

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining- Based Approach

*Ting Zhang**

Modern Postal College
Chongqing University of Posts and Telecommunications
Chongqing, 400065, China
Email: zt2826616176@163.com

*Corresponding author

Received 1 June 2023; Accepted 29 June 2023

Abstract. In recent years, driven by digital intelligence technologies such as big data, cloud computing, The Internet of Things and artificial intelligence, various industries are accelerating their digitalization and how to effectively use digital technologies to enhance the position of each industry in the global value chain has become a practical problem that needs to be studied. Although digital technologies are widely used in manufacturing industries, few studies have discussed in depth the application of digital technologies in logistics enterprises, especially how to provide guidance and support for the digital transformation of logistics enterprises with the help of the digital intelligence technology lexicon. In order to promote the digital transformation of logistics enterprises, the study constructs a digital intelligence technology dictionary for logistics enterprises based on text mining technology to improve the efficiency and quality of digital transformation of logistics enterprises. The study used Python crawler technology to crawl the annual reports of logistics enterprises from 2012-2021, as well as the relevant textual information about logistics from the State Council. Subsequently, the study extracted keywords closely related to digital and intelligent transformation through keyword extraction text mining technology, and at the same time, combined with experts' research results on digital transformation, further analyzed and summarized the keywords to build a digital intelligence technology dictionary for logistics enterprises. The content of this digital intelligence technology dictionary includes key technologies for the digital and intelligent transformation of logistics enterprises, aiming to provide support and reference for the digital transformation of logistics enterprises in order to promote digital upgrading and the intelligent transformation of the industry.

Keywords: logistics enterprises; digital intelligence technology; text mining; word frequency

1. Introduction

With the maturity and development of artificial intelligence technology, cloud computing

Ting Zhang

technology, blockchain technology, and big data technology, the digital economy can change the concept of consumers from the traditional pursuit of "possession" to the pursuit of shared "use", improve the efficiency of mining and reuse of existing resources, and reduce the opportunity cost of production factors [1]. During the two sessions of the National People's Congress, Xu Guanju, a deputy to the National People's Congress and chairman of Chuanhua Group, believes that digital intelligence technology is the winner of the future and will play an indispensable role in the construction of a high-standard market system. In this context, developed Western countries have proposed to vigorously promote the digital transformation of traditional industries, and the Chinese government has also introduced relevant policies to create a favorable macro environment. In particular, in May 2022, the General Office of the State Council released the "14th Five-Year Plan for Modern Logistics Development", which clearly sets the promotion of logistics quality, efficiency and cost reduction as the important task of modern logistics development in the current period. The digital revolution profoundly affects all industries, and the logistics sector, as an important link supporting supply chain operations, is not immune to the impact of digital technology. According to Wang et al., smart cities and logistics in China have grown significantly due to the influence of the situation [2]. With the increasing scale of manufacturing and trading operations, logistics companies are playing an increasingly important role in the supply chain but at the same time facing greater challenges. To reduce costs and improve efficiency, the use of digital intelligence has become an inevitable choice for logistics companies.

RFID, the Internet of Things, big data analysis, barcodes, automatic mobile carts and other advanced digital intelligence information technologies are widely used in logistics enterprise management. These technologies greatly improve the efficiency of all aspects of logistics, such as warehousing, sorting, loading and unloading, transportation, and processing and distribution [3]. In order to achieve the visualization, traceability and control of factor flow, logistics enterprises should continuously develop digital and intelligent technologies. They should also apply digital capabilities to improve the corporate environment, promote market development, enhance the effectiveness of information platforms and build social integrity mechanisms. Therefore, the article prepares to construct a small dictionary of digital intelligence technology in logistics. Its purpose is to explore the application of digital intelligence technology in logistics enterprises, assess their level of digital intelligence technology, and provide suggestions for achieving high-quality development.

In the study, we clarified the following objectives: (1) To obtain annual reports of logistics enterprises and textual data about logistics as the primary data source for the study. (2) Apply text mining technology to extract correlation information about logistics technology through keyword extraction methods. This will help reveal the current state of logistics enterprises in terms of their application of digital intelligence technology. (3) Analyze and summarize the keywords to build a dictionary of digital intelligence technology for logistics enterprises, aiming to promote the high-quality development of logistics enterprises. To ensure the reliability and accuracy of the study, we utilized the Jieba disambiguation tool during the dictionary construction process. We also conducted comparisons and verifications with existing standard dictionaries to ensure the precision and consistency of the vocabulary. At the same time, we employed screening and disambiguation strategies to handle multiple matching results and cases of ambiguity in

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

order to determine the final results of word separation. Through the disambiguation and word frequency analysis of annual reports from logistics companies, we gained valuable insights into the technological advancements in logistics companies.

