Briefly analyzing the impact of software versions on technical implementations

--an example of excel co-word analysis

Shu-min Chen[1*], An Hu[2a]

[1]School of Basic Medical Sciences, Jiangxi medical college, Jiangxi Province Shangrao, China
[2]Depaetment of Network Technology, Software Engineering Institute of Guangzhou, Guangzhou, China

[*]Corresponding author e-mail: chenoranger11@sohu.com
[a]e-mail: han@mail.seig.edu.cn

Abstract. The conventional wisdom that "higher version is better than lower version" has been challenged by recent research, which has demonstrated that this is merely a superficial assumption.In order to reveal the nature of the impact of software on technical implementation, this paper uses the co-word analysis method to examine the influence of different versions of Excel on technical implementation, delving into the intricacies of their mechanisms and efficiencies. Results: The processing of unidimensional data has been improving with version updates, and the processing efficiency has been increasing, but has not produced an order of magnitude difference; the processing of linked data is n times more efficient in both version 2021 and version 2016 than in version 2010. Conclusion:1st,The advantages of software version upgrades are the material basis for the influence they exert;2nd,Versions do not spontaneously influence the realisation of the technology, they can only come from the method of change; 3rd,Along with changes in methodology, advantages resulting from version upgrades are passed on to technology; 4th,Version of the relative efficiency ratio equation; 5th,Worst time complexity is the most appropriate measure of version efficiency; 6th, It can be reasonably deduced that a significant enhancement in overall efficiency can only be achieved through the implementation of an AI-enabled change in software; 7th,Software efficiency cannot be cumulative and depends only on the change in the highest order term of time complexity.

Keywords: software version; technical implementation; impact; co-word analysis; excel

1 Introduction

Software is the foundation of the new era of information technology, the cornerstone of digital economy development, and the crucial support for building a manufacturing powerhouse, a network powerhouse, and a digital China. [1] With the proliferation of computer applications, there has been a corresponding increase in the volume of

ICPDI 2024, September 06-08, Jinan, People's Republic of China
Copyright © 2024 EAI
DOI 10.4108/eai.6-9-2024.2353634

software. Simultaneously, as application usage becomes more sophisticated, software versions are continuously being updated. [2-3]The higher version software not only maintains compatibility with the functions and features of the lower version, but also introduces new functionalities, enhances original performance, addresses errors and vulnerabilities from previous versions, and optimises software operation. [4]Therefore, a qualitative conclusion can be made that the high version is superior to the low version. However, this conclusion does not address the impact of the version on technology implementation, and lacks relevant data to support it, relying solely on subjective intuition. A review of domestic and international databases revealed that existing research does not adequately address the aforementioned inquiries. Currently, software version research primarily focuses on software defects, maintenance, and related aspects, with a noticeable absence of literature pertaining to how software versions impact the realisation of specific technologies. [5-7]

In order to accurately assess the influence of software version on technology implementation, this study employs co-word analysis to investigate the impact of different versions of Excel software on co-word analysis implementation. The research findings are anticipated to address the fundamental question of how software versions impact technical implementation, and to establish a scientific foundation for future software version selection, procurement, design, development, and management. This will facilitate the identification of a more suitable version and the formulation of a more targeted version upgrade plan.

2 Excel Co-word Analysis

Co-word analysis is a prevalent research method in bibliometrics and has been extensively utilized in academic paper research. [8-10]The objective of co-word analysis is to provide a comprehensive overview of the knowledge composition and research trends within a specific subject area. This is achieved by utilising statistical data on representative keywords in individual papers and analysing the topic structure within that particular field or discipline. [11]

The Excel co-word analysis method employs the use of the Excel software tool to facilitate the conduct of co-word analysis.

3 Time Complexity

The relationship between the time required for execution and the size of the problem is referred to as the time complexity. The time spent by the computer during the execution of the algorithm is a measure of the algorithm's efficiency. The time spent by an algorithm is proportional to the number of executions of statements within the algorithm. Consequently, the algorithm that has undergone a greater number of executions of statements will necessarily have spent a longer period of time. [12] Consequently, the time complexity of an algorithm is typically quantified in terms of the number of fundamental operations incorporated into the algorithm. In a versioning analysis, the operational steps required to run the software are equated to the number of times the basic operations in the algorithm are performed. The time complexity is then employed to ascertain the efficiency of the disparate versions.

4 Experimentation

4.1 Subject of an Experiment

In recent years, the development of software technology has led to the widespread use of bibliometric analysis tool software in co-word analysis. This, on the one hand, facilitates the process of literature analysis, but on the other hand, it also reveals numerous issues. 1) Some analysis tool software is insufficiently multifunctional, and in practice it is often necessary to install several software packages to meet the demand. [13]Concurrently, the analytical tools employed employ disparate methodologies, support disparate data sources, and utilise disparate data formats. To illustrate, when utilising Bibexcel, Citespace and BICOMB analysis tools, solely Web of Science data sources are obtained. Consequently, it is either unfeasible or challenging to conduct an analysis of Chinese data sources, such as HowNet, Weipu, Wanfang. Consequently, in the practical application, it is necessary to undertake a process of iteration. Consequently, in practical applications, it is necessary to utilise different analysis tools, switch between different data sources and convert different data formats. This difficulty is particularly apparent when attempting to compare and analyse the differences between domestic and foreign research fields. 2) Some of the analysis tool software is more complex, and therefore less accessible to the novice user. It requires a significant investment of time before one can fully master the functionality of the software. To illustrate, consider the use of Visual Basic, Visual C++, and Visual Basic for Applications (VBA) to develop a program for the analysis of common words. [14]

In comparison to the aforementioned bibliometric analysis tools, Excel is relatively straightforward to operate, does not impose limitations on the data sources that can be utilised, and co-word analysis can be completed independently of other software. [15] Consequently, in order to ascertain the influence of the Excel co-word analysis method, it is sufficient to consider the impact of a single factor change in the version of Excel.

