A Creative Computing Approach to Analysis of Speaker Intention Using the DIKCW Model

Lin Zou[1], Hongji Yang[1,*], and Jintao Lin[2]
[1]University of Leicester, Leicester, UK
[2]Yunnan University, Kunming, China
lz179@leicester.ac.uk, hongji.yang@leicester.ac.uk , linjintao80@163.com
*Corresponding author

Abstract—The burgeoning field of text analysis has been significantly enriched by the integration of creative computing methodologies. This research introduces a novel approach to multi-dimensional text analysis through the development of a Data, Information, Knowledge, Creativity, and Wisdom (DIKCW) Model. The model is designed to facilitate speaker intention extraction by systematically processing and interpreting textual data across multiple layers—Data, Information, and Knowledge. The Data layer accommodates 13 distinct types. The Information layer is enriched by eight types of information. A universal function has been formulated to amalgamate these diverse information types, yielding an index that quantifies information density. Transitioning to the Knowledge layer, the model identifies six types of knowledge. Algorithms fed by the universal function and information types categorize the knowledge embedded in the text, setting the stage for ontology representation. This work serves as a foundational pillar for future research endeavors aimed at understanding the intricate relationships between data, information, and knowledge within the context of creativ- ity and wisdom. The research employs creative computing techniques, rule-based reasoning, and visualisation to offer a robust framework for extracting meaningful insights from complex data sets.

Keywords–Creative Computing; DIKW Model; Natural Language Processing

1. Introduction

The advent of the digital age has led to an unprece- dented proliferation of textual data, which has become an invaluable resource for various fields ranging from social sciences to artificial intelligence. The extraction of meaningful insights from this vast corpus of data necessitates the development of sophisticated models and techniques. While traditional text analysis methods have made significant strides, they often fall short in capturing the multi-dimensionality of human language, particularly when it comes to understanding speaker intention. This research aims to address this gap by introducing a novel approach that leverages creative computing methodologies and the Data, Information, Knowledge, Creativity, and Wisdom (DIKCW) Model.

The DIKCW Model is an extension of the well-established Data, Information, Knowledge, Wisdom (DIKW) hierarchy, incorporating the element of creativity to offer a more comprehensive framework. The model is designed to systematically process and interpret textual data across multiple layers—Data, Information, and Knowledge—to facilitate the extraction of speaker intention. Each layer is enriched by a variety of types, this research defined 13 distinct types at the Data layer and eight types at the Information layer. A universal function amalgamates these diverse types, yielding an index that quantifies information density. The Knowledge layer further categorises the knowledge embedded in the text, setting the stage for ontology representation.

Creative computing, an interdisciplinary field that inte- grates computing with creative arts, provides the method- ological backbone for this research. By employing creative computing techniques, this work transcends the limita- tions of conventional text analysis methods, offering a more nuanced understanding of textual data. Rule-based reasoning and visualisation techniques are also employed to enhance the robustness of the framework.

The significance of this research is manifold. Firstly, it serves as a foundational pillar for future research endeavors aimed at understanding the intricate relation- ships between data, information, and knowledge within the context of creativity and wisdom. Secondly, the model has practical applications in various domains such as sentiment analysis, natural language processing [1,2], and human-computer interaction, among others. Lastly, by focusing on speaker intention, the research addresses a critical yet often overlooked aspect of text analysis, thereby contributing to the broader discourse on human communication and artificial intelligence.

In summary, this research introduces a groundbreaking approach to multi-dimensional text analysis by developing the DIKCW Model. Through the integration of creative computing methodologies, rule-based reasoning, and vi- sualisation techniques, the research offers a robust frame- work for extracting meaningful insights from complex data sets, particularly in the realm of speaker intention.

2. Background and Related Research

The landscape of text analysis has been significantly influ- enced by various models and methodologies, among which the Data, Information, Knowledge, Wisdom (DIKW) hi- erarchy has been a cornerstone. The DIKW model has been

2693-9371/23/$31.00 ©2023 IEEE
DOI 10.1109/QRS-C60940.2023.00114

extensively used to understand the transformation of raw data into actionable wisdom. However, the model has been critiqued for its lack of attention to the creative aspects that often play a pivotal role in knowledge gener- ation and wisdom. This research extends the traditional DIKW hierarchy to include Creativity, resulting in the DIKCW Model. The addition of Creativity as a layer aims to capture the multi-dimensional facets of information processing, particularly in the context of speaker intention analysis.

