

---

Type of the Paper (Article, Review, Communication, etc.)

# Analysis of The Development Trend of Sports Research in China and Taiwan Using Natural Language Processing

Wei-Yuan, Shih <sup>1</sup>, Tu-Kuang, Ho <sup>2,\*</sup>

<sup>1</sup> Centre for Teacher Education and Career Development, University of Taipei City, Taiwan; weiyuan@go.utaipei.edu.tw

<sup>2</sup> Institute of Physical Education and Health, Yulin Normal University Yulin City, China; 30255670202@qq.com

\* Correspondence: 30255670202@qq.com; Tel: +886-9-811-38657

**Abstract:** A digital text abstract presents the essential information of an article, and we can find the trend and value of the research by analyzing it rigorously and digging up knowledge. Therefore, this study focuses on the abstracts of index journals in China and Taiwan from July 2010 to June 2020 (a total of 3,283 abstracts). Through the concepts of text mining and natural language processing (NLP), it constructs processes such as text retrieval, text segmentation and word cloud analysis, TF-IDF weight analysis, co-word analysis, network analysis, and trend analysis, and analyses a large amount of text data. The results show that the scope of research in China covers the fields of social sports and sports science, and research in Taiwan covers both natural and social sciences. The network diagram highlights the richness of sports-related research fields in the two regions, but research on sports philosophy is relatively rare. It is suggested that all disciplines/departments should re-allocate the same resources, so as to show a balanced development trend and help expand a new chapter in the sports academic field.

**Keywords:** word segmentation; word cloud analysis; TF-IDF weight analysis; co-word analysis; network analysis

---

## 0. How to Use This Template

The template details the sections that can be used in a manuscript. Note that each section has a corresponding style, which can be found in the “Styles” menu of Word. Sections that are not mandatory are listed as such. The section titles given are for articles. Review papers and other article types have a more flexible structure.

Remove this paragraph and start section numbering with 1. For any questions, please contact the editorial office of the journal or support@mdpi.com.

## 1. Introduction

China and Taiwan share the same origins in terms of their history, culture, language, and text, coupled with the development of network information and the prevalence of digital collections, so that sports academic exchanges between the two regions benefit from the unprecedented application of information technology (IT), and have changed from the closed mode of the past to an open mode to create a good and mutually beneficial academic developmental environment [1,2]. However, in the process of open academic exchange, it is often necessary to face limitations in processing big data, which promotes data mining technology and further extends the application of text mining [3-7]. Because the process of text mining is highly dependent on computation and model application, natural language processing (NLP) technology has cross-disciplinary application value, and sports-related academic research is no exception [8-11]. This also means that the content analysis method that traditional sports research institutes rely on must be improved accordingly in order to raise data processing efficiency [4,8,12]. In order to

---

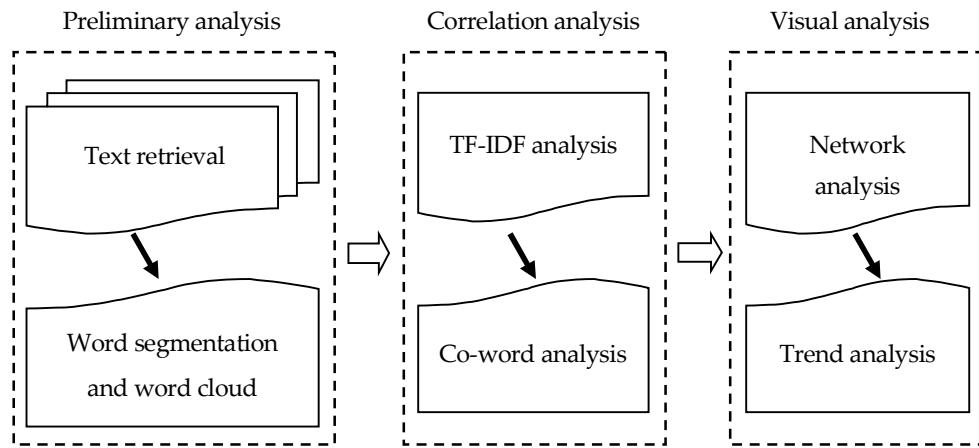
make up for the shortcomings of the content analysis method, such as being time-consuming, laborious, and not objective, this study proposes a program for analyzing and processing text data in sports-related fields based on the concept of NLP, which is different from the traditional content analysis method, with a faster automatic analytical process and the ability to systematically extract and analyze the importance and relevance of keywords in abstract content [5,13]. It integrates the methods of word segmentation, word cloud analysis, TF-IDF (term frequency-inverse document frequency) analysis, and co-word analysis, and then applies visual network diagrammatic drawing technology [14-18] to present the distribution of core keywords and their network relationships in sports research abstracts from China and Taiwan (a total of 3,283 articles). This reveals possible hidden knowledge structures and provides an important reference for evaluating future development trends in the two regions.