4.2 Experimental Environment

Marxists posit that the productive activity of human society progresses in a stepwise manner, evolving from a lower to a higher level. Consequently, they argue that people's understanding of both the natural world and the social aspects of reality also develops in a stepwise manner, progressing from a more superficial to a more profound understanding, encompassing a broader range of perspectives. [16-18]

The comprehension of software market demand is not a static phenomenon; rather, it is a dynamic and evolving process. It is essential that a software undergoes a series of processes, from awareness to practice, before it can be accepted by the market and ultimately mature. In order to guarantee the dependability and longevity of the software, subsequent versions of the software are not frequently updated and undergo minimal alterations. To illustrate, the most significant distinction between Excel 2019, Excel 2021 and Excel 2016 is the introduction of the icon feature, new functions and charts. It is noteworthy that the other features and operations have not undergone significant alterations. [19]

Following the release of Excel 2010, Microsoft has released four subsequent versions: The following versions of Excel have been released: Excel 2013, Excel 2016, Excel 2019, and Excel 2021. [20]Each iteration has seen an increase in features and enhancements. In light of the aforementioned considerations, it is more prudent to select distinct variants of the same temporal interval for the experiment, thereby enhancing the experimental outcome. The experimental environment was configured to utilise the Windows 10 operating system, with six versions of Excel 2000,2003,2007,2010, 2016 and 2021 installed. In addition, the literature on the application of AI in medicine (1981–2022) was retrieved from the China Knowledge Network and the literature records were exported to Excel as the experimental materials.

4.3 Experimental Methods

The processing of data utilising bibliometric analysis tools is categorised into two distinct approaches: unidimensional analysis and correlation analysis. [21]The data obtained from the co-word analysis is therefore divided into two categories: unidimensional analysis data and correlation analysis data.

In order to gain a deeper understanding of the impact of different software versions on the implementation of the co-word analysis method, a comparison of different versions of Excel on the processing of unidimensional analysis data and correlation analysis data was carried out. The aim of this comparison was to evaluate the time complexity of different versions of processing data. The objective of the study was threefold: firstly, to ascertain the impact of the version on efficiency; secondly, to compare the methods of different versions of data processing, with the aim of understanding the correlation between the version and the processing method; and thirdly, to compare the results of different versions of data processing, with the aim of understanding the effect of version on processing. The third objective was to compare the results of different versions of data processing, in order to understand the impact of version on the processing effect.

4.4 Experimental Procedure

Unidimensional analysis of data.

A unidimensional analysis is a method of examining a particular field of research, whereby the focus is on the econometric features that are present within that field. This approach allows for the identification of the specific dimensions of interest within a given collection of literature. [22]Accordingly, the processing step is divided into two stages. Initially, statistics are counted according to a specific field. Subsequently, charts are produced in accordance with the aforementioned statistics. To illustrate, the implementation of "year distribution" entails the enumeration of papers published in each year according to the "year" field. This is followed by the construction of a line graph that depicts the relationship between the year and the number of papers, as reflected by the statistical results.

In order to ascertain the impact of each iteration of Excel on the processing of unidimensional analysis data, the literature records pertaining to the application of AI in the medical field were initially processed using different versions of Excel for econometric analysis, including year distribution and journal distribution, among others. Subsequently, a comparative and analytical assessment was conducted to evaluate the processing capabilities of each version.

Correlation Analysis Data.

The association analysis is based on unidimensional analysis, whereby the co-word matrix of the analysed units is calculated. This further suggests the connection between the features of the research domain. [22]Thus, the generation of the co-occurrence matrix is a requisite step in the process of association analysis and represents the fundamental data set upon which this analysis is based.

The processing of correlation analysis data is divided into two stages: word frequency statistics and the construction of a co-word matrix. In order to ascertain the impact of different versions of Excel on the processing of correlation analysis data, we employed various versions of Excel to generate covariance matrices for literature records on the application of artificial intelligence in the field of medicine and to analyse the processing of each step in the main links of "word frequency statistics" and "constructing covariance matrices". We conducted a detailed analysis of the processing of each step in the main link "word frequency statistics" and "constructing co-word matrix".

4.5 Experimental Results

Processing of Unidimensional Analysis Data.