2.1 Research in DIKW Model and Creativity

The Data, Information, Knowledge, Wisdom (DIKW) hierarchy has been a foundational model in the field of information science, serving as a framework for under- standing the transformation of raw data into actionable wisdom [3]. Originating from the work of Zeleny [4], the model has been employed across various disciplines, including computer science, management, and healthcare [5, 6]. It delineates a linear progression from data, which is unprocessed and context-free, to information, which adds meaning to data, then to knowledge, which involves the application of information, and finally to wisdom, which encompasses judgment and decision-making [7]. While the DIKW model has been instrumental in shaping our understanding of information processing, it has been critiqued for its lack of attention to the creative aspects that often play a pivotal role in knowledge generation and wisdom [8]. Creativity, defined as the ability to generate novel and valuable ideas, is increasingly recognised as a crucial component in the information lifecycle [9]. It is in this context that the DIKW model is extended to include Creativity, resulting in the DIKCW Model. The addition of the Creativity layer aims to capture the multi-dimensional facets of information processing. Creativity is not merely an add-on but is integrated into each layer of the model, influencing how data is collected, how information is interpreted, and how knowledge is applied [10]. This extension aligns with the growing body of research that emphasises the role of creativity in problem-solving, innovation, and decision-making [11]. The DIKCW Model is particularly relevant in the context of text analysis and speaker intention extraction. Traditional text analysis methods often employ machine learning algorithms or linguistic rules but seldom capture the complexity and nuance of human intentions fully [12]. By applying the DIKCW Model, this research aims to offer a more comprehensive framework for understanding the intricate layers of human communication.

Integrating creativity into the Data-Information-Knowledge-Wisdom (DIKW) model can result in a new framework, which can term as the DIKCW Model. In this extended model, “Creativity” serves as an independent layer, positioned between “Knowledge” and “Wisdom.”

The rationale for this placement is that creativity often acts as a catalyst for transforming knowledge into wisdom by enabling innovative problem-solving, ethical considerations, and the generation of new knowledge.The layer descriptions as follows:

- Data: Raw facts and figures without context,
- Information: Data processed to be meaningful,
- Knowledge: Information understood and applied to contexts,
- Creativity: The ability to generate original ideas and solutions by synthesising existing knowledge in novel ways, and
- Wisdom: The judicious application of knowledge and creativity for problem-solving and decision-making

2.2 Layer Definition

Starting with the Data layer, the DIKCW model empha- sises the raw, unprocessed textual input [13]. In syntax analysis, this raw data is parsed into its grammatical components, such as tokens and parts-of-speech [14]. For semantic analysis, this data serves as the initial corpus that will be examined for word meanings and relationships [15]. For the purpose of extracting speaker intentions, the raw data is the substrate containing not just the words but also the context, tone, and even the cultural background in which these words are used. The Information layer processes this data into a more organised form [16]. In syntax analysis, this could involve part-of-speech tagging and sentence breaking. In semantic analysis, this might include identifying key terms, synonyms, and antonyms. In speaker intentions, this could mean identifying key phrases, entities, or even the sentiment expressed in the text.

Transforming the syntax data into structured information largely depends on the contents of text in research, in some cases, may only need select relevant methods to reach our goals [17]. Which means the choice of methods is context-dependent. It’s essential to align the methods with objectives, the nature of data, and the resources available. This structured information serves as the basis for further analysis and interpretation in the subsequent “Knowledge,” “Wisdom,” and “Creativity” layers. Through literature review, there are identified eight characteristics that involves in the information layer, This process is often facilitated by specific techniques [18]. They are Part-of-Speech Tags, Name Entities, Sentence Boundaries, Syntactic Relationships, Coreference Reso- lution, Sentiment Markers, Topic Keywords, Temporal Markers.

- Part-of-Speech Tages: Label each token with its corre- sponding part-of-speech (e.g., noun, verb, adjective).
- Named Entities: Identify and categorise named entities like organisations, dates, and locations.
- Sentence Boundaries: Determine the start and end of individual sentences within a text.

TABLE I
Data Layer

Data Type	Description	Example Tokens
Alphabetic	Standard letters from the alphabet	“The”, “stock”, “market”
Numerical	Numbers, including integers and decimals	“5.2”, “2023”
Punctuation Marks	Standard punctuation like commas, periods, exclamation marks, etc.	“,”, “!”
Special Characters	Characters with specific meanings in certain contexts like “@”	“@”
Whitespace	Spaces, tabs, and other forms of whitespace	“ ” (space)
Date and Time	Standard or custom date and time formats	“01/01/2023”, “12:30 PM”
Unicode Characters	Emojis or other non-standard characters	“ ”
Hyperlinks	URLs or other web links	“http://example.com”
Casing	Upper and lower case letters	“The” (initial capital), “market” (all lower case)
Language Markers	Indicators for language changes in multilingual texts	“EN:”, “ES:”
Abbreviations	Shortened forms of words or phrases	“Dr.”, “Ave.”
Acronyms	Initial letters of phrases	“NASA”, “UN”
Mathematical Symbols	Symbols used in mathematical equations	“+”, “=”, “<”

- Syntactic Relationships: Identify the stableyntactic roles of tokens and their relationships (e.g., subject, object).
- Coreference Resolution: Determine which tokens refer to the same entities across sentences.
- Sentiment Markers: Identify tokens or phrases that indicate sentiment (e.g., positive, negative).
- Topic Keywords: Extract keywords that indicate the main topics or themes of the text.
- Temporal Markers: Identify tokens that indicate time frames or sequences.