The main function of an abstract is to help readers quickly understand the main contents of the article, and to screen out the necessary important knowledge from numerous articles in the academic literature [19,20]. At present, thanks to the rapidly developing new information technology, many scholars have opened their research results up to the world through online academic database pages (such as Google Scholar), in order to lay a good foundation for academic development [21-24]. Therefore, systematic analysis of research abstracts accumulated over the years will enhance the understanding of the development process, research characteristics and future development trends of an academic field [19]. However, the speed of data generation and storage on various network platforms far exceeds the speed that people can analyze and digest, which also allows data mining and text mining technology to play an extremely important role in exploring the application of big data analysis [5,6,12,13,25]. Generally, data mining is mostly applied to the processing of structured data, just like a table with a fixed structure, in which each column has its own clear definition and numerical value. Based on the needs of researchers, these structured data can be used for clustering analysis, classification analysis, association rule analysis, sequential pattern analysis, time-series similarity analysis, and link analysis, in order to dig out valuable implicit knowledge and information [4,12]. Text mining is much more complicated, because its original input data are unstructured information in the form of text, unlike structured data, which have their specific structure to follow; however, in these seemingly messy text data, there is also a lot of information worthy of in-depth investigation, which needs to be explored by relevant researchers [26]. Therefore, NLP applies the language rules commonly used by people to carry out disassembly and reorganization operations for the words in the text content, i.e., it can grasp the importance and relevance of keywords, and thus can roughly understand the semantic content and become an important basis for exploring new knowledge [5,18].

One of the main media of network information transmission is text [8,17,27]. From the perspective of management, most studies have used NLP to analyze the structured knowledge and language relations related to management, establish the vocabulary system of this knowledge, and further construct the decision support system or retrieval system for management [5,13,28]. However, it is rare to apply NLP technology in the sports academic field to explore its knowledge structure and future development trends. In addition, due to various political factors, sports exchanges between China and Taiwan only focus on sports competitions and visits, while exchanges involving academic fields are few, which has become a problem worthy of attention from governments and academic circles. In view of this, with the rapid increase in the demand for electronic information interpretation, it is urgent for the two regions to expand their relevant academic fields through faster and more objective automatic data analysis to promote knowledge exchanges between them. Through the rigorous NLP process, this study presents the current situations and differential results of sports academics in the two regions, which will help to create new opportunities and new directions for research and physical education development in an online environment.

## 2. Methods

The process of this study was based on the concepts of text mining and NLP. For more, please refer to [5-7,14-16,18,27-30]. In the SCI (science citation index) and SSCI (social science citation index) articles of [31], through the steps of text retrieval, word segmentation, word cloud analysis, TF-IDF analysis, co-word analysis, network analysis, trend analysis, etc., the text data of the two regions are excavated for comparison and analysis of their similarities and differences.



**Figure 1.** Research flow chart

### 2.1. Text retrieval

In this study, the CSSCI and TSSCI comprehensive academic journals ranked in the top two for impact factor (IF) were selected. The text retrieved in China included that from China Sport Science (IF=0.326), which publishes research papers and reviews reflecting the latest achievements in sports research, and the Journal of Shanghai University of Sport (IF=0.221), which is a comprehensive journal of sports academic theory whose main task is to present the latest high-level research results in the field of sports science. The text selected from Taiwan was that from the Physical Education Journal (IF=0.663), which aims at publishing original research papers in the field of sports and exercises, and producing high-quality academic works, and from Sports & Exercises Research (IF=0.420), which contains excellent original articles on sports academics in Taiwan, hoping to contribute to the existing knowledge of mankind. Important messages, including titles and keywords, can be retrieved from the short abstract text to give full play to their value and utility [6,7,29]. Referring to the 10-year research interval of [16] and [31], this study retrieved the text abstracts of the two regions from July 2010 to June 2020, including 1,420 abstracts in China Sport Science and 1,142 abstracts in the Journal of Shanghai University of Sport, with a subtotal of 2,562 abstracts in China, 345 abstracts in the Physical Education Journal and 376 in Sports & Exercises Research. There was a subtotal of 721 abstracts in Taiwan, totaling 3,283 abstracts retrieved.

### 2.2. Word segmentation and word cloud analysis

Words, phrases, and sentences are the basic units in the process of language recognition, and their meanings will affect the understanding of the text. Accurate word segmentation is very important [27,30]. In this study, the CKIP Chinese word segmentation system [17,28] was used to disassemble the words and mark the parts of speech of each vocabulary automatically. Then, meaningful words were retrieved, such as nouns, verbs, and adjectives [6], while punctuation marks or meaningless words, prepositions, connectives, articles, etc., were not explored. Finally, through the procedure of estimating word frequency, the frequency of word occurrence could be learned for subsequent analysis. At the same time, word frequency patterns were compared by using a word cloud, where the more times a word appears, the greater the frequency of that word [15].

### 2.3. TF-IDF analysis

TF-IDF is a classical keyword retrieval method, which is simple and effective [6]. It can convert words into statistical data and calculate their frequency. TF indicates the number of times words appear in the text, while IDF can show the importance of words in all texts, and at the same time increase common words and expand rare words [5,18]. Therefore, if key words appear in more different texts and have a lower frequency in each text, they can obtain more weight. Refer to the explanations of TF-IDF conversion by [32] and [18]. In the general file (F) of all samples, taking the words (wi) of all words (Dterms) in a file (f1) as an example, the formula is as follows:

$$\text{TF-IDF } wi = \frac{wi \text{ occurrence times}}{\text{Total number of Dterms}} \times \left( \log \frac{\text{Total number of Fs}}{\text{Number of files with } wi} + 1 \right)$$

