Article

Implementation for Comparison Analysis System of Used Transaction Using Big Data

Byungjoon Park [1], Hasung Kim [2] and Byeongtae Ahn [3],*

[1]   Computer Science and Engineering, Sejong University, Seoul 04997, Korea; 1byungjoon412@sju.ac.kr
[2]   IT College, Suwon University, Seoul 04997, Korea; 13053019@suwon.ac.kr
[3]   Liberal & Arts College, Anyang University, Gyeonggi-do 13992, Korea
*   Correspondence: ahnbt@anyang.ac.kr

check for
updates

Abstract: With the recent increase in used trading sites that support used trading, users want to find various information in real time, and the development of the Internet consists of direct and indirect connections between businesses and consumers. This change created a new type of C2C (Commerce to Commerce) transaction. However, each used trading site has its own characteristics, making it difficult to standardize one. Therefore, in this paper, we construed a system that provides the user's used transaction data in real time and provides the desired information quickly. In this paper, we developed the crawler system needed to develop an integrated transaction system for second-hand goods through Internet e-commerce transactions, defined morphological analyzers, and described the service that users can employ in the web environment by using the system developed in the paper.

Keywords: big data; e-commerce; data analysis; used transaction; block chain

1. Introduction

A surge in Internet use has enabled online-based e-commerce in the form of B2C (Business to Commerce). The development of the Internet established a direct connection between businesses and consumers, and the link between consumers. This change created a new type of C2C transaction [1]. With these changes, consumers play the role of consumers as well as sellers online, and the scope of consumers is gradually expanding. Therefore, the way consumers purchase products is becoming more diverse, and the number of buyers using online trading sites for used products is increasing steadily [2]. In particular, the economy of second-hand goods is growing rapidly due to low growth and a prolonged downturn in consumption [3]. The CCSI (Consumer Composite Sentiment Index) is declining significantly, while the volume of used transaction sales is growing exponentially.

The second-hand market has a lot of transactions between individuals, so it is difficult to aggregate the size of the market accurately. However, the distribution industry estimates that the size accounts for around KRW 20 trillion, except for the used-car market, a typical second-hand market [4]. In this regard, it is considered that the market has a high potential, and there are more moves to generate new revenue in the market by strengthening services, such as by launching mobile apps and inspecting products.

Due to this trend, large platforms such as Gmarket and Auction, which have facilitated the previous online e-commerce transactions, developed the service for the second-hand market. As such, many platforms for the market have emerged, such as Dang-geun Market, which uses GPS (Global Positioning System) information to provide local-based second-hand products [5].

The following inconveniences also arise due to the increase in the size of second-hand market transactions and the advent of various trading services:

- Products distributed by various second-hand transaction services;

- It is difficult to compare the prices of the same products on various platforms;
- It is difficult for an individual to measure a product by placing one product on different platforms at different prices.

The analysis of various second-hand transaction services before this paper found that the most significant problem was the spread of goods by service. For instance, when searching for MacBook Pro 2019, one product was found only. Besides this, the price distribution had an average difference of KRW 100,000 for each service, and there was a case in which the price was different by up to KRW 300,000. To solve this problem, I established an integrated platform to view all second-hand transactions on a single website [6].

The primary business model of the system stores products from used trading sites on an integrated platform and displays them to users. Users are connected to the relevant used trading site through a link. In addition, users' search records and visit records are used later [7]. The system allows you to find all the goods from several used trading sites in one search. Therefore, it is easy to compare the prices of goods, and used trade sales were activated. As a result, the related product recommendation function and commission revenue generation also occurred [8].

A variety of problems have arisen due to a vacuum in the legal status of online second-hand traders. Nevertheless, the size of the online second-hand market is expanding day by day, and you can see the form of consumption as having unlimited market growth and potential for development in the future [9]. Therefore, in line with this growth, we have prepared a way to promote the transactions in the used market and meet the needs of consumers. In addition, used site data, which is not categorized, is classified into a product list desired by the user.

The system crawls the web crawling instance by searching for used trading sites, and classifying and storing the data crawled by the data processing module. Crawling was limited to laptops, PCs, refrigerators and TVs. The crawler bot was developed and automated by itself. The amount of crawled data was 600,000 per day, and it crawled three on sites.

Currently, many platforms for used transactions have been developed and used. However, no system provides standardized information by crawling data from various platforms in real time. In this paper is described a service that provides standardized information by web crawling data scattered on various platforms in real time.

Existing systems only provide product information for used transactions. However, this system supports Multi-platform Search by collecting data from all portal systems for used transactions through robot engines. In addition, detailed condition searching, which is not provided in the existing system, is supported. Therefore, this paper supports Specific Options Search and Multi-platform Search, which are not provided by the existing system.

