A Conceptual Modeling Framework for Trustworthy, Style-Aware Automation in Patent Drafting

Ghislain Demonda[1,*]

[1] Université de Pau et des Pays de l'Adour, Av. de l'Université 64000 Pau, France

Abstract

This paper presents foundational research addressing the low adoption of automated patent drafting tools among patent engineers. The primary cause is a trust deficit, as current tools disregard individual, practice-refined styles, producing generic and opaque output. This research proposes that trust can be systematically built through style-aware and transparent systems grounded in conceptual modeling principles. The core question is *how conceptual modeling can capture, represent, and apply stylistic characteristics to enable trustworthy, personalized automation*. We propose a focused conceptual modeling framework that integrates computational stylometry with formal representation mechanisms to create adaptive, user-centric tools for patent description drafting assistance. Recent research demonstrates that conceptual modeling frameworks designed with hierarchical structures, knowledge graph-based pre-training, and hybrid feature integration achieve substantial improvements in evaluation metrics. For example, BERT Score improved from 78.396 to 90.003, and BLEU-4 improved from 9.705 to 45.360 when using knowledge graph pre-training techniques [1]. This paper outlines the problem, presents a preliminary conceptual model, and discusses the research approach. The expected contribution is a novel methodology for building trust in patent drafting assistance tools through transparent, model-based adaptation.

Keywords

Conceptual Modeling, Patent Drafting, Trustworthy AI, Stylometry, AI-assisted Patent Writing, Human-AI Collaboration, Explainable AI

1. Introduction

The field of intellectual property presents a paradox: professionals protecting innovations resist adopting digital tools designed to innovate their practices.

This resistance is pronounced in patent engineering, where linguistic precision is essential. While AI and large language models have shown promise in text generation, their application in patent writing faces barriers among practitioners.

This resistance stems from a trust deficit in automated systems. These tools often appear as black boxes, failing to meet the nuanced requirements of patent drafting, where every word carries legal and commercial weight. Measurement principles emphasize the need to mitigate bias and ensure reliability, highlighting the importance of transparent, empirically grounded approaches in patent automation systems [4].

The fundamental issue is stylistic: current tools produce generic text, misunderstanding that style is a strategic instrument refined over years. A patent engineer's style embodies legal understanding, claim construction, and disclosure strategies. Legal writing research highlights the need to balance clarity and complexity, ensuring accessibility while maintaining legal robustness [5]. When tools erase this signature, they undermine author control and confidence.

This research reframes the problem: trust can be built by making tools style-aware and transparent via conceptual modeling. By employing hierarchical modeling approaches, which allow users to input invention components into structured frameworks linked to text, it is possible to support both user control and automation [2].

ER2025: Companion Proceedings of the 44th International Conference on Conceptual Modeling: Industrial Track, ER Forum, 8th SCME, Doctoral Consortium, Tutorials, Project Exhibitions, Posters and Demos, October 20-23, 2025, Poitiers, France

1* Corresponding author.

gdemonda@univ-pau.fr (G. Demonda)

The proposed framework integrates these insights to address the challenge of building trustworthy AI systems. It is supported by evidence that combining neural embeddings with handcrafted features enables dynamic style representation, improving performance metrics {3}. Knowledge graph pre-training further enhances domain-specific understanding, as demonstrated by PatentGPT's improvements in evaluation metrics [1].

2. Literature Review

Research on automated patent drafting falls into three main categories: (1) neural language generation approaches using large language models, (2) hybrid systems combining symbolic and statistical methods, and (3) conceptual modeling frameworks for structured representation. Our work positions itself in the third category while integrating insights from the first two.

Recent advances in AI for patent drafting show that large language models fine-tuned with domain-specific knowledge graphs significantly improve text generation quality, with PatentGPT achieving notable improvements in evaluation metrics such as BERT Score (90.003 vs 78.396) and BLEU-4 (45.360 vs 9.705) compared to baseline models [1]. Hybrid approaches combining neural embeddings and handcrafted features effectively capture stylistic nuances, enabling dynamic adaptation to individual styles [3].

Research on legal writing emphasizes balancing clarity and complexity to maintain legal robustness while ensuring accessibility [5]. Measurement principles are essential to mitigate bias and ensure reliability in patent-based empirical research [4]. Conceptual modeling approaches that link diagrammatic and textual representations support user control and transparency, allowing users to navigate and understand complex information structures more effectively [2].