Table 1. presents a comparative analysis of distinct approaches to unidimensional data processing.

excel version	**manipulate**	**time complexity**	**methodologies**	**results**
2000	statistical data and plot chart	Best time complexity $O(1)$ Worst time complexity $O(n)$ Average time complexity $O(n)$	countif function、pivot table	A paucity of statistical data; Single Chart
2003	ibid	ibid	ibid	ibid
2007	ibid	ibid	ibid	ibid
2010	ibid	ibid	ibid	ibid
2016	ibid	ibid	pivot table、left function	The integrity of statistical data; Multiple charts
2021	ibid	ibid	pivot table, Smart Fill	ibid

Table I illustrates the various iterations of single-dimensional data processing methodologies and the discrepancies in the statistical analysis of missing data. In the 2010 version (that is, the Excel 2010 version of the abbreviation, as referenced below), the application of the countif function or pivot table to all records of statistics will yield incomplete statistical results due to the absence of requisite fields. To illustrate, when enumerating the number of papers in each year, the "Year " field will be absent, thereby preventing the display of the year for all records. Conversely, the "PubTime-Publishing Time" field will not be absent and includes the year information. In the 2016 version (that is, the Excel 2016 version of the abbreviation, the same below), the "PubTime " field can be moved to the left of the four numbers in order to obtain the year data. In the 2021 version (that is, the Excel 2021 version of the abbreviation, the same below), the "smart fill" function can be triggered by pressing the Ctrl + E keys to separate the data by year. This is achieved by using the "PubTime " field.

The optimal approach is to work with a single dimension of data and dedicate time to statistical analysis and charting. In version 2010, the initial step is to establish the conditions for the first countif function. Subsequently, the function can be replicated in other records through the use of relative referencing. The completion of the statistics through the countif function requires five steps, while the completion of the same through the pivot table necessitates four steps. The charting process, on the other hand, requires only three steps. It can thus be concluded that the optimal time complexity for the 2010 version is $T(n) = (countif\ function\ or\ pivot\ table) + charting = (5\ or\ 4) + 3 = (8\ or\ 7) = O(1)$.The optimal time complexity for the 2016 version is $T(n) = left\ function + pivot\ table + charting$, where the left function has the same number of steps as the countif function. The resulting complexity is $T(n) = 5 + 4 + 3 = 12 = O(1)$. In the 2021 version, the optimal time complexity is $T(n) = Smart\ Fill + Pivot\ Table + Charting$. This is because the Smart Fill operation has two steps, resulting in $T(n) = 2 + 4 + 3 = 9 = O(1)$.

The most unfavourable scenario is that of processing n one-dimensional data. In the 2010 version, the processing of n one-dimensional data items is performed individually, with the same n one-dimensional data processing operations repeated. This results in the worst time complexity of $O(n)$ .The 2016 version is the first to employ the left function to replenish missing data and subsequently utilise pivot tables for the processing of one-dimensional data in a sequential manner, thereby generating charts. Consequently, the time complexity is represented by $T(n) = n + 5$, which is equivalent to $O(n)$. The 2021 version is the first to employ the "smart fill" technique to replenish missing data and subsequently utilise pivot tables for one-dimensional data processing, one by one, to generate charts. This results in the worst time complexity, $T(n) = n + 5 = O(n)$. The 2021 version represents the inaugural instance of utilising the 'smart fill' separation technique to supplement the missing data, subsequently employing pivot tables for the one-dimensional data processing, which is conducted one by one to generate charts. This results in the worst time complexity of $O(n) = n + 2 = O(n)$.

The average time complexity is identical for all versions and is given by $T(n) = \frac{(1+2+\cdots+n)}{n} = \frac{(n^2+n)}{2n} = O(n)$.

Data Processing for Correlation Analysis.

Table 2. A comparative analysis of the statistical processing of word frequency in different versions.

excel version	manipulate	methodologies	time complexity	results
2000	Keyword(s) parsed	Separator breakdown	Best time complexity $O(1)$ Worst time complexity $O(1)$ Average time complexity $O(1)$	In the process of data entry, instances have been observed where the input of invalid data has resulted in the creation of blank columns.
2003	ibid	ibid	ibid	ibid
2007	ibid	ibid	ibid	ibid
2010	ibid	ibid	ibid	ibid
2016	ibid	The search and replace function; Separator breakdown	ibid	It is not possible to create invalid blanks.
2021	ibid	ibid	ibid	ibid
2000	Keyword statistics	The initial step is to create a multi-column, one-by-one copy of the first and last entries, which should be connected to a column. Subsequently, any field values that are empty should be removed. Finally, a pivot table should be generated.	Best time complexity $O(1)$ Worst time complexity $O(n)$ Average time complexity $O(n)$	It is necessary to implement a manual and somewhat laborious process for the removal of empty field values.
2003	ibid	ibid	ibid	ibid
2007	ibid	ibid	ibid	ibid
2010	ibid	ibid	ibid	ibid
2016	ibid	inverse perspective column; pivot table	Best time complexity $O(1)$ Worst time complexity $O(1)$ Average time complexity $O(1)$	The device is straightforward to operate and is designed to automatically remove data from empty field values.
2021	ibid	ibid	ibid	ibid

A comparison of the statistical processing of word frequencies of the different versions yielded the results presented in Table II. Table II demonstrates that utilising the Excel 2010 version of keyword statistics will not only become increasingly onerous as the number of keyword columns increases, but will also retain a considerable number of empty field values. In the optimal scenario, there would be only two columns of keywords, requiring only a two-step copy and paste operation. The implementation of a "column copy" would entail arranging the two columns of keywords into a single column and then locating and deleting the two empty field values from the two-step operation. This would result in the best time complexity $T(n) = column\ copying\ +$

$locating\ and\ deleting\ the\ value\ of\ the\ empty\ field\ +\ pivot\ table = 2 + 2 + 4 = 8 = O(1)$. In the worst case, there are n columns of keywords. The first n-1 times, "column replication" must be performed in order to arrange all the keywords into a column. Then, the empty field values must be located and deleted. Finally, the pivot table statistics must be performed. Therefore, the worst time complexity is $O(n)$. Meanwhile, the average time complexity is calculated as $O(n)$. The inverse pivot column function in Excel 2016 and 2021 is not affected by the number of keyword columns. The number of steps in its operation remains constant. It can be connected to a keyword column without empty field values, and thus its time complexity is $O(1)$.