### 2.4. Co-word analysis

A single word cannot provide an understanding of the focus of a study, but the semantic features of words can be captured through the co-occurrence frequency of words to obtain the answer [6,14,33]. The most suitable index for evaluating the frequency of co-occurrence words is the equivalence index, which is also called association strength [34]. In other words, co-word analysis can identify the linkages between words, and deduce the research topic according to the number of co-occurrence phrases. See [14] and [35] for an explanation of co-word calculation. Taking the word (wi) and the word (wj) among all words (Dterms) of a sentence (s1) as examples, and the formula is as follows:

$$co(wi, wj) = \sum_{s1 \in D} (wi)_{s1} (wj)_{s1}$$

### 2.5. Network analysis

Visualization of network diagrams can enhance the understanding of the connection and cluster relationships between words [16], just like the concept of “look for a good steed according to a picture.” One can then explore the text knowledge. When exploring the group structure of various types of relationship, network analysis technology can build a radial social network, in which key words can be regarded as nodes in the network, and information such as network density, degree centrality, and K-Core can be obtained through the combination of strong and weak links [5,36]. Density is used to judge the connection status of nodes, and a high-density network also means that the nodes have stronger cohesion, while degree centrality and K-Core can provide an understanding of the distribution or concentration of the overall network diagram [36,37]. This study made good use of the NetDraw function of the UCINET software [16] to determine the aforementioned data and network diagrams, and to provide important analysis information.

## 3. Results

### 3.1. Preliminary analysis

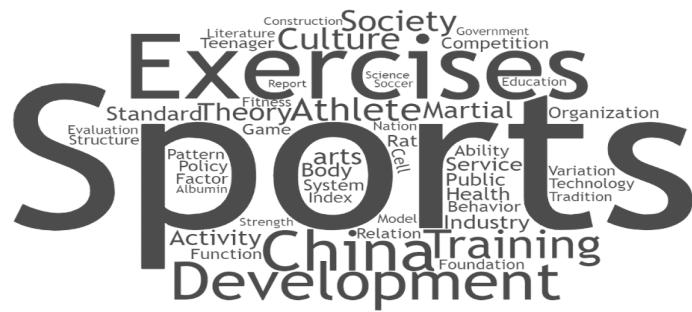
For China, 2,562 abstracts were retrieved from China Sport Science and the Journal of Shanghai University of Sport (853,357 words). For Taiwan, 721 abstracts were retrieved from the Physical Education Journal and Sports & Exercises Research (351,368 words). Then, the CKIP Chinese word segmentation system [17,28] was used for word segmentation to sort out meaningful words [6]. Meaningless characters and words were not analyzed, such as “and,” “also,” “as well as,” “with,” “because of,” etc. For China, 9,518 meaningful words were screened out. For Taiwan, 4,476 meaningful words were screened out. Then, the frequency of words was analyzed. The ranking results of each word are shown in Table 1. For China, the word ranking first was “體育” (sports) (10,388 times), followed by “運動” (exercises) (4,700 times), “中國” (China) (2,907 times), and “發展” (development) (2,736 times). For Taiwan, the word ranking first was “運動”

(exercises) (2,792 times), followed by “訓練” (training) (853 times), “動作” (action) (737 times), and “身體” (body) (673 times).

**Table 1.** Statistical analysis table of word frequency

Region	Ranking	Word (frequency)	Ranking	Word frequency	Ranking	Word frequency
China	1	sports (10,388)	6	athlete (1,638)	11	activity (1,107)
	2	exercises (4,700)	7	culture (1,533)	12	standard (996)
	3	China (2,907)	8	society (1,488)	13	industry (988)
	4	development (2,736)	9	theory (1,265)	14	services (983)
	5	training (2,074)	10	martial arts (1,191)	14	body (983)
Taiwan	1	exercises (2,792)	6	player (639)	11	behavior (394)
	2	training (853)	7	athlete (517)	12	ability (381)
	3	action (737)	8	muscle (454)	13	model (372)
	4	body (673)	9	activity (412)	14	study (369)
	5	performance (658)	10	scale (398)	14	sports (369)

Due to the abundant data, this paper can only present the statistical analysis of the top 15 words in each region, but with the thinking of looking at pictures to catch the meanings, the frequency analysis results of the top 50 words in each region are presented with the help of word cloud, as shown in Figures 2 and 3. The font size of a word in a word cloud represents how often it appears [15]. This result can provide a preliminary understanding of the research words with the highest occurrence rate in the two regions and their general situations. It was also found that besides the different word rankings, some words were no repeated, such as “文化” (culture), “武術” (martial arts), “水準” (standard), “肌肉” (muscle), “量表” (scale), and “學習” (study).



**Figure 2.** Word cloud of word frequency in China



**Figure 3.** Word cloud of word frequency in Taiwan

### 3.2. Correlation analysis

This stage focuses on the TF-IDF analysis and co-word analysis, and identifies the key words from the relative relations of the words for network analysis.

### 3.3. TF-IDF analysis

According to the TF-IDF calculation method proposed by [32] and [18], the word frequency statistics are imported to calculate the weight of key words. Once the TF-IDF moderates, the difference will serve as the basis for analysis and judgment [13]. As shown in Table 2, for China, the word ranking at no. 1 in terms of TF-IDF is “體育” (sports) (0.0044), followed by “運動” (exercises) (0.0041), “訓練” (training) (0.0039), and “武術” (martial arts) (0.0037), and no. 27 is “體系” (system) (0.0021). When the exploration continued, the word ranking no. 28 was “因素” (factor) (0.0019). After TF-IDF decreased sharply by 0.0002, the decreasing trend moderated, so only the words with rankings up to 27 were taken for China. For Taiwan, the word ranking no. 1 in terms of TF-IDF was “訓練” (training) (0.0032), followed by “動作” (action) (0.0029), “身體” (body) (0.0027), and “肌肉” (muscle) (0.0025), and no. 26 was “健康” (health) (0.0014). The word ranking at 27 was “關係” (relationship) (0.0012). Similarly, after TF-IDF decreased sharply by 0.0002, the decreasing trend moderated, so only the words with rankings up to 26 were taken for Taiwan. Based on these results, the co-occurrence relationships of key words in the two regions were explored.