However, this transition to the Knowledge layer is fraught with challenges that require careful consideration. One of the most significant challenges is the complexity involved in synthesising information into knowledge [19]. This often requires advanced algorithms and substantial computational resources, making it a non-trivial task. Another challenge is the element of subjectivity [20]. The transition from information to knowledge frequently involves human judgment or interpretive algorithms, which can introduce bias or variability into the results. The quality of the knowledge generated is also highly dependent on the quality of the underlying data and information, making data integrity a critical concern [21]. Scalability poses another challenge; as the volume of data grows, the methods used to synthesise it into knowledge must also scale, which is not always straightforward. Moreover, the interpretability of the knowledge generated can be an issue, especially when advanced algorithms are used. Finally, ethical considerations cannot be overlooked, as the Knowledge layer can sometimes reveal sensitive or personal information.

2.3 Extended DIKW Model to Speaker Intention

In the traditional DIKW (Data-Information-Knowledge-Wisdom) model, the Wisdom layer represents the pinnacle of data transformation, where accumulated knowledge is applied in a context-sensitive manner to make judi- cious decisions, solve complex problems, and understand broader implications. Wisdom in this context is often seen as the ability to discern the long-term consequences and ethical dimensions of actions. It is the layer where data and information are not just understood but are also applied meaningfully. Wisdom integrates cognitive, emotional, and ethical dimensions, providing a holistic framework for action and understanding. It is the culmination of a linear progression from raw data to actionable insight, often requiring a deep temporal and contextual under- standing that is built upon the underlying layers of data, information, and knowledge. In the traditional DIKW (Data-Information-Knowledge-Wisdom) model, each layer builds upon the previous one in a hierarchical manner [22]. If maintain this format, the placement of the Creativity layer would depend on its perceived relationship with the existing layers. Here are two plausible options:

- Option 1: Creativity as an Extension
  Data → Information → Knowledge → Wisdom → Creativity
  In this configuration, Creativity is seen as the ultimate application of Wisdom. Once one has moved from Data to Wisdom, the Creativity layer serves as the pinnacle where wisdom is not just applied but also used to generate novel ideas, solutions, or perspectives. This placement suggests that true creativity is the result of a deep understanding and wise application of knowledge.

TABLE II
Knowledge Layer

Information Type	Description	Example in Text Analysis
Tacit Knowledge	Difficult to transfer; deeply rooted in individual experiences, ideals, and values	Intuitive understanding of a creative process in text analysis.
Explicit Knowledge	Easily communicated and shared; often in the form of data, formulas, and specifications.	Metrics or features that can be directly extracted and analysed
A Priori Knowledge	Independent of experience; universal truths often associated with logic and mathematics.	Grammatical rules or syntactic structures universally accepted.
Procedural Knowledge	Skills or procedures required to perform a task ("know-how").	Algorithms or methods used for text analysis, such as NLP techniques.
Empirical Knowledge	Gained through personal experience and human interaction.	Learning from past cases to make future decisions in text analysis.
Meta-Knowledge	Knowledge about knowledge; understanding the limitations, scope, and reliability of what is known.	Understanding limitations of text analysis algorithms or data biases.

- Option 2: Creativity as an Intermediary Layer
  Data → Information → Knowledge → Creativity → Wisdom
  Here, Creativity is placed between Knowledge and Wisdom. In this configuration, Creativity serves as the mechanism by which Knowledge is transformed into Wisdom. It suggests that creative thinking is essential for the judicious application of knowledge, which is what constitutes wisdom. Creativity here acts as the catalyst that converts theoretical or practical knowledge into actionable wisdom.

Each option has its merits and would lead to different interpretations of the relationships between Data, Information, Knowledge, Wisdom, and Creativity. For the purpose of speaker intentions extraction, Option 2: "Creativity as an Intermediary Layer" would likely be more beneficial. In this configuration—Data → Information → Knowledge → Creativity → Wisdom—Creativity serves as the mechanism that transforms Knowledge into Wisdom. Speaker intentions inherently deals with the use of language in context, which often requires creative interpretation and application of rules to understand implied meanings, social cues, and intentions.
Placing Creativity between Knowledge and Wisdom allows for a more dynamic understanding of how context-specific knowledge (e.g., idiomatic expressions, cultural norms) can be creatively applied to generate wisdom in the form of sound judgments, ethical considerations, and effective communication strategies. This placement also aligns well with the notion that creative thinking is essential for the judicious application of knowledge, a key aspect of pragmatics. Moreover, in speaker intentions extraction, the ability to creatively interpret and generate language

Figure 1. DIKW Model

can significantly influence the quality of communication and the depth of understanding, which are crucial for developing wisdom in social interactions. Therefore, having Creativity as an intermediary layer provides a theoretical space to explore these complex, context-dependent aspects of language use, making it a more suitable option for speaker intentions extraction.

