Article

Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research

Binling Nie * and Shouqian Sun

Institute of Industrial Design, College of Computer Science, Zhejiang University, Hangzhou 310027, China; ssq@zju.edu.cn

* Correspondence: nbl1221@zju.edu.cn; Tel.: +86-131-0770-2260

Abstract: The research goal of this paper is to identify major academic branches and to detect research trends in design research using text mining techniques. In this paper, the information about scientific literature in design research isprocessed. A combination of clustering and bibliometric analysis led to shaping four academic branches and summarizing each academic branch. Then, research trends and the evolution for each academic branch are explored. We perform a two-dimensional text mining approach, including bibliometric and network analysis, in order to detect trends of major academic branches. Specifically, the bibliometric characterization aims to assess design research area outputs, while the network analysis intends to reveal research trends in each academic branch of design research and the evolution of core research themes.

1. Introduction

The volume and diversity of scientific literature are expanding every day. A basis for future development is recorded in detail by abundant archived scientific literature. Data mining tries to discover information hidden in scientific literature, which is not accessible by simple statistical techniques [1]. Text mining techniques area significant subset of data mining that aims to extract knowledge from unstructured or semi-structured textual data and haswidespread applications in analyzing and processing textual documents [2]. Hence, combining textual mining techniques and bibliometric analysis can be exploited to help us discover more unseen patterns in research fields than simple bibliometric analysis.

Great design has been the crucial point to make more aesthetically pleasing and more practical products and provide more adaptable services. It was revealed by the Design Management Institute that many design-centric companies like Apple and Target outperformed the S&P 500 by 224% from 2003–2013 [3]. Therefore, building a systematic view of design research becomes increasingly essential. Until now, analyzing the design research area from the perspective of scientometrics is still a largely unexplored area. To the best of our knowledge, there has been only one paper applying a quantitative approach to investigating the evolution and future trends of design research by analyzing citations of papers in the journal Design Studies [4]. Additionally, few existing studies [5,6] presented a qualitative method to analyze the design research area. Yet, there has been no in-depth study on keeping track of the current advances in the design research area.

Consequently, we use textual mining techniques to identify major academic branches and detect research trends in design research. Our work is perhaps the first such attempt of this kind to present a detailed bibliometric analysis in research fields.

This paper is organized as follows. We first gave a brief introduction in Section 1. Then, we explain thematerials and methods in Section 2. Next, experiment results show the identification of

major academic branches and their trends in Section 3. Finally, we discuss the results and make some conclusions in Sections 3 and 4.

2. Materials and Methods

In this section, we demonstrate how to implement the text mining process. First, we prepare all of the needed data and preprocess them into the needed format. Then, we discuss a combination of clustering and bibliometric analysis to identify major academic branches of design research and summarize each academic branch (Section 2.2). Finally, we describe how a two-dimensional text mining approach including bibliometric and network analysis isused to detect research trends from different perspectives (Section 2.3).

2.1. Data Preparation

To give a clear distinction to the research goal of this paper, the definition of design is clarified as follows: design may be a substantive referenceto the aesthetic, functional, economic and sociopolitical dimensions of both the design object and design process [7]. The data areprovided by Web of Science (WOS). Data retrieval strategy: the subject is "design research" or "design studies"; the publication period isconfined from "between year 2004 and 2015" (retrieved data areduring May 2016 and June 2016). Accurately retrieved reference type is "Article"; language is "English"; data category is refined as "engineering multidisciplinary or engineering manufacturing or engineering industrial". To better meet the requirements of the definition of design, we expand the number of papers in the main journals, which are discussed in Quality perceptions of design journals: The design scholars' perspective [8]. For obviously unrelated journals, for example Design Codes and Cryptograph, we excluded all of their papers. A total number of 20,218 publications isobtained. We further divide the data into two parts: BASE data and Citation data. BASE data store items containing: title, authors, journal name, keywords, abstract, publishing date. The items in BASE data are fed into the BASE table. Citation data (i.e., references cited by the articles included in the core dataset) store the following items: the cited article's title, the cited article's author, the cited article's journal, total number of the citing article, per year's number of the citing article during 2004–2015. Similarly, the data in the citing dataset are fed into the Citation table.

2.2. Identification of Major Academic Branches

In this subsection, we discuss how to perform clustering, which is considered as an important tool for data mining, to identify major academic branches of design research. Detailed steps of the algorithm are explained as follows,

1. Preprocess the document set, including word segmentation and removing stop words.
2. Use Latent Dirichlet Allocation(LDA) [9] to extract features and determine the optimal number oftopics (features)
3. Calculate the results of the k of 2, 3, 4, 5, 6,7 by the K-means algorithm [10], then combine the Sum of Squared Error(SSE) and inter-cluster distance to obtain the best classification result. We use the distance between cluster centroids to measure inter-cluster distance. The calculation formula of SSE is listed as Equation (1).

$$SSE = \sum_{i=1}^{k} \sum_{d_j \in C_i} |d_j - cen_i|^2 \tag{1}$$

where k is the number of clusters, C_i is the cluster i, d_j represents the document j that belongs to the cluster i and cen_i is the clustering centroid of the cluster i.

We briefly introduce LDA below. LDA is a hierarchical Bayesian model. It is used to modelcorpora of documents that can be represented by bags of words. The generative process of document sampling assumes a set of topics, where each document is sampled from a mixture of topics, and each topic is a

discrete probability distribution that defines how likely each word appears in a given topic. As the number of topics affects the fitting performance of the LDA model, therefore we use common criteria evaluation of perplexity to determine the optimal number of topics. Normally, the smaller the degree of perplexity, the better the LDA model performs. The calculation formula of perplexity is as follows,

$$perplexity(D) = \exp\left| -\frac{\sum\limits_{i=1}^{n} \log(p(d_i))}{\sum\limits_{i=1}^{n} N_i} \right| \tag{2}$$

where n is the number of documents, N_i represents the length of the document d_i and $p(d_i)$ is the probability that the LDA model generates the document d_i.

There are many methods to estimate the model parameters;we use the Gibbs sampling algorithm [11] in this paper. The standard approach to "smooth" the multinomial parameters is assigning positive probability to all vocabulary items whether or not they are observed in the training set.

Based on classification results, we calculate traditional bibliometric indicators such as total outputs, total citation and citation-based impact assessment [12] to provide an architecture overview of design research.

2.3. Detection of the Research Trends of Design Research

In this subsection, we describe how a two-dimensional text mining approach including bibliometric and network analysis isused to detect research trends from different perspectives. The bibliometric characterization aims to assess academic outputs trends and the development trends of the design research area. The network analysis intends to find out research trends in each academic branch of design research and the evolution of core research themes.