**Table 2.** TF-IDF analysis table

**Region**   **Ranking**   **Word**   **TF-IDF**   **Ranking**   **Word**   **TF-IDF**   **Ranking**   **Word**   **TF-IDF**

1	sports	0.0044	10	Rat	0.0032	19	health	0.0024	
2	exercises	0.0041	11	public	0.0030	20	theory	0.0024	
3	training	0.0039	12	service	0.0030	21	behavior	0.0023	
4	martial arts	0.0037	13	body	0.0030	22	society	0.0023	
China	5	athlete	0.0036	14	game	0.0029	23	organization	0.0022
	6	culture	0.0036	15	activity	0.0028	24	standard	0.0021
	7	China	0.0035	16	eenager	0.0027	25	ability	0.0021
	8	development	0.0034	17	policy	0.0026	26	function	0.0021
	9	industry	0.0033	18	competition	0.0025	27	system	0.0021
	1	training	0.0032	10	behavior	0.0021	19	coach	0.0017
	2	action	0.0029	11	scale	0.0021	20	model	0.0017
	3	body	0.0027	12	performance	0.0020	21	factor	0.0016
	4	muscle	0.0025	13	self	0.0020	22	index	0.0016
Taiwan	5	player	0.0025	14	sports	0.0020	23	Taiwan	0.0016
	6	athlete	0.0025	15	society	0.0019	24	development	0.0015
	7	study	0.0025	16	student	0.0018	25	effect	0.0014
	8	sports	0.0024	17	ability	0.0018	26	health	0.0014
	9	joint	0.0022	18	activity	0.0017	27	-	-

### 3.4. Co-word analysis

The frequency of co-words facilitates the understanding of an article's focus [14]. After analysis using the calculation methods proposed by [14], as well as [35], as shown in Table 3, for China, the co-words ranking first in the number of co-occurrences were “體育” (sports) – “發展” (development) (782 times), followed by “體育” (sports) – “社會” (society) (716 times), “發展” (development) – “社會” (society) (439 times), “體育” (sports) – “理論” (theory) (427 times), and “體育” (sports) – “運動” (exercises) (419 times). For Taiwan, the co-words ranking first in the number of co-occurrences were “運動” (exercises) – “表現” (performance) (193 times), followed by “訓練” (training) – “運動” (exercises) (163 times), “選手” (player) – “運動” (151 times), “運動” (exercises) – “運動員” (athlete) (137 times), and “身體” (body) – “運動” (exercises) (132). The co-word matrices (as shown in Table 4) of 27\*27 and 26\*26, respectively, for the two regions were produced for network analysis. Due to space limitations, only the first 20 pairs of key words in the ranking of co-occurrence times are presented. Here, we can firstly find out the different research styles of the two regions. China focuses on issues concerning sports, development, society, theory, China, and culture. Taiwan focuses on exercises, performance, training, athletes, the body, and other aspects. However, the closeness, linkages, cluster relationship, and overall structure of each key word still needed to be presented in the visual analysis stage.

**Table 3.** Statistics of co-occurrence times

Region	Ranking	Word	Word	Co-occurrence	Ranking	Word	Word	Co-occurrence
China	1	sports	development	782	11	exercises	standard	318
	2	sports	society	716	12	sports	system	302
	3	development	society	439	13	exercises	development	291
	4	sports	theory	427	14	development	theory	289
	5	sports	exercises	419	15	theory	society	245
	6	sports	China	415	16	culture	development	237
	7	exercises	athlete	394	17	China	society	233
	8	exercises	training	378	18	sports	ctivity	231
	9	China	development	327	19	sports	service	225
	10	sports	culture	325	19	sports	organization	225
Taiwan	1	exercises	performance	193	11	exercises	development	111
	2	training	exercises	163	12	exercises	ability	109
	3	player	xercises	151	13	exercises	factor	107

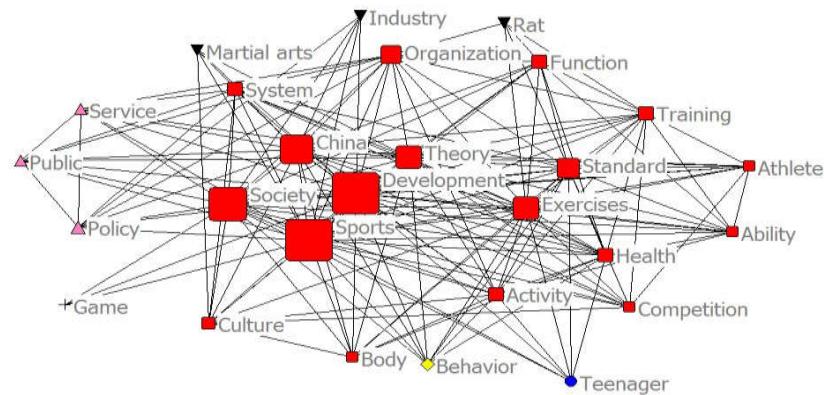
4	exercises	athlete	137	14	training	performance	98
5	body	sports	132	15	exercises	'index	97
6	exercises	effect	123	16	exercises	scale	86
7	action	exercises	121	16	exercises	Taiwan	86
8	exercises	activity	120	18	muscle	exercises	85
9	exercises	model	116	18	exercises	society	85
10	exercises	health	112	20	action	performance	82