Table 3. presents a comparative analysis of the processing of different versions of constructing co-word matrices.

excel version	manipulate	methodologies	time complexity	results
2000	High-frequency words are teamed up in pairs	The initial step is to categorise the second keyword as the primary keyword and eliminate any empty records within the second keyword. Subsequently, the keyword columns of the two groups should be merged and copied and pasted into the first and last two columns of keywords.	Best time complexity $O(1)$ Worst time complexity $O(n^2)$ Average time complexity $O(n^2)$	The operation is unwieldy, and with the expansion of the number of columns of keywords, the frequency of the initial and concluding links between the first and last columns of keywords has also increased.
2003	ibid	ibid	ibid	ibid
2007	ibid	ibid	ibid	ibid
2010	ibid	ibid	ibid	ibid
2016	ibid	Unpivot; Merge multiple worksheets	Best time complexity $O(1)$ Worst time complexity $O(n)$ Average time complexity $O(n)$	The device is straightforward to operate and is capable of simultaneously deleting both blank records and single-keyword entries. It can also be used to arrange the first column of keywords and other columns of keywords into two distinct columns.
2021	ibid	ibid	ibid	ibid

2000	Eliminate low-frequency words	The initial step is to create a new, empty column after each existing column. Subsequently, the VLOOKUP function should be employed to ascertain the high-frequency words that correspond to the	Best time complexity $O(1)$ Worst time complexity $O(n)$ Average time complexity $O(n)$	The process is unwieldy, and the number of removals rises in proportion to the number of keyword columns.
2003	ibid	ibid	ibid	ibid
2007	ibid	ibid	ibid	ibid
2010	ibid	ibid	ibid	ibid
2016	ibid	Advanced filtering	Best time complexity $O(1)$ Worst time complexity $O(1)$ Average time complexity $O(1)$	Regardless of the number of keyword columns, the actual number of keyword columns processed in the end is only two.
2021	ibid	ibid	ibid	ibid

Excel 2010 is employed for the purpose of clearing low-frequency words and high-frequency words in two-by-two teams, with the objective of constructing a matrix for the creation of a co-word matrix.[23] In the event that Excel 2016 or 2021 is utilised, it is necessary to alter the sequence of the aforementioned operations. Specifically, the order should be: in turn, high-frequency words two by two teaming, clear low-frequency words, build the matrix.

A comparative analysis of the construction of co-word matrices was conducted to obtain Table III. As evidenced in Table III, the utilisation of Excel 2010 for the identification of high-frequency words in two teams was accomplished through two distinct phases. In other words, the initial step involves identifying the second keyword as the primary sorting criterion, then deleting the second keyword from the remaining records, and finally deleting the remaining single-keyword records. This entire process necessitates six distinct steps. The subsequent step entails the repeated use of copy and paste operations, whereby the remaining keywords from the two groups are merged and appended to the end of the keywords in the two columns. [19] In the optimal scenario, only two columns of keywords, A and B, are merged into a single set, AB. This results in the best time complexity of $O(1)$. In the event that two columns of keywords are combined to create a total of $C_n^2 = \frac{A_n^2}{A_2^2} = \frac{n!}{2!(n-2)!}$ kinds, this process must be repeated $C_n^2 - 1$ times. The first and last of these columns are then merged to form two new columns of keywords. Consequently, the worst-case time complexity of the second step is $T(n) = \frac{n\times(n-1)}{2} = O(n^2)$, while the average complexity is $T(n) = \frac{\left(1+3+6+\cdots+\frac{n\times(n-1)}{2}\right)}{n-1} = O(n^2)$ .The 2010 version of the two-by-two teaming of high-frequency words exhibits the worst time complexity $T(n) = \frac{(n\times(n-1))}{2} + 6 = O(n^2)$. The average time complexity is also $O(n^2)$.

As illustrated in Table III, the Excel 2016 and 2021 high-frequency words two-by-two team is divided into two stages to achieve its objective. These stages are designated as "reverse pivot columns" and "multiple worksheet merger." In the optimal scenario, only two columns of keywords are required, eliminating the need for "reverse pivot columns" and "multiple worksheet merger." This direct composition of high-frequency words results in the best time complexity of $O(1)$. In the worst case, where there are n keywords, the first step requires n columns of keywords to undergo n-2 times the "reverse perspective columns" operation, in addition to the last two columns of keyword matching worksheets. This results in a total of n-1 worksheets, with the "reverse perspective columns" not being affected by the number of columns of keywords. The number of operation steps is therefore 8. The second step, "multiple worksheets merge," entails directly merging n-1 worksheets to form the first and last columns of keywords. This operation is not subject to the limitations of the number of worksheets and requires 14 steps. Consequently, the worst time complexity ($T(n)$) is given by the following equation: $T(n) = Inverse\ Pivot\ Columns\ +\ Multi - Worksheet\ Merge\ = (8 * (n-2)) + 14 = O(n)$, the average time complexity is $O(n)$.