Bibliometric analysis involves total outputs and citation-based impact assessment. Counting the number of research publications reflects the impact and usefulness of scientific research output. Additionally, thecitation count is a good measure of the influence power of a research paper. High citation indicates more effectiveness, usefulness and productiveness. A previous study [13] computed the values of the Citation Function (CF) and the Co-Citation Function (CCF) to identify the changes of thematic clusters over time. The profile of both functions is similar, and CCF is more expressive. However, a large data collection amount and computational amount are needed for CCF. Therefore, we apply an approach thatis based on CF, presented as follows,

$$CF_t = \sum_i^N C_{i,t} \left(\frac{C_i}{\max C} + \frac{P_i}{\max P} \right) \tag{3}$$

where N is the number of documents in the analyzed thematic cluster, $C_{i,t}$ is the number of citations of the document i in the year t, C_i is total number of citations of the document i, max C is the maximum value of the citations in the analyzed documents, P_i is the period when the document i has been cited and max P is the maximum period where the document can be cited for the data under analysis. The value of CF stands for a research domain's development level and is related to the number of citations only. It provides easy identification of the developing research areas and offers an alternative view on the subject development.

Network analysis is carried out to further analyze the social relations among the core themes. Keywords reflect some important information about research trends. In this paper, high frequency keywords are extracted from the abstract and title of scientific literature in each academic branch. After that, networks based on the keyword co-occurrence matrix are constructed to visualize the internal dynamics of major academic branches in the design research area by using Ucinet 6.0.

3. Results

Since this is the first quantitative study that analyzes the design research area in detail, it might be useful to reveal some general findings that provide an overview of the design research area. We list the locations of the authors and find out the journal's influence among the international design research community in Appendix A. Then, we combine text mining techniques and bibliometric analysis to obtain major academic branches of design research (Section 3.1) and identify the development trends of major academic branches (Section 3.2).

3.1. Major Academic Branches of Design Research

Major academic branches of design research are identified in this subsection. Firstly, the corpus of the LDA model consists of unique words obtained from the BASE table, which consists of 4893 unique words. To improve accuracy, a special stop word list just for academic articles in design research is created to remove the stop words. Next, we repeat the calculating of the perplexity of LDA model when the number of topic varies from 10–100, which aims to obtain an optimal topic set. Results are shown in Figure 1. When the number of topics is 30, the performance is the best. Thus, we apply LDA to remodel the preprocessed corpus by setting the number of topics to 30. Then, we get the topic distribution as the selected real-valued features for each article. Finally, we use K-means to perform classification with selected features and compare classification results, which is evaluated by SSE with different k values. Additionally, we compute inter-cluster distance with different k values, as listed in Table 1. In Figure 2, we can see that the value of SSE gets smaller as the value of k gradually increases from 2–8. The value of SSE achieves a relatively stable state when the value of k is greater than four. From Table 1, we can observe that the inter-cluster distances between Cluster 1 and Cluster 4, Cluster 2 and Cluster 3, Cluster 2 and Cluster 5, Cluster 3 and Cluster 5 are obviously smaller than other inter cluster distances when k is seven. We can merge Clusters 1, 4 and Clusters 2, 3, 5 into two clusters and obtain four clusters. The same analysis logically would apply to the circumstances when k is five and six.Thus, it is reasonable to classify all of the research articles into four clusters. Then, we calculate the four cluster centroids. According to top three topics in each cluster centroid in Table 2 and the top ten words of each topic in Table A3 of Appendix B, each cluster can be interpreted as an academic branch of design.
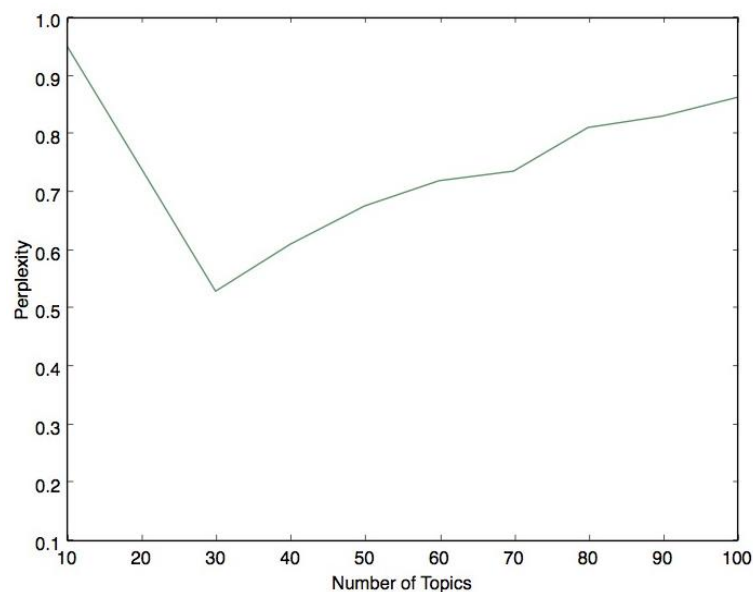
Cluster 1 includes these words, "interaction", "human-computer", "user", "interface", etc. They reveal a discipline of design thatpays attention tosatisfying the needs and desires of the majority of people who will use the product. Therefore, Cluster 1 is called interaction design. Interaction design is about users' behaviors. It focuses on creating engaging interfaces with understanding how users and the web communicate with each other [14]. User experience is vital to all kinds of products and services. On the web, user experience becomes even more important [15].

The most frequently-used words in Cluster 2 are "ergonomics", "musculoskeletal", "work", "physical", "back", "pain", etc. These words are involved in harmonizingthe functionality of tasks with the human requirements of those performing them. As a result, Cluster 2 is named ergonomic design. It is a branch of science drawing from physiology, engineering and psychology studies. It is often the most difficult variable to factor into the early stages of the design process. With poor ergonomic design, users may even be injured [16]. Ergonomic design is said to be human-centered design focusing on usability. It seeks to ensure that human restrictions and capabilities are met and supported by design options.
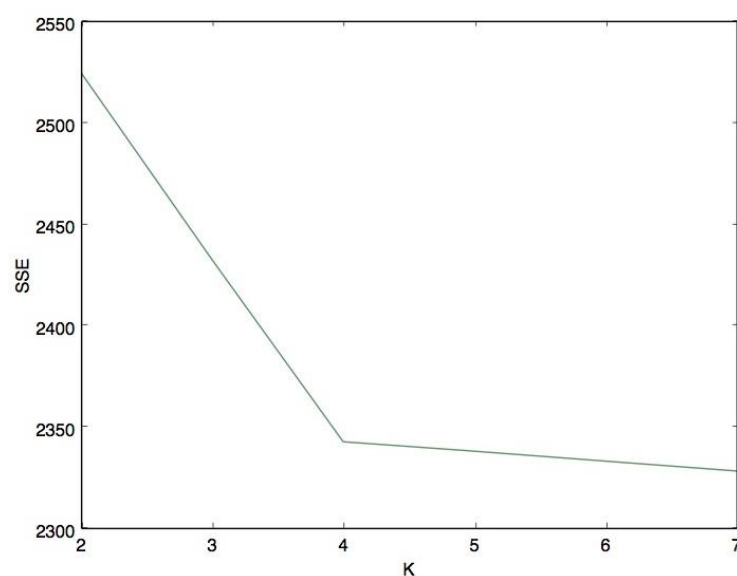
These words, "product", "design", "process", "theory", etc., are contained in Cluster 3. The content of studies includes the process of creating a new product. These words also demonstrate the methods, strategies, research, analysis and management of design. To conclude, Cluster 2 belongs to product design. It presents an in-depth study of structured design processes and methods, which has many benefits in education and industry. On the industry side, a structured design process is mandatory to effectively decide what projects to bring to market, schedule this development pipeline in a changing uncertain world and efficiently create robust delightful products. On the educational side,

the benefits of using structured design methods include concrete experiences with hands-on products, applications of contemporary technologies, realistic and fruitful applications of applied mathematics and scientific principles, studies of systematic experimentation, exploration of the boundaries of design methodology and decision making for real product development [17].