Section 2 of this paper compares and analyzes standards and related research for the composition of web services, and Section 3 presents cases at home and abroad. Section 4 proposes the configuration of the second-hand transaction-integrated platform service. Section 5 develops the system based on the system configuration diagram. Section 6 validates the operability of the system designed in this paper, and evaluates whether functional requirements are satisfied through comparison with other services previously employed. Finally, Section 7 presents the conclusion.

2. Related Research

The method proposed in this paper has developed a service that allows users to inquire about and search for goods on an integrated platform after crawling them into other second-hand transaction web services. An average of over 600,000 data was processed per day to implement a service that provides real-time product information to consumers [10].

To develop the system, clients and servers of the web system were configured using JavaScript language, and data analysis was implemented using Python language. Finally, MongoDB was used to support the JSON type as a database [11].

2.1. Web System Configuration

JavaScript was used to develop the web system in this system. JavaScript language was developed by Google in 2008 to improve performance with the advent of the V8 JavaScript engine, and they later attempted to develop JavaScript in various ways, including the use of Ajax to develop the web.

Advances in these technologies have led JavaScript, which has performed the ancillary function of data verification, to play a central role in system architecture. JavaScript libraries were used to enable the implementation of system logic within browsers and to solve problems caused by the operating web servers described in the introduction.

The client-based web system configuration method used servers in a server-based architecture in Node.j, which can operate on low-specification servers to replace web application servers such as Apache Tomcat, and web servers as Nginx [12].

Web application used View (Vue.js), a type of open-source progressive JavaScript framework utilized to create a user interface. View is very small, light and less complex than other JavaScript frameworks such as Angular, Backbone and React. It also has a structure that can be adopted gradually. The core library focuses on declarative rendering and component configuration, and can be embedded on existing pages. Advanced functions required for complex applications, such as routing, state management and build tools, are provided through officially maintained support libraries and packages.

View (Vue.js) uses a Model-View-ViewModel (MVVM) pattern unlike the Model-View-Control (MVC) pattern used by other common JavaScript frameworks. The MVVM pattern is a variation of the MVC pattern, and the key is to create an abstraction of View. The abstraction of View is reusable and easy to test [13].

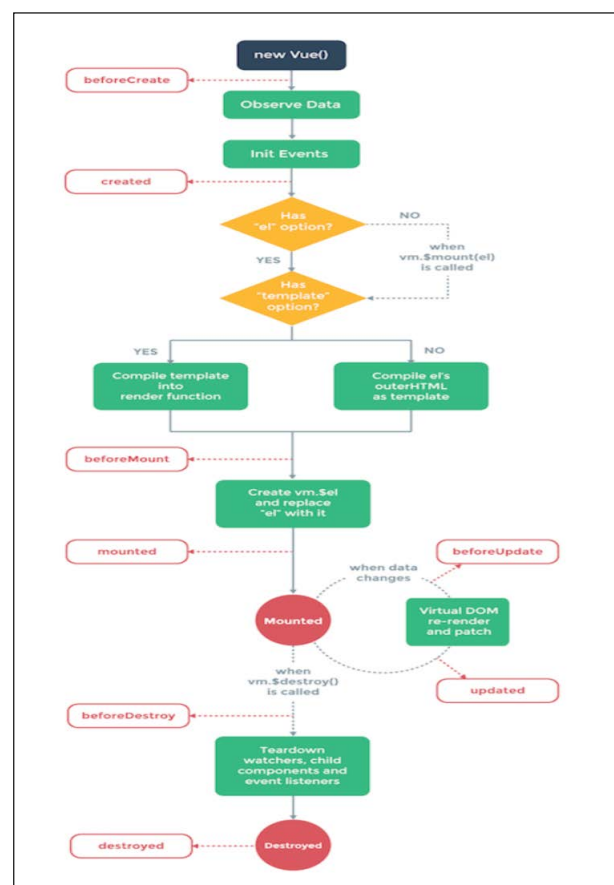Figure 1 below shows the tasks handled by the view model instance:

Figure 1. Flow Chart of View Model Instance.

2.2. JavaScript

In the traditional web page development method, JavaScript was used, based on HTML files, but in this paper, JavaScript libraries were executed as the basis. Then, the system combines the required libraries into JavaScript files for creating web pages, allowing them to share data, and it creates web pages that load HTML data that exist in template files to offer to users. This eliminates the need for the additional loading of individual libraries, thereby enabling the sharing of data and allowing web pages to be changed only by changes to template files [14]. This structure facilitates the combination of multiple JavaScript libraries, and allows them to operate within a single framework.