Clear the worst case of low-frequency keywords is the operation of n keywords. excel 2010 processing needs to be first in each keyword column after a new blank column (operation n steps to form n blank columns), and then each keyword column using the vlookup function to find out with the matching high-frequency words, and then sorted and remove the low-frequency words. Each vlookup function requires four steps to complete the lookup operation, that is, the sorting needs two steps to complete, delete the need for two steps to complete, so its entire removal of low-frequency keywords the worst time complexity of $T(n) = blank\ column\ operation\ +\ vlookup\ function\ operation\ +\ sorting\ to\ delete\ =\ n\ +\ 4n\ +\ 4n\ =\ 9n\ =\ O(n)$. And as mentioned earlier excel 2016 version in the construction of the co-word matrix operation order compared with excel 2010 has changed, so that at this stage (clearing the low-frequency keywords) the actual number of keyword columns processed only two columns, and then use the "Advanced Screening" (3 steps to complete) can be obtained from the high-frequency word columns. Therefore, its time complexity for all types of $T(n) = 3 = O(1)$.

4.6 Experimental Analysis

Mechanistic Analysis.

The advantages resulting from software version upgrades, including the introduction of new features, a reduction in the number of bugs and vulnerabilities, more optimised results, and the incorporation of more advanced algorithms, provide the material basis for the influence of versions on technical implementation. As an example, the 2016 version of Excel includes a "Power Query (query enhancement)" plug-in, which is not present in the 2010 version. This plug-in enables the more efficient execution of the "inverse pivot column" and "multiple worksheets merge" functions.

Nevertheless, the version does not have an immediate and direct impact on the implementation of the technique. A comparison of Tables II and III reveals that the Excel 2016 and Excel 2021 versions operate in a similar manner, exhibiting the same time complexity and resulting outputs. Therefore, it can be concluded that although the version has been upgraded, there has been no alteration in the implementation method of the technology, and thus no improvement in efficiency. That is to say, the efficiency of Linked Data processing in version 2021 is identical to that of version 2016. A similar phenomenon is observed in the processing of unidimensional analysed data across different versions, as illustrated in Table I. While newer versions of Excel offer enhanced support for multi-threaded computation and the utilisation of multiple processor cores, this does not directly translate to improved implementation of the underlying technology. [24]Consequently, the optimal computational engine remains unable to function effectively.

However, with the alteration of the technical implementation method, the advantages yielded by the version upgrade will be conveyed to the technology, thereby facilitating an enhancement in technical implementation efficiency. As illustrated in Table II, the 2010 version of the 'keyword statistics' necessitates the manual removal of the field value, which is represented by empty data. In contrast, the 2016 version of the operation steps is more concise, and the removal of the field value is automatically completed. The reason for this is that the Excel version from 2010 was upgraded to the 2016 version of the 'inverse perspective columns' feature, which is also known as 'keyword statistics'. The implementation of the 'columns to copy' and 'positioning to delete empty field values' functions has resulted in a change to the 'reverse perspective columns' feature. This alteration to the implementation method optimises the efficacy of the novel feature, reducing the time complexity from $O(n)$ to $O(1)$ and enhancing the efficiency of 'Keyword Statistics'. Further research revealed that the new function of the 2016 version is based on more sophisticated algorithms. The 2016 version of Excel has integrated neural network algorithms, allowing users to create and train neural network models for in-depth analysis of table data.[25]

It can be seen, therefore, that the purpose of methodological change is the realisation of technology. In order for this to occur, methodological change must be based on the advantages of versioning. Thus, technology drives the updating of versions, and the updating of versions facilitates the change in the method of technological realisation, which in turn facilitates technological progress.

Analysis of Efficiency.

Discussion of Efficiency Measures.

As previously stated, time complexity is used to identify different versions of efficiency, with time complexity categorised as best, worst, and average. The optimal time complexity represents the most ideal scenario, which is more extreme and has a smaller probability of occurrence. [26]In this instance, the aforementioned versions process the least amount of data, thereby rendering any discrepancy in their time complexity inconsequential. As illustrated in Table I, Table II and Table III, the optimal time complexity of each version is identical, with a value of $O(1)$. It is therefore evident that the optimal time complexity is not an appropriate basis for comparison of efficiency.

The average time complexity is defined as a scenario in which all potential executions are equally probable. [27]The inherent complexity of reality renders the analysis and quantification of the average time complexity of software versions a challenging endeavour. It can thus be concluded that average time complexity is not an appropriate basis for comparison of efficiency.

The worst-case time complexity represents the least favourable scenario, although it is also a more extreme situation. The probability of occurrence is relatively low, but it can reflect the maximum influence of the version on the technical implementation. This is the most suitable of the three as a measure of the efficiency of the version of the standard. The ratio of the worst time complexity of the two versions allows for a clear determination of the efficiency improvement ratio of the new version in comparison to the old one. This provides a scientific basis for the selection, procurement, design, development and management of software versions.