These words, "visualization", "information", "data", "mining", etc., explain the practice of presenting information in a way to display information effectively with graphic design. Overall, Cluster 4 is impliedto be information design. It is defined as the art and science of preparing information so that it can be used by human beings with efficiency and effectiveness. It aims to develop documents that are comprehensible, rapidly and accurately retrievable and easy to translate into effective action. Another objective is to design interactions with equipment that are easy, natural, and as pleasant as possible. This involves solving many problems in the design of the human-computer interface [18].

Figure 1. The relationship between the degree of perplexity and the number of topics.

Figure 2. The relationship between SSE and the value of *k* in K-means.

Table 1. Inter-cluster distance with different *k* values.

k Value	Inter-Cluster Distance
2	1–2: 0.1771
3	1–2: 0.1770; 1–3: 0.2720; 2–3: 0.1991
4	1–2: 0.1222; 1–3: 0.2166; 1–4: 0.2831; 2–3: 0.1992; 2–4: 0.1832; 3–4: 0.2780
5	1–2: 0.1246; 1–3: 0.2202; 1–4: 0.2855; 1–5: 0.1820; 2–3: 0.1509; 2–4: 0.1836; 2–5: 0.0165; 3–4: 0.2800; 3–5: 0.1582; 4–5: 0.2438
6	1–2: 0.1073; 1–3: 0.1412; 1–4: 0.0220; 1–5: 0.1823; 1–6: 0.0400; 2–3: 0.2401; 2–4: 0.1786; 2–5: 0.2838; 2–6: 0.1460; 3–4: 0.2196; 3–5: 0.3037; 3–6: 0.1906; 4–5: 0.2594; 4–6: 0.0312; 5–6: 0.2331
7	1–2: 0.2880; 1–3: 0.2234; 1–4: 0.0266; 1–5: 0.1854; 1–6: 0.2598; 1–7: 0.2591; 2–3: 0.0120; 2–4: 0.1838; 2–5: 0.0476; 2–6: 0.1395; 2–7: 0.1373; 3–4: 0.2815; 3–5: 0.0506; 3–6: 0.2318; 3–7: 0.2308; 4–5: 0.2454; 4–6: 0.3172; 4–7: 0.3176; 5–6: 0.2014; 5–7: 0.1970; 6–7: 0.2533

Table 2. Top three topics in cluster centroids.

Classification Number	Top Three Topics in Cluster Centroids
Cluster 1	Topic 20, Topic 21, Topic 1
Cluster 2	Topic 9, Topic 7, Topic 19
Cluster 3	Topic 12, Topic 3, Topic 29
Cluster 4	Topic 14, Topic 24, Topic 0

In addition, the research outputs of each branch are computed to preliminarily reflect the focus of design research. The result is listed in Table 3. The publication output of production design has been at the lowest level of the four branches. Thisis owing to the fact that there are less doctoral programs of product design atdomestic and foreign institutions. Correspondingly, there isa small number of publications referring to product design, not to mention articles which are selected by Science Citation Index (SCI). In general, students who specialize in product design prefer to be designers not researchers. In contrast, the other two branches, information design and ergonomic design, are more involved in engineering. More and more interdisciplinary researchers participate in the studies of information design and ergonomic design. Besides, it appears that the developing branch "interaction design" has attracted more research attention than the veteran branch called "product design". It owns the rapid development of networks. Towards the end of the 1980s and early 1990s, the introduction of computer networks had a large impact on interaction design. As graphics and computational capabilities increased, it was possible to interactively visualize large-scale data. With the awareness of the importance of the usability of its products in a competitive environment, interaction design has grown gradually over the years [14]. It can be seen that technology has broadened the boundaries of design research. Additionally, "ergonomic design" has a great impact on other academic branches. "Interaction design" and "information design" affect each other. Much of the experience of "product design" is used for reference by "ergonomic design".

Table 3. Research outputs of major academic branches.

Branch	TP	TC	ACPP	CDS		
Interaction design	7272	71,482	9.83	Ergonomic	Product	Information
				1091	1874	10,200
Ergonomics design	5293	46,368	8.76	Interaction	Product	Information
				12,084	4002	8291
Product design	2617	22,076	8.44	Interaction	Ergonomic	Information
				4911	14,044	3167
Information design	5112	44,750	8.12	Interaction	Ergonomic	Product
				19,793	1454	2612

TP(Total Publications); TC(Total Citations); ACPP(Average Citation Per Paper); CDS(Cross-Domain Citations).

3.2. Trends of Major Academic Branches

To detect the trends of major academic branches, we perform a two-dimensional approach including bibliometric and network analysis. The bibliometric characterization aims to assess design research area outputs' trends. Additionally, the network analysis intends to find out research trends in each academic branch of design research and the evolution of core research themes.

3.2.1. Identifying Research Trends with the Bibliometric Characterization

The time-trend analysis among four branches is displayed in this subsection. Trends about publication outputs and citations in design research are listed in Figures 3 and 4. Firstly, an obvious rise can also be seen in the annual publication output related to interaction design. There is no doubt that the popularity of the Internet motivates the rapid development of interaction design. The value of CF grows until 2009, revealing that interaction design has shown tremendous growth. After 2009, it keeps flat without two local minima near 2010 and 2013. This indicates continued development of interaction design. Secondly, it also appears that the annual publication output of the traditional branch, product design, has grown steadily. The value of the CF of product design grows until 2015. However, there are temporary drops of the values between 2009 and 2010 and between 2013 and 2014, understood as temporary interruptions in the development of product design. Thirdly sometimes the annual publication output of ergonomic design slightly increases, and sometimes, it slightly decreases. Overall, it is growing. According to Figure 4, the value of CF grows at a steady rate, which indicates stable development of ergonomic design. Finally, the annual publication output of information design keeps increasing year by year. Nevertheless, the value of CF firstly increases and then decreases. Information design seems to become stagnant after 2009.

