

Competitive Intelligence Text Mining: Words Speak

A. Zarei*, M. Maleki, D. Feiz and M. - A. Siah sarani Kojouri

Faculty of Economic, Management and Administrative Sciences, Semnan University, Semnan, Iran.

Received 20 December 2016; Revised 22 February 2017; Accepted 10 May 2017

*Corresponding author: a_zarei@semnan.ac.ir (A. Zarei).

Abstract

Competitive intelligence (CI) has become one of the major subjects for researchers in the recent years. The present research work is aimed to achieve a part of CI by investigating the scientific articles on this field through text mining in three inter-related steps. In the first step, a total of 1143 articles released between 1987 and 2016 are selected by searching the phrase "competitive intelligence" in the valid databases and search engines; then through reviewing the topic, abstract, and main text of the articles as well as screening the articles in several steps, the authors eventually selected 135 relevant articles in order to perform the text mining process. In the second step, pre-processing of the data is carried out. In the third step, using non-hierarchical cluster analysis (k-means), 5 optimum clusters are obtained based on the Davies–Bouldin index, for each of which a word cloud is drawn; then the association rules of each cluster are extracted and analyzed using the indices of support, confidence, and lift. The results obtained indicate the increased interest in research works on CI in the recent years and tangibility of the strong and weak presence of the developed and developing countries in formation of the scientific products; further, the results show that information, marketing, and strategy are the main elements of CI that along with other prerequisites can lead to CI and, consequently, the economic development, competitive advantage, and sustainability in the market.

Keywords: *Competitive Intelligence, Text Mining, Association Rules Mining, Word Cloud.*

1. Introduction

Today, the organizations are increasingly paying attention to the competitive intelligence (CI) because it supports the needs of an organization for collecting, interpreting, and publishing the external information [1]. CI supports the strategic process in an organization and, like a sensor, shows the managers whether the organization is competitiveness [2]. It is a vital element of the company's planning and management process that puts together the data and information from a very broad and strategic perspective and allows the company to foretell what would happen in the competitive environment [3]. This indicates that the CI has become the latest weapon in the economic war so that many of the emerging economies view it as a way to overcome the larger and more industrialized countries in the economic war [4]. CI does not mean industrial espionage, and its key principle is that 90% of all the information that a company requires for making important decisions and understanding the market

and competitors is available in the general environment of the industry or can be systematically developed from the data of the industry's general environment [5]. The reason for emergence of this new scientific field is the explosion of information due to the increased access to information as well as the rapid and widespread proliferation of information in the commercial databases; besides, the broad social and political changes, increased pace of business, increased global competition due to emergence of new competitors, and technological changes can be mentioned as other reasons [6]. Data mining has become one of the major tools of business intelligence for discovering knowledge so that several extensive research works [7-9] have focused on the potential tools of this technique. One of the major branches of data mining is text mining, which generally deals with extraction of the hidden knowledge from unstructured textual documents [10]. One of the main applications of

text mining is to help CI that the activists of the business world are seeking to acquire. In the scientific-academic area, numerous articles have been written on CI by various authors in different countries in successive years; however, further exploration in this regard can result in valuable applicable knowledge, which seems uninvestigable without using the data mining techniques. For example, searching the term “competitive intelligence” in the Science Direct and Emerald databases results in 1248 and 15549 accessible articles, respectively, the investigation, study, and summation of which would be an exhausting, time-consuming, and, in some cases, impossible task for researchers, indicating the necessity of the use of data mining to achieve the initial exploratory vision of the research work. The present research work is seeking to achieve the following objectives:

- Choosing the most relevant articles in the field of CI through screening the articles by investigating the titles, abstracts, and texts.
- Clustering the articles in the field of CI, extracting the central words, and drawing the word cloud of each cluster.
- Discovering the association rules for each cluster and extracting the application rules in accordance with the support, confidence, and lift indices as well as the experts’ opinions.

2. Literature review

The CI concept embraces the two main keywords intelligence and competitiveness. The concept of intelligence is a part of the marketing strategy, which causes the increase in the company’s competitive power and processes of its strategic plans [11]. Today the concept of intelligence is considered as a process that improves the competitiveness and process of strategic planning [12]. For instance, according to the UNCTAD’s definition, competitiveness in the field of export refers to diversifying the export baskets, maintaining higher values of the export growth over time, improving the content of export activity in terms of skill and technology, and expanding the base of the domestic companies in order to achieve competitiveness in the international arena so that the competitiveness becomes sustainable and continuous and, subsequently, the level of living standard in the country is promoted [13]. Regarding the above-mentioned definitions for intelligence and competitiveness, CI referring to the intelligence of a company or a business is the art of absorbing and transferring the knowledge from the extra-organizational elements and the environment through specific rules and certain

standards to the organization in order to protect the competitive threats, identify and utilize the potential opportunities, and build the future, and thus it can influence the competitive position of the company [14-16].

According to Rouach and Santi (2010), in the knowledge-based organizations, which emphasize on managing the organizational knowledge, rearing the learning organizational culture, and implementing the supportive organizational structures in order to achieve the organizational intelligence, five classes of intelligence play fundamental roles, and are considered as the antecedents of intelligence in the age of knowledge. The five classes include market intelligence, competitor intelligence, technology-based intelligence, human intelligence, and structural intelligence [16]. Some researchers believe that the first two types of intelligence, namely market intelligence and competitor intelligence, are sub-categories of the competitive intelligence, and the other three including human intelligence, structural intelligence, and technological intelligence are sub-categories of the organizational intelligence [17]. Champs and Nayak (1995) has presented CI in three classes [18], as what follow. 1) Market intelligence: this type of intelligence is responsible for providing and supplying the current and future trends of the requirements and preferences of the customers for the new markets; in this intelligence, the information of the customers, suppliers, buyers, and distributors is collected and analyzed. 2) Competitors’ intelligence: this type of intelligence is responsible for evaluating the evolution of the competitors’ competitive strategy through continuous and systematic investigation of the competitors’ structure and providing new alternative products for the industry. Competitor intelligence is a continuous effort to evaluate the behaviors and capabilities of competitors to develop the competitive advantage [19]. Competitor intelligence is aimed to refine and analyze the actions and performance of the competitors in order to achieve their vision [20]. 3) Technological intelligence: this type of intelligence is responsible for the cost-benefit analysis of the new and current technologies, prediction of the technologies that will disappear in the future as well as the technologies that will dominate in the future.