3. Proposed Approach

3.1 Design of DIKCW Model

Data-Information-Knowledge-Wisdom (DIKW) hierarchy like Figure 1, integrating “Creativity” into the pipeline (Figure 2). This added dimension is instrumental for understanding how creative problem-solving skills and innovative thinking can augment the capacity of knowledge and wisdom within a system. In the context of text analysis and semantic analysis, the DIKCW model offers a well-defined approach for capturing different layers of understanding—ranging from syntactic to semantic, and further to pragmatic levels.

Figure 2. DIKCW Model

In typical Natural Language Processing (NLP) models, semantic analysis often focuses on the representation and understanding of meaning from a syntactical point of view. However, achieving pragmatics—that is, the ability to understand context, intent, and consequence—is a far more complex task. Here, the DIKCW model provides an integrated approach for this transition. The model suggests that semantic analysis, primarily residing in the “Information” and “Knowledge” stages, can be improved through creative algorithms at the “Creativity” stage, subsequently leading to wise decisions or wisdom.
The DIKCW model also enables the inclusion of semantic features that are often ignored or underutilised. Features like co-reference resolution, anaphora resolution, or identification of rhetorical structures can be effectively processed using creative computing techniques, thereby augmenting the semantic interpretation of text. The end result is a richer, more context-aware semantic representation.

3.2 Regex Representation for Data Layer

Extracting various data types from text is an essential step in syntax analysis, particularly in the context of the DIKCW model. Below is a code snippet that demonstrates how to extract the 13 different data types summarised from a chat input. The code uses regular expressions for pattern matching and the langid library for language identification.

- Alphabetic Data to match alphabetic characters:[a-zA-Z]+
- Numerical Data: -?+(+)?
- Punctuation: [.,;!?]
- Special Characters: [@%&],
- Whitespace: \s
- Date and Time: 4-2-2
- Unicode Characters: \pL
- Hyperlink: $\https?://[^\s]+$
- Casing: [A-Z]+
- Language Markers: [lang=“([a-zA-Z-]+)”]
- Abbreviations: (Dr|Mr|Mrs|etc)
- Acronyms: [A-Z]2
- Mathematical Symbols: [+*/=()]

3.3 Transition of Layers

The SIE mechanism can be integrated into the DIKCW model as Information layer that interacts with both the “Information” and “Knowledge” components. This layer would be responsible for: Extracting syntactic information from raw text data. Transforming this information into a structured format that can be utilized for creative computing tasks. Feeding this structured information back into the DIKCW model to enrich its “Knowledge” layer. SIE contains. There are eight types of information has been summarised when this approach trying to define information layer, along with the formulas that applied to extract each type from the 13 data types. Each of these algorithms can be adapted to consider the 13 data types in DIKCW Model-based syntax analysis. For example, punc- tuation and special characters can be crucial for sentence boundary detection, while Unicode characters may be relevant for named entity recognition in multilingual texts. Mathematical symbols and numerical data can be essential for identifying named entities like equations or dates, respectively. By integrating these algorithms, one can build a robust information layer that effectively transforms raw data into structured, actionable information. Below, this paper outlines the eight types of information that defined, along with the mathematical formulas used to extract each type from the 13 data types that extracted. This research proposes equations (1) to (8) to represent 8 types of information.

Part-of-Speech Tags

$$P(W,T) = \prod_{i=1}^{n} P(w_i \mid t_i) \times P(t_i \mid t_{i-1}) \quad (1)$$

Where W is the sequence of words, T is the sequence of

POS tags, $P(w_i|t_i)$ is the likelihood of word w_i, give tag t_i, and $P(t_i|t_{i-1})$ is the transition probability from tag t_{i-1} to t_i.

Named Entities

$$P(y \mid x) = \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(y, x)\right) \quad (2)$$

Where y is the sequence of named entity tags, x is the sequence of words, $Z(x)$ is the normalisation factor, fi are feature functions and λi are the feature weights.

Sentence Boundaries

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \quad (3)$$

Here $f(x)$ is the classification functions, x is the feature vector for a given token, N is the number of training samples, Yi is the lable of the i-th training sample, $K(xi, x)$ is the knernel function, αi are Lagrange multipliers, and b is the bias .

Syntactic relationships

$$S = \max_{T \in Y} \sum_{(h,m) \in T} \text{score}\,(h, m) \quad (4)$$

Where S is the optimal parse tree, T is a candidate tree, Y is the set of all possible trees, h is the head word, m is the modifier and $score(h,m)$ is the score of the edge between h and m.

Coreference Resolution

$$P(a, a') = \sigma\big(w \cdot \phi(a, a')\big) \quad (5)$$

where a and a' are mentions, *sigma* is the sigmoid function, w is the weight vector, and ϕ is the feature vector.

Sentiment Markers

$$P(c \mid d) = \frac{P(d|c) \times P(c)}{P(d)} \quad (6)$$

Where c is the class label (e.g., positive, negative), and d is the document.

