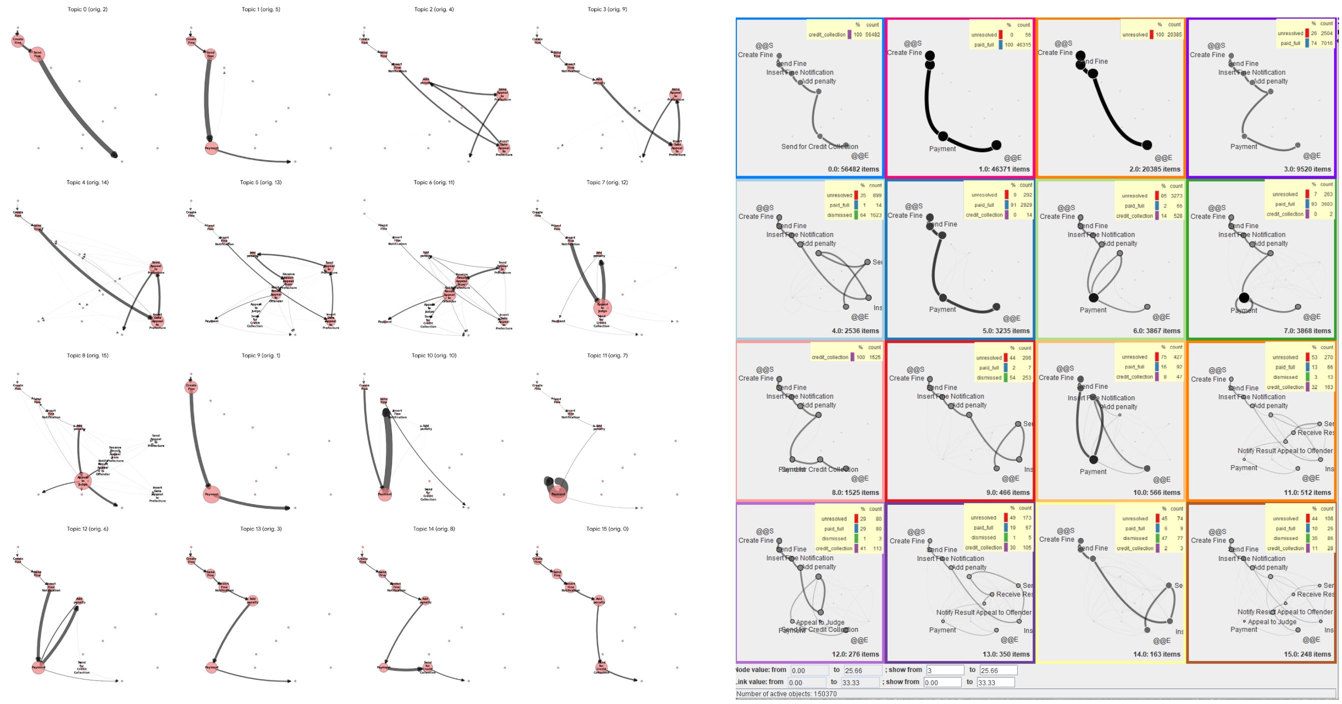

**Figure 4:** *Topic views by topic semantics and case aggregation. Left: topic–term weights mapped onto the activity/2-gram graph, highlighting semantic signatures of each topic. Right: DFG-like graphs aggregated over cases by dominant topic, with node sizes and edge widths proportional to observed frequencies; insets show case counts and outcome distributions.*

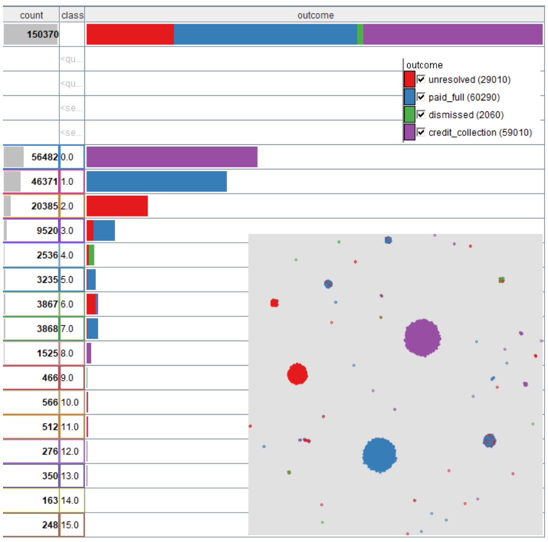


**Figure 5:** *Case counts and outcome proportions by dominant topic; inset: embedding of all cases in topic space, colored by outcome.*

`unresolved` outcomes, reflecting cases that stall at the notification stage (Topic 2).

