https://doi.org/10.31489/2025Ec1/52-62

JEL: C87, D40, L16 Received: 24.11.2024 | Accepted: 08.01.2025

ORIGINAL RESEARCH

The evolution of business ecosystems: A text mining-based analysis of innovation and competition (1993–2023)

Toktar Shuren[1]*, **Assylbek Nurgabdeshov**[2]

Abstract
Discussions about ecosystems are mostly relevant in current time, due to the fact that ecosystem approach is becoming more recognized and more applied in the modern world. In this article the attempt was made to identify the basic directions of ecosystem development in the context of business, innovation and competition. The main trends of this phenomena by time intervals were determined. As an analytical tool the authors' application called "Friendly text mining" which is based on the "NLTK" package of programming language python was put into service. From the Scopus database journals 600 articles for the period from 1993 to 2023 were selected and processed. These findings have shown a significant increase in ecosystem related research, especially since 2014, digitalization, platform-based professional models and variations have been associated with the growing renowned strategy. Analysis also highlights the reduction of traditional cluster-based research, indicating that ecosystems are gradually changing clusters as an impressive structure to understand inter-firm cooperation and competition. In addition, this study recognizes statistically significant relationships between key words such as "ecosystem", "platform", "digitalization", "digitalization", and "innovation", which underscores each other's interconnected nature in contemporary business research. This study contributes to literature by showing the effectiveness of text mining methods, providing a scalable and systematic approach to the evolution of the educational discourse. Future research should find the structural mobility of ecosystems, their impact on industry change, and the role of emerging technologies in their evolution.

Keywords: ecosystem, business, digitalization, diversification, text mining, innovation, python.

Introduction

In this article, we intend to explore the genesis of ecosystems in the context of business and competition through bibliographic analysis, namely using text mining technology, since this technology is relatively recent and, at the time of writing, the ecosystem phenomenon has not been investigated in this way. To do this, the authors want to find out which topics have been most relevant in the above context over the past 30 years, as well as to identify how and which of them have changed their relevance with the advent of ecosystems in the economic scientific works.

Text mining has emerged as a valuable tool in scientific research (Fluck, 2005). It enables the automatic retrieval and extraction of information from scientific articles, which is crucial given the vast volume of literature in these fields. This is particularly evident in the field of mobile learning, where text mining techniques have been used to analyze and extract information from a large number of research articles (Salloum, 2018). The potential of text mining in science and technology research is further underscored by the development of a text mining tool to support complex tasks (Korhonen, 2012; Cockburn, 2018). However, there is a need for further integration and refinement of text mining technology to fully realize its potential in scientific research (Losiewicz, 2003; Chesbrough, 2003).

Since the word "ecosystem" has roots in biological sciences, studying the genesis of ecosystems in a business context may lead to irrelevant articles. To avoid this risk, we decided to use "business", "competition", and "innovation" as keywords instead of "ecosystem" during the article search. To conduct the study, 600 articles published from the journals of the Scopus database were analyzed, according to the keywords mentioned above. The selection of articles covered a fairly large period from 1993 to 2023. Using bibliographic analysis, the authors examined the state and evolution of business processes, competition in innovative research, and the role of ecosystems within them.

The goal of the article is to identify the main directions of ecosystem development in scientists' research, through the prism of business, innovation and competition, in order to determine the trends of these

[1]*Faculty of Economics, Karaganda Buketov University, Karaganda, Kazakhstan, shuren_t@buketov.edu.kz (corresponding author)
[2]Edinburgh Business School, Heriot-Watt University, UK, a.nurgabdeshov@hw.ac.uk

phenomena in time intervals, as well as to find anomalous values, which can provide a prerequisite for further research of these values and connections. Additionally, the authors aim to explore the relationship between clusters and ecosystems in the context of competition and test the hypothesis that as interest in ecosystem studies grows, the focus on clusters declines. The authors have contributed to the improvement of text mining technology by combining the most important functions from the NLTK python package, as well as by adding their own lines of code, which are presented later in the article. For their own convenience, as well as possibly for other interested researchers, the above functions were combined into one user-friendly web interface program, which the authors called “Friendly text mining”. With the help of “Friendly text mining”, the user, having no programming knowledge, will be able to perform various operations (lemmatization, removal of stop words, search for the most used words in the context of periods, etc.) with a large volume of text.

This study showed that in a competitive business environment, the popularity of ecosystem research is growing. Jacobites et al. support the conclusions of this study in the context of creating ecosystems as a mechanism for sustainable competitive advantage of companies. The authors use an ecosystem approach to strategic management and analyze the mechanisms of building an ecosystem of partners and clients (Jacobides et al., 2018).

We agree with scientists (Boschma, 2015; Andersson et al., 2010; Lakhani, 2007) in the opinion that an ecosystem approach can stimulate innovation and entrepreneurship development. The authors explore the roles of various ecosystem participants, including startups, venture investors, accelerators and educational institutions, and analyze the relationship between them to create a favorable environment for innovation.