Environmental changes have caused changes in the commercial companies’ decision-making and competition methods. The leaders in modern companies are facing decisions that require merging, analysis, and summarization of the

internal and external information from various resources. The information technologies such as extensive organizational systems, data mining, and text mining are responsible for collecting and merging the interactions from different perspectives in order to support the decisions [21, 22]. The data mining process is a process of knowledge discovery since it embraces the human knowledge [23]. One of the fields of data mining is text mining [24]. Text mining refers to the process of deriving or extracting high quality information from a text. Text mining is based, on the one hand, upon the computational linguistics and, on the other hand, upon the mathematical-statistical computations and data analysis. The technologies used in text mining include information retrieval, information extraction, subject tracking, summarization, categorization, conceptual relationship, information visualization, and question responding [24, 25]. The main advantage of using the text mining technologies for CI is the high capability of text mining for rapid processing, objectification, and exclusion of large amounts of text data [26].

3. Methodology

Approach of the present research work is close to the three-step model of the process of knowledge discovery in text (KDT) presented by Karanikas and Theodoulidis (2002) [27]. The KDT model is composed of three steps: 1) data collection, 2) text pre-processing, and 3) text mining. The present research work passed the above-mentioned three steps. In the first step, a total of 1143 articles

published between 1987 and 2016 were selected by searching the phrase "*competitive intelligence*" in the databases such as Science Direct, IEEE, Springer, Wiley, Taylor & Francis, Sage, Emerald, and Jstor as well as the search engines such as Google and Google scholar. By reviewing the titles, abstracts, and original texts as well as various steps of screening, the authors finally selected 135 articles on CI in order to perform the text mining process. In the second step of this research work, pre-processing of the data was performed. The tasks such as segmentation, conversion of the uppercase letters to lowercase letters, and removal of the stop word were performed in this step. Tokenization refers to the process of breaking a sequence of strings into the pieces called segments. The purpose of tokenization is to discover the words in a sentence and identify the meaningful keywords [28]. The stop words are the non-verbal words with no information load. The prepositions, prefixes, and pronouns can be considered as stop words. Etymology is the process of converting words into their roots. Many words in the English language are indeed the different forms of the same words (Liau and Tan, 2014) [28]. In the third step, using the non-hierarchical cluster analysis (k-average), the number of optimum clusters was obtained based on the Davies–Bouldin index, the word cloud of each cluster was drawn, and, finally, the association rules of each cluster were extracted and analyzed. Figure (1) shows the steps covered in the present research work.

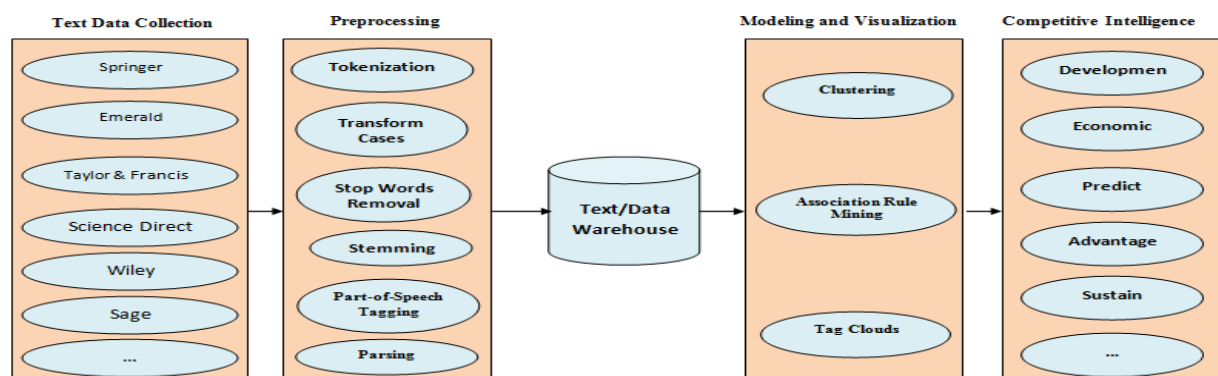


Figure 1. Steps covered in present research work.

4. Data analysis

4.1. Descriptive statistics

Investigating the selected articles in terms of the publication year and their producing continent can be notable from various aspects. Analysing the slope of production of the articles on CI indicated that taking this subject into consideration is

increasing with a slight slope so that the importance of this field has reached its peak with publication of 26 articles in 2015. Diagram (1) shows the selected articles of the present research work based on their publication year.

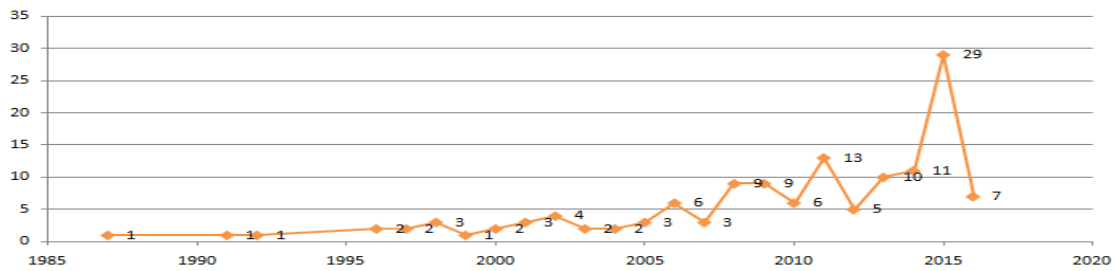


Diagram 1. Articles of present research work based on their publication year.

Investigating the published articles on CI based on the continent producing the articles indicates that this subject is of special position and importance among the Asian researchers, and Europe and South America rank the second and third, respectively. The interesting point is that among the Asian, European, and South American countries, China, Spain, and the United States have published the highest number of published

articles on CI, respectively, indicating the research orientations in these countries. Reviewing the countries producing the articles on the field of CI obviously shows an outstanding gap between the developed and developing countries as well as tangibility of strong and weak presence of both groups in formation of the scientific products. Diagram (2) represents the article-producing countries based on their continent.

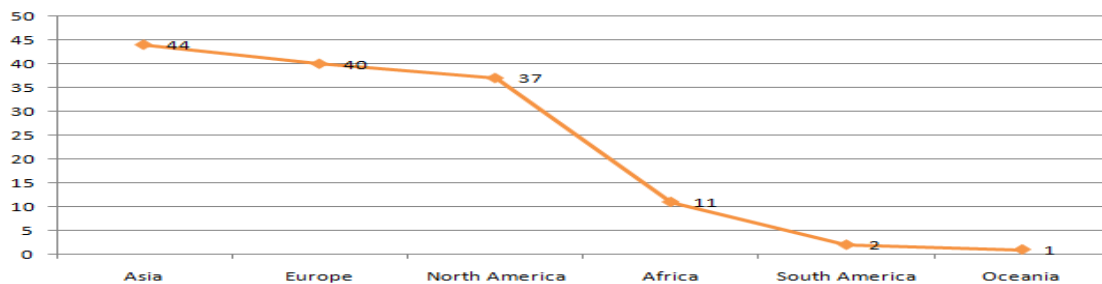


Diagram 2. Article-producing countries based on their continent.