2. Literature review

Digital technology is a combination of information, computing, communication and connectivity technologies [4]. Through further development, Sebastian et al. provide a more explicit elaboration of digital technologies, such as big data, cloud computing, blockchain, Internet of Things, artificial intelligence, and virtual reality technologies [5]. Digital intelligence technology, i.e., digital technology and intelligent technology, includes both digital technologies such as cloud computing, big data technology, Internet of Things, blockchain, 5G communication, and intelligent technology represented by intelligent robots, image and language recognition, natural language processing, machine learning, and neural networks. In academic circles, foreign scholars often refer to technologies such as big data, block chain, Internet of Things and artificial intelligence as digital technologies, while domestic scholars sometimes translate these technologies as digital intelligence technologies since the concept of "digital intelligence" was first proposed by a group at Peking University in 2015, but there is no significant difference between the two terms.

Digital technology was once seen as a simple tool for improving productivity and providing better data management and information services. But as digital technology has innovated and evolved, people have begun to realize that its importance extends far beyond that. Digital technology has brought about tremendous impact and change. Digital technology has become a transformative technology that impacts all aspects of production methods, management styles, and organizational structures. Not only that, the widespread application of digital technology has also triggered extensive social, economic and cultural changes and reconfigurations, driving the intelligence and transformation and upgrading of different industries. Therefore, an increasing number of scholars believe that digital technology is the "engine" that drives the whole social, economic and cultural changes rather than just a simple tool. As an important part of modern information technology, digital technology is leading human society towards a more digital, intelligent and sustainable future, and its impact and development will certainly reveal a new chapter in history. Many researchers have confirmed that digital technologies can pave the way for functional changes that bring operational benefits [6] and competitive advantages to companies through monitoring and optimization [7]. IoT technologies enhance the connection between items through information exchange and improve the transparency of information, enabling efficient management and control of logistics processes [8]. With the increasing popularity of artificial intelligence technology, many modern logistics companies are trying to use it to optimize logistics processes and improve logistics efficiency [9]. As blockchain is considered a solution to reliably connect and manage IoT devices, logistics may be one of the most promising application areas because of the large number of possible IoT objects (e.g., vehicles, goods, etc.) in the logistics environment. Obukhova et al. state that companies should introduce new technologies and use digital technologies to change existing business processes, including communication, distribution, business relationship management, and other related processes [11].

In order to accurately assess the role and contribution of digital intelligence technology in enterprise development, you can refer to the enterprise's digitalization degree

Ting Zhang

index. Li and other scholars used principal component analysis to synthesize the three dimensions of digital investment, digital technology and business model transformation to form the enterprise digital transformation index [12]. Chen-Yu Zhao extracted word frequencies from the annual reports of manufacturing listed companies and the sample data of companies with more successful digital transformation, and after standardizing the word frequency data, he used the entropy value method to determine the weights of each index to obtain the total digital development index of manufacturing listed companies [13]. Wu Fei et al. used the content of the annual reports of enterprises as the research object, used the text mining method to construct a dictionary of digital transformation in the manufacturing industry, and used the natural logarithm of the summed word frequencies as a measure of the digital transformation index of enterprises [14]. Tu and He et al. scholars measured the level of digital transformation by calculating the frequency of each word appearing in the annual reports and taking the natural logarithm [15].

The findings of these research scholars show that with the wide application of digital intelligence technology in logistics activities, logistics enterprises have an increasingly urgent need for digital intelligence technology terminology, yet there is no dictionary of digital intelligence technology specifically for this industry. In addition, the study also found that applying text mining technology to evaluate enterprise digital intelligence technology has feasibility in practice, which provides useful guidance and support for the digital intelligence process of logistics enterprises. In view of this, the study takes logistics enterprises as the research object and uses text-mining technology to analyze their annual reports in order to build a digital intelligence technology lexicon applicable to the logistics field.

3. Data acquisition

3.1. Select data

(1) Enterprise annual report

The annual report of an enterprise contains rich information and data, such as the enterprise's development plan and business strategy. By analyzing this information, we can find out the future development direction and focus of the enterprise and can predict the future development trend of the enterprise, which provides an important reference for enterprise management and investment decisions. Therefore, taking the A-share listed logistics companies in China's Shanghai Exchange and Shenzhen Exchange as the research objects, we adopt the combination of the annual report text analysis method and policy text data to explore the metrics of digital intelligence technology of logistics enterprises.

The information in annual reports can be divided into hard information and soft information [16]. Hard information is usually recorded in numerical form and is quantitative, such as financial information. While soft information is usually recorded in textual form and is qualitative, such as management discussion and analysis information (MD&A). Durnev et al. [17] argue that the content in management discussion and analysis (MD&A) reports can improve the efficiency of investment. Therefore, we use the content of the (MD&A) part of the company's annual report as the target of analysis. In 2012, IBM proposed digital transformation, emphasizing the use of digital technology to reinvent customer value and enhance customer interaction and collaboration, and the application of digital technology began to be studied by scholars, so the sample time of this paper is set from 2012-2021, covering the last decade of data.