The results of our research coincide with the analytical approaches of West and Boyers, in the context of business process diversification. Scientists have presented an overview of research related to open innovation and the use of external sources to stimulate innovation in ecosystems that transform traditional business models. They also consider the challenges that organizations face when working in ecosystems (West and Boyers, 2014).

Also, the results of our study coincide with the results of Parker et al., who study the role of the digital platform in building an ecosystem in the modern economy. The authors analyze how platforms are changing business models and how companies can use an ecosystem approach to achieve success (Parker et al., 2016).

We disagree with Ketels in the opinion that clusters are becoming more relevant in current world, and we argue that ecosystems are more preferable form of interactions organization compared to a cluster. In his researches, the author discusses various strategies and tools of industrial policy that can be used for the development of clusters and their benefits (Ketels, 2008; Ketels, 2013; Bresnahan, 2001).

Methodology

To achieve the goal of this article, a special application “Friendly text mining” was developed in the python programming language, using the NLTK package. The NLTK library assists the computer in analyzing, preprocessing and comprehending the written text. We have implemented the concepts of text mining using this library. The main function of the “Friendly text mining” is to analyze a large number of thematic articles and identify trends by finding and comparing the most used words in each period, illustrating all the results in the form of figures and tables. For completeness of the analysis, it also allows searching for words and phrases separately.

To develop this application, the authors took the following libraries in the python programming language as a basis:

NLTK is one of the most popular NLP libraries in Python (Field, 2017). Allows to perform operations on the text such as tokenization, lemmatization, removal of stop words and others.

EEL is a small library for creating interactive web applications, with full access to Python features and libraries. It is thanks to EEL that the authors managed to automate the entire process and make the program more understandable for users who are not familiar with programming.

Matplotlib is a Python library for plotting graphs, which allows the creation of a variety of graphs, with their subsequent customization.

Results

Loading the corpus.

We begin by uploading the files, that is, the “corpus” that we have prepared in advance. These are 600 articles in total from the “Scopus” database with keywords “business”, “competition” and “innovation”. To investigate the indicative dynamic of changes we chose 100 articles in any period from 1993–1998, 1999–2003, 2004–2008, 2009–2013, 2014–2018, 2019–2023.

```
@eel.expose
def extract(quantity, periods):
    periods = int(periods) + 1
    quantity = int(quantity) + 1
    data = ""
    if os.path.isfile('./txt/1.txt'):
        for k in range(1, periods):
            f = open(f'./txt/{k}.txt', 'r', encoding="utf-8")
            text = f.read()
            data = str(data) + str(text) + "/n"
            f.close()
            print(f"{k}/6 files are reloading")
            storage[k] = data
            data = ""
    else:
        for k in range(1, periods):
            for i in range(1, quantity):
                text = extract_text(f"./web/corpus/{k}/{i}.pdf")
                print(f"{i}\{quantity-1} in period {k}")
                data = str(data) + str(text) + "/n"
                f = open(f'./txt/{k}.txt', 'w', encoding="utf-8")
                f.write(data)
                f.close()
                storage[k] = data
                data = ""
    return "Uploading has been completed!"
```

Figure 1. Code for loading of the corpus into the program.

Note — compiled by the authors in the programming language "Python"

Figure 1 reports the lines of code for executing the first step of the program. The data on the number of periods and articles are calculated by this function. The corpus is loaded by extracting all text content from PDFs and saving it in TXT format and in the variable "storage". To do this, we used the ready-made extract function from the PDFminer library.

Cleaning the data and finding the frequency.

In Figure 2, the main lines of code for text analysis are provided. Feature of this function is that it's made up from various mathematical and scientific analysis python libraries such as "FreqDist", "WordNetLemmatizer", "MatPlotLib" and etc. All these libraries can be used at the same time with the help of authors' codes, which make text mining more comprehensive and comfortable.

```
1  @eel.expose
2  def creation_table(i):
3      found_words = []
4      lem = WordNetLemmatizer()
5      stop_words = set(stopwords.words("english"))
6      all_stop_words = stop_words.union(additional_stop_words)
7      filtered_words = []
8      lemmed_words = []
9      text = storage[i]
10     # remove numbers
11     text_nonum = re.sub(r'\d+', '', text)
12     # remove punctuations and convert characters to lowercase
13     text_nopunct = "".join([char.lower() for char in text_nonum if char not in string.punctuation])
14     # substitute multiple whitespace with single whitespace
15     text_no_doublespace = re.sub('\s+', ' ', text_nopunct).strip()
16     cleaned_text = text_no_doublespace
17     tokenized_text = word_tokenize(cleaned_text)
18     for w in tokenized_text:
19         if w not in all_stop_words:
20             filtered_words.append(w)
21         for w in filtered_words:
22             lemmed_words.append(lem.lemmatize(w))
23     frequency = FreqDist(lemmed_words)
24     frequency.plot(30, cumulative=False)
25     frequency_words = frequency.most_common(30)
26     a = frequency.most_common(30)
27     x = ' '
28     common_words_local = []
29     for w in range(30):
30         x = x +  f' {w+1}. {a[w][0]} : {a[w][1]} <br>'
31         common_words_local.append(a[w][0])
32     plt.show()
33     common_words[i] = common_words_local
34     print("common_words were added successfully")
```

Figure 2. Code for preparing the corpus and finding the frequency.