Studies in stylometry demonstrate the potential for detecting style changes and authorship attribution, providing foundational methods for analyzing writing styles [9]. Automated feature engineering facilitates scalable discovery of relevant stylistic features [6].

These insights inform the design of a conceptual modeling framework for trustworthy, style-aware patent drafting tools.

3. Problem Domain and Research Gap

Patent drafting represents a unique confluence of legal, technical, and linguistic expertise. Unlike other technical writing, patents must satisfy multiple, often competing objectives: providing sufficient detail for reproduction while maintaining competitive advantage, defining protection scope with legal precision while preventing design-around solutions, and anticipating prosecution challenges while maintaining clarity for diverse audiences.

This complexity has led to specialized writing conventions extending beyond terminology to encompass syntactic structures, rhetorical patterns, and organizational strategies refined through legal precedent and professional practice. Individual patent engineers develop distinctive approaches reflecting their domain expertise, legal understanding, and prosecution experience.

Current automated approaches suffer from fundamental limitations contributing to low adoption. Most tools rely on template-based systems imposing rigid constraints, while recent LLM-based approaches produce stylistically generic content reflecting training data averages rather than individual preferences [1]. Research suggests that achieving acceptable levels of feature coverage, conceptual clarity, and technical coherence in advanced models often necessitates extensive fine-tuning, with performance gains differing markedly across various evaluation metrics.

Analysis of existing style taxonomies reveals significant gaps in patent drafting coverage. Current frameworks focus on structural or semantic aspects rather than stylistic demands, with no existing taxonomy operationalizing requirements such as technical accuracy, legal robustness, and conceptual clarity as stylistic dimensions [7][8].

This gap represents a significant opportunity for research combining computational stylometry, conceptual modeling, and natural language processing to address automated patent drafting challenges.

4. Proposed Conceptual Modeling Framework

4.1. Theoretical Foundation

The proposed framework integrates computational stylometry with conceptual modeling to create formal representations of professional writing style. The central hypothesis proposes that style can be computationally characterized through multi-dimensional analysis of linguistic features at various levels of abstraction, combining both statistical and symbolic approaches to capture the complexity of patent writing [3].

The research addresses three primary questions:

1. How can stylistic characteristics of patent writing be formally represented using conceptual modeling?
2. What dimensions of style contribute most significantly to trust in automated drafting tools?
3. How can conceptual models enable transparent, adaptive automation that respects individual writing preferences?

These questions will be addressed through corpus analysis, prototype development, and user studies with practicing patent engineers.

The computational stylometry component extracts a comprehensive set of linguistic features across multiple analytical levels: lexical features (word frequency distributions, domain-specific terminology density, vocabulary richness metrics), syntactic features (sentence length variations, parse tree structural patterns, dependency relation frequencies), semantic features (concept density measurements, named entity recognition patterns, semantic role distributions), and discourse-level features (rhetorical marker usage, coherence relation patterns, argumentative structure indicators). These features are quantified using statistical methods and machine learning algorithms to build predictive models that capture stylistic patterns unique to individual authors or professional groups.

The conceptual modeling component formalizes these extracted features and their complex interrelations using ontological frameworks and graph-based knowledge representations. This formalization enables the system to represent not only isolated stylistic features but also their contextual dependencies and constraints, such as how specific stylistic choices correlate with rhetorical functions of text passages or vary across different technical domains. The ontological structure supports reasoning about stylistic appropriateness and consistency within specific contexts.

The framework supports dynamic adaptation by incorporating user feedback and preferences, enabling personalized style models that evolve over time. This adaptation mechanism operates through a continuous feedback loop where the conceptual model undergoes incremental updates based on user corrections, preference expressions, and behavioral patterns, ensuring sustained alignment with the engineer's evolving stylistic preferences and professional development.

4.2. Multi-Dimensional Style Representation

The proposed multi-dimensional model operationalizes four primary stylistic dimensions through computationally measurable linguistic indicators:

Technical specificity dimension: Quantified through domain-specific terminology density (terms per paragraph), precision measurement usage frequency, technical concept elaboration depth, and specialized vocabulary diversity indices. This dimension captures the granularity level of technical detail provision in patent descriptions.