Topic Keywords

$$p(\boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\prod_{d=1}^{D} p(\theta_d \mid \alpha) \prod_{n=1}^{N} p\big(z_{d,n} \mid \theta_d\big) p\big(w_{d,n} \mid z_{d,n}, \beta\big)\right) \quad (7)$$

Here, D is the number of documents, N is the number of words in each document, w represents the observed words, z represents the topic assignments for each word, and θ represents the topic distributions for each document. α and β are Dirichlet priors for the document-topic and topic-word distributions, respectively. This joint probability function encapsulates the generative process of LDA in a single equation, summarising how topics are generated for documents and how words are generated for topics.

Temporal Markers

$$P(X, O) = \prod_{t=1}^{T} P(x_t \mid x_{t-1}) P(o_t \mid x_t) \quad (8)$$

Here, X is the sequence of hidden states representing whether a token is part of a temporal marker or not, O is the sequence of observed tokens, T is the length of the sequence, $P(xt|xt-1)$ is the state transition probability, and $P(ot|xt)$ is the emission probability.

3.4 Universal Information Formula

Combining the eight different mathematical formulas into a single "universal" formula is a non-trivial task because each formula is designed to solve a specific type of problem in natural language processing. However, our approach to integrate these different types of information is to define a composite score or function that takes into account all these individual metrics.
Here's a simplified conceptual representation of a universal formula (9) that combines the eight different types of information:

$$\begin{aligned} U = \alpha \cdot P_{W|T} + \beta \cdot P_{y|x} + \gamma \cdot P_{T|W} + \\ \delta \cdot P_{a_i|a_j} + \epsilon \cdot P_{C|F} + \zeta \cdot P_{w|d} + \\ \eta \cdot P_{h_i|m_i} + \theta \cdot P_{t_i|t_{i-1}} \end{aligned} \quad (9)$$

Where: U is the universal function, W is the sequence of words, T is the sequence of tags or tree structure, y is the sequence of labels (entities), x is the sequence of observed features, C is the class label (e.g., positive, negative), $F$are the features (words or phrases), d is a document, $z$is a topic, h and m are head and modifier words, a are mentions, $\alpha,\beta,\gamma,\delta,\epsilon,\zeta,\eta,\theta$ are weights for each component. Each type of information has a corresponding function that calculates a probability or score for that type of information based on the paragraph text. These are then used to calculate the Universal Score for each paragraph. This dynamic thresholding based on the mean score provides a relative measure of information density, which can be more informative than using an arbitrary fixed threshold. Paragraphs with a Universal Score above the mean are labeled as "High Information Density", while those below are labeled as "Low Information Density".

3.5 speaker intention extraction

Data and information has been extracted from previous stage of research, next step need to identify the knowledge type in knowledge layer for input, one of our approach is to use a weighted scoring system that takes into account the different types of data and information have been gathered. This system can be designed as a combination a set of rules and scoring system that classify the knowledge into one of the six types: Tacit, Explicit, A Priori, Procedural, Empirical, and Meta-knowledge.

3.6 Knowledge Identification Rules (3)

Our research proposed a Rule based Knowledge Identification (RKI) algorithm to facility our objectives. First the universal score Uk can be calculated using the given data and information types. Then use Uk to calculate fk(Uk) for each knowledge type k. Finally, to identify K = aargmaxkfk(U).

The rule-based system to classify the knowledge type based on the types of information and data. For example: If the information contains a lot of part-of-speech tags and syntactic relationships, it may be more aligned with Procedural Knowledge. If the information contains name entities and topic keywords, it may be Explicit Knowledge. First in equation (10) Let I be the set of all information types and D be the set of all data types. Let WI and WD be the weight vectors for information and data types, respectively. The universal score for knowledge U_k can be calculated as:

$$U_k = \sum_{i \in I} W_{I_i} \times I_i + \sum_{d \in D} W_{D_d} \times D_d \tag{10}$$

The knowledge type in equation (11) K can then be identified as:

$$K = \arg\max_{k \in \{\text{ Tacit, Explicit, A Priori, Procedural, Empirical, Meta }\}} \tag{11}$$

where *fk* is a function that maps the universal score *Uk* to a likelihood score for each knowledge type *k*.

To identify the knowledge type, this approach proposes use a weighted scoring system that takes into account the different types of data and information. According to previous research, the relationships has been summarised between knowledge and information as:

- Explicit Knowledge: Presence of Name Entities, Topic Keywords
- Tacit Knowledge: Presence of Sentiment Markers, Coreference Resolution
- A Priori Knowledge: Presence of part-of-Speech Tags, Syntactic Relationships
- Procedural Knowledge: Presence of Sentence Boundaries, Temporal Markers
- Empirical Knowledge: High Universal Score
- Meta-Knowledge: Presence of multiple types of information

And to use the universal score to gauge the information density. For example:

- High Density (>30): Explicit, high Universal Score and High Information Density as it often represents facts and well-defined information.
- Medium Density (15-20): Empirical. Low to Medium Universal Score, presence of numerical data or terms that are often used in observations or reports.
- Low Density (<15): Tacit, low Universal Score and "Low Information Density", possibly due to a presence of pronouns which often indicates subjective or experience- based knowledge.