Figure 2 shows the overall conceptual diagram of the system. JavaScript was used throughout the system and was supported by the server itself.
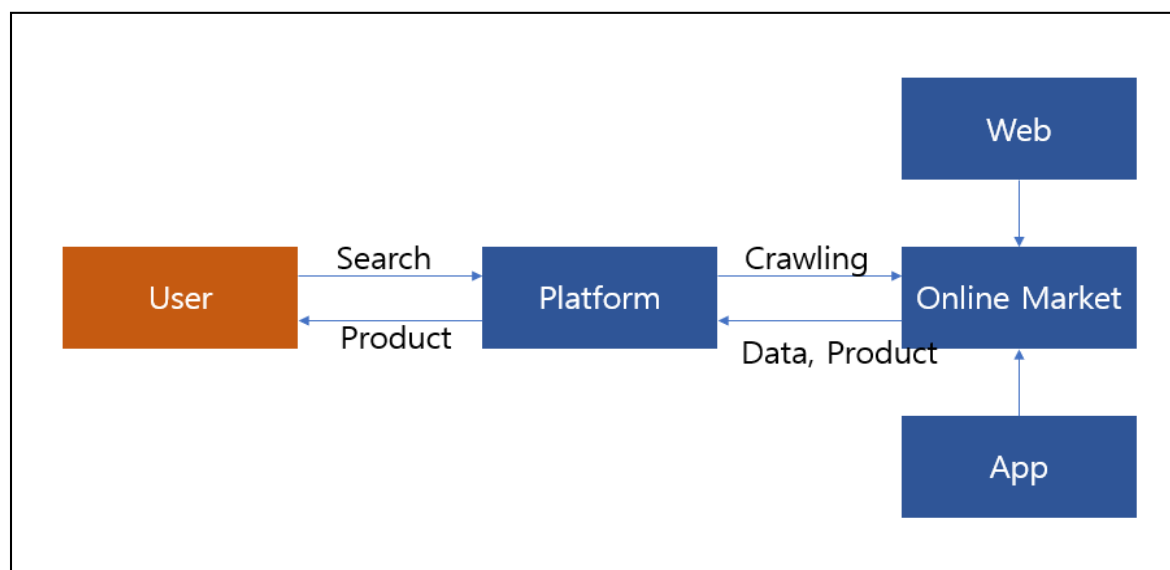

Figure 2. Concept Diagram of System.

2.3. Crawling Implementation

Recently, the popularity of and demand for big data-related jobs have also been on the rise. The range of data collection, which is a material for data analysis, varies greatly depending on capability. A typical TextMining process starts with collecting analysis data, and visualizes the results of the analysis easily through the pre-treatment process and noun treatment process to extract meaningful words only, and remove unnecessary words [15].

Even though there would be no problem if there was account information from which one could quickly get data with a single line of instructions by accessing the database of the companies, in general, corporate data is protected for security and customer information protection. Furthermore, these data have various characteristics.

The first characteristic is that big data means the collection of data in PeataByte or ExaByte, which are beyond GigaByte and TeraByte in terms of volume. The actual data size and loadable data range of the main memories that can be processed by the computer are significantly different. In this regard, most studies handle the data by dividing the available capacities of the main memories.

The second characteristic is the vast data processing speed (velocity), beyond the data size. This indicates that the rate of data analysis is very fast, as Internet users who generate actual data are collected and processed in real time after preprocessing to obtain quality information. This background is caused by the increased prevalence of mobile devices, network speed and computer power that enables real-time analysis. What large computers and supercomputers can process in the past can now be handled by PC-level servers.

The third characteristic is that it collects various data on computers. In other words, unstructured data, such as reviews, sensing data generated by multiple sensors and video data from photos and videos, are studied through various analysis methods, including the analysis of traditionally structured data from relational databases in the past [2–4]. Data collection, a tool for data analysis, is a significant challenge in a society wherein such characteristics have been used to analyze various forms of mass data quickly. The technology for the automatic collection of information from a web environment is called web crawling. The importance of web crawlers is highlighted by the increased contents of various media forms, such as social media.

This paper seeks to establish a second-hand product crawling system. It also intends to open up the source codes in the environment developed by the open-source software Python, and even build web pages that automatically crawl actual data by accessing the real trading platform for second-hand goods [16].