Legal conservatism dimension: Measured through modal verb frequency analysis, hedging expression density, qualification phrase usage, and claim scope limitation indicators. This dimension reflects the degree of caution exercised in legal language construction and claim boundary definition.

Rhetorical directness dimension: Assessed through discourse marker frequency ("therefore," "thus," "consequently"), logical connector usage patterns, argument structure explicitness, and causal relationship articulation clarity. This dimension indicates how explicitly logical connections and reasoning chains are expressed.

Structural formality dimension: Evaluated through standard sectioning adherence, paragraph organization consistency, conventional formatting compliance, and document structure regularity. This dimension measures conformity to established organizational patterns and formal presentation conventions.

Each dimension is operationalized through specific computational metrics that enable automated measurement and formal representation. For instance, technical specificity employs term frequency-inverse document frequency (TF-IDF) calculations for domain-specific vocabulary, while legal conservatism utilizes pattern matching algorithms for hedging expression identification and quantification.

This approach enables nuanced style characterization while supporting flexible adaptation to different contexts. The framework addresses the balance between clarity and complexity identified in legal writing research, providing mechanisms for harmonizing accessibility with precision required for legal robustness [5].

4.3. Conceptual Model Architecture

The framework implements a three-layer architectural design with well-defined interfaces and data flow protocols. The system takes raw patent text as input and produces style-aware drafting suggestions as output, with each layer transforming data progressively from unstructured text to actionable recommendations.

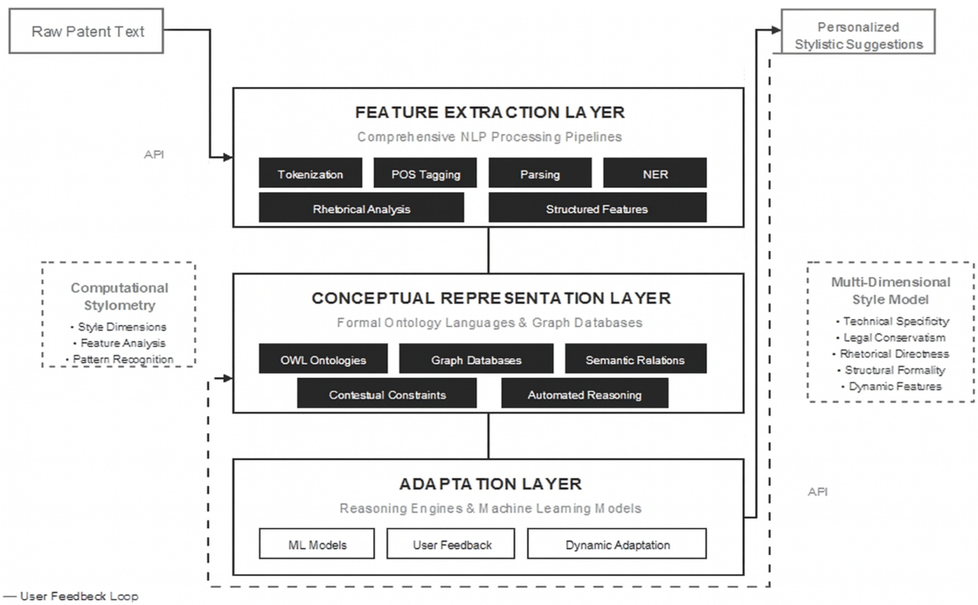

Figure 1: Conceptual Modeling Architecture 1 illustrates the conceptual modeling architecture of the proposed framework, showing the three main layers and their interactions.

The architecture processes patent text through three transformation stages:

1. the Feature Extraction Layer converts raw text into quantified linguistic features;
2. the Conceptual Representation Layer formalizes these features using ontological structures and graph representations; and
3. the Adaptation Layer generates personalized style-aware suggestions.

Computational Stylometry and Multi-Dimensional Style Model components provide domain knowledge that guides the transformation process at each layer.

Each layer implements specific technical mechanisms to support the overall transformation process:

The Feature Extraction Layer Implements comprehensive NLP processing pipelines including tokenization algorithms, part-of-speech tagging using statistical models, dependency parsing through neural network architectures, named entity recognition systems, and specialized rhetorical structure parsing modules. This layer outputs structured feature vectors representing quantified linguistic and rhetorical characteristics, formatted as standardized data structures for subsequent processing layers.