4. Case Study

This case study will first explain the dataset to understand its structure and contents. After that, a in-depth analysis will be processed, including creative visualisations and advanced statistical analyses,

4.1 Sample Analysis

This research approach creates a novel algorithm to seek the speaker's intentions based on DIKCW model. By applying the algorithm into python code, this research gen- erates a list of result. The dataset comprises 200 records, each containing a sentence along with a Universal Score and various categorical attributes such as Information Density, Knowledge Type, and Speaker Intention. The aim of the in-depth analysis was to explore the relationships among these variables, employing both conventional and creative analytical techniques.

The result contains 6 columns which reflect this the main approach:

- Sentence: This contains textual data, presumably a set of sentences that were analysed.
- Data Types: A dictionary-like string that appears to list the different types of data present in each sentence (e.g., Alphabetic words).
- Universal Score: A numerical score assigned to each sentence, potentially indicating its importance or relevance.
- Information Density: Categorised as either "Low Information Density" or "High Information Density," this column presumably indicates the amount of information packed into the sentence.
- Knowledge Type: This categorises the type of knowledge conveyed by the sentence (e.g., Empirical, Explicit, Procedural).
- Speaker Intention: Specifies the intent behind the sentence, such as being informative or speculative.

4.2 Visualisation

The first visualisation result is to complete the sentiment anlaysis in Figure 3. Sentiment polarity was calculated for each sentence, and its correlation with Universal Score was explored. The analysis revealed a weak correlation, suggesting that sentiment polarity and Universal Score are largely independent.

The 3D scatter plot in Figure 4 visualises the relation- ship between "Universal Score," "Knowledge Type," and "Speaker Intention." The color gradient represents the Universal Score, with darker colors indicating lower scores and lighter colors indicating higher scores.

This creative visualisation provides a multidimensional view of the dataset and may prompt further investigations into the relationships between these variables and the Universal Score seems to have a range of values across all encoded categories of Knowledge Type and Speaker