(2) Policy Text

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

Policy texts can reflect government management priorities, policy orientations and decision-making directions, etc., and are an important basis for analysis and research in a certain field or in the development of a certain enterprise or industry. At the same time, the policy text contains a large amount of information and data, which can be used as a data source for analysis. Therefore, policy texts can also provide important reference data for governments, enterprises and researchers to guide decision-making and research directions. Since the data period of annual reports of enterprises is from 2012 to the end of 2021, the release of national policies is superior to that of enterprises and will be based on the feedback from enterprises, so we chose the policy text with a deadline of one year later than the annual reports of enterprises, i.e., until the end of 2022.

3.2. Data acquisition

Specifically, the data collection process consists of two main steps. First, during the data collection process, we conduct a large-scale search and crawl of the Shanghai and Shenzhen exchanges and Chinese government websites using a series of Python web crawling techniques, such as Selenium and other frameworks. By setting relevant keywords and search terms, we obtain as large a sample of data as possible to ensure the breadth, adequacy, and representativeness of the data sources [25], using some of the code shown in Figure 1.

Next, we employ a series of integration and pre-processing operations, including steps like data filtering and de-weighting, to standardize and make the data more usable. In the process of integrating and processing the data, we perform a large number of statistical analyses, such as calculating the total amount and frequency distribution of the data, to gain insights into the overall characteristics and trends of the data. The results of these analyses provide a solid foundation and basis for subsequent research work and also help us better understand the meaning and patterns in the data. By continuously applying various technical means and tools, such as data cleaning scripts and data processing libraries, we end up with a highly accurate and consistent dataset that allows us to better analyze and interpret the acquired data. This data collection and processing process makes our study more reliable and scalable and provides strong data support for further applications and research.

3.3. Data results

The data acquisition method used in this paper is based on the Selenium framework, a Python web crawler technology. The aim is to search and crawl the annual reports of listed companies in the logistics industry from 2012 to 2021 and store them according to specific classifications. By specifying relevant keywords and conditions, we obtained as many data samples as possible to ensure the adequacy, representativeness, and accuracy of the dataset. Specifically, we obtained the 10-year annual report data of 117 listed logistics companies, which included a total of 967 PDF documents. The Shanghai Exchange accounted for 648 annual report data from these PDF files, while the Shenzhen Exchange provided 319. To enrich our data acquisition, we also searched for the term "logistics" on the State Council Policy Document Library on the Chinese government website to extract logistics-related policy texts. As of December 15, 2022, we have collected 15 policy texts, which serve as important references for understanding policy positions and forms.

In conclusion, this paper categorically stores data from the 2012-2021 annual reports of

Ting Zhang

listed companies in the logistics industry and ensures the accuracy and usability of the data through steps such as data integration and pre-processing. Additionally, we obtained logistics-related policy texts, which provide important information for studying the logistics industry's development trends and policy forms. This is of significant help and importance for our subsequent research and analysis work.

```
def pdf_down(self, json_str):...
    url_data = json.loads(json_str)
    # Iterate through the container
    j = 0
    while j < len(url_data):...
        # Get the dictionary
        name = url_data[j]["title"]
        # Filter available data
        if name.rfind('Report Summary') == -1 and name.rfind('Canceled') == -1 and name.rfind('English') ==
            time_out = (random.random() * 3 + 3)
            time.sleep(time_out)
            # Remove special characters
            if name.rfind('*') != -1:...
                name = name.replace('*', 'a', 1)
            # splice download pdf path
            url = "http://www.szse.cn/api/disc/info/download?id=" + url_data[j]["id"]
            myPdf = requests.get(url, allow_redirects=True)
            # saveFile: download to the local path
            saveFile = "C:/Users/86156/Desktop/Shenzhen Stock Exchange/" + name + ".pdf"
            open(saveFile, "wb").write(myPdf.content)
            print(name + ",download complete")
        j += 1
```

Figure 1: Code snippet showing the part of the code used for downloading annual reports from the Shenzhen Stock Exchange

4. Data pre-processing

4.1. Data cleaning

In the data processing process of this article, we used the Xunjie PDF converter to extract the text from PDF files and convert them into a usable text format. For the annual reports of listed companies from 2012-2014, 2015, and 2016-2017, we needed to process different disclosure formats to extract the sections "Report of the Board of Directors", "Management Discussion and Analysis", and "Operating Conditions Discussion and Analysis". We converted these sections into plain text files in TXT format for storage. During the conversion process, we encountered certain issues that could result in abnormal text, such as partial blank or excessively short text lengths, due to problems during the crawling process. These abnormal texts do not convey specific semantic information and may affect the accuracy of text mining results. Therefore, we processed these abnormal texts to improve the accuracy and reliability of text mining and analysis.

We performed a series of processing steps on these abnormal texts, including text filtering, cropping, repairing, and merging. We used the text-completion method for the abnormally short texts to repair the missing parts by combining them with the preceding and following text. Blank texts were removed using a filtering tool, while excessively long texts were trimmed to an appropriate length using the text-trimming method. Texts affected by page segmentation errors were also processed using text merging to ensure consistency and coherence in text semantics.