Development software tools have a number of programmable languages, such as Java, C, C++, C#, Python, JavaScript, VisualBasic.NET, R, and PHP. However, the language that has recently become the most popular is Python. According to the Popularity of Programming Language (PYPL) published by Github.io in September 2018, Python has become the most popular language in the programming world in the last five years, with the greatest loss seen for PHP. The actual indicators are shown in Figure 3 below; while Python grew 14.5%, PHP decreased 6.5%. In addition, Python grew by 5.7%, compared to 2017 [17].
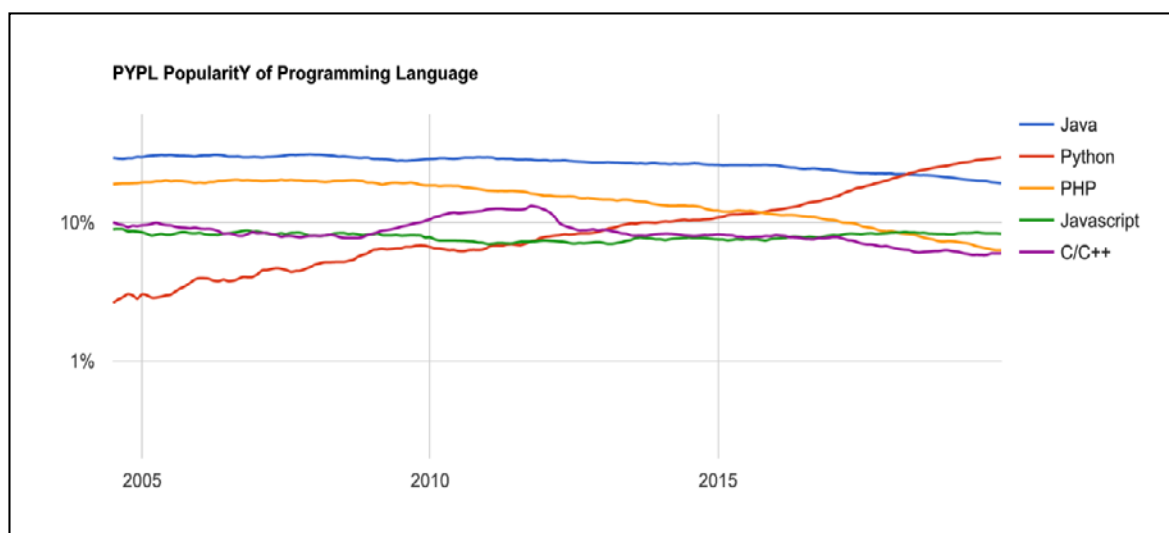

Figure 3. Indicators of Popularity of Programming Language.

Python has the advantage of having a simplified grammatical structure over C and Java, making it easier for users to learn quickly and easily. In real coding, it is made highly legible by using "indent" to distinguish repetitive statements and conditional statements, and can be easily implemented by providing a large number of libraries.

Python has two representative libraries that offer web crawling, including Beautifulsoup for static crawling and Selenium for dynamic crawling.

Beautifulsoup is a Python-based library that makes it easy to import data from HTML or XML, which are imported from web pages. Data contained in HTML or XML are helpful for extracting data to a specific format. Selenium is a web test framework. HTML and XML using dynamic data can be imported by using Chrome or Internet Explorer, which are available for the Selenium framework.

Second-hand product data are a dynamic web page, but there is no need to use Selenium because static HTML and XML are provided to users. Since the Internet has to be brought up virtually and continuously due to the characteristics of the Selenium Library, the loading time here is longer

than with Beautifulsoup. Therefore, due to the nature of this paper, which requires the updating of extensive data in real time to show second-hand goods to consumers, this was conducted through the Beautifulsoup library.

2.4. Data Analysis Implementation

Individual consumers write most of the texts on second-hand goods, so they are usually unstructured text. In this regard, a natural language processing technique is needed to process these texts. In order to identify technical entities in postings written in Korean, it is necessary to consider the features of the Korean language that have the characteristics of agglutinative language. Besides, the performances of Korean morpheme analyzers (such as OKT, MeCab, Komoran, and KKma) applicable to second-hand goods were compared and analyzed in this paper, because a rapid processing speed appropriate for real-time service was needed [18]. Figure 4 displays the results in diagram form. The left numbers represent the time in seconds.
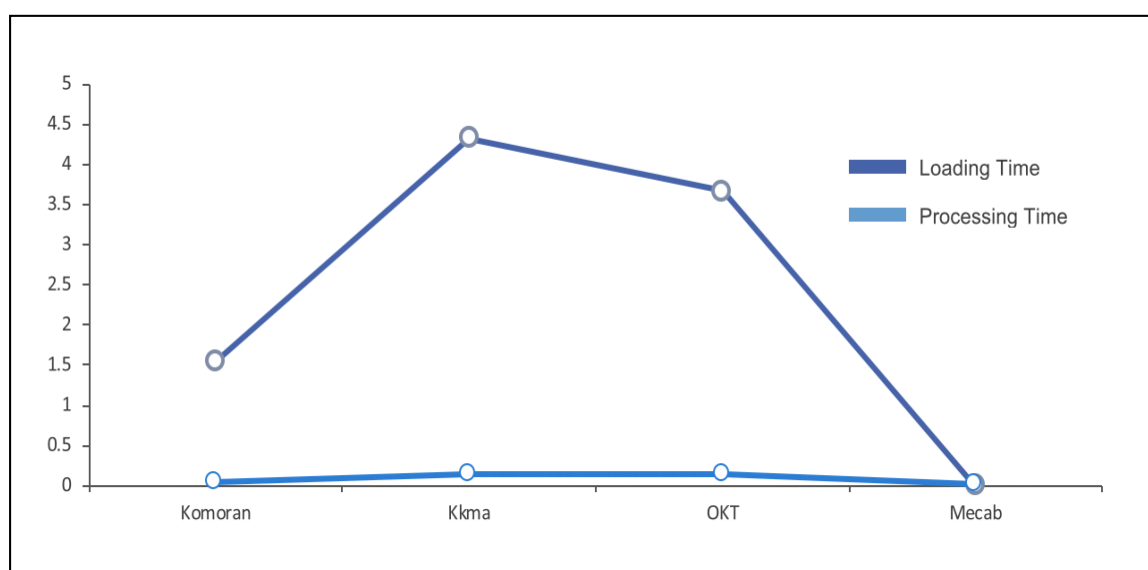