Note — compiled by the authors in the programming language "Python"

The first function is called "Remove numbers". The name of this function speaks for itself. The program removes all the numbers that are present in the corpus, since we seek for the most popular words but not the numbers.

The next function "Remove punctuation" removes punctuation marks, symbols, double spaces and other signs that do not contain semantic load.

The "Convert to Words" function splits the text into word tokens. This can be done by the function word_tokenize which is included in the NLTK package. This was made in row 17. Once the tokenization is completed, we will be able to learn valuable and useful information from the tokens. Frequency distribution is one of them. However, before that we need to find and delete stop words. Stop words are words and sentences that do not play any role in the intellectual analysis of the text. Usually "am, is, are, this, a, an" are treated as stop words. However, depending on the purpose of text mining, different words can be considered as stop words. Stop words in NLTK are the most common words in the data. These are words that you don't want to use to describe the topic of your content. They are pre-defined and cannot be deleted. This function removes the basic stop words that are written in the program code. After this process, all punctuation and non-informative stop words will be removed. This operation is done from row 18 to 20.

The lexicon normalization process takes into account another form of text noise. For example, the words "connection", "connected" and "connecting" can be combined into one word "connect". It does this by reducing all derivative-related variants of a word to their common base term. There are usually two ways to normalize the lexicon. These are stemming and lemmatization. Stemming is a method of linguistic normalization that either reduces words to their root word, or cuts out word-forming affixes. This process lowers the meaning of words to their root word. Lemmatization is the process of bringing words to their root word, which are lemmas that are linguistically correct. It does this by using techniques such as vocabulary and morphological analysis to change the base term. In most cases, lemmatization is considered as a more advanced process than stemming. Stemmer analyzes each word independently, without taking into account the surrounding text. For example, the word “better” comes from the word “good”, which serves as its lemma. This object will not go through the process of creating the basis for lemmatization, because this requires a preliminary search in the dictionary. NLTK has built-in function to make this procedure called lemmatize. In row 21-22 we use this function to all filtered words from previous steps. Since the corpus is ready, the text has passed all stages of processing and text analysis operations can be performed.

We have studied the ready-made vocabulary units and can find the distribution of the 30 most popular words in all periods. To find the most popular words in our word units we use another python library called “FreqDist”. From row 23 to 31 the function most_common of this library is implemented. Moreover, one more library called “matplotlib” was included into current research. This library helps to visualize textual information into the graphs and figures. We can get useful and sometimes unexpected information that would be difficult to notice without the use of text mining technology.

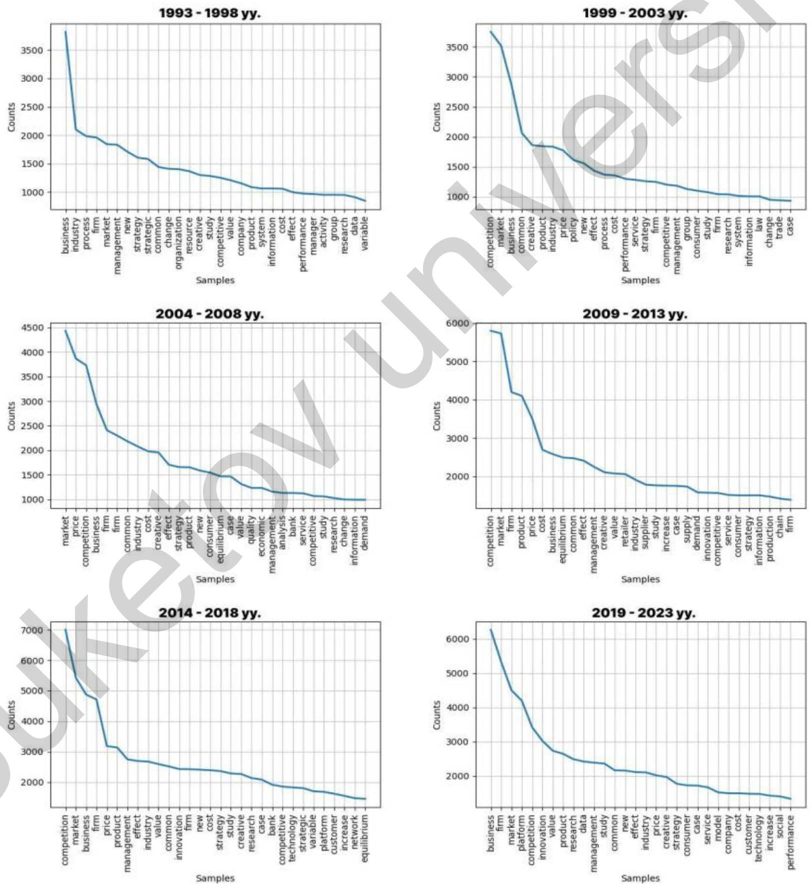

Figure 3. Distribution of the 30 most popular words in all periods.