Additionally, the topic evolution and citations of main topics in each branch are computed to reflect the influence of each branch. In Figures 5–12, we can learn that the influence of each branch increases yearly. As can be observed from Figures 5 and 6, Topic 9 is ascending with a rapid growth, while Topic 7 and Topic 19 keep up with and surpass Topic 9. The tendency reveals that ergonomic design is concerned with the scientific study of products and the environment, to be safe, healthy and comfortable for human beings. As shown in Figures 7 and 8, three main topics alternately increase, and Topic 3 keeps expanding rapidly. It shows that product design tends to be focused on developing the understanding of the design process. By comparing Figures 9 and 10, Topic 21 keeps its predominance, and Topic 1 intends to catch up. This indicates that research about the interaction of social media is a global trend in interaction design. According to Figures 11 and 12, Topic 14 maintains a competitive advantage with a slight drop, and Topic 0 has grown rapidly. It elaborates that information design presents how data are processed and what data you see.
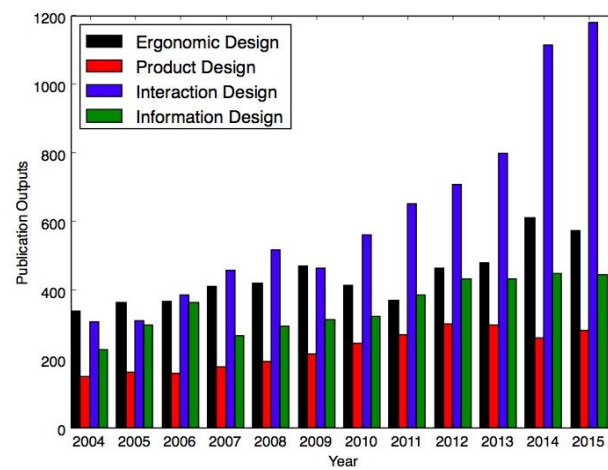
Figure 3. Publication outputs of major academic branches in design research.
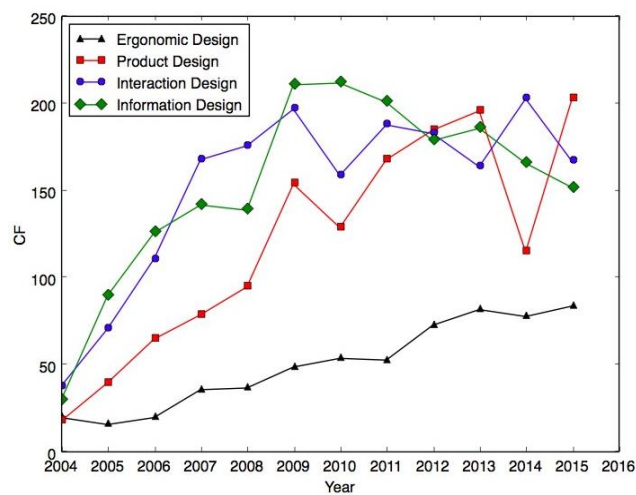

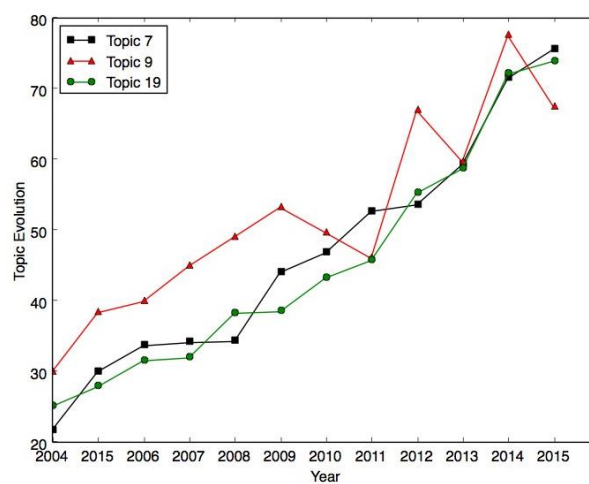**Figure 4.** Trend of the Citation Function (CF) in design research.
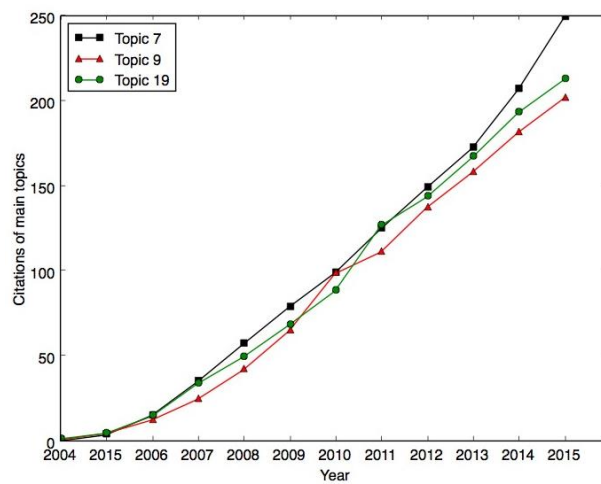

Figure 5. The main topics' evolution in ergonomic design.

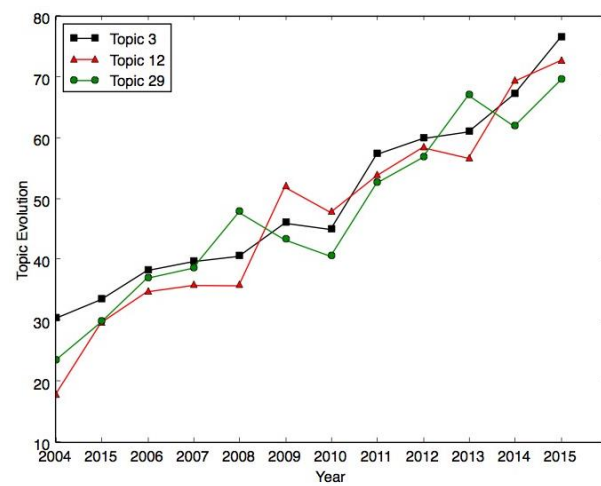**Figure 6.** The citations of the main topics in ergonomic design.

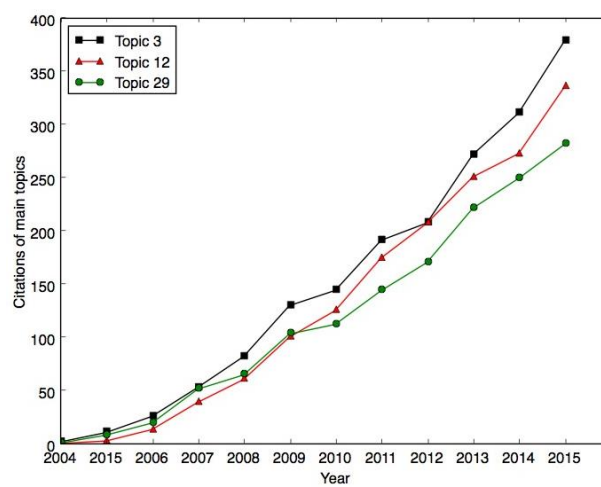**Figure 7.** The main topics' evolution in product design.

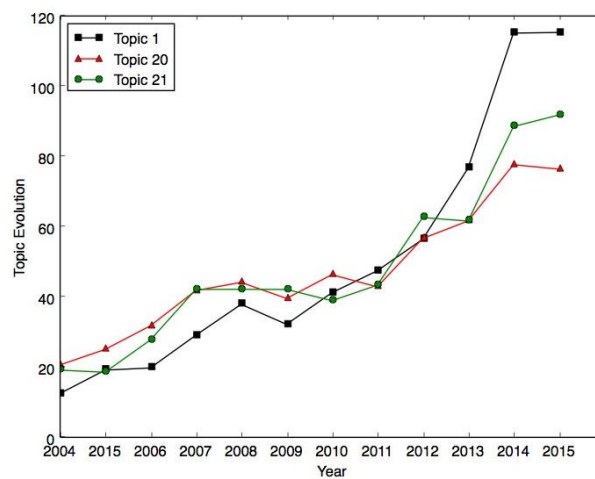**Figure 8.** The citations of the main topics in product design.