For example, comparing the worst time complexity of "high-frequency word two-by-two teaming" in Table 3, we can see that the 2010 version is $O(n^2)$, and the 2016 version is $O(n)$. From "The higher the time complexity order of magnitude the lower the efficiency "[28] we know that the 2016 version is more efficient than the 2010 version and it is n times more efficient than the 2010 version; Comparing the worst time complexity of the "keyword statistics" in Table 2, we can see that the 2010 version is $O(n)$ and the 2016 version is $O(1)$, which means that the processing efficiency of the 2016 version is n times higher than that of the 2010 version. The preceding analysis allows us to derive the following version of the relative efficiency ratio equation:

$$\frac{p_1}{p_2} = \frac{T_2(n)}{T_1(n)} = k \tag{1}$$

In this context, p_1 represents the efficiency of version 1, p_2 denotes the efficiency of version 2, $T_1(n)$ signifies the worst time complexity of version 1, $T_2(n)$ is the worst time complexity of version 2, and n is the size of the problem to be processed. Consequently, the inverse of the ratio of their worst time complexities is equal to the ratio of relative efficiencies of the versions. Once the size of the problem to be processed, denoted by n, is known, the relative efficiency ratio, k, of the versions is also known and non-zero. As the size of the problem to be processed, represented by the variable n, increases, so too does the time complexity. Consequently, the efficiency difference between the versions in question also increases.

Discussion of Method Changes.

From Table I, we can see that the worst time complexity ratio of the 2021, 2016, and 2010 versions is $O(n){:}\,O(n){:}\,O(n) = 1{:}\,1{:}\,1$, i.e., there is no order-of-magnitude difference between the three versions in terms of processing efficiency of unidimensional analyzed data. In terms of the overall number of operations, the 2016 version exhibited a greater increase in the number of operating steps and a more extensive left function processing than the 2010 version. However, it also demonstrated an enhanced integrity of the entire statistical data. The 2021 version ensures the integrity of statistical data on the basis of a simpler operation than that of the 2016 version, which reduced the number of operating steps from 12 to 9. In conclusion, the upgrading of the version has resulted in an enhanced processing method for single-dimensional data, improved data integrity, a reduction in the number of processing steps, and enhanced processing efficiency. Nevertheless, this alteration in methodology has not altered the manner in which unidimensional data must be processed manually, one by one. Despite an alteration in the number of steps in the processing process, once the quantity of data to be processed becomes considerable, the impact of this alteration on the overall efficiency is minimal and insufficient to produce an order of magnitude difference.

Table II and Table III illustrate the reduction of steps in the operation of four processes:(1) "reverse pivot columns" instead of "column copy", "positioning to delete empty field values"; (2) "Inverse Pivot Columns" instead of "delete blank records and a single keyword", "two by two"; (3) "multiple worksheets combined "instead of" first and last "; (4)" advanced screening "instead of" vlookup function ". Table 2 illustrates the repetition of the "keyword statistics" in the "column copy" and "locate and delete the empty field value" operations, which are then replaced by the "reverse pivot columns". This automated processing is replaced by the "reverse pivot column" time complexity from $O(n)$ to $O(1)$. Table 3 illustrates how the "high-frequency words two by two" can be employed to replace the manual operations that are currently performed consecutively. These include "deleting blank records and single keywords", "combining two by two", and "first and last together". The "first and last together" operation can be replaced by the "reverse pivot columns" and "multiple worksheets merge" automated processes, which reduce the time complexity from $O(n^2)$ to $O(n)$. It is evident that a series of repetitive manual operations can be transformed into an automated process. The sequence of operations and the number of steps involved in the automated processing are predetermined. The processing time is independent of the problem size, n, and the time complexity undergoes a transition from a high order to a low order or even to a constant order. The introduction of artificial intelligence (AI) into software represents a significant opportunity for enhancing overall efficiency. By automating numerous time-consuming manual processes, AI can potentially increase efficiency by an order of magnitude. This is because AI enables the intelligent processing of data, which in turn improves overall efficiency. [29]In other words, the integration of AI into software can lead to a substantial improvement in overall efficiency.

Discussion of Efficiency Improvement Methods.

In the majority of cases, optimising time complexity is of greater importance than optimising space complexity, given that the time spent running the software cannot be recovered, whereas the space can be recovered. It is therefore recommended that efforts be directed towards reducing time complexity as a means of enhancing efficiency. Table III illustrates that the time complexity of the "Constructing Co-word Matrix" operation is $T(n) = Two\ high - frequency\ words\ are\ grouped\ together\ + Clearing\ the\ low - frequency\ words\ +\ Constructing\ the\ matrix$. The various iterations of the "build matrix" operation entail the same nine steps, resulting in a time complexity of $T(n) = T(9) = O(1)$. It can thus be concluded that the worst time complexity of the 2010 version of "Constructing the covariance matrix" is $T(n)$= $O(n^2) + O(n) + O(1) = O(n^2)$. The most computationally expensive operation in the 2016 version is $T(n)$=$O(n) + O(1) + O(1) = O(n)$. The 2021 version exhibits the same time complexity as the 2016 version. In other words, the efficiency of "constructing the co-occurrence matrix" in each version is contingent upon the alteration of the highest-order term in the time complexity.

Table II illustrates that the time complexity of "word frequency statistics" T(n)= keyword parse + keyword statistics. It can be demonstrated that the worst time complexity of the 2010 version of 'word frequency statistics' is $T(n)$= $O(1) + O(n) = O(n)$, and that the worst time complexity of the 2016 version of 'word frequency statistics' is $T(n)$=$O(1) + O(1) = O(1)$. The 2021 version exhibits the same time complexity as the 2016 version. In other words, the efficiency of the 'word frequency statistics' in each version is contingent upon the alteration of the highest-order term in the time complexity.