Finally, by processing the anomalous texts, we obtained a high-quality text dataset that enables us to conduct text mining and analysis better. This data processing step serves

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

as an important guarantee and foundation for subsequent research and analysis.

To ensure the high quality of the collected text data, this paper includes a series of processing steps. During this process, we adopted a search method to directly eliminate blank texts and texts shorter than 5000 characters in length. This step aims to remove text passages with no substantial content or that are too short, ensuring the accuracy and reliability of subsequent text mining. These processing steps eliminate the interference of abnormal texts on analysis results and improve the effectiveness of text mining. Simultaneously, retaining longer texts with actual content helps extract more meaningful information and provides a more reliable basis for subsequent data analysis and mining.

When processing policy documents, each of the 15 policy documents is converted into plain text files in TXT format while retaining the independence of each policy document. This approach allows for a better analysis of differences and commonalities among the policy documents, furthering our understanding of policy forms, content, and implementation. Furthermore, converting policy documents into TXT format facilitates subsequent text mining and analysis, enabling better exploration of the information and principles contained within the policy documents. Relevant code snippets used can be found in Figure 2.

```
# Performs logic based on year.
# 2012-2014 Board of Directors Report - Section Important Matters
if time <= 2014 and time >= 2011:
    # First grouping
    s1 = content.split("Section Board Report I")
    if len(s1) == 2:
        content_1 = s1[1]
        # Second grouping
        s2 = content_1.split("Section Important Matters I")
        # Output txt
        with open(save_url, "w", encoding="utf-8") as file_w:
            file.write(s2[0])
    else:
        txt_contain.append(txt_name)
```

Figure 2: Partial code for board report content extraction

4.2. Chinese word separation

After the data cleaning of the text, it is necessary to carry out the word separation process to segment the text into individual words or phrases based on word boundaries so that the information expressed in the text content can be more easily extracted through the identification of keywords. In this paper, we have chosen a string-matching method for Chinese word segmentation. To perform string matching, a dictionary containing many words or phrases must be prepared in advance. The size of the dictionary is usually large to cover as many words as possible. The string statements to be segmented are compared with the words in the dictionary one by one. Starting from the beginning of the statement, consecutive character fragments are compared with words in the dictionary. If a matching

Ting Zhang

word is found, it is considered a segmentation result.

Jieba is a commonly used Chinese word segmentation library in Python, which adopts more advanced word segmentation algorithms. It has higher efficiency and better stability compared to other word segmentation libraries, enabling quick and accurate Chinese word segmentation. In order to improve the recall rate of the search engine word ranking, this paper introduces an improvement to the exact mode, which involves re-segmenting long words (word count >2). Additionally, the 'cut_all' parameter of the 'jieba.cut' method is utilized. The default value of the 'cut_all' parameter is False, indicating the exact mode, and it attempts to provide the most accurate results for word segmentation. This processing approach helps capture key information in long words, generates more segmentation results, resolves some ambiguities, improves the accuracy and comprehensiveness of word segmentation, and better meets the requirements of search engines. In this paper, we employ the search engine model of the 'jieba' word segmentation library to aggregate the text collections from 15 policy documents, 648 SSE annual reports, and 319 SZSE annual reports, respectively, utilizing the string matching method."

4.3. Sub-word processing

4.3.1. Deactivation of words

During the data processing, in order to avoid the over-processing of English participles, we only filter out the Chinese participles and exclude the processing of English participles. The purpose of doing so is to preserve the characteristics of the English language during the development of natural language processing algorithms in order to prevent the incorrect processing or rejection of English subwords. At the same time, we have observed that there are many high-frequency words that have no actual meaning in the text data. These words occupy significant computational resources and storage space during natural language processing, which reduces both processing efficiency and accuracy. Therefore, we adopt a common text preprocessing method, which removes frequently used words that lack meaning.

By adopting this method, we successfully preprocessed the data and eliminated stop words from the text. Such data preprocessing operation can effectively reduce computation and processing time, while improving the accuracy and efficiency of data processing. Additionally, the processed data can better represent the original text and contribute to the successful implementation of subsequent text mining and data analysis.

Common stop words include prepositions, dummy words such as the verbs "is", "have", and "in", conjunctions, "the", "had", and other interjection words, pronouns such as "he", "she", and other personal pronouns, common numbers, symbols, and other words used to connect semantics. Jieba's word splitter comes with different Chinese stop words lists, which cover common Chinese stop words, including prepositions, conjunctions, verbs, pronouns, and common symbols and numbers. Since these words are very common, removing them does not change the content of the original text. On the contrary, it can better emphasize the topic of the text. Part of the code used for filtering custom stop words is shown in Figure 3. Therefore, this paper chooses to use Jieba's predefined stop word list to filter these meaningless stop words when performing the word separation process.