4.2. Non-hierarchical cluster analysis (k-means)

Cluster analysis is aimed for decomposing a dataset into groups so that the data within a group has the highest similarity to each other and the highest differences with other groups [29]. The K-means cluster method was used to achieve the Euclidean distance between the clusters. During clustering, the number of groups (K) would be discovered, which is not clear at the beginning of the heuristic approach of clustering [30, 31]. K-means algorithm is a well-known clustering algorithm. In spite of its advantages such as high speed and ease of employment, this

algorithm suffers from the problem of local optima [32]. In order to find the number of optimum k, the k-means algorithm with entry of k = 2 to k = 10 was performed, and then the Davies–Bouldin index was used to determine the number of optimum clusters. The Davies–Bouldin index calculates the average similarities between each cluster and the cluster that is most similar to it so that the lower the index value, the better the produced clusters. Table (1) shows the values for the Davies-Bouldin index based on the k (number) of different clusters.

Table 1. Values for Davies-Bouldin index based on number of clusters.

Number of clusters	2	3	4	5	6	7	8	9	10
Davies–Bouldin index	3.281	3.082	2.70	2.432	2.598	2.543	2.501	2.471	2.762

As shown in Table (1), the Davies–Bouldin index considers 5 clusters as appropriate for clustering. In case of 5 clusters, the Davies–Bouldin index has obtained the lowest value that is 2.432. The results of cluster analysis showed that the statistical population of the present research work could be clustered in five clusters, each of which is different from other clusters in terms of the studied dimensions. Afterwards, 10 words were presented in each cluster along with their final center; besides, the Knime software was used to draw the word cloud of each cluster. These word clouds are unique to each cluster, demonstrating its specific conditions. The fact that the word “information” is centralized in all tag clouds indicates its importance in the domain of competitive intelligence. Other words surrounding this one differ among the tag clouds in terms of importance, distance, and size, and these word clouds must not be considered equal. These figures show the details of the dispersion of words in each cluster on the one hand, and the distance of each word from cluster centers on the other. Moreover, they demonstrate the importance of each word in each cluster through the size of the word (more important words in each tag cloud are larger).

Therefore, since these word clouds offer good visual representations of each cluster, the readers can more easily understand each cluster through these word clouds.

First cluster: This cluster accounted for 37% of the statistical sample’s articles, and the pivotal subjects in this cluster included marketing, information, and corporation. Marketing, as the most important keyword, was located at the center of this cluster, and other words such as information, corporation, and competitors were at longer distances from this word. Analysis of the words in the first cluster shows the importance of taking into account the internal (companies, products, marketing) and external (competitors, customers) environments considering the strategy and applying the information to accomplish industrialization and development. The importance of marketing in CI is such that it is referred to as the marketing intelligence. Marketing intelligence has been defined as the capability to achieve and use the commercial resources and the processes that “the companies learn about their markets and use the market knowledge” [33]. Table (2) presents the words and final centers of the first cluster.

Table 2. Words and final centers of first cluster.

Words	Marketing	Information	Firm	Competitor	Strategy	Analysis	Development	Product	Industrial	Customer
Cluster Center	0.394	0.343	0.161	0.146	0.146	0.134	0.120	0.106	0.104	0.085

In this cluster, the word “information” ranked first in terms of frequency with 2795 times of repetition, and appropriated the center of the word cloud, indicating the importance of this word in

the cluster. The words “marketing” and “strategy” acquired the next ranks with 1456 and 1197 times of repetition, respectively. Figure (2) shows the first word cloud.



Figure 2. Word cloud of first cluster.

Second cluster: This cluster accounted for 30% of the statistical sample articles, and the pivotal subjects of the cluster included data, information, knowledge, and technology. This cluster embraced the various subjects from analysis and

path of data to knowledge, and it seemed that the secret of excellence (superiority) was rooted in being equipped and armed with knowledge, awareness of the competitors’ status, awareness of the environment, and being updated in terms of

technology. The information that is somewhere between data and knowledge would help the decision-makers to understand their competitors

and make quick and intelligent strategic decisions [34, 35]. Table (3) represents the words and final centers of the second cluster.

Table 3. Words and final centers of first cluster.

Words	Information	Data	Analysis	Competitor	Marketing	Product	Knowledge	Environment	Development	Technology
Cluster Center	0.679	0.143	0.136	0.115	0.106	0.106	0.105	0.102	0.091	0.084

In this cluster, the word “information” ranked first in terms of the frequency with 2177 times of repetition, and appropriated the center of the word cloud, indicating the importance of this word in

this cluster. The words “marketing” and “strategy” were in the next ranks with 1209 and 1179 times of repetition, respectively. Figure (3) shows the word cloud of the second cluster.



Figure 3. Word cloud of second cluster.

Third cluster: This cluster accounted for 18% of the statistical sample articles, and the pivotal subjects included knowledge, information, environment, and network. In this cluster, on one hand, special importance was put on knowledge and analysis and, on the other hand, the environmental factors were also taken into consideration. Knowledge is a flowing combination of the experiences, values, ground

information, and expert vision, providing a framework for evaluation and application of the experiences and new information [36]. Acquisition of knowledge requires technology, environmental exploration, and strategy sharing it with the internal and external units. Table (4) represents the words and final centers of the third cluster.

Table 4. Words and final centers of third cluster.

Words	Knowledge	Information	Strategy	Development	Analysis	Environment	External	Network	Marketing	Share
Cluster Center	0.304	0.264	0.141	0.120	0.112	0.102	0.079	0.077	0.076	0.075

In this cluster, the word “information” ranked first in terms of the frequency with 1575 times of repetition, and appropriated the center of the word cloud, indicating the importance of this word in

this cluster. The words “marketing” and “knowledge” were in the next ranks with 830 and 557 times of repetition, respectively. Figure (4) shows the word cloud of the third cluster



Figure 4. Word cloud of third cluster.