The time complexity of the 'correlation analysis data' is given by $T(n) = word\ frequency\ statistics\ +\ construct$ $co - word\ matrix$. The 2010 version of the "correlation analysis data" exhibits the worst time complexity, $T(n)$= $O(n) + O(n^2) = O(n^2)$, while the 2016 version displays the worst time complexity of $T(n)$=$O(1) + O(n) = O(n)$. The 2021 version exhibits the same time complexity as the 2016 version. In accordance with the formula (1), $\frac{p_{2016}}{p_{2010}} = \frac{T_{2010(n)}}{T_{2016}(n)} = \frac{O(n^2)}{O(n)} = n$.Similarly, the ratio of the efficiency of the three versions (2021, 2016, and 2010) is n:n:1. That is to say, the efficiency of "correlating and analyzing data" in each version depends on the change of the highest order term of time complexity.

In conclusion, the efficiency of software running cannot be cumulative; rather, it is contingent upon the change of the highest order term of time complexity. This is due to the fact that when the problem size tends to infinity, the asymptotic relationship between the software running time and the problem size of the highest order term has the greatest impact on the time complexity. Consequently, the other can be almost ignored, or even ignored entirely with the highest term multiplied by the constant [30]. Consequently, a comprehensive examination of the relative efficiency of the version, with the objective of identifying the highest order term of time complexity in the method change, enables the discovery of the "bottleneck" and inefficient links in the original version. This, in turn, facilitates the implementation of optimization measures pertaining to automated processing, thereby achieving an efficiency improvement of the version by an order of magnitude.

5 Conclusion

Software is the fundamental support for scientific and technological innovation. It is the primary factor that facilitates the comprehensive integration of informationization and industrialization, enabling its implementation in specific industries, specific products, and plays a pivotal role in the development of manufacturing capabilities, network capabilities, and the digitalization of China. This paper employs the Excel copula analysis method as its experimental object and examines the impact of different versions on the technical realization of the copula analysis method. The findings indicate that with the advancement of the version, the processing methodology for single-dimensional data has consistently improved, and the processing efficiency of the version has also increased. However, there is no discernible difference in the magnitude of these improvements. Furthermore, the processing of correlated data in the 2016 version is more efficient than that in the 2010 version, with the efficiency increasing as the size of the problem to be dealt with increases. However, the difference in efficiency between the versions becomes larger as the problem size increases. Notably, the efficiency of Linked Data processing in version 2021 is the same as in version 2016.

In the course of investigating the aforementioned experiments, the central inquiry of "how software version affects technology implementation" was scrutinized in terms of mechanism and efficiency. The ensuing analysis yielded the following conclusions:

(1) The advantages that arise from software version upgrades provide the foundation for understanding how the version affects the realization of technology. It is not the version itself that directly influences the realization of technology; rather, it is the change in the realization method of technology that gives rise to the version. As a consequence of a change in the method of technology realization, the advantages generated by a version upgrade will be transferred to the technology, thereby improving the efficiency of technology realization.

(2) The most appropriate metric for measuring versioning efficiency is the worst time complexity. The equation for calculating the relative efficiency ratio of versions is: $\frac{p_1}{p_2} = \frac{T_2(n)}{T_1(n)} = k$

(Once the value of n has been established, the value of k is also known with certainty and is not equal to zero)

(3) It is only through the implementation of AI-enabled software updates that an order of magnitude improvement in overall efficiency can be achieved.

(4) It is not possible to achieve cumulative operational efficiency in software; rather, it is contingent upon the alteration of the highest-order term of time complexity.

If we conceptualize software operation efficiency as a bucket of water, then each update to the software version can be viewed as a means of compensating for the initial deficit in efficiency, and each alteration to the technical realization method can be seen as a way of strengthening the overall efficiency. It is my hope that readers will be inspired by this article to identify a more suitable version and to develop a more targeted version of the upgrade program.

Acknowledgement

Funding: Science and Technology Research Project of Jiangxi Provincial Department of Education (GJJ213413)