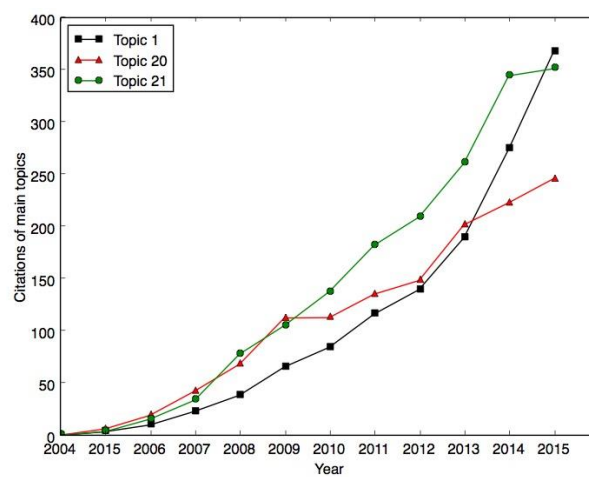**Figure 9.** The main topics' evolution in interaction design.

Figure 10. The citations of the main topics in interaction design.
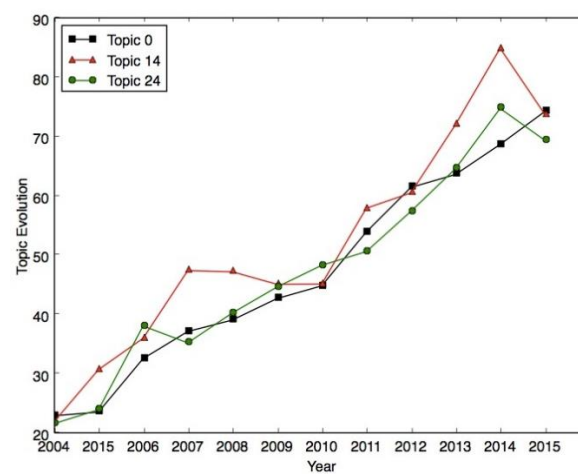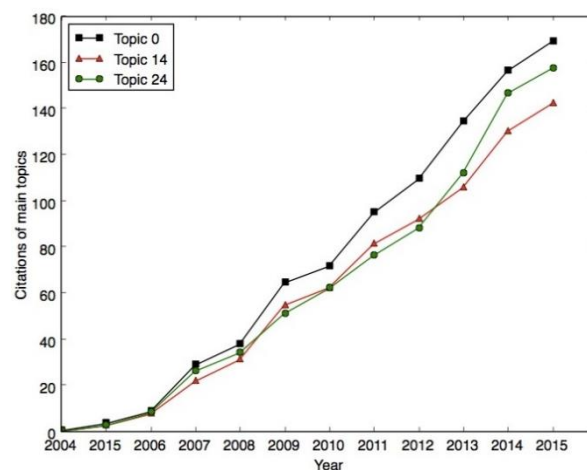

Figure 11. The main topics' evolution in information design.

Figure 12. The citations of the main topics in information design.

3.2.2. Identifying Research Trends with Network Analysis

We further analyze keywords to reveal the trend of core themes in each academic branch. We divide the full period (2004–2015) into three periods. These are 2004–2007, 2008–2011 and 2012–2015. For each branch, the co-occurrence relationships among the top 50 high-frequency keywords in each period are listed in Figures 13–16. The connection relationship between two words is represented by the lines. The thicker the line, the stronger the connection. Additionally, the size of nodes marks the degree of the core or edge. The colors represent *k* values listed in Figures 13–16. The bigger the k value, the more important the node. Core themes in each of the four years are represented by red nodes.

As can be observed from Figure 13, "social media" is the biggest node during 2004–2015, except for "human-computer interaction". It explains that social network sites are paying more attention to exploring the interaction between the user and the user interface, with the popularization of computer networks [19]. The color of nodes, such as "personality", "gender", "motivation", "emotion", "technology" and "e-learning", is red, and their size remains stable in the three periods, which states that researchers continue to study the use of computers from a psychological perspective [20]. Some researchers address human interactions with computers [21]. Others emphasize the psychological effects of computers on phenomena such as learning, cognition, personality and social interactions [22–25]. Additionally, the k values of "virtual reality" nodes in the three periods are 12, 13 and 17. The rising tendency reveals that several researchers have increased research interest in the design of 3D interaction [26]. "Mobile phone" is an emerging theme during 2012–2015. With the increasing penetration of mobile phones, quite a few studies focus on the interactive pattern of smart phones users and examine factors that influence problematic smart phone use [27,28].

As shown in Figure 14, the biggest node is "ergonomics", which is the core theme during 2004–2015. Most of the research in ergonomic design applies ergonomics/human factors in the design, planning and management of technical and social systems at work or leisure [29,30]. The second frequently-used word is "biomechanics". It states that some studies in ergonomic design assess quantitatively the impact of the biomechanical factors on the risk of injury to control occupational injuries [31]. The nodes such as "musculoskeletal disorders", "postures", "fatigue", "mental workload", "low back pain", "comfort", "heart rate" and "safety" continuously appear as red nodes in co-word networks, and their size have kept stable in the past 12 years, which reveals that they are the core themes during 2004–2015. It suggests that due to the increasingly complex technology, the increment of work activities required of people is leading to a broader range of physical stress, as well as psychological stress. Most of the studies hold the view that the principle of ergonomics design is to ensure the comfort, health and safety of people at work. To yield ergonomic design principles, a host of researchers leverage fundamental knowledge of human capabilities and limitations and the basic

understanding of cognitive, physical, physiological and social aspects of human behaviors [32,33]. Furthermore, "situation awareness" is an emerging theme during 2012–2015. An increasing number of studies explore how situation awareness affects human interaction [34].

By comparing the three pictures in Figure 15 side by side, "product design", "design theory", "design process", "design methodology", "design cognition", "design activity", "design research", "design tools" and "design models" are the most frequently-used words during 2004–2015. It reveals that a multitude of researcher have been focused on the theory of design. It presents in-depth studies of the methods, strategies, research and analysis of design [35]. Besides, it is related to the management of the design strategy, process and implementation [36]. The focus of much publications includes basic theoretical advances, case studies, new methodologies and procedures; as well as empirical studies. The red nodes such as "design theory", "design process", "design research" and "design cognition" continuously appear in co-word networks in the past 12 years, and their size remains stable, which represents the core themes during 2004–2015. It demonstrates that the IT evolution has influenced the design professions. Many researchers invent new tools, methods and techniques to design and manufacture products [37]. The core theme "design education" during 2008–2015 shows a rising importance. There are more and more researchers developing more innovative approaches to design education. The design research area affects the development of the economy and society. "Problem solving", "collaborative design", "design management" and "protocol analysis" are emerging themes during 2012–2015. It demonstrates that several researchers provide insight into process structure and the organization of design to enhance the quality of products [38]. Additionally, the research method "protocol analysis" captures researchers' attention [39]. Nevertheless, a few themes, such as "consumption" and "ethnography", have become obsolete with the passage of time.