Fourth cluster: This cluster accounted for 8% of the statistical sample articles, and the pivotal subjects included technology, data, and analysis. In this cluster, the two words “technology” and “competitive advantage” attract the attention, and it seems that, in this cluster, in addition to other factors such as data, information, and strategy, the technology has been specially taken into consideration, and this word has been assumed as the main focus of the keywords of the cluster, which can pave the ground for achieving the

competitive advantage. Technology is considered as the most important element in many industrialized companies, and means any kind of information that can affect development or production of the products [37]. The importance of technology in formation of CI is to the extent that Champs and Nayak (1995) have referred to it as the technological intelligence [18]. Table (5) represents the words and final centers of the fourth cluster.

Table 5. Words and final centers of fourth cluster.

Words	Technology	Data	Analysis	Information	Development	Industry	Marketing	Strategy	Technique	Advantage
Cluster Center	0.435	0.263	0.257	0.252	0.145	0.103	0.083	0.077	0.075	0.075

In this cluster, the word “information” ranked first in terms of frequency with 406 times of repetition, and appropriated the center of the word cloud, indicating the importance of this word in this

cluster. The words “marketing” and “knowledge” were in the next ranks with 318 and 275 times of repetition, respectively. Figure (5) shows the word cloud of the fourth cluster.



Figure 5. Word cloud of fourth cluster.

Fifth cluster: This cluster accounted for 7% of the statistical sample articles, and the central subjects included competitors, information, market, and industry. In this cluster, the subject of “competitors” has been specially taken into consideration, representing the great importance of this subject in CI. Legally and ethically, CI

uses the public resources to achieve the strengths and weaknesses of the competitors. CI would enable the companies to develop the corporation in order to achieve a larger share of the market and have a successful competition against the international competitors [38]. The companies competing at the global level face numerous

challenges since the consumers use the internet to compare products and prices, and thus these companies need to be ahead of their competitors

[39]. Table (6) represents the words and final centers of the fifth cluster.

Table 6. Words and final centers of fifth cluster.

Words	Competitor	Information	Market	Industry	Strategy	Product	Develop	Data	Plan	Technology
Cluster Center	0.454	0.316	0.290	0.172	0.167	0.146	0.138	0.131	0.128	0.090

In this cluster, the word “information” ranked first in terms of frequency with 783 times of repetition, and appropriated the center of the word cloud, indicating the importance of this word in this

cluster. The words “market” and “strategy” were in the next ranks with 454 and 381 times of repetition, respectively. Figure (6) shows the word cloud of the fifth cluster.



Figure 6. Word cloud of fifth cluster.

4.3. Association rules mining

Data mining is the implicit, valid, and potential extraction of the useful knowledge from a raw dataset [40], and is widely used for discovering the hidden rules from the data with the aim of extracting the knowledge for making decisions such as predicting the costumers’ behaviors [41, 42]. The “association rules” is a data mining technique, which helps discovering the interesting relationships between the variables in massive data [43, 44]. The “association rules” is a sign of the cooperation of occurrence of two variables, meaning that if two variables frequently concur, there will be a strong relationship between them [40]. Since the association rules have the potential to produce many patterns and rules, the three indices of support, confidence, and lift as well as the researchers’ opinions were used to evaluate and extract the rules. The support index has a numerical value between zero and one so that the greater values indicate that two objects are more inter-related. By determining a threshold for this criterion, the user can obtain only the rules, the support of which is greater than the threshold value; accordingly, by reducing the search space, the time required to find the association rules can be minimized. The confidence criterion has also a

numerical value between zero and one so that the greater the number, the higher the quality of the rules. The use of this criterion along with the support index would be an appropriate complement for assessment of the association rules but there still remains the problem that there might be a rule with high confidence that is not valuable from our viewpoint. The lift criterion represents the rate of independence between the objects A and B, which can be a numerical value between zero and infinity. In fact, the lift criterion considers the value of the concurrence between the features and imports the value of individual occurrence of the consequent section of the rule, namely the object B, in its calculations. The values close to 1 imply that A and B are independent from each other and, accordingly, do not represent the lift rule. For this criterion, the values lower than 1 indicate a negative relationship between A and B, while the values higher than 1 indicate that A provides more information about B, and thus the rule of $A \Rightarrow B$ is evaluated to be of higher lift. Table (7) presents a summary of the definitions along with the formula of the three above-mentioned indices [45].

Table 7. Summary of definitions along with formula of three measures [45].

Measure	Description	Formula
Support	Usefulness of discovered rule $A \rightarrow B$	$P(A \cap B)$
Confidence	Certainty of discovered rule $A \rightarrow B$	$P(B A)$
Lift	Correlation between the occurrence of items in discovered rule $A \rightarrow B$	$\frac{P(B A)}{P(B)}$

In the present work, the three above-mentioned indices along with the experts' opinions were used to extract the practical and interesting rules, the

results of which are presented for each cluster in Table (8).

Table 8. Practical and interesting rules for each cluster.