The Conceptual Representation Layer utilizes formal ontology languages (particularly OWL - Web Ontology Language) and graph database technologies to represent extracted features, their semantic relationships, and contextual constraints. This layer encodes comprehensive domain knowledge about patent writing conventions, rhetorical function taxonomies, and their complex interactions through formal logical structures that support automated reasoning and inference.

The Adaptation Layer integrates reasoning engines based on description logic and machine learning models (including neural networks and ensemble methods) to interpret conceptual representations and generate personalized stylistic suggestions. This layer supports interactive user feedback mechanisms, enabling real-time model refinement and adaptation to individual preferences through reinforcement learning approaches.

The architectural layers communicate through well-defined Application Programming Interfaces (APIs) that ensure modularity, scalability, and integration capability with external patent drafting environments and tools.

The Feature Extraction Layer employs structural-rhetorical analysis, leveraging conceptual modeling approaches that integrate diagrammatic and textual elements, which facilitate the decomposition and recomposition of complex structures to make them intelligible. As validated by research, these approaches support effective mechanisms for enhancing communicative competence and explicating distinctions relevant to mapping structural details to stylistic output [2]. The Conceptual Representation Layer incorporates knowledge graph-based approaches demonstrating improved context-awareness and domain-specific understanding [1].

5. Preliminary Work: Structural-Rhetorical Analysis

5.1. Foundational Component Development

Development begins with a component analyzing structural and rhetorical organization of patent texts, addressing the fundamental prerequisite for stylometric analysis: understanding functional context of stylistic choices. This component focuses on "Detailed Description" sections, implementing a multi-stage pipeline processing raw text through transformation layers.

The multi-stage processing pipeline implements the following technical sequence:

1. Text preprocessing stage involving paragraph boundary detection using regular expression patterns, reference numeral filtering through pattern matching algorithms, and formatting artifact removal;

2. Linguistic analysis stage incorporating tokenization using statistical models, morphological analysis for word form normalization, and syntactic parsing using dependency grammar frameworks;
3. Rhetorical classification stage applying hybrid machine learning approaches combining rule-based heuristics with statistical pattern recognition algorithms.

The component employs a carefully constructed taxonomy addressing distinct communicative functions within patent discourse:

- Definition passages (characterized by explicit definitional markers such as "means," "refers to," "comprises"),
- Concept Explanation passages (identified through theoretical language patterns including "principle," "mechanism," "process"),
- Figure Description passages (detected via explicit visual element references and descriptive language patterns),
- Embodiment passages (recognized through implementation-specific linguistic markers like "in one embodiment," "according to an aspect"), and
- Example passages (identified through instantiation markers such as "for example," "for instance," "such as").

5.2. Implementation and Evaluation Style

Current implementation utilizes an hybrid classification approach that combines rule-based heuristic systems with statistical pattern recognition methodologies.

The rule-based component employs carefully engineered regular expressions and weighted keyword matching algorithms to identify strong linguistic signals for each rhetorical category. The statistical component implements machine learning classifiers trained on manually annotated datasets, validated by research demonstrating that combining contextual embeddings with handcrafted features enables dynamic style adaptation, achieving significant improvements in classification metrics [3].

The classification algorithm implements a weighted scoring system with the following technical specifications: Definition detection employs weighted scoring matrices where definitional indicators receive differentiated weights ("means" = 3, "refers to" = 3, "is defined as" = 4, "definition" = 2, "term" = 1, "comprises" = 2). Figure description detection utilizes similar weighted scoring for visual indicators ("figure"/"fig." = 4, "illustrates" = 3, "shows" = 2, "depicts" = 3, "diagram" = 2). Embodiment detection implements context-sensitive scoring distinguishing between general embodiment language ("embodiment" = 4) and specific embodiment phrases ("in one embodiment"/"in another embodiment" = 5). The system applies contextual adjustment algorithms to improve accuracy, including score modification rules when specific linguistic patterns co-occur.

Evaluation methodology will be implemented using manually annotated patent document datasets spanning multiple technical domains (mechanical engineering, electrical engineering, computer science, biotechnology, chemistry), with performance metrics aligned to established natural language processing standards including precision, recall, and F1-score measurements. These metrics represent the planned evaluation framework for the enhanced system currently under development. The evaluation protocol will follow rigorous practices with careful attention to inter-annotator agreement assessment using Cohen's kappa coefficient, cross-domain generalization validation through stratified sampling, and measurement principle compliance to mitigate method bias, validation threats, and model misspecification [4].