Figure 4. Morphological Analyzer Performance Comparison Graph.

The experiment showed the average library loading time and processing time for postings on a second-hand product. The *x*-axis is the library of each morphological analyzer, and the *y*-axis is the time for part-of-speech tagging and library loading.

The system developed in this paper is a system that needs to update data in real time to show to users, and the processing speed is most important. As a result of comparing and evaluating the performances of the morphological analyzers, MeCab showed an overwhelmingly fast processing speed, and it was found that the analyzer has similar results compared to other analyzers in comparing extracted index words [19]. However, as with other analyzers, they are not able to properly analyze newly coined words, or sentences or proper nouns. Figure 5 shows the MeCab analysis results.

The required data was "맥북에어 2018 256기가" (Macintosh laptop year 2018 & HDD 256 GB), but the words, "-에 and "-어" are categorized as adverbs and verbs respectively because of the characteristics of Korean grammar. Mecab is suitable for most search engines and algorithms, but it is not suitable for the used product data search and analysis used in this study.

맥북에어 2018 256기가 판매 합니다.

Mecab

[
('맥', 'NNG'), ('북', 'NNG'), ('에', 'JKM'), ('어', 'VV'), ('어', 'ECS'),

('2018', 'NR'), ('256', 'NR'), ('기가', 'NNG'), ('판매', 'NNG'), ('하', 'XSV'),

('ㅂ니다', 'EFN'), ('.', 'SF')
]

Figure 5. MeCab Analysis Results.

2.5. Database

The traditional method uses the storage of structured data in general. However, since big data environments need to be able to store large amounts of data and have unstructured data, they need new technologies other than traditional data storage and management technologies. To resolve this matter, the NoSQL (Non relational Structure Query Language) database has emerged as an alternative to conventional databases [11].

NoSQL is a new type of database that does not define relationships between data and does not have a fixed schema, and it is a lightweight database of RDBMS (Relational DataBase Mamagement Systme). The NoSQL system is referred to as DBMS in all other forms that are created by removing the features of relational databases. This is also called "Not Only SQL" in terms of highlighting the fact that SQL query languages are available.

The system developed in this paper used both NoSQL and RDMBS. The data storage of products used NoSQL, and other systems such as members used RDBMS. The product data on trading websites of second-hand products are all unstructured data, and there is no need to join tables between products. However, the system was developed using NoSQL because large data needs to be processed quickly.

The big data analysis system must have a large capacity for information storage, as well as a fast information collection and processing speed. Therefore, in this paper, it is more suitable to use NoSQL, such as Mongo DB and Casandra DB, than RDBMS, such as Oracle DB and PostgreSQL. NoSQL is easy to store and processes large amounts of data by distributing them to multiple servers. Furthermore, in terms of server processing power, a scale-out method that distributes processing through multiple servers is more suitable than scale-up. Therefore, in the paper, NoSQL suitable for a big data analysis system was used.

3. Cases

3.1. Domestic Cases

There are similar domestic cases to this system, including Joonggonara. The following are representative cases similar to the system.

3.1.1. Joonggonara

Joonggonara is the largest trading platform for second-hand products in Korea, with a total of 21 million members. Until 2015, it did not have its own platform, and had operated using the café service of Naver, Korea's largest portal service company. It developed its own app and introduced the app through product sharing between platforms.