Rule	Premises (If)	Conclusion (Then)	Confidence	Support	Lift
Cluster 1					
C1-1	data, develop, technology	specify, advantage	0.769231	0.2	3.496503
C1-2	strategy, marketing, data	develop, specify, advantage	0.769231	0.2	3.496503
C1-3	strategy, public, develop, critic	challenge, advantage	0.833333	0.2	3.472222
C1-4	strategy, marketing, technology, critic	product, develop	1	0.22	3.333333
C1-5	strategy, market, data, specify	develop, advantage	1	0.22	3.333333
C1-6	strategy, data, critic	marketing, advantage	0.833333	0.2	3.205128
C1-7	strategy, product, plan	economic, generate	0.666667	0.2	3.030303
C1-8	market, technology	develop, critic, future	0.666667	0.2	3.030303
C1-9	strategy, market, data, critic	develop, advantage	0.909091	0.2	3.030303
C1-10	strategy, technology	market, develop, advantage	0.714286	0.2	2.97619
C1-11	plan, relationship	develop, knowledge	0.769231	0.2	2.95858
C1-12	technology, critic, relationship	market, develop	1	0.2	2.941176
C1-13	marketing, knowledge	develop, relationship	0.909091	0.2	2.840909
C1-14	data, analysis	technology	0.75	0.3	1.63043478
C1-15	information	product	0.68	0.34	1.36
C1-16	analysis	information	0.64	0.32	1.28
C1-17	environment	information	0.625	0.3	1.25
C1-18	information	place	0.6	0.3	1.30434783
C1-19	strategy, information	product, economic	0.714286	0.2	2.55102
C1-20	data, analysis	information, technology	0.6	0.24	2.142857
Cluster 2					
C2-1	scan, demand	economic	1	0.325	2
C2-2	competitor, strategy	marketing	1	0.325	2
C2-3	scan, information	strategy	0.928571	0.325	1.954887
C2-4	Scan	knowledge, economic	0.65	0.325	1.857143
C2-5	Data	pattern, inform	0.65	0.325	1.857143
C2-6	import	inform, competitor	0.65	0.325	1.857143
C2-7	scan, knowledge	economic	0.928571	0.325	1.857143
C2-8	information, data	pattern	0.928571	0.325	1.857143
C2-9	information, import	competitor	0.928571	0.325	1.857143
C2-10	aware	knowledge	0.8125	0.325	1.625
C2-11	marketing, competitor	public	0.722222	0.325	1.520468
C2-12	information	knowledge	0.75	0.375	1.5
C2-13	demand	knowledge	0.736842	0.35	1.473684
C2-14	Data	custom	0.65	0.325	1.444444
C2-15	Scan	knowledge	0.7	0.35	1.4
C2-16	Scan	information	0.7	0.35	1.4
C2-17	knowledge	economic	0.7	0.35	1.4
C2-18	Scan	demand	0.65	0.325	1.368421
C2-19	environment	information	0.65	0.325	1.3
C2-20	environment	knowledge	0.65	0.325	1.3
Cluster 3					
C3-1	people, communication	aware, culture	0.888889	0.333333	2.666667
C3-2	communication, know	aware, knowledge	0.8	0.333333	2.4
C3-3	intellectual	sustain, industry	0.727273	0.333333	2.181818
C3-4	secondary, communication	know	1	0.333333	2.181818
C3-5	secondary, knowledge	world	1	0.333333	2.181818
C3-6	human, employee	intellectual	1	0.333333	2.181818
C3-7	communication, knowledge	know	1	0.333333	2.181818
C3-8	culture, train	intellectual	1	0.333333	2.181818
C3-9	secondary, aware, primary	know	1	0.333333	2.181818
C3-10	social, promotion, commit, environment	consumer	1	0.333333	2.181818
C3-11	employee	sustain, intellectual	0.8	0.333333	2.133333
C3-12	employee	sustain, advantage	0.8	0.333333	2.133333
C3-13	purchase, market	profit	0.888889	0.333333	2.133333
C3-14	secondary	inform, anticipate	0.666667	0.333333	2
C3-15	relationship	experience, collaboration	0.666667	0.333333	2
C3-16	advantage, employee	sustain	1	0.333333	2

Continued Table 8. Practical and interesting rules for each cluster.

C3-17	marketplace, finance	promotion	1	0.333333	2
C3-18	relationship, quality	consumer	0.888889	0.333333	1.939394
C3-19	human, intellectual	industry	0.888889	0.333333	1.939394
C3-20	market, employee	profit	1	0.333333	2.4
Cluster 4					
C4-1	social, network	relationship	1	0.454545	2.2
C4-2	communication	knowledge, customer	1	0.454545	2.2
C4-3	knowledge	communication	1	0.454545	1.833333
C4-4	opportunity, network, economic, develop, competitor	future	1	0.454545	2.2
C4-5	network	global	0.833333	0.454545	1.833333
C4-6	strategy	opportunity, capable	0.833333	0.454545	1.833333
C4-7	strategy, data	capable	1	0.454545	1.833333
C4-8	extern, capable	opportunity	1	0.454545	1.833333
C4-9	market, data	capable	1	0.454545	1.833333
C4-10	custom, communication	knowledge	1	0.454545	1.833333
C4-11	strategy, opportunity, extern	capable	1	0.454545	1.833333
C4-12	opportunity, economic, competitor	development	1	0.454545	1.833333
C4-13	network, economic, competitor	development	1	0.454545	1.833333
C4-14	knowledge, economic, competitor	development	1	0.454545	1.833333
C4-15	firm, environ, economic	development	1	0.454545	1.833333
C4-16	opportunity, network, economic, competitor	development	1	0.454545	1.833333
C4-17	environment	knowledge	0.833333	0.454545	1.527778
C4-18	customer	knowledge	0.833333	0.454545	1.527778
C4-19	competitor	knowledge	0.833333	0.454545	1.527778
C4-20	information	competitor	0.833333	0.454545	1.527778
Cluster 5					
C4-1	network, efficiency, distribution, data	customer	1	0.4	2.5
C4-2	Strategy	way, sustain, product	1	0.4	2.5
C4-3	strategy, marketing, external	industry	1	0.4	2.5
C4-4	systematic, strategy, external	world, expert	1	0.4	2.5
C4-5	vision, plan, future, experience	share	1	0.4	2.5
C4-6	systematic, strategy, extern	expert, development	1	0.4	2.5
C4-7	strategy, market, force, external	industry	1	0.4	2.5
C4-8	systematic, strategy, external, development	world, expert	1	0.4	2.5
C4-9	vision, strategy, plan, external	sustain, practitioner	1	0.4	2.5
C4-10	strategy, logic, extern, employee	product, culture	1	0.4	2.5
C4-11	Skill	world	1	0.4	2
C4-12	Relationship	experience	1	0.4	2
C4-13	technology, marketing	global	1	0.4	2
C4-14	strategy, logic, external, employee, culture	product	1	0.4	2
C4-15	people, skill, world	develop	1	0.4	2
C4-16	world, strategy, external	expert	1	0.4	2
C4-17	vision, strategy, practitioner, plan, external	sustain	1	0.4	2
C4-18	way, vision, transfer, plan, employee	culture	1	0.4	2
C4-19	network, efficiency, data, custom	distribution	1	0.4	2
C4-20	practitioner, information	expert	1	0.4	2

In the first cluster, the notable point is that on one hand, the competitive advantage is achieved through development and knowledge so that the above-mentioned keywords have been obtained in many subsequent rules, which can show us appropriate prerequisites in this regard. For example, the rule C1-5 has assumed that development and acquisition of the competitive advantage are achieved through taking into consideration the strategy, marketing, data, and specialization; thus the support number of this rule (that is 0.22) indicates that the probability of occurrence and relationship is antecedent and consequent, and the number 1 in the confidence index implies the high quality of this rule and its 100% repetition in all the articles in this cluster. On the other hand, in this cluster, the methods of information acquisition (C1-16 and C1-17) as well as its applications (C1-15 and C1-18) have been discovered. In fact, two methods of information

acquisition, and analysis as well as paying attention to the surrounding environment have been considered in this cluster and application of the information obtained in the production and distribution of the products have been shown.