Note — compiled by the authors in application “Friendly Text Mining”

We can see that all graphs have signs of hyperbole. This means that the words standing closer to the axis of the abscissa coordinates have the greatest number of repetitions and, accordingly, vice versa.

For example, in Figure 3, in the period from 1993 to 1998 we can observe that the word business is the most popular and occurs 3821 times, surpassing the next most frequent word industry (2100) almost by 2 times. The words process, market, management, new, strategy, strategic are also quite common. These words met in this period from 1500 to 2000 times. The last seven positions were occupied by the following words: performance, manager, activity, group, research, data, analysis and competition, which met less than 1000 times. Over the course of 30 years, the frequency of use of these words has varied from graph to graph. Later in the article, we will combine all the data from these graphs into one table for a more visual comparison.

Comparing the results

We have identified the top 30 popular words of each period and now we can observe the trend of changing words in all periods (Table 1). In this table, we have added all the words that were in the top 30 popular words in at least one of the periods. Comparing the absolute meanings of these words in each of the periods would not be entirely correct, since their general meanings differ. Therefore, the authors decided to change the values as a percentage relative to the total number of each word. For example, the word bank appeared only 4,913 times in all periods, and 7.47% of that number, or 367 words, were used between 1993 and 1998. Further, to make the table more visual, conditional formatting of the table was applied in a gradient from red (lowest values) to green (higher values).

If a word is monotonously colored in all periods, then it is equally distributed between periods. We should pay special attention to those words where there is a sharp change in the gradient. In this case, we are talking about the following words: bank, chain, data, innovation, law, manager, model, network, organization, platform, retailer, social network, supplier, supply, technology.

Table 1. The list of all popular words among 6 periods.

Words	**1993–1998**	**1999–2003**	**2004–2008**	**2009–2013**	**2014–2018**	**2019–2023**	**Total words**
1	2	3	4	5	6	7	8
Activity	24,39 %	16,48 %	9,24 %	9,60 %	22,05 %	18,23 %	3895
Analysis	13,21 %	11,50 %	17,72 %	18,38 %	20,06 %	19,13 %	6372
Bank	7,47 %	2,32 %	22,94 %	24,38 %	38,90 %	3,99 %	4913
Business	16,37 %	12,28 %	12,58 %	11,06 %	20,88 %	26,84 %	23341
Case	9,24 %	10,62 %	16,70 %	20,02 %	23,75 %	19,67 %	8758
Chain	10,66 %	5,75 %	12,85 %	39,45 %	14,94 %	16,35 %	3602
Change	20,71 %	13,93 %	14,65 %	14,56 %	18,93 %	17,21 %	6798
Common	11,21 %	16,07 %	16,99 %	19,28 %	19,58 %	16,88 %	12840
Company	20,68 %	11,31 %	17,40 %	10,11 %	13,57 %	26,93 %	5570
Competition	3,32 %	15,30 %	15,21 %	23,66 %	28,55 %	13,96 %	24523
Competitive	15,87 %	15,24 %	13,55 %	19,89 %	23,52 %	11,94 %	7875
Consumer	2,06 %	14,97 %	21,03 %	20,54 %	17,83 %	23,57 %	7336
Cost	9,63 %	12,34 %	18,02 %	24,57 %	21,77 %	13,66 %	10972
Creative	11,33 %	16,23 %	17,06 %	18,40 %	19,76 %	17,22 %	11457
Customer	11,25 %	9,86 %	12,28 %	15,84 %	26,47 %	24,30 %	6106
Data	12,42 %	11,77 %	10,73 %	12,98 %	18,95 %	33,15 %	7313
Demand	4,64 %	12,26 %	17,65 %	28,28 %	20,72 %	16,45 %	5604
Economic	10,73 %	15,39 %	21,00 %	15,00 %	18,10 %	19,77 %	5861
Effect	8,78 %	12,62 %	15,00 %	21,23 %	23,72 %	18,65 %	11350
Equilibrium	0,44 %	4,96 %	21,73 %	36,93 %	21,48 %	14,46 %	6751
Firm	11,57 %	7,35 %	13,57 %	8,21 %	27,83 %	31,48 %	16937
Group	15,19 %	18,00 %	9,19 %	19,74 %	16,69 %	21,20 %	6256
Increase	6,42 %	9,05 %	14,51 %	26,01 %	22,86 %	21,15 %	6762
Industry	16,53 %	14,45 %	16,34 %	15,06 %	21,03 %	16,59 %	12703
Information	15,68 %	14,85 %	14,63 %	22,27 %	14,98 %	17,59 %	6767
Innovation	2,64 %	7,69 %	4,61 %	19,06 %	29,38 %	36,62 %	8260
Law	11,33 %	44,09 %	12,30 %	10,98 %	12,65 %	8,65 %	2277
Management	15,86 %	10,22 %	10,01 %	19,46 %	23,75 %	20,70 %	11552
Manager	32,25 %	15,34 %	11,65 %	9,58 %	16,58 %	14,60 %	2986
Market	7,23 %	13,82 %	17,42 %	22,52 %	21,32 %	17,69 %	25453
Model	8,64 %	7,78 %	10,01 %	13,74 %	28,89 %	30,93 %	4933
Network	12,51 %	13,31 %	11,86 %	9,30 %	34,15 %	18,86 %	4299
New	15,84 %	14,45 %	14,71 %	12,62 %	22,33 %	20,05 %	10776
Organization	30,89 %	15,14 %	11,74 %	11,61 %	17,52 %	13,11 %	4532
Performance	14,24 %	19,04 %	13,17 %	13,77 %	20,16 %	19,62 %	6819
Platform	0,17 %	0,17 %	1,33 %	5,32 %	26,55 %	66,45 %	6319
Policy	13,15 %	29,05 %	13,53 %	13,51 %	15,46 %	15,31 %	5545