**Table 4.** Co-word matrix table

China	A.	B.	C.	D.	E.	Taiwan	A.	B.	C.	D.	E.
A. sports	0	419	168	93	149	A. training	0	70	45	52	80
B. exercises	419	0	378	20	394	B. action	70	0	42	37	69
C. training	168	378	0	15	174	C. body	45	42	0	25	27
D. martial arts	93	20	15	0	1	D. muscle	52	37	25	0	23
E. athlete	149	394	174	1	0	E. player	80	69	27	23	0

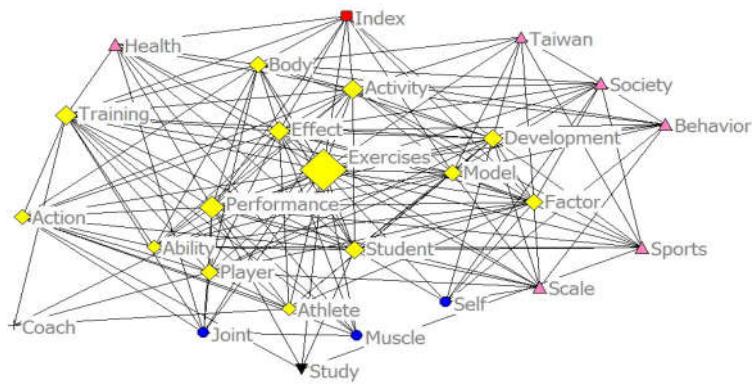
### 3.5. Visual analysis

Firstly, the co-word matrix table was converted to a UCINET-readable format [38], and then the network density, degree centrality, and K-Core were converted [5,36,37], thereby presenting nodes and clusters with different appearances. The bigger and more central the nodes are, the more important they are. The density of the network diagram for China was 0.501, which means that 50.1% of the possible links actually appeared. Figure 4 also shows that there are six clusters in the network diagram. Cluster 1 (square, K-Core 10) takes social sports development as the main core, has the most nodes, is in the center of the overall structure, and occupies the largest area, which can be said to be the trend of sports research in China. Cluster 2 (diamond, K-Core 9) only has behavior nodes, but it is located at the center point directly below the overall structure diagram, showing that various sports behavior issues have also received attention. Cluster 3 (regular triangle, K-Core 8) has three nodes, which are close to the left side of the network diagram, indicating that sports public policy and other research accounts make up a considerable proportion. Cluster 4 (round, K-Core 7) only has the node of teenager, suggesting a focus on research at the teenager level, e.g., teenager sports activities, teenager physicality, mental health, etc. Cluster 5 (inverted triangle, K-Core 5) is directly on the top of the network diagram, including research on rats and martial arts, which is less than the first four clusters. Finally, Cluster 6 (cross, K-Core 3), which takes sports games as an independent node, is located at the lower left of the network diagram, and focuses on topics related to sports games.



**Figure 4.** Network diagram for China

The density of the network diagram for Taiwan was 0.508, which means that 50.8% of possible links actually appeared. It can be seen from Figure 3 that there are also six clusters in the network diagram for Taiwan. Cluster 1 (diamond, K-Core 10) apparently takes exercises as the center point of the whole network diagram, with the largest node, and the node colors of natural science and social science can be seen from the left and right nodes, respectively; so, the focus of sports research in Taiwan is around the key words of exercises. Cluster 2 (square, K-Core 9) only has the index node, which is located at the center point directly on the top of the overall structural diagram. According to its research text, the research field of this node is inclined toward physiology-related numerical indicators, constituting a certain proportion of research papers. Cluster 3 (regular triangle, K-Core 8) mainly surrounds the right side of the network diagram, focusing on sociological issues, with Taiwan as the main research area. Cluster 4 (round, K-Core 7) focuses on related influencing factors on the muscles and joints as well as psychological aspects of student athletes and players. Cluster 5 (inverted triangle, K-Core 5) only has the node of learning, and the research that has attracted attention is about motor skills learning, learning performance, etc. Cluster 6 (cross, K-Core 4) is located at an edge corner of the network diagram, exploring coaches' knowledge and related training abilities. This cluster is not a main trend node. Although there are some identical nodes between the two regions, they are far from each other in the cluster sequencing and structure, which shows that due to the differences in people, local conditions and policies, each region has its own characteristics, which leads to completely different research patterns.