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

```
import jieba.posseg as posseg

# Custom stopwords table
stopwords = ['preposition', 'is', 'has', 'in', 'of', 'had', 'he', 'she'] # Custom stopwords list

def remove_stopwords(text):
    words = posseg.cut(text) # Use lexical annotation to filter the stopwords
    result = []
    for word, flag in words:
        if word not in stopwords and flag != 'uj':
            result.append(word)
    return result
```

Figure 3: A snippet of the code used to filter custom stop words.

4.3.2. Processing missing values

During the data processing of this paper, we discovered missing values in the obtained annual report data, and some of the annual reports were published in image format, which could not be directly extracted or converted into text. To address this issue, we manually checked the annual report data and identified the non-textual content to process and fill in the missing values accordingly. Specifically, we manually observed and examined the annual report data to identify non-textual parts, such as images, charts, and other non-textual elements. For these non-textual parts, we utilized a specific text string, “picture missing,” to mark the absent picture content. This allowed the subsequent analysis and mining processes to identify and handle the missing values. In doing so, we avoided inaccurate or incomplete data analysis and mining while ensuring data integrity.

To ensure data accuracy and usability, we created special markers to indicate missing non-textual content. This approach facilitated the effective identification and handling of missing values during subsequent analysis and mining. Specifically, during the data processing in this paper, we created a special string, “picture missing,” to represent missing picture content. Part of the code utilized for this purpose is shown in Figure 4. Through this approach, we accurately identified and processed the missing values, ensuring the integrity and reliability of the data throughout the classification, storage, and text-mining processes.

In summary, this paper employed a method of manually checking the annual report data and identifying non-textual content. It utilized special tags to handle missing values and employed the specific text string “picture missing” to mark missing picture content. This data preprocessing method effectively ensured data accuracy and usability. Additionally, in subsequent text mining and data analysis, we accurately identified and processed missing values, resulting in more accurate and representative outcomes.

4.3.3 Screening and disambiguation

In text mining, screening and disambiguation are very critical and complex processes. During the disambiguation process, there may be multiple matching disambiguation schemes or ambiguities. This means that the best disambiguation result needs to be selected from multiple candidate schemes. Therefore, when performing screening and disambiguation, we need to comprehensively analyze the text in order to determine the final disambiguation result, eliminate ambiguity, and ensure the accuracy and semantic consistency of the disambiguation.

Ting Zhang

```
import pandas as pd

# data containing missing values has been retrieved from the annual report data
data = {'company name': [], 'annual report content': []}
df = pd.DataFrame(data)

# marker missing value as a specific text string
missing_value_marker = 'picture missing'
df_filled = df.fillna(missing_value_marker)

# Output the processed data
print(df_filled)

import re
```

Figure 4: Part of the code for processing missing values of non-textual content

Generally speaking, after word separation, we need to filter and disambiguate the obtained word separation results. The specific steps include the following:

- (1) Determine the best candidate word separation results through context analysis and understanding.
- (2) For the disambiguation results, determine the correct disambiguation result based on the context and background.
- (3) If multiple matches occur, select the longest word as the final result. Longer words generally have clearer semantic meanings and can reflect the text content more accurately.
- (4) Filter and remove punctuation marks and stop words appearing in the results. These punctuation marks and stop words do not substantially help the text mining and analysis process but only add noise and interference.
- (5) Process and correct any ambiguities in the disambiguation results through manual or automatic methods, ensuring the accuracy and reliability of the disambiguation results.

Through the aforementioned screening and ambiguity processing steps, we can better address the challenges that arise in the word separation process, making the results more accurate and meaningful. This provides a more reliable and accurate foundation for subsequent text mining and analysis.

4.3.4. Word classification result dimensionality reduction

Based on the aforementioned methods, the numbers of subwords obtained after applying the aforementioned text information are 3144, 60762, and 31912, respectively. Since the number of subwords is too large, we refer to a series of classical academic literature to form a custom dictionary and study only the number-wise technology within the custom dictionary. Among them, Chen-Yu Zhao screened and summarized keyword combinations from digital technology applications, Internet business models, smart manufacturing, and modern information systems and used Python for text analysis and word frequency statistics [13]. On the other hand, Wu Fei et al. constructed a list of keywords related to artificial intelligence, blockchain, cloud computing, and big data by reviewing policy documents and research reports [14]. This academic literature provides theoretical support for the digital transformation of enterprises and important references for identifying the

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

characteristic words of digital intelligence technologies.

4.3.5. Writing and importing dictionaries

It is not enough to obtain a custom dictionary using the above method; it is also necessary to supplement the crawled policy text dictionary and filter and select the words in it according to the characteristics and needs of the logistics digital intelligence technology field. In terms of policy documents, we supplement them with a series of technology-oriented policy documents to further enhance the characteristic keywords. These policy documents provide guidance and support for the development of physical logistics enterprises, which contain important requirements and promotion measures for digital intelligence technologies. By referring to these policy documents, we can obtain a more comprehensive and practical list of characteristic words for digital intelligence technology that suits the needs of the actual logistics business. The screened words are compiled into the dictionary document following the corresponding dictionary document format specification. In determining the characteristic words related to the number-wise technology for logistics enterprises, this paper takes into account the perspectives of both the academic and industrial fields and forms the final sub-word dictionary based on the following classification principles. Finally, the prepared dictionary file is imported into the text analysis tool to enable recognition and utilization of the words in it.