3.1.2. Dang-geun Market

Dang-geun Market was launched in 2015 with the model of a direct transaction market for used products near you. Unlike previous trading platforms for second-hand products, this service is based on the location of users [20]. Registering an area where users reside will allow them to identify second-hand items that are traded in real time in the area. The indicator of "manner temperature", to prevent impolite behavior in a transaction, can confirm the reliability of users. If a professional business wants to promote its products, the business can register as a local company and advertise itself in the area.

3.1.3. Bunjang

In 2010, Bunjang launched its mobile app, which was faster than Joonggonara. It includes a service used by professional shops and merchants, including individual sellers. This service has increased accessibility through "Lightning Talk", which allows users to chat within the app. It targeted the needs of users who were reluctant to disclose personal information, such as names and phone numbers [21]. It also developed a system whereby users can check each other's personal information when making a purchase, and introduced a safe transaction system for used products for the first time in Korea.

3.1.4. Danawa

In 2000, Danawa started with a service that offers a price comparison service for digital cameras. Since then, the service has also begun to provide information on computer parts, cars and second-hand products. It contains all the market prices of the open markets that consumers usually use, making it easier for users to figure out the costs before purchasing the products. Unlike other open markets, it has a specific filtering function, so users can find the product that they want to buy.

3.1.5. Naver Shopping

Naver Shopping is a product search and price comparison service of Naver, Korea's largest portal site. In addition to products from the existing open market, the service offers a smart store service where users can sell goods by entering the Naver Store. The number of smart store businesses increased from 100,000 in 2016 to 240,000 in 2018, and the turnover reached KRW 10 trillion in 2018.

3.2. Overseas Cases

The followings are representative cases similar to the system developed in this paper.

3.2.1. Craigslist

This site provides not only second-hand goods, but also houses and job postings. It started its first service in San Francisco in 1995 and expanded the service to other U.S cities in 2000, and now operates in 50 countries [22]. Users can use the service in to find desired items by selecting an area and searching for items in the area.

3.2.2. Amazon

Amazon provides users with a system that resells and discounts like-new, open-box and pre-owned products returned by buyers. Buyers can purchase products with more confidence, as such products are sold by the global giant company Amazon. A 30-day return guarantee is provided.

3.2.3. eBay

As a multinational e-commerce company, eBay brokers C2C and B2C sales. This company provides auction-style sales and instant buying-style sales. Buyers do not need to pay their charge, but sellers have to pay the commission when they sell more than a certain number of goods [23].

4. System Configuration Diagram

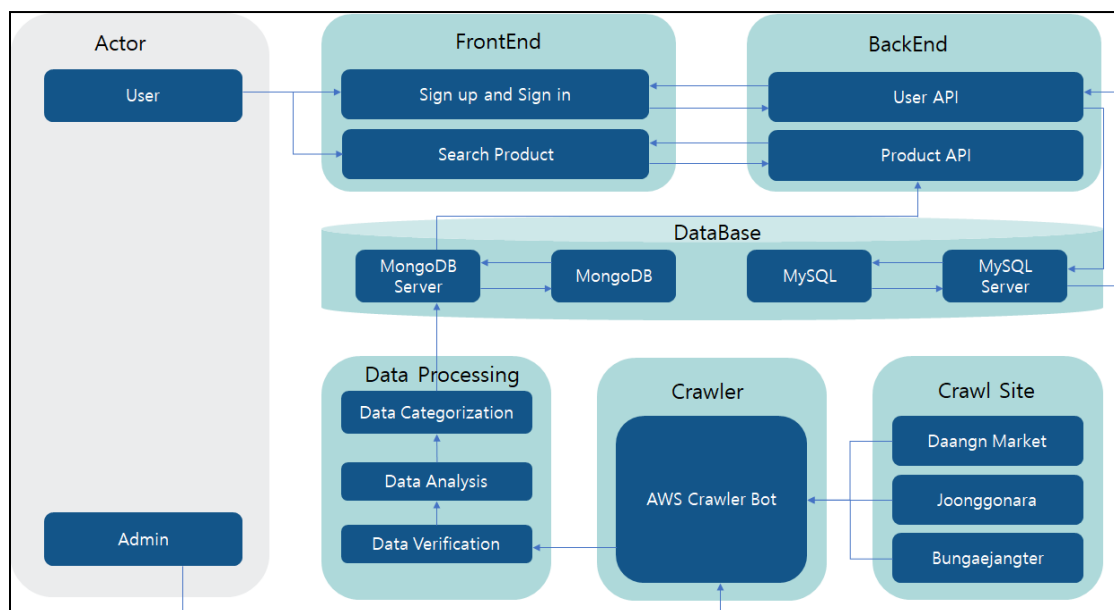Figure 6 shows the first-stage system configuration of project planning.

Figure 6. System Structure Diagram.