**Figure 5.** Network diagram for Taiwan

## 4. Discussion

Based on NLP technology, this study retrieved a total of 3,283 abstracts from target journals (2,562 in China; 721 in Taiwan); 53 key words (27 in China; 26 Taiwan) and their related matrices were then retrieved. Then, the NetDraw function of the UCINET software was used to convert the key words and their network relations into graphical

structures. According to the network relationship structure of each node in Cluster 1 (square) in Figure 4, most of the related studies in China in the last 10 years focused on sports development and actively expanded the 2 major fields of social sports and sports science in China. Social sports is not only the bridge to implementing national sports and competitive sports, but also the foundation for improving various social developments [39]. At present, China's development in this field is in a critical period of creating a well-off society in an all-round way, and its strategic goals cannot be separated from the government's support and guidance, including deepening the reform of the sports field, enhancing the core competitiveness of sports, and improving the health level of all people [40]. Therefore, according to the keyword network relationship structure and attributes of the left half of Cluster 1 (square) (organization, system, China, theory, development, society, sports, culture, body) and Cluster 3 (regular triangle), it was found that the topics discussed by the target journals in the development of social sports covered related research such as public service systems and sports policy reform. The results of this study highlight that against the macroscopic background of social transformation and the transformation of government functions in China, we should re-plan the social sports organization system and development policy in the new era while maintaining stability, and thus create a well-off social environment that can ensure the physical and mental health of all people [39,40].

On the other hand, according to the keyword network relationship structure and attributes of the right half of Cluster 1 (square) (function, training, level, athlete, sports, ability, health, activity, and competition) and Cluster 2 (diamond) and Cluster 4 (circle), it was found that sports science research actually occupied a considerable proportion, and the topics discussed were mainly the improvement of sports standards and the training efficiency of competitive athletes. In addition, some research topics were related to sports physiology (e.g., physiological function experimental research) and sports behavior (e.g., promoting healthy sports behavior by teenagers). Since the 1980s, the reform and opening-up has brought great vitality to the development of China's society and economy, and also injected new vitality into the development of Chinese sports science. In October 1981, China Sport Science set up sports theory (physical fitness research), sports training, sports medicine, sports biomechanics, sports psychology, and other columns when it was first published. In the article "An Analysis of Sports Science System" published in the fourth issue of 1982, the field of sports science was more clearly subdivided into sports sociology, sports economics, sports management, sports anatomy, sports medicine, sports biomechanics, sports training technology, sports comition technology, and other disciplines [39]. Therefore, sports science is a discipline between social science and natural science, and related research should not only concern social science, but also integrate it with natural science [40]. At present, dozens of branches and marginal branches (such as sports humanities, sports, nature, etc.) have gradually formed in the field of sports science in China, and in the future, sports science in China will move toward the development trend of the subdivision of disciplines and integration of theories [39].

Although the history, culture, language, and characters of China and Taiwan are of the same origin, their differences in sports policy orientation make the academic development of the two regions different. It can be inferred from the research results shown in Figure 5 that most of the related studies in Taiwan in the last 10 years have focused on sports-related issues and actively developed two major fields: natural science and social science. The main research objects were players, athletes, and students; how to promote the development of sports performance; training efficiency; and related social behaviors. According to the network relationship structure between the left half of Cluster 1 (diamond) (body, activity, effect, training, exercises, performance, movement, students, abilities, athlete, and contestants) and Cluster 2 (square) and Cluster 4 (circle), it can be seen that research in natural science should cover sports psychology, sports physiology, and sports biomechanics. At present, natural science research in Taiwan focuses on sports psychology, sports physiology, and sports biomechanics, and tries to improve the sports performance, training efficiency, physical ability, and action display of players or athletes

through various scientific methods, knowledge, and principles [41]. Then, according to the network relationship structure between the right half of Cluster 1 (diamond) (development, mode, and factors) and Cluster 3 (regular triangle), we can get a glimpse of social science issues. In the process of sports academic development in Taiwan, the discipline development, key topics, and theoretical viewpoints of social sciences are indispensable knowledge bases [42]. Its core value lies in trying to critically examine the role, function, and significance of exercises/sports in society, and even exposing people's myths and psychological reactions to exercise/sports through relevant research, thus promoting the normal development of healthy behavior, lifestyles, social systems, and school education [43].

## 5. Conclusions

In conclusion, due to the improvement in people's health awareness and competitive sports standards in China, improving the public sports service system, policies, and the foundation of sports science is the driver to realize the vision of a well-off society, and it is also an important element for the diversified expansion of sports academic fields in the future to enrich the connotation of Chinese social sports [39,40]. Compared with China's sports academic development trend, the Sports Department of the Ministry of Education of Taiwan explained the core spirit and direction of national sports development from 2013 to 2023 with the slogan "Healthy Nationals, Excellent Competition, Vibrant Taiwan" in the White Paper on Sports Policy, and revealed that scholars in relevant academic fields will continue to pay attention to the development of natural science and social science. Looking at the attributes of the keywords in Figures 4 and 5, it is not difficult to see that research on the "philosophy of sport" in the two regions is slightly weak. It is suggested that all disciplines/departments should develop in a balanced way, so as to realize the connotation of diversity in the sports academic field.

Although the results of this study clearly point out the research focuses and trends of mainstream journals in the two regions, due to various political and economic obstacles and restrictions, the academic circles and industries of both sides have no way to communicate through sports, thus eliminating the tense and embarrassing atmosphere on both sides of the Taiwan Strait. In view of this, only by breaking the long-standing prejudice and estrangement between the people of the two regions and actively promoting dialogues between the people and the academic circles with mutual trust and mutual benefit can we create the conditions for sustainable sports academic development. In terms of research limitations, NLP can only mine and analyze text data messages, but when faced with structured data messages, there are limitations in its processing procedures. It is suggested that related research can be supplemented by data mining technology to obtain more macroscopic research results. Finally, this study only analyzed two famous academic journals each for China and Taiwan, which is really only a part of the whole. In addition, it is suggested that researchers expand the scope of exploration and analysis according to the research purpose and demands so as to increase the breadth and depth of the research.