After obtaining the custom small dictionary, to ensure the accuracy and reliability of the index measurement, further processing is conducted when calculating the word frequency of the number-wise technology in this paper. Negative words such as “no” and “none” are excluded from the keywords. Figure 5 shows some of the codes used for excluding negative expressions before the keywords. This step is taken to exclude cases that are related to digital intelligence technology but express negative meanings, in order to ensure the accuracy of the index measurement.

```
def remove_negations(text):
    negations = ['none', 'none', 'no'] # Define a list of negations
    pattern = r'\b(?:{})\s*'.format('|'.join(negations))
    processed_text = re.sub(pattern, '', text)
    return processed_text

# Keep the path
fi = input("set save name")
```

Figure 5: Some codes for excluding negative expressions before keywords

5. Keyword extraction and dimensional structure

5.1. Keyword extraction process

In order to extract keywords related to digital intelligence technology and construct the dimensional structure, this study adopts a custom dictionary-based approach to build a digital intelligence technology dictionary for logistics enterprises. Firstly, we collected and organized a large number of common words related to logistics digital intelligence technology through relevant literature and real-world knowledge, and compiled a custom dictionary comprising these words. The custom dictionary includes terms and domain-

Ting Zhang

specific terms closely associated with digital intelligence technologies, such as big data, blockchain, cloud computing, artificial intelligence, and Internet of Things.

Next, using this custom dictionary, we cross-referenced it with the crawled annual reports of logistics companies. We screened the annual reports for terms that matched those in the custom dictionary, in turn obtaining specialized terms and keywords related to the logistics domain. Finally, we incorporated the words from the crawled logistics policy text and categorized and organized the words in the dictionary based on their subject features. This enriched the content of the Digital Intelligence technical dictionary, and the keyword visualization is displayed in Figure 6-1. This process ensured that the extracted keywords related to the number-wise technology were accurately and professionally identified. A portion of the code used for this purpose is demonstrated in Figure 6-2.



Figure 6-1: Keyword Visualization

5.2. Construction of dimensional structure

Although the keyword dictionary obtained through the aforementioned method reflects important words and terms of digital intelligence technology in the logistics field, it merely presents a list of terms without a corresponding structured and systematic description. To gain a clearer understanding and organize the application and connotation of digital intelligence technology in the logistics field, we intend to construct a simple dimensional structure using the extracted keywords. This will allow for a better understanding and organization of the application scenarios and roles of digital intelligence technology in logistics enterprises.

To build a dictionary of digital intelligence technologies in the logistics field, we employ keyword extraction techniques for classification and organization. During the classification process, we thoroughly consider the nature, application scenarios, and technical characteristics of digital intelligence technologies. We also seek input from relevant experts to divide digital intelligence technologies into different dimensions. By doing so, we construct a hierarchical and structured scheme that aids in comprehending and managing the applications of digital intelligence technologies in the logistics field.

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

```
custom_dict = []
# Iterate through the contents of the chronicle for word matching
matched_keywords = []
}for report in annual_reports:
    # Use jieba splitting to split the content of the annual report
    words = jieba.lcut(report)
    # Determine if the results are in the custom dictionary
    matched_words = [word for word in words if word in custom_dict]
    # Add the matched keywords to the list
}    matched_keywords.append(matched_words)
# The content of the logistics policy text
policy_text = "The content of the logistics policy text"
# Use jieba to load a custom dictionary
jieba.load_userdict(policy_text)
# Iterate through the content of the annual report for word matching
matched_keywords = []
}for report in annual_reports:
    # Use jieba to split the annual report content into words
    words = jieba.lcut(report)
    # Determine if the results are in the custom dictionary
    matched_words = [word for word in words if word in custom_dict]
    # Add the matched keywords to the list
}    matched_keywords.append(matched_words)
```

Figure 6.2: Part of the code of the keyword extraction process

Based on the distinct roles and characteristics of digital intelligence technologies in logistics, we divide them into the following five main dimensions: intelligent sensing and identification, automation and control, data and communication, intelligent algorithms and technologies, and Internet of Things and cloud computing. Under each dimension, we further categorize different technologies, application scenarios, and business functions, which results in a more detailed description (refer to Table 1).

This classification scheme aims to provide a more professional and standardized content, catering to the actual needs of the logistics field, and promoting the application and adoption of digital intelligence technologies in logistics.

(1) The key technologies under the dimension of “intelligent sensing and recognition” are primarily used for sensing and recognizing information in the logistics environment. This involves identifying items, goods, and people, among other things. These technologies provide logistics companies with more accurate data support and process optimization. Smart sensors, image recognition, and face recognition contribute to automation, accuracy, and traceability within logistics systems by providing precise data and information. RF systems are utilized for wireless communication and identification technologies in logistics. Smartphones typically possess a range of sensing and

Ting Zhang

identification functions, enabling interaction with other logistics systems. Environmental monitoring technologies monitor parameters such as temperature, humidity, and gases in the logistics environment.