The system developed herein collects data by invoking crawler bots at three sites at regular intervals. It performs data analysis after validating the data. In this process, predefined categories are classified into each category, and necessary information is extracted and stored in the database.

Figure 7 indicates the overall data flow diagram of the system. Specific explanations are as follows.
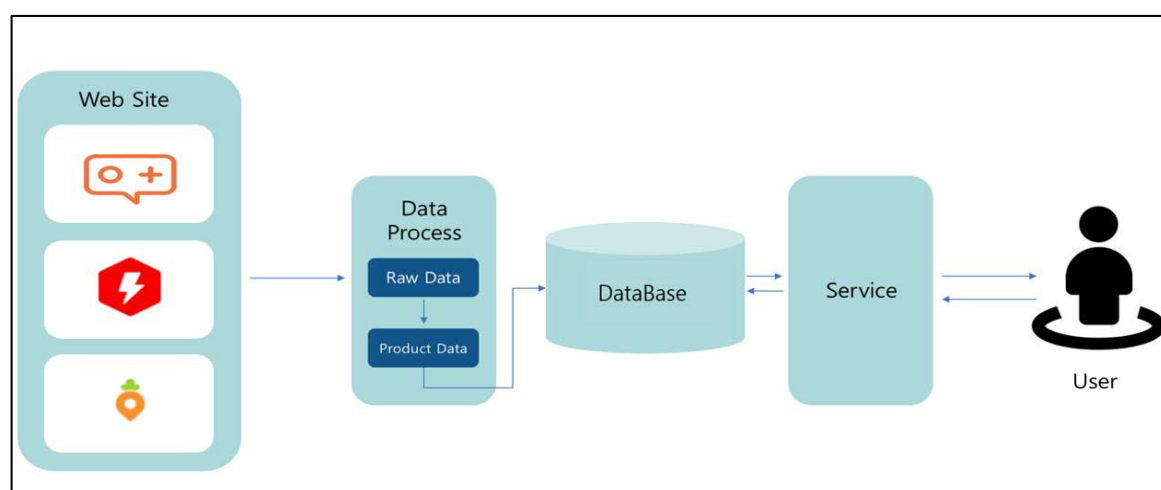

Figure 7. Portal System Flow Diagram.

4.1. Crawler Bot

Crawler bots crawl each site every minute and import data from the previous postings to the most recent postings. One crawler bot is responsible for a single website and records and terminates the most up-to-date information. The crawler bot first checks for valid data to automatically delete information about purchases and free shares. Therefore, it does not pass all of the crawled text to the data processing process.

4.2. Data Processing

The data processing process filters out unnecessary data through the further verification of primary filtered data, and classifies the data into each category by matching the data set for classification. The categorized items store meaningful information and raw data in MongoDB by extracting detail options.

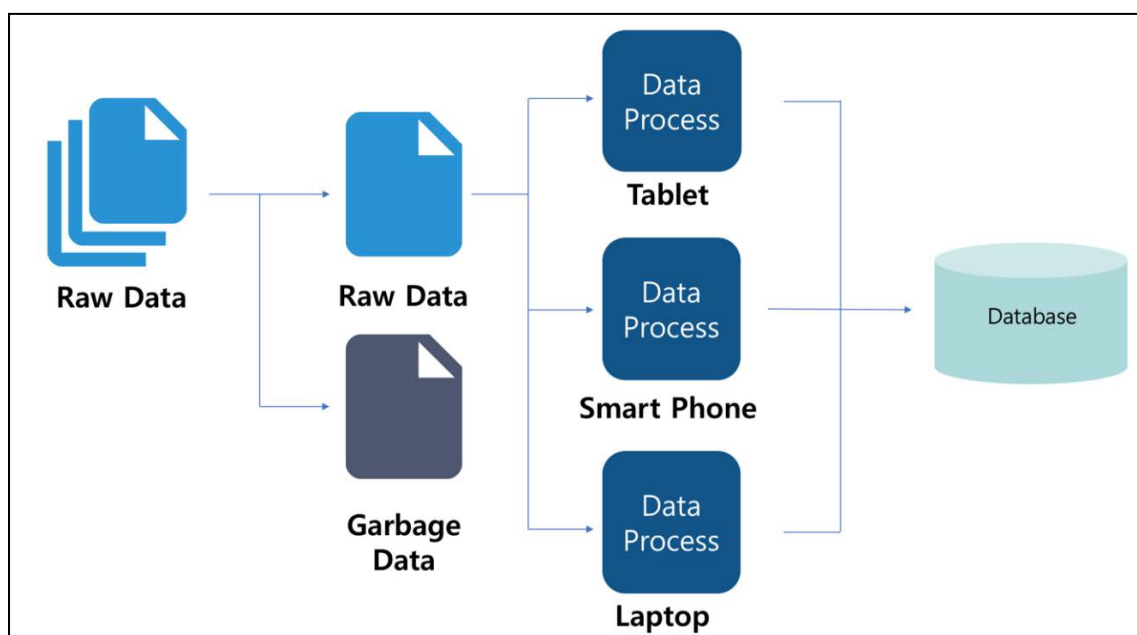Figure 8 indicates the configuration of the data storage system.