These regimes demonstrate that topic-based trace representation captures structurally distinct process variants whose behavioral characteristics systematically relate to business-relevant outcomes such as resolution status, duration, and cost. Rather than merely encoding event occurrence, the proposed representation isolates interpretable behavioral archetypes that differentiate compliance, enforcement, contestation, delayed, and incomplete resolution dynamics within the process.

## 5. Discussion and Outlook

We proposed a text-based trace encoding combined with NMF topic modeling that replaces high-dimensional boolean vectors by low-dimensional, interpretable topic weights. A domain-specific vocabulary of activities and $n$-grams preserves basic control-flow structure while keeping feature spaces compact; the resulting topics and topic-weight embeddings support clustering, classification, and dimensionality reduction in a semantically meaningful space.

This first study is limited to a single, well-known public event log and to two simple encodings (activities and 2-grams). Although the framework accommodates richer vocabularies, including lagged-follows and attribute-augmented tokens, these lead to more complex topic structures that demand more sophisticated visual representations, which we leave for future work. We have applied the same methodology to several additional real-world process mining data sets, but omit these case studies due to space constraints.

Topic modeling also entails design choices (vocabulary, discretization, number of topics $K$, ensemble parameters) and non-negligible computation for very large logs and richer encodings. In future work, we aim to refine these choices, develop visual encodings for richer, attribute-aware vocabularies, and integrate topic-based embeddings more tightly with coordinated multi-view exploration, including resource and performance facets.


Proceedings published by Eurographics - The European Association for Computer Graphics.

## References


[AAAW22] Andrienko N., Andrienko G., Adilova L., Wrobel S.: Visual analytics for human-centered machine learning. *IEEE Computer Graphics and Applications 42*, 1 (2022), 123–133. doi:10.1109/MCG.2021.3130314. 2

[AAH24] Andrienko G., Andrienko N., Hecker D.: Topic modelling for spatial insights: Uncovering space use from movement data. *Computers & Graphics 122* (2024), 103989. doi:10.1016/j.cag.2024.103989. 3

[AHK01] Aggarwal C. C., Hinneburg A., Keim D. A.: On the surprising behavior of distance metrics in high dimensional space. In *Database Theory — ICDT 2001* (Berlin, Heidelberg, 2001), Van den Bussche J., Vianu V., (Eds.), Springer Berlin Heidelberg, pp. 420–434. doi:10.1007/3-540-44503-X_27. 1

[BvdA09] Bose R. P. J. C., van der Aalst W. M. P.: Context aware trace clustering: Towards improving process mining results. In *Proc. Int. Conf. Data Mining, SDM* (2009), SIAM, pp. 401–412. doi:10.1137/1.9781611972795.35. 2

[CAA*20] Chen S., Andrienko N., Andrienko G., Adilova L., Barlet J., Kindermann J., Nguyen P. H., Thonnard O., Turkay C.: Lda ensembles for interactive exploration and categorization of behaviors. *IEEE Transactions on Visualization and Computer Graphics 26*, 9 (2020), 2775–2792. doi:10.1109/TVCG.2019.2904069. 3

[CFDS24] Colonna J. G., Fares A. A., Duarte M., Sousa R.: Process mining embeddings: Learning vector representations for petri nets, 2024. arXiv:2404.17129. 2

[DCM22] Di Ciccio C., Montali M.: *Declarative Process Specifications: Reasoning, Discovery, Monitoring*. Springer International Publishing, Cham, 2022, pp. 108–152. doi:10.1007/978-3-031-08848-3_4. 3

[DKvBDW18] De Koninck P., vanden Broucke S., De Weerdt J.: act2vec, trace2vec, log2vec, and model2vec: Representation Learning for Business Processes. In *Business Process Management* (Cham, 2018), Weske M., Montali M., Weber I., vom Brocke J., (Eds.), Springer Int. Publishing, pp. 305–321. doi:10.1007/978-3-319-98648-7_18. 2

[dLM15] de Leoni M., Mannhardt F.: Road traffic fine management process. Eindhoven University of Technology, Dataset, 2015. Dataset ID 284. doi:10.4121/uuid:270fd440-1057-4fb9-89a9-b699b47990f5. 2, 3

[EMK*21] Espadoto M., Martins R. M., Kerren A., Hirata N. S. T., Telea A. C.: Toward a quantitative survey of dimension reduction techniques. *IEEE Trans. Vis. Comp. Graph. 27*, 3 (2021), 2153–2173. doi:10.1109/tvcg.2019.2944182. 1, 2

[JRSW19] Jablonski S., Röglinger M., Schönig S., Wyrtki K. M.: Multi-perspective clustering of process execution traces. *Enterp. Model. Inf. Syst. Archit. Int. J. Concept. Model. 14* (2019), 2:1–2:22. doi:10.18417/EMISA.14.2. 2

[LNC*17] Luo M., Nie F., Chang X., Yang Y., Hauptmann A., Zheng Q.: Probabilistic non-negative matrix factorization and its robust extensions for topic modeling. In *Thirty-first AAAI conference on artificial intelligence* (2017). 2, 3