(2) The key technologies under the dimension of “automation and control” are primarily employed to automate and control the logistics process, thereby realizing automated operation and information flow within the logistics system. The main technologies used to automate and control the logistics process include automatic control systems, automated equipment, and mobile application technologies. Robots, AGV, and driverless technology play a significant role in replacing humans for logistics operations and transportation tasks, thereby improving efficiency and accuracy. Data processing and data exchange serve as crucial links in achieving logistics automation and control. Through the processing and exchange of data, automated operation and information flow within the logistics system can be achieved. Smart warehousing utilizes automation technology and intelligent control systems to manage and optimize warehousing operations, including automated storage and retrieval systems and intelligent shelving. Milkrun cycle pickup, on the other hand, is a logistics pickup method that employs automation technology and intelligent control systems to optimize the pickup and delivery process of goods.

(3) Under the “Data and Communication Dimension,” these technologies are primarily used for processing, transmission, and management of logistics information. The internet, GPS, GIS, and other technologies enable global transmission and real-time positioning of logistics information. WMS is utilized for warehouse management and inventory control. Technologies such as data analysis, data mining, and big data analysis extract useful information to optimize logistics decision-making and operations. ICT (Information Communication Technology) encompasses the technology used to acquire, process, store, transmit, and display information through electronic means. Data transmission and data sharing are key technologies for achieving the flow and sharing of logistics information. APP applications involve data transmission, communication, and user interaction. Users can remotely control smart homes and view device statuses, among other features. POS systems are used for sales and payment transactions. They collect and process transaction data. IC cards utilize chip technology for storing and transmitting data and are commonly employed for identity verification, payment processing, and storage of personal information.

(4) Under the dimension of “intelligent algorithms and technologies,” these technologies are primarily employed for intelligent optimization of logistics processes. This includes the application of intelligent algorithms, artificial intelligence technologies, and augmented reality technologies. They can be utilized for logistics monitoring, decision support, and process optimization. AIOT refers to the application of artificial intelligence technology to the Internet of Things (IoT) for achieving intelligence and optimization within the logistics system through the connection and data interaction of IoT.

(5) Under the dimension of “IoT and cloud computing,” key technologies are primarily employed for interconnection and data processing within logistics systems. These technologies include IoT technology, cloud computing, blockchain, and edge computing. They support the interconnection of logistics devices, offer efficient data storage and computing capabilities, enable agile data processing, enhance collaboration and communication among logistics stakeholders, and ensure data security and trustworthiness.

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

With the aforementioned approach, we have developed a customized small dictionary and a simplified dimensional structure. These tools enable a comprehensive depiction and understanding of the applications and contributions of digital intelligence technologies in the logistics field. They also facilitate the advancement of digital and intelligent logistics technologies. Additionally, the construction of this dictionary and dimensional structure provides a foundation and reference for future research and serves as a basis for exploring the application of digital intelligence technology in the logistics field.

Table 1: Key words of digital intelligence technology in logistics enterprises

Keyword dimensions	Keyword quantifiers
Intelligent Sensing and Recognition Class	Intelligent sensors, image recognition, face recognition, barcode recognition, RFID, sensors, infrared sensing, RF systems, smartphones, environmental monitoring
Automation and Control	Automatic control, automatic detection, automatic identification, robot, AGV (automatic guided vehicle), autonomous mobile robot, driverless, unmanned aircraft, unmanned forklift, unmanned sorting vehicle, unmanned tractor, railed automatic running vehicle (RGV), mobile application technology, monitoring technology, data processing, data exchange, intelligent storage, Milkrun cycle pickup
Data and Communication	ICT (information and communication technology), Internet, interconnection, WMS (warehouse management system), GPS, GPRS, GIS (geographic information system), ETC (electronic toll collection system), ITS (intelligent transportation system), electronic data exchange, data analysis, data sharing, data management, data mining, data transmission, data visualization, big data analysis, information technology, APP, POS, IC card
Intelligent Algorithm and Technology	Intelligent algorithm, augmented reality (AR), virtual reality (VR), mixed reality (MR), 5G, 3D printing, intelligent algorithm, artificial intelligence technology, logistics monitoring technology
Internet of Things and Cloud Computing	Internet of Things, Cloud Computing, Blockchain, Edge Computing, AIOT (Artificial Intelligence Internet of Things)

6. Conclusion and prospect

In this paper, we have constructed a dictionary of digital intelligence technologies for logistics enterprises using text mining technology. This dictionary sheds light on the key technologies and applications of digital intelligence transformation in logistics enterprises. By conducting word frequency statistics, we can identify the prevalent digital intelligence technologies utilized in logistics enterprises, and thereby discern the trends in digital intelligence technology adoption. Moreover, by analyzing the correlation and emergence of keywords, we can provide targeted suggestions and guidance for digital transformation to help logistics enterprises grasp the essence of digital and intelligent transformation and enhance their digital transformation efforts.