Words	1993–1998	1999–2003	2004–2008	2009–2013	2014–2018	2019–2023	Total words
1	2	3	4	5	6	7	8
Price	4,93 %	11,73 %	25,62 %	23,27 %	21,06 %	13,38 %	15102
Process	27,62 %	19,04 %	10,91 %	10,72 %	16,90 %	14,81 %	7176
Product	7,50 %	12,73 %	11,42 %	28,36 %	21,68 %	18,31 %	14464
Production	6,89 %	7,94 %	15,23 %	31,42 %	21,10 %	17,43 %	4688
Quality	8,52 %	11,93 %	25,12 %	16,95 %	15,52 %	21,96 %	4904
Research	10,55 %	11,56 %	11,41 %	14,99 %	23,74 %	27,76 %	8978
Resource	23,80 %	14,57 %	14,05 %	10,24 %	19,07 %	18,27 %	5731
Retailer	2,18 %	4,94 %	11,15 %	54,73 %	7,67 %	19,33 %	3766
Service	9,54 %	16,54 %	14,48 %	19,60 %	18,24 %	21,60 %	7735
Social	5,99 %	12,12 %	14,25 %	12,83 %	22,53 %	32,29 %	4358
Strategic	21,52 %	11,27 %	9,75 %	18,77 %	24,52 %	14,17 %	7341
Strategy	15,79 %	12,37 %	16,29 %	14,83 %	23,23 %	17,48 %	10159
Study	13,04 %	10,92 %	10,77 %	17,98 %	23,23 %	24,06 %	9824
Supplier	7,15 %	5,47 %	15,97 %	42,86 %	17,58 %	10,97 %	4165
Supply	3,04 %	7,98 %	14,75 %	47,15 %	13,63 %	13,44 %	3682
System	17,07 %	16,26 %	12,85 %	18,97 %	16,22 %	18,63 %	6216
Technology	9,55 %	10,84 %	14,64 %	8,16 %	31,35 %	25,46 %	5812
Trade	11,82 %	28,95 %	16,85 %	14,81 %	12,84 %	14,72 %	3240
Value	11,17 %	8,07 %	12,14 %	19,24 %	24,00 %	25,38 %	10789
Variable	12,53 %	11,66 %	12,71 %	20,60 %	25,20 %	17,30 %	6734
Firm	5,94 %	9,32 %	21,59 %	37,66 %	21,71 %	3,77 %	11156

Note — compiled by the authors in application "Friendly text mining"

From Table 1, we can easily see the spread of words in each period, and coloring helps to better notice how evenly words were used in all periods. If the color shades of a certain word differ slightly from each other from period to period and have a red shade, this indicates that this word was equally relevant throughout the entire period under study and has a high probability of continuing this trend. In contrast, if the colors in the periods differ, and some periods have a more yellow or green hue, it shows that these words were most relevant in these years, unlike others.

Of the total number of activity words used in all periods, which amounted to 3,895, approximately 24,4% of this number was used in the period from 1993–1998, that is, in the first period under consideration. The distinctive periods are periods 2004–2008 and 2009–2013, where there is a significant decrease in the frequency of this word.

The word bank was used in approximately 86% of cases in the 3rd, 4th and 5th periods and about 14% in the rest, with the highest value of 38.9% in the 2013–2018. In this period, the following words also showed the greatest importance: network — 34.15%, competition — 28.55%, technology — 31.35%.

The word chain showed the greatest relevance in the period from 2009–2013, scoring slightly less than 40%. The following words had the similar trend: cost — 24.57%, demand — 28.28%, equilibrium — 36.93, product — 28.36%. production — 31.42, retailer — 54.73, supplier — 42.86. supply — 47.15, firm — 37.66%. Most of the above words are related to each other. For example, the words product, production and chain are related to production topics. At the same time, demand, supply, supplier, retailer, equilibrium, firm describe market participants and market relationships. Therefore, the main object of research in the period from 2013 to 2018 can be suggested as market, as well as industrial production.