In the second cluster, the methods of knowledge acquisition, knowledge application, necessity of intelligence in monitoring, and economization method have been shown. Analysis of the rules in this cluster shows that acquiring the knowledge requires awareness, information, considering the demands, monitoring, and considering the surrounding environment since the acquired knowledge can lead to economization. For instance, the C2-12 rule shows the pre-requisiteness of information for knowledge as well as the difference between knowledge and information so this rule with support of 0.375 indicates the appropriate relationship of these two variables; on the other hand, the value of 0.75 for

confidence of this rule indicates the repetition of this rule in the articles and the importance of this subject. In this cluster, it is worth taking into consideration the C2-6 rule since it shows us two applications of import, namely acquisition of information and awareness about the competitors. In other words, according to this rule, import is also the origin of CI for successful companies.

In the third cluster, the emphasis has been put on the importance of the staff and human capital. Education and culture, which can lead to the intellectual capital; furthermore, the necessity of paying attention to the primary and secondary data for acquisition of knowledge has been taken into account. For example, the rules C3-6 and C3-8 show that the staff, human capital, education, and culture are considered as the appropriate and essential tools for obtaining the intellectual capital, which can result in favorable outcomes such as sustainability and durability in the market and industrialization (C3-3 rule). In the C3-9 rule, the results of using the primary and secondary data have been shown, which should be particularly considered in acquisition of the CI. Further analysis of the rules of this cluster reveals the special importance of the staff in acquisition of the competitive advantage and the intellectual capital (rules C3-11 and C3-12).

In the fourth cluster, the methods of communicative development, development, and foresight have been presented. In this cluster, the emphasis is put on the social intelligence and communicational networks for expanding the relationships, and the customer, environment, and competitor have been referred to as the resources and prerequisites, and their relevant rules have been presented. In the field of business, success and superiority belong to the pioneer and futuristic companies, and there are various ways to achieve this goal. In the C4-4 rule, some of the prerequisites for reaching a successful future have been discovered including opportunity seeking, network relations, paying attention to the competitors, and developed and dynamic economy. This rule with support of 0.45, confidence of 1 and lift of 2.2 can provide the appropriate factors for the futuristic companies.

In the fifth cluster, some of the prerequisites of industrialization, export, sustainability in market, globalization, culture, and production have been presented. For instance, the C4-7 rule shows that industrialization requires strategy, marketing, force, and paying attention to the environment. Further, the C4-10 rule reminds us that our products and culture reflect our strategy, laws, environment, and staff; besides, the rules C4-12

and C4-13 show the methods of gaining experience and globalization, respectively. It should be noted that the value for the confidence index in this cluster is 1 for all the rules, indicating their 100% repetition in all the articles.

5. Discussion and conclusion

Changes in the structure of global economic system resulting from application of the modern technologies, response to the changes in the global demand patterns, strict competition at the global level, etc. have caused many alterations in the competition patterns and factors affecting the competitiveness among the countries around the world. The prerequisite for success and excellence in the current highly competitive market is being equipped with CI, which has been specially taken into consideration in the current era. In the present research work, the patterns and rules hidden in the research articles were investigated using the text mining techniques. The present work offered a precise review and analysis in the domain of CI using text-mining techniques (clustering and association rules), neglected by previous works. The major findings of the current work can be categorized into descriptive and inferential statistics. In the former, a good visual representation was offered to the readers and researchers in the domain of CI using the clustering technique and tag clouds. In the latter, the most important finding of the work led to association rules, which can be valuable and practical for researchers in the domain of CI. The results of the present research work are notable in four categories including descriptive statistics, clustering, word cloud, and association rules, which will be discussed now.

Analysis of the articles on CI in terms of descriptive statistics indicates the ascending slope and more interest in this field in the academic studies so that the use of it in the companies and various production-service organizations such as medicine [46, 47], stores [48], hotels, restaurants [49], e-commerce, schools and universities [50], shipping companies [47], multi-national corporations [51], libraries [52], small and medium enterprises [53, 54], knowledge-based companies, etc. has been investigated; on the other hand, it shows the gap between the countries in terms of production of knowledge in the field of CI so that the gap between the developed and developing countries is tangibly obvious.

Clustering and drawing the word cloud in the present research work led to discovering significant results. In the field of CI, the keywords (data, information, knowledge, strategy,

marketing, technology, communications, network and knowledge sharing, competitors, product, customer) form its foundation and basis, without which CI would not be achieved; thus they should be regarded as a systematic view consistent with the organizational goals and strategies. Analysis of the words resulted in each cluster is also notable. The emphasis is put in the first cluster on the internal and external environments; in the second cluster, on data path, information, and knowledge; in the third cluster, on knowledge, information, network, and knowledge sharing; in the fourth cluster, on technology and competitive advantage; and in the fifth cluster, on competitors. Drawing the word cloud in each cluster showed that the word “information” should be specially taken into consideration as the heart of CI; furthermore, the words like marketing, strategy, and knowledge are of especial importance in the formation and effectiveness of CI. The use of the association rules in this research work led to two important results. First, it led to the discovery of the prerequisites of the CI dimensions such as knowledge, technology, products, distribution, etc., and showed us the abilities required for achieving these capabilities as well as the field on which the investment should be focused; second, putting together these dimensions would lead to more significant results, which can be referred to as the consequence of CI. Some of these results include development, acquisition of competitive advantage, sustainability, economization, future, and production at global level. In other words, the first result shows the business activists the path to achieve CI, and the second result draws the outcome of achieving CI.