[MAAS25] Moussavi L., Andrienko G., Andrienko N., Slingsby A.: Visually-supported topic modeling for understanding behavioral patterns from spatio–temporal events. *Computers & Graphics 129* (2025), 104245. doi:10.1016/j.cag.2025.104245. 3

[MCSW24] Miksch S., Ciccio C. D., Soffer P., Weber B.: Visual analytics meets process mining: Challenges and opportunities. *IEEE Comp. Graph. Appl. 44*, 6 (2024), 132–141. doi:10.1109/MCG.2024.3456916. 1

[MHSG18] McInnes L., Healy J., Saul N., Grossberger L.: Umap: Uniform manifold approximation and projection. *Journal of Open Source Software 3*, 29 (2018), 861. doi:10.21105/joss.00861. 4

[RAS25] Rullo A., Alam F., Serra E.: Trace encoding techniques for multi-perspective process mining: A comparative study. *WIREs Data. Mining. Knowl. Discov. 15*, 1 (2025). doi:10.1002/WIDM.1573. 2

[Rob07] Roberts J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Fifth Int. Conf. Coordinated and Multiple Views in Exploratory Visualization* (2007), pp. 61–71. doi:10.1109/CMV.2007.20. 1

[SAD*26] Schuster D., Aigner W., Di Francescomarino C., Turkay C., Zerbato F.: KAVA-PM: Knowledge-assisted visual process mining. *Information Systems 137* (2026), 102638. doi:https://doi.org/10.1016/j.is.2025.102638. 2

[Sv08] Song M., van der Aalst W. M.: Towards comprehensive support for organizational mining. *Decision Support Systems 46*, 1 (2008), 300–317. doi:https://doi.org/10.1016/j.dss.2008.07.002. 1

[TOBC23] Tavares G. M., Oyamada R. S., Barbon S., Ceravolo P.: Trace encoding in process mining: A survey and benchmarking. *Eng. Appl. Artificial Intelligence 126* (2023), 107028. doi:10.1016/j.engappai.2023.107028. 2

[vdA16] van der Aalst W. M. P.: *Process Mining – Data Science in Action, Second Edition*. Springer, 2016. doi:10.1007/978-3-662-49851-4. 1

[vdEAA*23] van den Elzen S., Andrienko G., Andrienko N., Fisher B. D., Martins R. M., Peltonen J., Telea A. C., Verleysen M.: The flow of trust: A visualization framework to externalize, explore, and explain trust in ml applications. *IEEE Computer Graphics and Applications 43*, 2 (2023), 78–88. doi:10.1109/MCG.2023.3237286. 2

[vdEAK*25] van den Elzen S., Andrienko G., Kerren A., Resinas M., Weber B., Yu P.: Coordinated projections : A new approach to multi-faceted process exploration. In *Process Mining Workshops : ICPM 2025 International Workshops, Montevideo, Uruguay, October 20–24, 2025, Revised Selected Papers* (2025), Lecture Notes in Business Information Processing. To be published in March 2026. 1, 2, 3

[vJM*25] van den Elzen S., Jans M., Martin N., Pieters F., Tominski C., Villa-Uriol M.-C., van Zelst S. J.: Towards multi-faceted visual process analytics. *Information Systems 133* (2025), 102560. doi:10.1016/j.is.2025.102560. 1

[VK20] Vayansky I., Kumar S. A.: A review of topic modeling methods. *Information Systems 94* (2020), 101582. doi:10.1016/j.is.2020.101582. 2

[VPvM25] Vitale F., Pegoraro M., van der Aalst W. M., Mazzocca N.: Control-flow anomaly detection by process mining-based feature extraction and dimensionality reduction. *Knowledge-Based Systems 310* (2025), 112970. doi:10.1016/j.knosys.2025.112970. 1

[VZSW23] Völzer H., Zerbato F., Sulzer T., Weber B.: A fresh approach to analyze process outcomes. In *5th International Conference on Process Mining, ICPM 2023, Rome, Italy, October 23-27, 2023* (2023), IEEE, pp. 97–104. doi:10.1109/ICPM60904.2023.10271968. 3

[ZRSH20] Zandkarimi F., Rehse J.-R., Soudmand P., Hoehle H.: A generic framework for trace clustering in process mining. In *2020 2nd International Conference on Process Mining (ICPM)* (2020), pp. 177–184. doi:10.1109/ICPM49681.2020.00034. 2

[ZSW21] Zerbato F., Soffer P., Weber B.: Initial insights into exploratory process mining practices. In *Business Process Management Forum* (2021), Polyvyanyy A., Wynn M. T., Looy A. V., Reichert M., (Eds.), vol. 427 of *LNBIP*, Springer, pp. 145–161. doi:10.1007/978-3-030-85440-9_9. 2