The word data was used approximately the same way from the 1st to the 4th periods, slightly increasing the frequency in the 5th period, increasing almost twice to the value of 33.15%. The words firm, innovation, model, research, social, study, value had a similar tendency. The data from the table shows that the study of the above concepts over time becomes more interesting for the authors of publications. It becomes clear that the authors are starting to mention more about research and models, pay more attention to social issues, and there is also a constant increase in interest in innovation.

The word platform had the same tendency. However, its uniqueness lies in the fact that it showed the greatest growth among the words under consideration in the entire study. Approximately 26% of this word was used in the period from 2014–2018 and about 66% in the period from 2019–2023. This suggests that research on platforms is just beginning to gain momentum in scientific circles. This can be caused for several reasons. One of them is that large enterprises are increasingly realizing the importance of building their own ecosystem for their further growth (Asheim, 2011). In this regard, there is a growing interest in the platforms on the basis of which these ecosystems are being built. Later in the article, we will use the word search function to confirm or refute this assumption.

There is also a relationship between the words law, policy, trade. All of them were most relevant in the period from 1999–2003.

The words manager, organization, process tend to decrease. In the first period, all had the highest values (approximately 30%) and after a while they began to occur less frequently, up to about 14% in the last interval. This may indicate that the study of management and operational activities of the enterprise interested researchers in the early 1990s. However, this interest slowly waned throughout the rest of the period.

Studying the genesis of ecosystem.

In this part, we want to check when, in the context of business, innovation and competition, scientists' interest in ecosystem research began, and how this interest changed during the period from 1993 to 2023. Having determined this trend, we want to compare it with the distributions of words that were studied by the authors in previous steps and, possibly, identify statistically interrelated groups of words that in a certain way affect the growth or decline of each other's relevance.

Up to this stage, we have determined that the word ecosystem has not been used often enough to enter the top 30 from any period. However, we have noticed the rapid growth of the word platform, and as we found out from the works of Isckia's and others (Isckia, 2014), the ecosystem is built on the basis of a specific platform. These results are encouraging and give us a reason to continue the study. As a continuation, the authors decided to determine the exact number of occurrences of the word "ecosystem" in each period and compare it with the words of interest — those that show a sharp contrast from period to period — in Table 1. Additionally, in this step we also investigated words "diversification", "digital" and "cluster" to find out their correlation with the word "ecosystem", and to check these results with the conclusions of other scientists.

We can see how our assumption about the relationship between platforms and ecosystems has been confirmed. Both of these words in percentage terms were used in the first period less than 0.5 percent and remained approximately at this level until the third period. A significant increase was noticeable in the fourth period, where the values sharply exceeded the mark of 25–30 percent, and in the sixth period were approximately equal to 65 percent. This proves that ecosystems and platforms have become one of the main topics of researchers and this interest began to appear from 2014–2018. In the theoretical part, we mentioned that platforms are the basis for creating ecosystems, and the relationship of these words is obvious.

However, what really deserves attention is that the word digital has almost the same tendency of use in percentage and absolute terms. The sharp increase in the study of digitalization has also dramatically increased the relevance in the study of ecosystems. This may indicate that digitalization has a positive impact on the emergence and development of ecosystems. In the future, you can review articles from the sixth period and draw more accurate conclusions using the manual method of text processing.

Table 2. The spread of the word ecosystem and related words among 6 periods.

Words	1993–1998	1999–2003	2004–2008	2009–2013	2014–2018	2019–2023	All words
Cluster	8,95 %	5,42 %	29,00 %	23,83 %	9,08 %	23,71 %	793
Data	12,42 %	11,77 %	10,73 %	12,98 %	18,95 %	33,15 %	7313
Digital	2,21 %	1,39 %	0,74 %	1,39 %	22,93 %	71,33 %	1221
Diversification	5,03 %	2,17 %	3,09 %	7,55 %	15,56 %	66,59 %	874
Ecosystem	0,40 %	0,48 %	0,48 %	2,15 %	31,92 %	64,57 %	1253
Firm	11,57 %	7,35 %	13,57 %	8,21 %	27,83 %	31,48 %	16937
Innovation	2,64 %	7,69 %	4,61 %	19,06 %	29,38 %	36,62 %	8260
Model	8,64 %	7,78 %	10,01 %	13,74 %	28,89 %	30,93 %	4933
Network	12,51 %	13,31 %	11,86 %	9,30 %	34,15 %	18,86 %	4299
Platform	0,17 %	0,17 %	1,33 %	5,32 %	26,55 %	66,45 %	6319
Research	10,55 %	11,56 %	11,41 %	14,99 %	23,74 %	27,76 %	8978
Study	13,04 %	10,92 %	10,77 %	17,98 %	23,23 %	24,06 %	9824
Technology	9,55 %	10,84 %	14,64 %	8,16 %	31,35 %	25,46 %	5812
Value	11,17 %	8,07 %	12,14 %	19,24 %	24,00 %	25,38 %	10789

Note — compiled by the authors in application "Friendly text mining"

The word diversification has a similar connection. Despite the fact that this term appeared in the circles of scientists studying economics much earlier than our studied periods, it began to gain the greatest popularity only in the last two periods. This suggests that the creation and development of ecosystems along with digitalization makes it easier for companies to diversify their activities.