Figure 16 shows that "information visualization", "data visualization", "visual analytics", "volume visualization", "dimensionality reduction" and "feature extraction" have been core themes during 2004–2015. It reveals that information design is concerned with visualizing information with informational graphics, such as charts and diagrams. Information design is associated with making complex data easier to understand and to use, as well. Therefore, information design can be integrated into the research process by processing data into a visual format [40]. Beyond that, the k values of red nodes in three periods are 15, 16 and 17, revealing that the relationship between core themes is increasingly tight in information design. The core theme "visual analytics" shows increasing importance. Quite a few studies concentrate on solutions that give the user more freedom to visually represent information, rather than standard solutions, for which the structure of the visualization is fixed. Still other researchers address the challenge of visual analytics for big data with the enormous growth of data in the last few decades [41,42]. "GIS" continuously appears during 2004–2015. "Flow visualization" and "user interface" are emerging themes during 2008–2012. It states that information design is widely used in different fields. A part of the researchers charts the workflow of information to increase the efficiency of data curation tasks [43]. The others are devoted to studying the impact of information design on the user's focus of attention [44].

In conclusion, the size of red nodes in Figures 13, 14 and 16 has changed little in the three periods. We can infer that these research hotspots have a sustainable development. From Figure 15, the size of "design practice" gradually becomes the biggest one, and other nodes get much smaller, which reveals that "design practice" gradually occupied a dominant position in product design.

a during 2004-2007(red nodes,k=18;yellow nodes,k=16;blue nodes,k=15;dark green nodes,k=14;green nodes,k=13;gray nodes,k=12;purple nodes,k=7;black nodes,k=5)

b during 2008-2011(red nodes,k=18;blue nodes,k=16;gray nodes,k=13;purple nodes,k=7;dark blue nodes,k=5;dark green nodes,k=4)

c during 2012-2015(red nodes,k=18;dark blue nodes,k=17;purple nodes,k=15;green nodes,k=13;dark green,k=11;gray nodes,k=5;black nodes,k=4)

Figure 13. Co-word network of interaction design during 2004–2015.

a during 2004-2007(red nodes,k=19;dark blue nodes,k=5;black nodes,k=17;gray nodes,k=18)

b during 2008-2011(red nodes,k=22;dark blue nodes,k=21;black nodes,k=5)

c during 2012-2015(red nodes,k=23;dark blue nodes,k=5;black nodes,k=20;purple nodes,k=19;gray nodes,k=13)

Figure 14. Co-word network of ergonomic design during 2004–2015, emg: Electromyography.

Figure 15. Co-word network of product design during 2004–2015.
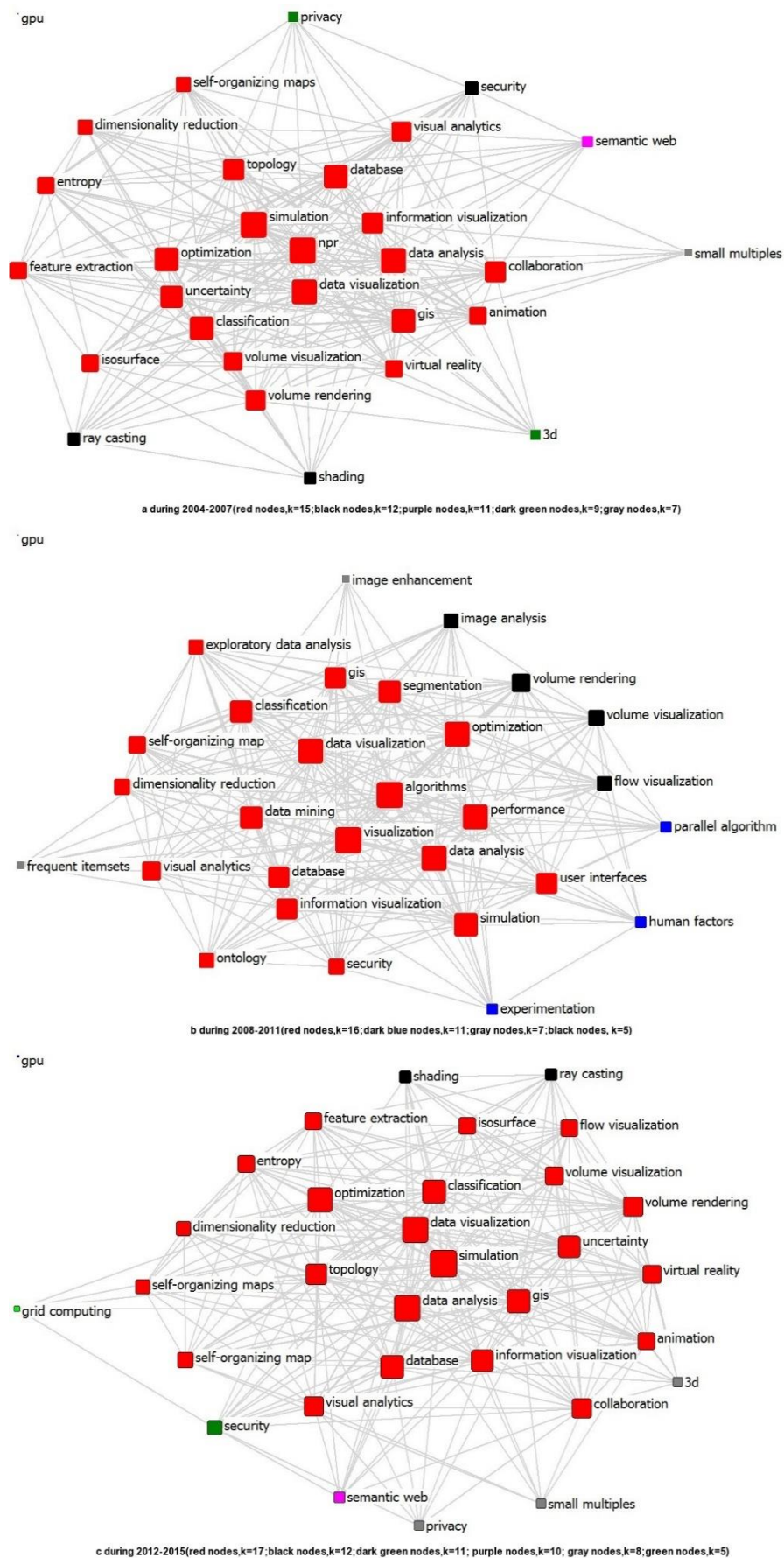
Figure 16. Co-word network of information design during 2004–2015.

4. Conclusions

This study presents a bibliometric, network-theoretic and text-based analysis of the design research area during the last 12-year period (2004–2015). Our work is the first in-depth study on keeping track of the current advances in the design research area by using text mining techniques. Furthermore, the result shows that the developed methods are universal and could be applied to manage the knowledge of various research fields. Additionally, text mining techniques used in this study could help researchers have a comprehensive understanding about the knowledge of a certain field hidden in a large amount of scientific literature. The clustering method provides an architecture overview of a certain field in more detail. Additionally, social network analysis further explores the core themes to help researchers with a better understanding of the development gain of a certain field.

For further studies, we will apply the author-topic model, a probabilistic model to link authors to observed words in the scientific literature of the research fields (a case study of design research), which will provide a general framework for exploration, discovery and query-answering in the context of the relationships of author and topics.