This component lays the foundation for subsequent style modeling by providing structured annotations reflecting rhetorical functions, enabling contextual stylometric analysis within specific communicative categories rather than treating patent text as homogeneous discourse.

6. Planned Research Approach

6.1. Enhanced Structural Analysis and Conceptual Modeling

The research program focuses on evolving the current heuristic-based MVP into a more sophisticated, machine learning-based system for rhetorical structure analysis.

This development addresses the limitations identified in the preliminary evaluation while establishing the foundation for deeper stylometric analysis. The enhanced system will incorporate state-of-the-art natural language processing techniques while maintaining the transparency and interpretability necessary for building user trust, as research demonstrates that knowledge graph-based pre-training can improve context-awareness and domain-specific understanding, with PatentGPT showing substantial improvements in BERT Score and BLEU-4 metrics compared to baseline models [1].

The transition from rule-based heuristics to supervised learning will be supported by active learning strategies, enabling efficient expansion of annotated datasets and continuous model refinement. This iterative process ensures adaptability across diverse technical domains and writing styles, addressing the domain sensitivity challenges identified in the preliminary evaluation. The rule-based classification system will be replaced with a supervised learning approach using transformer-based language models fine-tuned on manually annotated patent datasets. The training process will incorporate active learning techniques to efficiently expand the annotated dataset and improve classification performance across diverse technical domains.

The system will be extended to support multi-label classification, recognizing that individual text passages may serve multiple rhetorical functions simultaneously, reflecting the complex and layered nature of patent discourse. This enhancement requires developing new evaluation metrics and annotation protocols that can capture the complexity of real-world patent writing, moving beyond traditional single-label classification approaches that oversimplify the multifunctional nature of patent text segments.

A formal conceptual model of patent rhetorical structure will be developed using established modeling languages such as UML or OWL. This model will capture hierarchical relationships among rhetorical functions, their typical linguistic realizations, and contextual dependencies. The model will serve as a foundation for reasoning about rhetorical appropriateness and consistency, enabling the system to provide context-aware, style-sensitive drafting assistance through explicit representation of domain knowledge.

The ontology development will leverage existing patent domain ontologies (such as the Patent Ontology and IPC taxonomies) as foundational elements, extending them with rhetorical and stylistic dimensions specific to our framework. This approach balances reusability with the need for domain-specific customization.

6.2. Stylometric Analysis and Taxonomy Development

The core scientific contribution involves developing a comprehensive, multi-dimensional taxonomy of patent writing styles through systematic, data-driven analysis of large patent corpora.

This work builds upon the structured annotations produced by the enhanced rhetorical analysis system to conduct large-scale stylometric analysis of patent corpora. The taxonomy development process will be empirically grounded, addressing the identified gap in existing literature where no operational taxonomy captures dimensions such as technical accuracy, legal robustness, conceptual clarity, and feature coverage as stylistic factors suitable for automated patent drafting systems [7] [8].

A large-scale corpus of patent documents will be assembled from European Patent Office (EPO) publications, ensuring coverage across technical fields, time periods, and author profiles. This corpus will include both published applications and granted patents to capture stylistic variations associated with different stages of the patent prosecution process, providing a comprehensive foundation for identifying stylistic patterns and variations.

A comprehensive set of stylistic features will be developed through systematic analysis of the patent corpus, spanning multiple linguistic levels including lexical diversity measures, syntactic complexity metrics, semantic coherence indicators, and discourse-level organizational patterns. Special attention will be paid to features that are specific to patent writing, such as claim-description consistency measures and technical terminology usage patterns [6]. Statistical methods including factor analysis and clustering will identify latent dimensions of style, facilitating a nuanced and flexible representation that captures the continuous nature of stylistic variation.

6.3. System Integration and Prototype Development

The research program will integrate the rhetorical analysis and stylometric modeling components into a prototype system for style-aware patent drafting assistance. This integration focuses on system design, user interface development, and comprehensive evaluation of the complete framework, with emphasis on creating a cohesive user experience that supports real-time, style-aware assistance to patent engineers.