## 6. Patents

This section is not mandatory but may be added if there are patents resulting from the work reported in this manuscript.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1: title, Table S1: title, Video S1: title.

### Author Contributions:

"Conceptualization, Tu-Kuang, Ho and Wei-Yuan, Shih; methodology, Tu-Kuang, Ho and Wei-Yuan, Shih.; software, Tu-Kuang, Ho.; validation, Tu-Kuang, Ho.; formal analysis, Wei-Yuan, Shih.; investigation, Tu-Kuang, Ho.; resources, Tu-Kuang, Ho.; data curation, Wei-Yuan, Shih.; writing—original draft preparation, Tu-Kuang, Ho.; writing—review and editing, Tu-Kuang, Ho.; visualization, Wei-Yuan, Shih.; supervision, Tu-Kuang, Ho.; project administration, Wei-Yuan, Shih.; funding acquisition, X,X . All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

**Funding:** Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX" and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at <https://search.crossref.org/funding>. Any errors may affect your future funding.

**Data Availability Statement:** In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at <https://www.mdpi.com/ethics>. You might choose to exclude this statement if the study did not report any data.

**Acknowledgments:** In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

**Conflicts of Interest:** Declare conflicts of interest or state "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

## Appendix A

The appendix is an optional section that can contain details and data supplemental to the main text—for example, explanations of experimental details that would disrupt the flow of the main text but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

## Appendix B

All appendix sections must be cited in the main text. In the appendices, Figures, Tables, etc. should be labeled starting with "A"—e.g., Figure A1, Figure A2, etc.

## References

1. deLisle, J.; Goldstein, A; Yang, G. *The Internet, social media, and a changing China*. Philadelphia, University of Pennsylvania Press: PA, USA, 2016.
2. Teng, E. *Taiwan and modern China*. Oxford Research Encyclopedia of Asian History: available at: <https://oxfordre.com/asianhistory/view/10.1093/acrefore/9780190277727.001.0001/acrefore-9780190277727-e-155>, 2019.
3. Hammou, B. A.; Lahcen, A.A.; Mouline, S. Towards a real-time processing framework based on improved distributed recurrent neural network variants with fast text for social big data analytics. *Information Processing & Management*. **2020**, 57(1), 1-15.