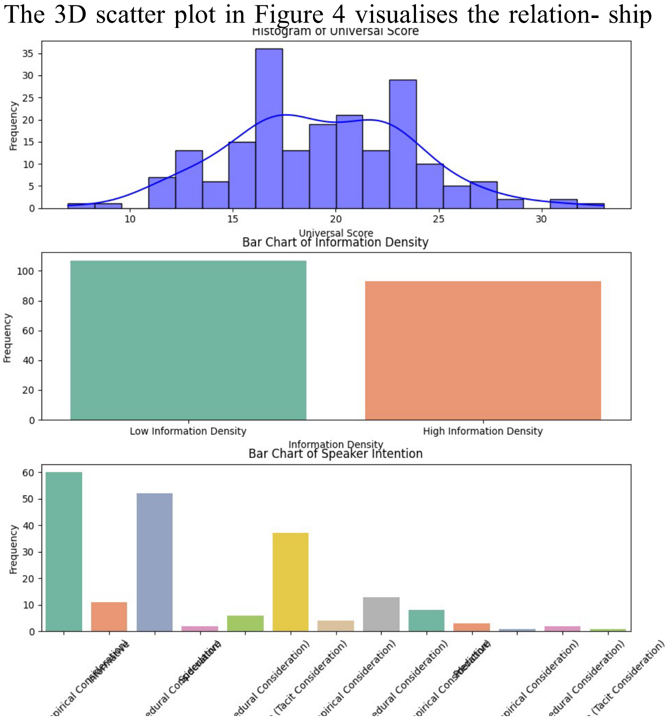

Figure 3. Sentiment Analysis

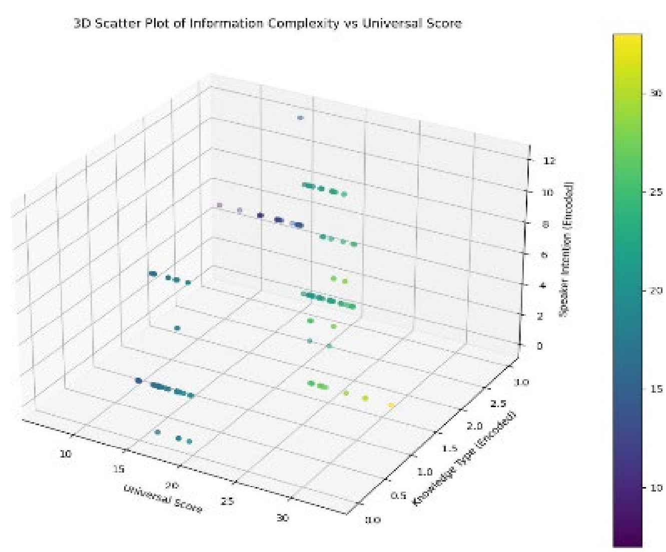

Figure 4. 3D Scatter Plot

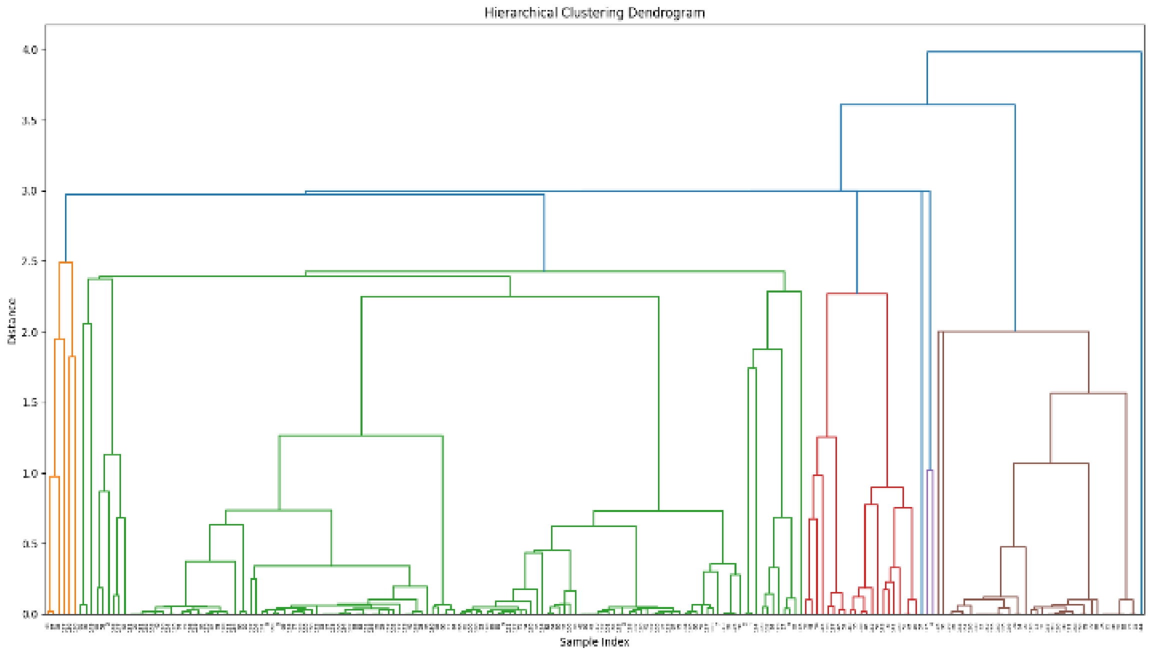

Figure 5. Hierarchical Clustering

TABLE III
Chi-Square Analysis

	Chi-Square Statistic	p-value
Information Density	196.0005	1.55832e-44
Knowledge Type	184.8022	8.09707e-40
Speaker Intention	185.6966	2.88373e-33

TABLE IV
Correlation Analysis

	Correlation Coefficient	p-value
KnowledgeType	-0.1221	0.0850126
SpeakerIntention	0.0210	0.768214
Sentiment	0.0479	0.500241

Intention. Whereas Knowledge Type and Speaker Intention: appear to have a complex relationship with Universal Score, as evidenced by the dispersion of points in the 3D space.

The dendrogram in Figure 5 provides a hierarchical representation of the clustering of sentences based on their "Universal Score," "Knowledge Type," and "Speaker Intention." The y-axis represents the distance between clusters, with lower values indicating similar clusters and higher values indicating dissimilar clusters.The x- axis represents individual samples (or sentences) in the dataset. The numbers in the brackets indicate the number of samples included in each cluster.

4.3 Validation

In all the combinations tested, the p-values are significantly below the commonly used alpha level of 0.05, suggesting that the variables are not independent and are indeed related to each other.

The Chi-Square statistics also indicate a very strong relationship, further confirming that these variables are not independent. The Table 3 summarises the Chi-Square test results between the Universal Score (binarized around its

mean) and the categorical attributes:
The Table 4 summarizes the point-biserial correlation coefficients between the Universal Score and the encoded categorical variables, along with Sentiment:
Universal Score and Information Density: The correlation coefficient of -0.817 indicates a strong negative relationship. The extremely low p-value suggests that this correlation is statistically significant. Universal Score and Knowledge Type: The correlation coefficient of -0.122 indicates a weak negative relationship, and the p-value of 0.085 suggests that this is not statistically significant at the 0.05 level. Universal Score and Speaker Intention: The correlation coefficient of 0.021 indicates a very weak positive relationship, and the p-value of 0.768 suggests that this is not statistically significant.
These analyses indicate strong relationships between the variables "Information Density," "Knowledge Type," and "Speaker Intention," as supported by the Chi-Square tests. The Universal Score also shows a strong negative correlation with Information Density but has weak or negligible relationships with the other categorical variables.
The in-depth analysis of the dataset using both conventional statistical methods and creative computing techniques revealed complex relationships among the variables. While Universal Score demonstrated a strong negative correlation with Information Density, its relationship with other variables was not as definitive. Creative analyses like sentiment analysis and hierarchical clustering offered novel perspectives but also confirmed the complexity of these relationships. The findings warrant further investigation and could be instrumental in understanding the intricate dynamics of text-based data in various applied domains.