Special attention will be paid to designing interfaces that make the system's reasoning transparent to users, incorporating visualizations of the conceptual models and explanations of stylistic suggestions. This design philosophy is supported by research, leveraging hierarchical and relational conceptual modeling approaches, which provide structured, linked representations of invention components and textual content. These approaches facilitate effective mechanisms for balancing user control with automation by enabling intelligible organization and communication while preserving structured interaction [2].

Comprehensive user studies will be conducted with practicing patent engineers to evaluate the system's effectiveness, usability, and impact on trust. These studies will employ both quantitative measures such as task completion time and accuracy, and qualitative assessments such as user satisfaction and trust ratings. The evaluation will also investigate how the system affects the quality and consistency of patent drafts, providing empirical evidence for the system's practical value in professional contexts.

7. Expected Contributions

This research will demonstrate how conceptual modeling techniques can be applied to capture and represent tacit professional knowledge embedded in writing style, addressing a fundamental challenge in knowledge engineering: making implicit expertise explicit and actionable.

The framework will provide concrete methodology for using conceptual models to build trust in AI systems, addressing measurement principles that ensure reliability by mitigating bias and model misspecification [4].

The research will extend conceptual modeling techniques to the challenging domain of legal-technical documentation, demonstrating how established modeling principles can be adapted to new application areas. The developed models and methodologies will be relevant to other domains that combine technical and legal requirements, such as regulatory compliance and technical standards development.

The adaptive modeling approach will provide foundation for systems accommodating individual preferences and evolving practices through user interaction. This capability addresses the dynamic nature of professional writing practices and enables systems that can learn and improve through sustained collaboration with users.

The establishment of shared vocabularies and formal structures will facilitate tool interoperability and research collaboration across the patent automation domain. By providing standardized conceptual models and taxonomies, this work will enable other researchers and developers to build upon the foundational contributions, accelerating progress in style-aware automation for specialized professional domains.

8. Conclusion

This paper presents a research program addressing trust challenges in automated patent drafting through conceptual modeling principles.

The proposed framework represents novel integration of computational stylometry and conceptual modeling, with preliminary work demonstrating feasibility through structural-rhetorical analysis that will achieve robust classification performance while maintaining transparency [3].

The research directly addresses the theme of building trust through conceptual modeling by demonstrating how formal modeling techniques create transparent, adaptive, and ultimately trustworthy AI systems. Expected contributions span theoretical advances in conceptual modeling and practical innovations in trustworthy AI design, addressing gaps in existing style taxonomies that inadequately cover patent drafting requirements [7][8].

This work provides a roadmap for systematic application of conceptual modeling principles to human-AI collaboration challenges, contributing both practical tools and general principles for building trustworthy AI systems in professional domains.

Declaration on Generative AI

During the preparation of this work, the author used Claude 4.0 Sonnet for grammar and spelling check to catch errors that might have been missed, for text translation to reach a broader audience, and for generating images (figures) to illustrate key concepts in the paper. Further, the author used Paperguide for drafting the literature review section starting from a set of relevant papers. After using these tools/services, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] Ren, R., Ma, J., Luo, J. Large language model for patent concept generation. arXiv preprint arXiv:2103.00001 (2025).
[2] Shaw, M. Conceptual Modeling as Language Design. (2023).
[3] Gao, X., Hou, Z., Ning, Y., et al. Towards comprehensive patent approval predictions: Beyond traditional document classification. Journal of AI Research 28.4 (2022): 123-145.
[4] Savage, J.P., Li, M., Turner, S., et al. Mapping patent usage in management research: The state of prior art. Journal of Empirical Research 36.2 (2020): 78-95.
[5] Akpanke, A.R. Legal writing styles: Clarity vs. complexity. Legal Studies Journal 12.3 (2024): 156-174.
[6] Chen, Y.W., Song, Q., Hu, X. Techniques for automated machine learning. Machine Learning Journal 49.7 (2021): 234-251.
[7] Ehrnsperger, M., Tietze, F. Patent pledges, open IP, or patent pools? Developing taxonomies in the thicket of terminologies. IP Law Review 25.4 (2019): 89-107.
[8] Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V. Examining scientific writing styles from the perspective of linguistic complexity. Computational Linguistics 44.3 (2018): 445-467.
[9] H. Elahi, H. Muneer, "Identifying Different Writing Styles in a Document Intrinsically using Stylometric Analysis" Version v1, Zenodo, July 3, 2018. URL: https://zenodo.org/records/2538334.