4. Han, J.; Kamber, M.; Pei, J. *Data mining: concepts and techniques*, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2012.
5. Kang, Y.; Cai, Z.; Tan, C.W.; Huang, Q.; Liu, H. Natural language processing (NLP) in management research: a literature review. *Journal of Management Analytics*, **2020**, 7(2), 1-34.
6. Li, J.; Huang, G.; Fan, C.; Sun, Z.; Zhu, H. Key word retrieval for short text via word2vec, doc2vec, and textrank. *Turkish Journal of Electrical Engineering & Computer Sciences*, **2019**, 27(3), 1794-1805.
7. Makhija, V.; Ahuja, S. Rule-based text retrieval from a bibliographic database. *DESIDOC Journal of Library & Information Technology*, **2018**, 38(1), 5-10.
8. Eisenstein, J. *Introduction to natural language processing*. The MIT Press: Massachusetts, MA, USA, 2019.
9. Fukuoka, Y.; Lindgren, T.; Dov, Y.; Hooper, J.; Aswani, A. Applying natural language processing to understand motivational profiles for maintaining physical activity after a mobile app and accelerometer-based intervention: the mPED randomized controlled trial. *JMIR mHealth and uHealth*, **2018**, 6(6), 1-11.
10. Patel, D.; Shah, D.; Shah, M. The intertwine of brain and body: a quantitative analysis on how big data influences the system of sports. *Ann Data Sci*, **2020**, 7(1), 1-16.
11. Zeng, Y. Evaluation of physical education teaching quality in colleges based on the hybrid technology of data mining and hidden Markov model. *International Journal of Emerging Technologies in Learning*, **2020**, 15(1), 4-15.
12. Sammut, C.; Webb, G.I. *Encyclopedia of machine learning and data mining*, 2nd ed. Springer: New York, NY, USA, 2017.
13. Chen, Y.T.; Chen, L.J.; Wu, T.Y. An investigation on the images of hakka culture tourism destinations by using natural language processing. *Journal of Outdoor Recreation Study*, **2016**, 29(2), 81-111.
14. Cecchini, F.M.; Riedl, M.; Fersini, E.; Biemann, C. A comparison of graph-based word sense induction clustering algorithms in a pseudoword evaluation framework. *Lang Resources and Evaluation*, **2018**, 52(3), 733-770.
15. Cobb-Walgren, C.J.; Pilling, B.K.; Barksdale, H.C.Jr. Does marketing need better marketing? A creative approach to understanding student perceptions of the marketing major. *e-Journal of Business Education and Scholarship of Teaching*, **2017**, 11(1), 97-117.
16. Dobermann, D.; Hamilton, I. S. Publication patterns in developmental psychology: trends and social networks. *International Journal of Psychology*, **2017**, 52 (4), 336-347.
17. Ma, W.Y.; Chen, K.J. Introduction to CKIP Chinese word segmentation system for the first international Chinese word segmentation bakeoff. *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, **2003**, 17, 168-171.
18. Vaidya, P.; Harinarayana, N.S. Social semantics and similarities from user-generated keywords to information retrieval: a case study of social tags in marine science. *DESIDOC Journal of Library & Information Technology*, **2018**, 38(1), 11-15.
19. Harvey, A.; Banerjee, A.; Tong, G.; Brezovjakova, H.; Rees, S.; Byrne, M. Anatomy of an abstract: a guide to writing a scientific abstract. *Journal of the National Student Association of Medical Research*, **2019**, 1(2), 54-60.
20. O'Donoghue, T. *Planning your qualitative research thesis and project: an introduction to interpretivist research in education and the social sciences*, 2nd ed. Routledge: New York, USA, 2018.
21. Harzing, A-W. Two new kids on the block: how do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?. *Scientometrics*, **2019**, 120 (1), 341-349.
22. Kousha, K.; Thelwall, M. Can Google Scholar and Mendeley help to assess the scholarly impacts of dissertations?. *Journal of Informetrics*, **2019**, 13(2), 467-484.
23. Martín-Martín, A.; Orduna-Malea, E.; Thelwall, M.; López-Cózar, E. D. Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, **2018**, 12(4), 1160-1177.
24. Zientek, L.R.; Werner, J. M.; Campuzano, M. V.; Nimon K. The use of Google Scholar for research and research dissemination. *New Horizons in Adult Education and Human Resource Development*, **2018**, 30(1), 39-46.
25. Qiu, Y.; Ji, W.; Zhang, C. A hybrid machine learning and population knowledge mining method to minimize makespan and total tardiness of multi-variety products. *Applied Sciences*, **2019**, 9(24), 5286.
26. Navathe, S.B.; Ramez, E. (2000). Data warehousing and data mining. In *Fundamentals of Database Systems*. Pearson Education: Singapore, SG, 841-872.
27. Newberry, K.M.; Bailey, H.R. Does semantic knowledge influence event segmentation and recall of text?. *Memory & Cognition*, **2019**, 47(6), 173-1187.
28. Chen, Y.J.; Liou, W.C.; Chen, Y.M.; Wu, J.H. (2019). Fraud detection for financial statements of business groups. *International Journal of Accounting Information Systems*, **2019**, 32,1-23.
29. Hughes, G.; Musco, P.; Caine, S.; Howe, L. Lower limb asymmetry after anterior cruciate ligament reconstruction in adolescent athlete: a systematic review and meta-analysis. *Journal of Athletic Training*, **2020**, 55(8), 811-825.
30. Singh, A.P.; Kushwaha, A.K. Analysis of segmentation methods for Brahmi script. *DESIDOC Journal of Library and Information Technology*, **2019**, 39(2), 109-116.
31. Yang, D.H.; Wang, Y.; Yu, T.; Liu, X. Macro-level collaboration network analysis and visualization with essential science indicators: a case of social sciences. *Malaysian Journal of Library & Information Science*, **2020**, 25(1), 121-138.
32. Aizawa, A. On the quantitative representation of term specificity based on terms and documents co-occurrences. *Journal of Information Processing Society of Japan*, **2000**, 41(12), 3332-3343. (In Japanese)
33. Hong, Y.; Yao, Q.; Yang, Y.; Feng, J-J.; Wu, S-D.; Ji, W-X.; Yao, L.; Liu, Z-Y. Knowledge structure and theme trends analysis on general practitioner research: a co-word perspective. *BMC Family Practice*, **2016**, 17(10), 1-15.
34. van Eck, N.J.; Waltman, L. Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **2007**, 15, 625-645.

---

- 35. Ohsawa, Y.; Benson, N.E.; Yachida, M. KeyGraph: automatic indexing by segmenting and unifying co-occurrence graphs. *The Transactions of the Institute of Electronics, Information and Communication Engineers*, **1999**, 82(2), 391-400.
- 36. Kim, K.; Kim, J. How does local partners network embeddedness affect international joint venture survival in different subnational contexts?. *Asia Pacific Journal of Management*, **2017**, 35, 1055-1080.
- 37. Jacobs, W.; Goodson, P.; Barry, A.E.; McLeroy, K.R.; McKyer, E.L.; Valente, T.W. (2017). Adolescent social networks and alcohol use: variability by gender and type. *Substance Use & Misuse*, **2017**, 52(4), 477-487.
- 38. Maluleka, J.R.; Onyancha, O.B. Research collaboration among library and information science schools in South Africa (1991–2012): an informetrics study. *Mousaion*, **2017**, 34(3), 36-59.
- 39. Wang, X.K.; Liu, Y.Z. On the subject structure of the sports science. *Journal of Physical Education*, **2002**, 9(1), 4-8. (In Chinese)
- 40. Han, H.; Zheng, J. Development of social sports organization on the 70th anniversary of the founding of the People's Republic of China: course review, reality consideration and future trend. *China Sport Science*, **2019**, 39(5), 3-12. (In Chinese)
- 41. Shih, W.Y.; Ho, T.K. Application of the text mining method to analyze the development trends of sports research in the era of big data. *Physical Education Journal*, **2020**, 53(4), 439-452. (In Chinese)
- 42. Hsu, S.Y. *A sociology of sport and sports*. Lucky Book Store: Taipei, ROC, 2012. (In Chinese)
- 43. Coakley, J. (2007). *Sports in society: issues and controversies*. McGraw-Hill: London, LON, 2007.