Reference

- [1] Strauss, A. C. and du Toit, A. S. A. (2010). Competitive intelligence skills needed to enhance South Africa’s competitiveness, *Aslib Proceedings: New Information Perspectives*, vol. 62, no. 3, pp. 302-320.
- [2] Havenga, J. and Botha, D. (2003). Developing competitive intelligence in the knowledge-based organization, available at: www.saoug.org.za/archive/2003/0312a.pdf (accessed 7 February 2008).
- [3] Bose, R. (2008). Competitive intelligence process and tools for intelligence analysis. *Industrial Management and Data Systems*, vol. 108, no. 4, PP. 510-528.
- [4] Kahaner, L. (1996), *Competitive Intelligence: From Black Ops to Boardrooms – How Businesses Gather, Analyse, and Use Information to Succeed in the Global Marketplace*, Simon and Schuster, New York, NY.
- [5] Thompson, S. H., Choo, W. Y. (2001). Assessing the Impact of Using the Internet for Competitive Intelligence, *Information & Management*, Vol. 30 .pp. 67-83.
- [6] Combs, R. E., Moorhead, J. D. (1992). *The Competitive Intelligence Handbook*, Metuchen, MJ: Scarecrow Press.
- [7] Wang, J., Hu, X. and Zu, D. (2007). Diminishing downsides of data mining, *International Journal of Business Intelligence and Data Mining*, vol. 2, no. 2, pp. 177-96.
- [8] Pechenizkiy, M., Puuronen, S. and Tsymbal, A. (2005). Why data mining research does not contribute to business?, in Soares, C. et al. (Eds), *Proc. of Data Mining for Business Workshop DMBiz (ECML/PKDD’05)*, Porto, Portugal, pp. 67-71.
- [9] Seo, W., Yoon, J., Parkc, H., Coh, B. Y., Jae-Min Lee, J. M., and Kwond, O. J. (2016). Product opportunity identification based on internal capabilities using text mining and association rule mining, *Technological Forecasting & Social Change*, vol. 105, pp. 94–104.
- [10] Yoon, B., Park, I., Coh, B.-Y. (2014). Exploring technological opportunities by linking technology and products: application of morphology analysis and text mining. *Technol. Forecast. Soc. Chang.*, vol. 86, pp. 287–303.
- [11] Juhari, M and Stephens, K. (2006). Tracing the Origins of Competitive Intelligence throughout History, *Journal of Competitive Intelligence and Management*, vol. 3, no. 4, pp. 61-82.
- [12] Saayman, A. Pienaar, J. Pelsmacker, P. Viviers, W. Cuyvers, L. Muller, M. and Jegers, M. (2008). Competitive intelligence: construct exploration, validation and equivalence, *Adlib Proceedings: New Information Perspectives*, vol. 60, no. 4. pp. 383-411.
- [13] UNCTAD (2002). *Trade and Development Report*, pp. 87-89.
- [14] Fourie, L. H. (1999). World-Wide Web as an Instrument for Competitive Intelligence in a Tertiary Educational Environment, *South African Journal of Information Management*, vol. 1, no. 2, pp. 137-151.
- [15] Fuld, L. (1985). *Competitor Intelligence: How to Get it? How to Use it?*, Wiley, New York.
- [16] Rouach, D., and Santi, P. (2001). Competitive intelligence adds value, *European management journal*, vol. 19, no. 5, pp. 200-224.
- [17] Alon, I. & Higgins, M. (2005). Global Leadership Success Trough Emotional & Cultural intelligence, *Bussiness Horizons*, vol. 48, pp. 501-512.
- [18] Des champs, J. P. & Nayak, P. R. (1995). *Product Juggernauts: How Companies Mobilize to Generate A Stream of Market Winner*. Harvard Business Press.

- [19] Adidam, P. T., Gajre, S. & Kejriwal, S. (2009). Cross-cultural competitive Intelligence Strategies, *Marketing Intelligence & Planning*, vol. 27 no. 5, pp. 666-680.
- [20] Johns, P., Van Doren, D. C. (2010). Competitive intelligence in service marketing A new approach with practical application, *Marketing Intelligence & Planning*, vol. 28, no. 5, pp. 551-570.
- [21] Johnson, G., Scholes, K. & Whittington, R. (2008). *Exploring Corporate Strategy: Text & Cases*, FT Prentice Hall Financial Times, Harlow, U.K.
- [22] Sauter, V. L. & Free, D. (2005). Competitive intelligence systems: qualitative DSS for strategic decision making, *The DATA BASE for Advances in Information Systems*, vol. 36, no. 2, pp. 43-57.
- [23] Brachman, R. J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G. & Simoudis, E. (1996). Mining business databases, *Communications of the ACM*, vol. 39, no. 11, pp. 42-48.
- [24] Oliveira, J. P. M. D., Loh, S. & Wives, L. K. (2004). Applying text mining on electronic messages for competitive intelligence, *Proceedings of the 5th International Conference on Electronic Commerce and Web Technologies*, Zaragoza, Spain.
- [25] D. Sullivan (2004). The need for text mining in business intelligence, *Information Management Special Reports*, available at: <http://www.informationmanagement.com/specialreports/20040210/8100-1.html>.
- [26] Dai, Y., Kakkonen, T. & Sutinen, E. (2010). Min EDec: a decision support model that combines text mining with competitive intelligence, 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), pp. 211-216.
- [27] Karanikas, H. & Theodoulidis, B. (2002), Knowledge discovery in text and text mining software”, technical report, Centre for Research in Information Management (CRIM), Department of Computation, UMIST, Manchester. key issues and future trends, *Marketing Intelligence & Planning*, vol. 19, no. 4, pp. 245-53.
- [28] Liau, B., Y. & Tan, P., P. (2014). Gaining customer knowledge in low cost airlines through text mining, *Industrial Management & Data Systems*, vol. 114, no. 9, pp. 1344 – 1359.
- [29] Hand, D., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*, MIT Press, Cambridge, MA.
- [30] Bacher, J., Pöge, A. & Wenzig, K. (2010). *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*, 3. erg., vollst. überarb. und neu gestaltete Aufl, Oldenbourg, Munich.
- [31] Izenman, A. (2008). *Modern Multivariate Statistical Techniques: Regression Classification and Manifold Learning*, Springer, New York, NY.
- [32] Lashka, M. & Moattar, M. H. (2017). Improved COA with Chaotic Initialization and Intelligent Migration for Data Clustering, *Journal of AI and Data Mining*, vol. 5, no. 2, pp. 293-305.
- [33] Vorhies, D. W. & Morgan, N. A. (2005). Benchmarking marketing capabilities for sustainable competitive advantage, *Journal of Marketing*, vol. 69 no. 1, pp. 80-94.
- [34] Tan Tsu Wee, T. (2001). The use of marketing research and intelligence in strategic planning: key issues and future trends, *Marketing Intelligence & Planning*, vol. 19, no. 4, pp.245 – 253.
- [35] Simkin, L. & Cheng, A. (1997). Understanding competitive strategies: the practitioners-academic gap, *Marketing Intelligence & Planning*, vol. 15, no. 3, pp. 124-34.
- [36] Davenport, T. H., & Prusak, L. (2003). *What’s the Big Idea?* Boston, Harvard Business School Press.
- [37] Benczúr, D. (2006). Competitive Intelligence and IT, towards a Knowledge-based approach, 37-44, 3-7 April 2006, 1-4244-0176-3/06/\$20.00 (C) 2006 IEEE.
- [38] Pellissier, R. & Kruger, J. P. (2011). Understanding the use of strategic intelligence as a strategic management tool in the long-term insurance industry in South Africa. *South African Journal of Information Management*, vol. 13, no. 1, pp. 1-13.
- [39] Sewdass, N. & Toit, A., D. (2014). Current state of competitive intelligence in South Africa, *International Journal of Information Management*, vol. 34, pp. 185-190.
- [40] Han, J. & Kamber, M. (2001). *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers Academic Press, San Francisco.
- [41] Forcht, K. A. & Cochran, K. (1999). Using data mining and data warehousing techniques, *Industrial Management & Data Systems*, vol. 99, no. 5, pp. 189-196.
- [42] Çiflikli, C. & Kahya-Özyirmidokuz, E. (2012). Enhancing product quality of a process, *Industrial Management & Data Systems*, vol. 112, no. 8, pp. 1181-1200.
- [43] Kim, C., Lee, H., Seol, H. & Lee, C. (2011). Identifying core technologies based on technological cross-impacts: an association rule mining (arm) and analytic network process (anp) approach. *Expert Syst. Appl.* Vol. 38, no. 10, pp. 12559-12564.
- [44] Shih, M.-J., Liu, D.-R. & Hsu, M.-L. (2010). Discovering competitive intelligence by mining changes in patent trends. *Expert Syst.* vol. 37, no. 4, pp. 2882-2890.
- [45] Kim, C., Lee, H., Seol, H. & Lee, c. (2011). identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach, *Expert*