Author Contributions: Binling Nie conducted the experiment and wrote the paper. Shouqian Sun offered useful suggestions for writing the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Top 15 research countries/territories in design research.

Country	TA	%	Single-Country		Collaborative-Country		
			SA	%	ICA	%	MC
USA	5445	26.93	4736	86.98	709	13.02	South Korea
England	1605	7.94	1280	79.75	325	20.25	Australia
Canada	1056	5.22	804	76.14	252	23.86	Finland
Peoples R China	1010	5.00	655	64.85	355	35.15	USA
The Netherlands	937	4.63	750	80.04	187	19.96	USA
Australia	861	4.26	688	79.91	173	20.09	USA
Germany	798	3.95	720	90.23	78	9.77	USA
South Korea	670	3.31	526	78.51	144	21.49	England
Spain	656	3.24	497	75.76	159	24.24	Scotland
France	640	3.17	516	80.63	124	19.38	England
Sweden	621	3.07	484	77.94	137	22.06	USA
Italy	480	2.37	394	82.08	86	17.92	England
Finland	391	1.93	304	77.75	87	22.25	USA
India	341	1.69	302	88.56	39	11.44	USA
Japan	257	1.27	214	83.27	43	16.73	USA

TA, Total Articles; SA, Single-country Articles; ICA, Internationally-Collaborative Articles; MC, Major Collaborator.

Table A2. Top 25 active journals in design research.

No.	Journal	TA	ISSN	Impact Factor	Edition
1	Computers In Human Behavior	2977	0747-5632	2.694	SSCI
2	Ergonomics	1500	0014-0139	1.556	SCIE and SSCI
3	Applied Ergonomics	1352	0003-6870	2.023	SCIE and SSCI
4	Computer-Aided Design	1345	0010-4485	1.801	SCIE
5	International Journal Of Industrial Ergonomics	1047	0619-8141	1.070	SCIE and SSCI
6	Human Factors	847	0018-7208	1.290	SCIE and SSCI
7	International Journal Of Human-Computer Studies	800	1071-5819	1.293	SCIE and SSCI
8	Behavior& Information Technology	698	0144-929X	0.891	SCIE and SSCI
9	Interacting With Computers	529	0953-5438	1.268	SCIE and SSCI
10	International Journal Of Human-Computer Interaction	529	1044-7318	0.850	SCIE and SSCI
11	Journal Of Documentation	487	0022-0418	0.43	SSCI
12	Journal Of Engineering Design	392	0954-4828	1.036	SCIE
14	Design Studies	359	0142-694X	1.295	SCIE
15	Design Issues	339	0747-9360	-	-
16	Computer Graphics Forum	330	0167-7055	1.642	SCIE
17	Research In Engineering Design	197	0934-9839	1.233	SCIE
18	International Journal Of Design	190	1991-3761	0.732	SCIE and SSCI
19	Computers & Geosciences	173	0098-3004	2.054	SCIE
20	Information Visualization	165	1473-8716	0.541	SCIE
21	Design Journal	147	-	-	-
22	Journal Of Design History	125	0952-4649	-	-
23	Computers In Industry	112	0166-3615	1.287	SCIE
24	ACM Transactions On Computer-Human Interaction	90	1073-0516	1.079	SCIE
25	Journal Of Visual Languages And Computing	77	1045-926X	0.893	SCIE

SSCI: Social Sciences Citation Index, SCIE: Science Citation Index Expanded.

Appendix B

Table A3. Top Ten Words in Topics.

Topics	Top Ten Words in Topics
Topic 0	data visualization clustering network fuzzy networks reduction map neural classification
Topic 1	social media network communication sites theory networks Internet digital support
Topic 2	user evaluation usability interface testing design studies experience interaction methods
Topic 3	design process cognition case collaborative problem knowledge solving practice engineering
Topic 4	mobile input computing speech touch law computer interaction technology movement
Topic 5	web visual search display information semantic displays navigation perception reading
Topic 6	mental cognitive workload team memory attention working measurement work stress
Topic 7	human health ergonomics factors safety error work occupational care construction
Topic 8	learning collaborative education cognitive machine computer motivation technology web-based ICT
Topic 9	musculoskeletal ergonomics back pain disorders low posture upper work assessment
Topic 10	driving driver tracking eye safety control traffic behavior attention distraction
Topic 11	information knowledge management security communication human acquisition behavior technology
Topic 12	design product service engineering experience system quality evaluation method aesthetics
Topic 13	technology perceived Internet acceptance behavior use satisfaction trust theory quality
Topic 14	data visualization visual mining interactive time clustering information spatial series
Topic 15	virtual reality interaction augmented 3D simulation environment environments modeling spatial
Topic 16	hand force tool time ergonomics grip ergonomic surgery tools movement
Topic 17	cognitive handling manual transfer training load function materials lifting control
Topic 18	decision support control system awareness automation making adaptive intelligent situation
Topic 19	work physical stress rate thermal energy heart comfort heat job
Topic 20	interaction user interface human-computer interfaces computer human computing graphical
Topic 21	computer Internet video games use gender differences anxiety social communication
Topic 22	motion muscle fatigue load activity falls electromyography gait kinematics carriage
Topic 23	design research theory methods innovation education history science thinking process
Topic 24	visualization information flow data field uncertainty context security extraction dynamic
Topic 25	vibration response comfort pressure whole-body seat finite posture noise method
Topic 26	3D human modelling body digital anthropometry shape simulation children measurement
Topic 27	image computing visualization optimization surface imaging processing reconstruction remote statistical
Topic 28	design change management simulation risk engineering factor assessment environmental space
Topic 29	design product software development engineering management process architecture new planning

References

1. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: Burlington, MA, USA, 2011.
2. Choudhary, A.K.; Oluikpe, P.I.; Harding, J.A.; Carrillo, P.M. The needs and benefits of text mining applications on post-project reviews. *Comput. Ind.* **2009**, *60*, 728–740. [CrossRef]
3. Jeneanne, R. What is the real value of design? *Des. Manag. Rev.* **2013**, *24*, 30–37.
4. Chai, K.H.; Xiao, X. Understanding design research: A bibliometric analysis of design studies (1996–2010). *Des. Stud.* **2012**, *33*, 24–43. [CrossRef]
5. Friedman, K. Theory construction in design research: Criteria: Approaches, and methods. *Des. Stud.* **2003**, *24*, 507–522. [CrossRef]
6. Liu, S.; Boyle, I.M. Engineering design perspectives, challenges and recent advances. *J. Eng. Des.* **2009**, *20*, 7–19. [CrossRef]
7. Design. Available online: https://en.wikipedia.org/wiki/Design (accessed on 14 April 2017).
8. Gemser, G.; de Bont, C.; Hekkert, P.; Friedman, K. Quality perceptions of design journals: The design scholars' perspective. *Des. Stud.* **2012**, *33*, 4–23. [CrossRef]

9. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

10. Hartigan, J.A.; Wong, M.A. A k-means clustering algorithm. *Appl. Stat.* **1979**, *28*, 100–108.

11. Heinrich, G. Parameter Estimation for Text Analysis. Available online: http://www.arbylon.net/publications/text-est2.pdf (accessed on 14 April 2017).