Ting Zhang

While this study has yielded preliminary insights into the digital intelligence technology landscape in logistics enterprises, providing valuable reference and support for their digital transformation and intelligent upgrading, it still encounters certain limitations and challenges. Firstly, the study relies on a single data source, which may not fully capture the diversity and complexity of industry development. Therefore, future endeavors should involve more extensive data collection and integration from diverse sources to enhance the lexicon's scope and inclusiveness. Secondly, text mining techniques may exhibit deficiencies in text analysis and recognition, potentially leading to lower accuracy rates in vocabulary and keyword recognition. Hence, future research needs to fortify data collection and analysis, integrating artificial intelligence and other technical methods to enhance the accuracy and precision of text mining technology, thereby improving the credibility and practicality of constructing a digital intelligence technology dictionary.

In conclusion, the future development of logistics enterprises in data mining, big data analysis, and number-wise applications is a systemic issue that requires continuous adjustment and optimization. This is necessary to adapt to the changes in the digital era and enhance scientific rigor and practicality. We can further optimize the construction of the digital intelligence technology dictionary, expand the scale and time frame of data sources, and improve the accuracy and practicality of keyword extraction. By applying these advancements to various fields within the logistics industry, such as logistics production, logistics warehousing, and logistics distribution, we can explore the sustainability of digital intelligence transformation for logistics enterprises. Additionally, we can explore how to effectively grasp the financial and market risks associated with digital transformation in the digital era. These efforts will provide reliable decision-making and support for future digital transformations.

Moreover, we will strive to strengthen international cooperation in related fields and embrace various developmental opportunities with an open attitude. This collaborative approach aims to enhance overall efficiency within the logistics supply chain and foster the healthy development of the industry.

REFERENCES

1. Y.Li, X.Yang, Q.Ran, et al., Energy structure, digital economy, and carbon emissions: evidence from China, *Environmental Science and Pollution Research*, 28 (2021) 64606-64629.
2. X.Wang, X.Le and Q.Lu, Analysis of china's smart city upgrade and smart logistics development under the COVID-19 epidemic, *Journal of Physics: Conference Series*, 1570(1) (2020) 012066.
3. A.Lagorio, G.Zenezini, G.Mangano, et al., A systematic literature review of innovative technologies adopted in logistics management, *International Journal of Logistics Research and Applications*, 25(7) (2022) 1043-1066.
4. A.Bharadwaj, O.A.El Sawy, P.A.Pavlou, et al., Digital business strategy: toward a next generation of insights, *MIS quarterly*, (2013) 471-482.
5. I.Sebastian, J.Ross, C.Beath, et al., How big old companies navigate digital transformation, *MIS Quarterly Executive*, 16(3) (2017) 197-213.
6. L.Koh, G.Orzes, F.J.Jia, The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management, *International Journal of Operations & Production Management*, 39(6/7/8) (2019) 817-828.

Measuring the Level of Digital Intelligence Technology in Logistics Enterprises: A Text Mining-Based Approach

7. I.Rubbio, M.Bruccoleri, A.Pietrosi, et al., Digital health technology enhances resilient behavior: evidence from the ward, *International Journal of Operations & Production Management*, 40(1) (2020) 34-67.
8. P.Tadejko, Application of Internet of Things in logistics—current challenges, *Ekonomia i Zarządzanie*, 7(4) (2015) 54-64.
9. Y.Zhang, The application of artificial intelligence in logistics and express delivery, *Journal of Physics: Conference Series*, 1325(1) (2019) 012085.
10. E.Tijan, S.Aksentijević, K.Ivanić, et al., Blockchain technology implementation in logistics, *Sustainability*, 11(4) (2019) 1185.
11. A.Obukhova, E.Merzlyakova, I.Ershova, et al., Introduction of digital technologies in the enterprise, *Web of Conferences*, 159 (2020) 04004.
12. R.Li, J.Rao and L. Wan, The digital economy, enterprise digital transformation, and enterprise innovation, *Managerial and Decision Economics*, 43(7) (2022) 2875-2886.
13. C-Y.Zhao, Digital development and service transformation-empirical evidence from listed manufacturing companies, *Nankai Management Review*, 24(2) (2021) 149-163.
14. F.Wu, H.Hu, H.Lin and X.Ren, Corporate digital transformation and capital market performance-empirical evidence from stock liquidity, *Management World*, 37(07) (2021) 130-144+10.
15. W.Tu and J.He, Can digital transformation facilitate firms' M&A: Empirical discovery based on machine learning, *Emerging Markets Finance and Trade*, 59(1) (2023) 113-128.
16. J.Liu and C.Wang, Can investors decipher textual information tone, *Nankai Management Review*, 24(5) (2021) 105-117.
17. A.Durnev and C.Mangen, The spillover effects of MD&A disclosures for real investment: The role of industry competition, *Journal of Accounting and Economics*, 70(1) (2020) 101299.