You can also observe the relationship between the words “ecosystem” and “cluster”. Based on the bibliographic analysis, it can be noted that scientists recognize that in many respects the terms “ecosystem” and “cluster” have common characteristic features (Adner, 2017). However, in determining the differences between these terms from each other, the opinions of economists differ. There is a point of view in the scientific literature that an ecosystem is a more perfect form of interaction organization compared to a cluster (Moore, 2006; Sherwani, 2018; Autio, 2021; Arthur, 2021). Moreover, scientists emphasize that the ecosystem is the next stage in the evolutionary development of cluster models in the context of digitalization. This fact is confirmed by the results of the study. Table 2 shows that clusters began to become relevant for researchers in periods three and four (29% and 24%, respectively, in percentage terms of universal use for the entire period), before the beginning of mass interest of scientists in the topic of ecosystems. In the fifth period, the popularity of the use of the word cluster decreased to 9%, at the same time, the use of the concept of ecosystems increased from 2% to 32%. In the final period, interest in the research of clusters increased again — about 24%, however, this does not compare with the increase in relevance for ecosystems. In the sixth period, the word “ecosystem” accounted for 65% of occurrences, which is three times more frequent than the use of word clusters.

We can also build a correlation table from this data to quantify the relationship. So, using the Microsoft Excel program, and based on the following formula for calculating the Pearson correlation coefficient, we can get Table 3.

Table 3. A table of correlation coefficients.

	cluster	***data***	***digital***	***diversification***	***ecosystem***	***firm***	***innovation***	***model***	***network***	***platform***	***research***	***study***	***technology***	***value***
Cluster	1,00													
Data	0,95	1,00												
Digital	0,78	0,90	1,00											
Diversification	0,83	0,93	0,99	1,00										
Ecosystem	0,79	0,91	0,99	0,98	1,00									
Firm	0,93	0,99	0,90	0,91	0,92	1,00								
innovation	0,91	0,98	0,92	0,93	0,94	0,98	1,00							
Model	0,93	0,99	0,90	0,91	0,92	1,00	0,99	1,00						
Network	0,90	0,96	0,82	0,83	0,85	0,98	0,95	0,98	1,00					
Platform	0,80	0,92	1,00	0,99	1,00	0,92	0,94	0,92	0,84	1,00				
Research	0,95	1,00	0,88	0,90	0,90	0,99	0,98	1,00	0,98	0,90	1,00			
Study	0,95	0,99	0,85	0,87	0,87	0,98	0,97	0,99	0,98	0,87	1,00	1,00		
technology	0,92	0,98	0,86	0,87	0,89	0,99	0,97	0,99	0,99	0,88	0,99	0,98	1,00	
Value	0,96	0,99	0,86	0,88	0,88	0,99	0,98	0,99	0,98	0,88	1,00	1,00	0,98	1,00

Note — compiled by the authors in application Microsoft Excel

Table 3 shows that the word “ecosystem” has a strong connection with the words digital, diversification and platform, since in all cases the correlation coefficient is higher or equal to 0.98 and their relationship is statistically significant. The word platform can be especially noted. Their correlation coefficient is equal to the maximum value of 1.

Summarizing the results obtained from Tables 1 and 2, we can conclude that in the context of business, competition and innovation, ecosystems were not particularly popular among scientists until 2014. As interest in topics such as innovation, digitalization, diversification, and platforms has developed, there has also been an increase in interest in ecosystems. You can also see how the word cluster is noticeably different from all the words. It has the most minimal correlation in comparison with others. It can be concluded that as diversification and digitalization grow, the cluster approach is giving way to the ecosystem approach. These connections serve as a prerequisite for further in-depth analysis of their causes and potential consequences. However, within the scope of this study, we will limit our discussion to these results.

Conclusions

The study confirmed the hypothesis that interest in ecosystem research is growing in scientific circles and the number of articles studying the cluster approach is decreasing. This is primarily due to the development of digital platforms that contribute to the evolution of the organization of business processes from clusters to ecosystems. In the context of digitalization, enterprises are striving to introduce innovative business methods. Ecosystems are priority areas of business development and promotion.

Statistically interrelated groups of words have been identified. These are words related to the topic of ecosystems, such as diversification, digital, cluster. (which in a certain way affect the development of ecosystems).

The development of digital platforms and networks creates new opportunities for businesses and consumers, accelerates the processes of interaction between all participants, and most importantly changes traditional business models. The diversification of enterprises' activities is accompanied by the creation, implementation and development of ecosystems.