References

1. Xia Jing, Cao Qing, and Yan Yongcai. “A study on the output effect of funding and policy of China's torch program software industry.” Business Watch 10.11(2024):76-83+95.
2. Wei, Yuanhua. Design and Implementation of a Web-based Network Management System. 2006. Tongji University, MA thesis.
3. Florian Fittkau, Alexander Krause, and Wilhelm Hasselbring. "Software landscape and application visualization for system comprehension with ExplorViz." Information and Software Technology 87. (2016): 259-277.
4. Shi Qiang. Design and Application of Automation Standard in Automotive Industry Based on Siemens SiCar Platform [D]. Shanghai Jiao Tong University, 2019. DOI: 10.27307/d.cnki.gsjtu.2019.004122.
5. Li P. Research on incremental code defect detection method for software version based on graph neural network [D]. Northwest University, 2021. DOI: 10.27405/d.cnki.gxbdu.2021.000340.
6. Liu W. J.. Research on severity prediction of software defect reports [D]. Dalian University of Technology, 2020. DOI: 10.26991/d.cnki.gdllu.2020.001677.
7. ZHANG Chi-Yu, LI Quan-Shuang, CAO Wei, CHENG Jing, YU Qian, SUN Jin-Yan, WEI Ji-Fu. Effect of different versions of RNAstructure software on FORS-D analysis [J]. Journal of Jiangsu University (Medical Edition), 2006 (04): 294-297+309. DOI: 10.13312/j.issn.1671-7783.2006.04.006.
8. [] WANG Xukun, HOU Jianhua, YAN Chao, BAI Changchuan. From the study of thinking in traditional Chinese medicine toward the science of thinking in traditional Chinese medicine [J]. Chinese Medicine Culture, 2021, 16(03): 217-226. DOI:10.16307/j.1673-6281.2021.03.003.
9. JIANG Xiaoman, Xue Yong. Research on education informatization based on visualization analysis to help education fair development[J]. Digital Education,2022,8(01):25-31.
10. WANG Qian, ZHANG Kun, LIANG Qiaomei. A review of carbon tax effect research: based on bibliometric analysis[C]//China Association for Science and Technology, People's Government of Guangdong Province. Proceedings of the 17th Annual Conference of the Chinese Association for Science and Technology - Sub 15 International Symposium on Science and Technology Innovation and Investment and Financing in Emerging Green Industries. [Publisher unknown],2015:12.
11. HU Guopeng, YANG Jusheng, CAI Lan. Bibliometric-based research on public service motivation in China: current status and innovation[J]. Hubei Social Science, 2022(02): 37-48. DOI: 10.13660/j.cnki.42-1112/c.015818.
12. Huang, Y. H.. Single/multi-target tracking algorithm based on unit correlation matching[D]. Nanjing University of Science and Technology, 2022. DOI: 10.27241/d.cnki.gnjgu.2020.001566.
13. Man Junlin. Design and implementation of literature information analysis system[D]. Dalian University of Technology, 2013.
14. LUO Lingyun, XU Lingyun. Visual empirical analysis of research hotspots in Chatham project [J]. Journal of Jintu,2014(05):65-69.
15. WANG Yuankai, ZHONG Ruizhu. Application of Excel software in establishing enterprise cost index database [J]. Engineering Cost Management, 2022(04): 86-91. DOI: 10.19730/j.cnki.1008-2166.2022-04-086.
16. Qi Zhitai. Mao Zedong's Philosophical Contributions in the Yan'an Period[C]//Qinghai Yan'an Spirit Research Association, Qinghai Yellow River Culture and Economic Development Research Association, Qinghai Jiangheyuan Culture Research Association. Pro-

ceedings of the Theoretical Symposium on the Yan'an Spirit and China's Modernization: The Yan'an Spirit and the Practice of Socialist Modernization. Qinghai People's Publishing House, 1999:6.

17. Pu Liangxian. Introduction to the contents of Theory of Practice[J]. Pragmatism, 1985(S1): 100-101.
18. Mao Zedong. The Theory of Practice[J]. Selected Works of Mao Zedong, 1951, 1(1).
19. Li Rui. Did you know about these new Excel 2019 features over Excel 2016? [EB/OL]. https://www.sohu.com/a/373802048_695443,2020-02-17.
20. Popular Science China. Encyclopedia of Science. Microsoft Office [EB/OL]. 2021 [2021-10-23]. https://baike.baidu.com/item/Microsoft%20Office/481476?fromtitle=Office%E5%8A%9E%E5% 85%AC%E8%BD%AF%E4%BB%B6&fromid=9373660&fr=aladdin.
21. Liu Zuoyi, Chen Xiaotian. The nature, role, method and procedure of scientific research evaluation [J]. Scientific Research Management, 2002(2): 33-40.
22. Yang Qian. A comparative study of the analytic functions of common bibliometric tools[J]. Intelligence Exploration,2021(10):87-93.
23. Chu Jiewang, and Guo Chunman. “EXCEL's method of realizing co-occurrence analysis - an example of knowledge management research in the domestic library intelligence field.” Journal of Intelligence 30. 03 (2011): 45-49.
24. Yang Jia-Ping. Research on the Structural Impact of Big Data and Artificial Intelligence Technology Application on Online Labour Market. 2019. Beijing University of Posts and Telecommunications, MA thesis.
25. Gongzi Shiwu Shuang. Neural network algorithms in Excel: intelligent analysis of tabular data [EB/OL]. [2023-10-9]. https://developer.baidu.com/article/details/1851822.
26. Cheng Jiao. Research on fast construction method of high-dimensional multi-objective Pareto non-dominated solution set[D]. Wuhan University of Science and Technology,2018.
27. Zhou Jiangtao. Research and Implementation of Efficient and Accurate String Matching Algorithm[D]. Harbin Institute of Technology,2012.
28. Yin, Shaohong, Dong, Qingxia, and Yuan, Chunmiao. Data Structures [M]. May 2004, 1st Edition. Tsinghua University Press, Beijing Jiaotong University Press, May 2004.
29. Maryam Asadi Aghbolaghi, et al."Learning generalizable AI models for multi-center histopathology image classification."npj Precision Oncology 8.1(2024):151-151.
30. Liu, Chang, Yao, Xuefeng, Hu, Bing, Zhang, Yongliang, and Yang, Jingjie. Data structures [M]. 3rd printing, August 2019. Shanghai Jiao Tong University Press, 2019:8-9.