5. Conclusion

The research successfully introduces a novel Data, Information, Knowledge, Creativity, and Wisdom (DIKCW) Model for multi-dimensional text analysis. This comprehensive model serves as a robust framework for understanding and interpreting textual data from differ- ent dimensions—Data, Information, and Knowledge. The study's results substantiate the model's efficacy by revealing strong associations between Information Density, Knowledge Type, and Speaker Intention. These empirical findings confirm that the variables are intricately interlinked and not independent of each other, thus validating the model's underlying hypothesis. The utilisation of creative computing techniques, rule-based reasoning, and visualisation contributes to the model's robustness, laying a foundational pillar for future research in understanding the complex relationships between data, information, and knowledge in the context of creativity and wisdom. Chi-square tests revealed strong associations between Information Density, Knowledge Type, and Speaker Intention. Particularly, the extremely low p-values (< 0.05) signify that these variables are not independent and share a complex relationship. Point-biserial correlation analysis between Universal Score and the categorical variables showed a strong negative relationship with Information Density, though it demonstrated weak to negligible corre- lations with Knowledge Type and Speaker Intention.The next phase should involve extending the DIKCW model to encompass pragmatic analysis. The goal is to introduce algorithms that can extract the speaker's social or cultural context, aiming for a more holistic understanding of the text. This will enable the model to not only categorise knowledge but also to interpret the conditions or motivations that led to the formation of that knowledge.

References

[1] B. Nandwalkar, S. Pardeshi, M. Shahade, and A. Awate, "Descriptive Handwritten Paper Grading System using NLP and Fuzzy Logic", International Journal of Performability Engineering, vol. 19(4), pp. 273-282, 2023.
[2] M. Yadav, and I. Kumar, "Image Processing-Based Transliteration from Hindi to English", International Journal of Performability Engineering, vol. 19(5), pp. 334-341, 2023.
[3] J. Rowley, "The Wisdom Hierarchy: Representations of the DIKW Hierarchy", Journal of Information Science, vol.33(2), pp.163-180, 2007.
[4] M. Zeleny, "Management Support Systems: Towards Integrated Knowledge Management", Human Systems Management, vol.7(1), pp.59-70, 1987.
[5] T. H. Davenport, and L. Prusak, "Working knowledge: How Organizations Manage What They Know", Harvard Business Press, Boston, 1998.
[6] D. L. Sackett, W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson, "Evidence-Based Medicine: What it is and What it isn't", British Medical Journal(BMJ), vol.312(7023), pp.71-72, 1996.
[7] R. L. Ackoff, "From Data to Wisdom", Journal of applied Systems analysis, vol,16(1), pp.3-9, 1989.
[8] M. A., Runco, and G. J. Jaeger, "The Standard Definition of Creativity", Creativity Research Journal, vol.24(1), pp.92-96.2012.
[9] P. B. Paulus, and N. B. Aijstad, "Group Creativity: Innovation Through Collaboration", Oxford University Press, Oxford, 2003.
[10] A. Cropley, "In Praise of Convergent Thinking", Creativity Research Journal, vol.18(3), pp.391-404, 2006.
[11] R. J. Sternberg, and T. I. Lubart, "The Concept of Creativity: Prospects and Paradigms", Handbook of Creativity, vol.1, pp.3-15, 1999.
[12] B. Pang, and L. Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, vol.2(1–2), pp.1-135, 2008.
[13] J. Wang, and Z. Wang, "Data Layer in the DIKW Pyramid: A Comprehensive Survey", Journal of Data and Information Management, vol.3(2), pp.49-60,

2018.
[14] D. Jurafsky, and J. H. Martin, “Speech and Language Processing”, 3rd ed., Stanford University, pp.123-156, 2019.
[15] C. D. Manning, and H. Schütze, “Foundations of Statistical Natural Language Processing”, MIT Press, Cambridge, MA, pp.205-230, 1999.
[16] J. L. Mey, “Pragmatics: An Introduction”, 2nd ed., Blackwell Publishers, Oxford, pp.90-113, 2001.
[17] D. M. Blei, A. Y. Ng, and M. I. Jordan,“Latent Dirichlet Allocation”, Journal of Machine Learning Research, vol.3, pp.993–1022, 2003.
[18] P. C. Tetlock, and B. A. Mellers, “The Imperfectly Rational Individual”, Behavioral and Brain Sciences, vol.41(1), pp.32-57, 2018.
[19] P. Domingos, “A Few Useful Things to Know About Machine Learning”, Communications of the ACM, vol.55(10), pp.78-87, 2012.
[20] B. Mittelstadt, and etc., “The Ethics of Algorithms: Mapping the Debate”, Big Data and Society, vol.3(2), pp.1-21, 2016.
[21] H. Witten, and etc., “Data Mining: Practical Machine Learning Tools and Techniques”, 4th ed., Morgan Kaufmann,Burlington, pp.340-370, 2016.
[22] N. Chambers, and D. Jurafsky, “Unsupervised Learning of Narrative Event Chains”, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp.789-799, 2008.