Systems with Applications, vol. 38, no, 10, pp. 12559-12564.

[46] Canongia, C. (2007). Synergy between Competitive Intelligence (CI), Knowledge Management (KM) and Technological Foresight (TF) as a strategic model of prospecting - The use of biotechnology in the development of drugs against breast cancer, *Biotechnology Advances*, vol. 25, pp. 57–74.

[47] Tuan, L. T. (2013). Knowledge sharing and competitive intelligence, *Marketing Intelligence & Planning*, vol. 32, no. 3, pp. 269-292.

[48] Rapp, A., Agnihotri, R. & Baker. T. L. (2011). Conceptualizing Salesperson Competitive Intelligence: An Individual-Level Perspective, *Journal of Personal Selling & Sales Management*, vol. 31, no. 2, pp. 141-155.

[49] Köseoglua, A., Rossc, G. & Okumus, F. (2016). Competitive intelligence practices in hotels Mehmet, *International Journal of Hospitality Management*, vol. 53, pp. 161–172.

[50] Garcia-Alsina, M., Cobarsí-Morales, J. & Ortoll, E. (2016). Competitive intelligence theoretical framework and practices, *Aslib Journal of Information Management*, vol. 68, no. 1, pp. 57 - 75.

[51] Tuan, L. T. (2015). Entrepreneurial orientation and competitive intelligence: cultural intelligence as a moderator, *Journal of Research in Marketing and Entrepreneurship*, vol. 17, no. 2, pp. 212–228.

[52] Nitse, P. S. & Parker, K. R. (2002). Library Science, Knowledge Management, Competitive Intelligence, *The Reference Librarian*, vol. 38, no.79-80, pp. 395-407.

[53] Wright, S., Bisson, C. & Duffy, A. P. (2012). Applying a behavioral and operational diagnostic typology of competitive intelligence practice: empirical evidence from the SME sector in Turkey, *Journal of Strategic Marketing*, vol. 20, no. 1, pp. 19-33.

[54] Smith, J., R., Wright, S. & Pickton, D. (2010). Competitive Intelligence programmes for SMEs in France: evidence of changing attitudes, *Journal of Strategic Marketing*, vol. 18, no. 7, pp. 523-536.

متن کاوی هوشمندی رقابتی: کلمات سخن می گویند

عظیم زارعی*، مرتضی ملکی، داود فیض و محمد علی سیاه سرانی کجوری

دانشکده اقتصاد، مدیریت و علوم اداری، دانشگاه سمنان، سمنان، ایران.

ارسال ۲۰۱۶/۱۲/۲۰؛ بازنگری ۲۰۱۷/۰۲/۲۲؛ پذیرش ۲۰۱۷/۰۵/۱۰

چکیده:

هوشمندی رقابتی به یکی از مباحث اصلی پژوهشگران در سال‌های اخیر تبدیل شده و کاربرد آن در شرکت‌ها و سازمانهای تولیدی-خدماتی مورد بررسی قرار گرفته است. هدف پژوهش حاضر رسیدن به بخشی از هوشمندی رقابتی از طریق بررسی مقالات علمی در این حوزه از طریق متن‌کاوی است که این امر در سه گام مرتبط باهم صورت پذیرفت. در گام اول با استفاده از جستجوی عبارت "هوشمندی رقابتی" در پایگاه‌های داده و موتورهای جستجو معتبر در مجموع ۱۱۴۳ مقاله بین سالهای ۱۹۸۷ تا ۲۰۱۶ انتخاب شدند. محققین از طریق مرور عنوان، چکیده و متن اصلی مقالات و در مراحل مختلف غربال مقالات، در نهایت ۱۳۵ مقاله مرتبط با موضوع هوشمندی رقابتی را به منظور انجام فرآیند متن‌کاوی انتخاب کردند. در گام دوم پیش پردازش داده‌ها صورت پذیرفت، در گام سوم با استفاده از تحلیل خوشه‌ای غیر سلسله مراتبی (k میانگین)، تعداد ۵ خوشه بهینه بر حسب شاخص دیویس بولدین بدست آمد و برای هر خوشه ابرواژه ترسیم شد، سپس با استفاده از شاخص‌های پشتیبان، اطمینان و جذابیت قواعد انجمنی هر خوشه استخراج و مورد تحلیل قرار گرفت. نتایج نشان داد از یک سو علاقه به پژوهش در زمینه هوشمندی رقابتی در سالهای اخیر افزایش یافته و از سوی دیگر حضور پررنگ و کم‌رنگ کشورهای توسعه یافته و در حال توسعه در شکل‌گیری تولیدات علمی محسوس است، همچنین اطلاعات، بازاریابی و استراتژی از عناصر اصلی هوشمندی رقابتی بوده که می‌توانند در کنار سایر پیش‌نیازها، منجر به هوشمندی رقابتی و متعاقب آن توسعه اقتصادی، مزیت رقابتی و ماندگاری در بازار شوند.

کلمات کلیدی: هوشمندی رقابتی، متن کاوی، قواعد وابستگی و ابر واژه.