12. Singh, V.K.; Uddin, A.; Pinto, D. Computer science research: The top 100 institutions in india and in the world. *Scientometrics* **2015**, *104*, 529–553. [CrossRef]

13. Sitarz, R. Identification of Research Trends in the Field of Separation Processes. Application of Epidemiological Model, Citation Analysis, Text Mining, and Technical Analysis of the Financial Markets. Available online: http://www.doria.fi/bitstream/handle/10024/93330/isbn9789522654847.pdf?sequence=2 (accessed on 14 April 2017).

14. Saffer, D. *Designing for Interaction: Creating Innovative Applications and Devices*; New Riders: Berkley, CA, USA, 2010.

15. Garrett, J.J. *Elements of User Experience: User-Centered Design for the Web and Beyond*; New Riders: Berkley, CA, USA, 2010.

16. Reed, M.P. Modeling ascending and descending stairs using the human motion simulation framework. *SAE Tech. Pap.* **2009**. [CrossRef]

17. Otto, K.N.; Wood, K.L. *Product Design: Techniques in Reverse Engineering and New Product Development*; Prentice Hall: Upper Saddle River, NJ, USA, 2003.

18. Horn, R.E. Information design: Emergence of a new profession. In *Information Design*; MIT Press: Cambridge, MA, USA, 1999; pp. 15–33.

19. Chung, M.; Munno, G.J.; Moritz, B. Triggering participation: Exploring the effects of third-person and hostile media perceptions on online participation. *Comput. Hum. Behav.* **2015**, *53*, 452–461. [CrossRef]

20. Buccafurri, F.; Lax, G.; Nicolazzo, S.; Nocera, A. Comparing twitter and facebook user behavior: Privacy and other aspects. *Comput. Hum. Behav.* **2015**, *52*, 87–95. [CrossRef]

21. Stiegler, H.; de Jong, M.D. Facilitating personal deliberation online: Immediate effects of two considerit variations. *Comput. Human Behav.* **2015**, *51*, 461–469. [CrossRef]

22. Sun, J.N.; Hsu, Y.C. A study of learners' perceptions of the interactivity of web-based instruction. In Proceedings of the International Conference on Human-Computer Interaction 2007, Beijing, China, 22–27 July 2007; pp. 351–360.

23. Nogueira, P.A.; Teófilo, L.F.; Silva, P.B. Multi-modal natural interaction in game design: A comparative analysis of player experience in a large scale role-playing game. *J. Multimodal User Interfaces* **2015**, *9*, 105–119. [CrossRef]

24. Brown, E.T.; Ottley, A.; Zhao, H.; Lin, Q.; Souvenir, R.; Endert, A.; Chang, R. Finding waldo: Learning about users from their interactions. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1663–1672. [CrossRef] [PubMed]

25. Zhang, F.; Kaufman, D. The impacts of social interactions in mmorpgs on older adults' social capital. *Comput. Hum. Behav.* **2015**, *51*, 495–503. [CrossRef]

26. Yoon, S.Y.; Choi, Y.J.; Oh, H. User attributes in processing 3d vr-enabled showroom: Gender, visual cognitive styles, and the sense of presence. *Int. J. Hum. Comput. Stud.* **2015**, *82*, 1–10. [CrossRef]

27. Fazal-e-Amin. Characterization of web browser usage on smartphones. *Comput. Hum. Behav.* **2015**, *51*, 896–902.

28. Wang, J.L.; Wang, H.Z.; Gaskin, J.; Wang, L.H. The role of stress and motivation in problematic smartphone use among college students. *Comput. Hum. Behav.* **2015**, *53*, 181–188. [CrossRef]

29. Wilson, J.R. Fundamentals of systems ergonomics/human factors. *Appl. Ergon.* **2014**, *45*, 5–13. [CrossRef] [PubMed]

30. Carayon, P.; Wetterneck, T.B.; Rivera-Rodriguez, A.J.; Hundt, A.S.; Hoonakker, P.; Holden, R.; Gurses, A.P. Human factors systems approach to healthcare quality and patient safety. *Appl. Ergon.* **2014**, *45*, 14–25. [CrossRef] [PubMed]

31. Lucas-Cuevas, A.G.; Perez-Soriano, P.; Priego-Quesada, J.I.; Llana-Belloch, S. Influence of foot orthosis customisation on perceived comfort during running. *Ergonomics* **2014**, *57*, 1590–1596. [CrossRef] [PubMed]

32. Wang, T.C.; Chuang, L.H. Psychological and physiological fatigue variation and fatigue factors in aircraft line maintenance crews. *Int. J. Ind. Ergon.* **2014**, *44*, 107–113. [CrossRef]

33. Ferguson, S.A.; Allread, W.G.; Le, P.; Rose, J.; Marras, W.S. Shoulder muscle fatigue during repetitive tasks as measured by electromyography and near-infrared spectroscopy. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2013**, *55*, 1077–1087. [CrossRef] [PubMed]

34. Jeon, M.; Walker, B.N.; Gable, T.M. The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. *Appl. Ergon.* **2015**, *50*, 185–199. [CrossRef] [PubMed]

35. Özcan, A.C. An overview of the early foundations and development of contemporary industrial design in turkey. *Des. J.* **2009**, *12*, 267–287. [CrossRef]

36. Grierson, H. Detailed empirical studies of student information storing in the context of distributed design team-based project work. *Des. Stud.* **2013**, *34*, 378–405. [CrossRef]

37. Kalay, Y.E. The impact of information technology on design methods, products and practices. *Des. Stud.* **2006**, *27*, 357–380. [CrossRef]

38. Nielsen, S.L.; Christensen, P.R. The wicked problem of design management: Perspectives from the field of entrepreneurship. *Des. J.* **2014**, *17*, 560–582. [CrossRef]

39. Vallet, F.; Eynard, B.; Millet, D.; Mahut, S.G.; Tyl, B.; Bertoluci, G. Using eco-design tools: An overview of experts' practices. *Des. Stud.* **2013**, *34*, 345–377. [CrossRef]

40. Emerson, J. Visualizing Information for Advocacy: An Introduction to Information Design. Available online: https://msu.edu/~jmonberg/415/Schedule_files/infodesign.pdf (accessed on 14 April 2017).

41. Claessen, J.H.; Van Wijk, J.J. Flexible linked axes for multivariate data visualization. *IEEE Trans. Vis. Comput. Graph.* **2011**, *17*, 2310–2316. [CrossRef] [PubMed]

42. Fischer, F.; Fuchs, J.; Mansmann, F.; Keim, D.A. Banksafe: Visual analytics for big data in large-scale computer networks. *Inf. Vis.* **2013**, *14*, 51–61. [CrossRef]

43. Maguire, E.; Rocca-Serra, P.; Sansone, S.A.; Davies, J.; Chen, M. Visual compression of workflow visualizations with automated detection of macro motifs. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2576–2585. [CrossRef] [PubMed]

44. Jakobsen, M.R.; Hornbaek, K. Transient or permanent fisheye views: A comparative evaluation of source code interfaces. *Inf. Vis.* **2011**, *11*, 151–167. [CrossRef]