The assumption about the relationship between platforms and ecosystems was confirmed. This proves that ecosystems and platforms have become one of the main topics of researchers. The development of digitalization processes has also influenced the increasing relevance of ecosystem research. This may indicate that digitalization has a positive impact on the emergence and development of ecosystems.

As a recommendation for future research, we suggest studying the content and composition of ecosystems as a means of enhancing business process efficiency. Additionally, increasing the state's role in ecosystem management could foster ecosystem growth and development. This could involve measures such as establishing innovation parks, providing tax incentives, financing and supporting startups, and other strategic initiatives.

The review of the articles demonstrates the trend of changing the focus of scientific research of scientists and shows how, as time passed, some words began to lose relevance and be replaced by others. Identifying these changes, especially those that have a sharp rise or decline, as well as further studying their causes, can help advance research in the field of business, innovation and competition. This review and recommendations can enrich the methodological landscape of innovative research and allow the community to use the opportunities provided by digital technologies.

Text mining applications provide access to the processing of a large amount of information and allow you to identify the main trends in research. Therefore, the role of text mining technology in scientific research is quite high.

References

Adner, Ron. (2017). Ecosystem as structure: An actionable construct for strategy. *Journal of management, 43,* 1, 39–58.

Andersson, T., Curley, M.G., & Formica, P. (2010). Embracing Business Ecosystems to Enable Sustainable and Accelerated Innovation. *Knowledge-Driven Entrepreneurship: The Key to Social and Economic Transformation*, 71–78.

Arthur, W. Brian. (2021). Foundations of complexity economics. *Nature Reviews Physics, 3,* 2, 136–145.

Asheim, Bjørn T., Ron Boschma, & Philip Cooke. (2011). Constructing regional advantage: Platform policies based on related variety and differentiated knowledge bases. *Regional studies, 45,* 7, 893–904.

Autio, Erkko & Arthur W. Brian. (2021). Foundations of complexity economics. *Nature Reviews Physics, 3,* 2, 136–145.

Boschma, R. (2015). Towards an evolutionary perspective on regional resilience. *Regional studies,* 5, 733–751.

Bresnahan, Timothy, Alfonso Gambardella, & AnnaLee Saxenian. (2001). "Old economy" inputs for "new economy" outcomes: Cluster formation in the new Silicon Valleys. *Industrial and corporate change,* 4, 835–860.

Chesbrough, Henry William. (2003). *Open innovation: The new imperative for creating and profiting from technology.* Harvard Business Press.

Cockburn, Iain M., Rebecca Henderson, & Scott Stern. (2018). *The impact of artificial intelligence on innovation: An exploratory analysis. The economics of artificial intelligence: An agenda.* University of Chicago Press.

Fluck, J., Deneke, H., & Gieger, C. (2005). Text mining in life sciences. *WIT Transactions on State-of-the-art in Science and Engineering,* 17.

Isckia, T., & Lescop, D. (2014). Platform- based ecosystems: leveraging network-centric innovation. *Understanding Business Ecosystems: How Firms Succeed in the New World of Convergence, De Boeck,* 97–111.

Jacobides, Michael G., Carmelo Cennamo, & Annabelle Gawer. (2018). Towards a theory of ecosystems. *Strategic management journal,* 8, 2255–2276.

Ketels, Christian HM, & Olga Memedovic. (2008). From clusters to cluster-based economic development. *International journal of technological learning, innovation and development, 1,* 3, 375–392.

Ketels, Christian. (2013). Recent research on competitiveness and clusters: what are the implications for regional policy? *Cambridge Journal of Regions, Economy and Society,* 2, 269–284.

Korhonen, A., Ó Séaghdha, D., Silins, I., Sun, L., Högberg, J., & Stenius, U. (2012). Text Mining for Literature Review and Knowledge Discovery in Cancer Risk Assessment and Research. *PLoS ONE, 7*(4).

Lakhani, Karim R. & Jill A. Panetta. (2007). The principles of distributed innovation. *Innovations: Technology, Governance, Globalization Summer*, *2*, 3.

Losiewicz, P.B., Oard, D.W., & Kostoff, D.N. (2003). *Science and Technology Text Mining Basic Concepts.* Air Force Research Lab.

Moore, J.F. (2006). Business Ecosystems and the View from the Firm. *The Antitrust Bulletin, 51*, 31–75.

Parker, Geoffrey G., Marshall W. Van Alstyne, & Sangeet Paul Choudary. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you.* WW Norton & Company.

Salloum, S.A., Al-Emran, M., Monem, A.A., & Shaalan, K.F. (2018). Using Text Mining Techniques for Extracting Information from Research Articles. *Intelligent natural language processing: Trends and Applications,* 373–397.

Sherwani, H.U. & Tee, R. (2018). Innovation and Value Creation in Business Ecosystems. *In Learning and Innovation in Hybrid Organizations: Strategic and Organizational Insights* (pp. 13–32). Cham: Springer International Publishing.

West, Joel & Marcel Bogers. (2014). Leveraging external sources of innovation: A review of research on open innovation. *Journal of product innovation management*, 4, 814–831.