Figure 8. Data Storage System.

4.3. Database

The system allows data storage in order to extract many options, and expands fields through further development after storing primary data. However, if new data are stored every hour, the RDBMS experiences performance degradation or instability problems [6]. Therefore, the system uses MongoDB, the NoSQL database that manages documents in JSON format. MongoDB solves the performance degradation of the JSON-type documents in the collection unit, rather than in the table of RDBMS. DB, which manages user information, uses MySQL, the existing RDBMS.

4.4. Server

Node.js was used to create REST API servers [9,10]. The system uses this framework because many low-cost activities import data from DB. As it often reads the list of posts, data in JSON format was used.

4.5. Front

Vue.js was used to develop the front part of the web. This allows users to search for specific options by selecting the category of products that they want to find, and then choosing the desired specifications and price range.

5. Implementation

Based on the design of this system, AWS was used as a server to operate the system. Mecab was used for crawling and data processing, and MongoDB was utilized for the efficiency of data storage.

5.1. AWS

AWS was used as a server to operate the system. AWS consists of three types: the instance for crawling and data processing, the DB instance, and the instance for web and api. The crawling instance consists of instances where both network performance and computer power are high because it requires to crawl websites and process data.

5.2. Crawler Bot

A crawler bot crawls data at each site for second-hand product transactions every minute [11]. At the end of the URL, a post number is placed, and each post has a unique number, which increases with the most recent information. In other words, the system can import the most recent post by increasing the number, and it stores the number by deciding it is the most recent posting and stopping the crawler bot if it is not able to import the posting repeatedly. When the next crawler robot is activated, it starts crawling again based on the number. To reduce throughput in the process of data processing, the primary data to be imported during crawling are filtered. The posts on second-hand product trading are classified as purchases because they include texts for purchase and free sharing, as well as for sale. Further, the text aimed at free sharing, with the sales amount "0", is not handled in the process. This also imports information on the title, text, preparation time and URL of the postings, and provides price information because most of the sites specify the price separately [24].

5.3. Data Processing

This is the process of dividing postings into each category and extracting specific options for products based on the raw data that have been crawled out. Since 600,000 new data are uploaded per day from a crawled site, and these texts are collectively posted at a certain time, it is necessary to crawl data quickly. To handle this matter, the system used the most efficient Mecab among other analyzers.

Komoran, Kkma and OKT were excluded from the review due to their long recall time, in spite of their proper data processing speed [25]. Mecab offers fast loading and processing, but has a problem with breaking down morpheme. To solve this problem, I developed a method to use the Mecab in the existing way in order to extract the words needed to filter data for significant products into the categories to be classified after crawling.

There are several types of garbage data defined in this paper. Articles for advertisement, articles for product purchase, articles for fishing and duplicate articles with the same content are garbage data. Sites that crawl and collect data have portals or platforms for used trading sites, but the amount of data is large because community sites such as social networks also collect data. In the case of the used product data of social network service sites, in particular, much garbage data come out. Due to the nature of social network services, the more recommended they are by other users or the more views, the more inflow of other users, so there are many posts that try to attract attention by exploiting this, and in order to sell their products faster and expose more, they constantly upload the same products. This effects not only the reliability of the site, but is also a n unnecessary waste of resources. Various methods were used to filter and process these garbage data [26]. First, the title of the used product posting data was hashed and saved in the DB. This is implemented so that users who use the

site cannot see duplicate data by comparing the hash value in the DB when a product from the same post comes in, and removing the previously stored data if they exist in the DB. Secondly, price data with non-outlier values were separated. These data were removed because most of the articles that attract the users' interest rather than selling products do not write normal prices. Finally, there are cases in which the number of views and recommendations of products is manipulated by using a specific company. Since the system described in this paper fetches products every minute in real time, the number of views and recommendations is meaningless, so the data are not collected together. In this way, much garbage data were also removed.

5.4. MongoDB

NoSQL was used, which is less restrictive than relational databases because there is no need to join between each category, and the information to be stored for each data is flexible. After the above data processing process, the system stores the data in the collection that fits each category. When additional options are extracted in the process of data processing, they are efficient because they do not have to modify the overall database structure or previous data, but only need to add new documents to the store [13].

5.5. Node.js

Node.js was used to handle many requests because its main purpose is to deliver data requested by the web without the high cost. Furthermore, many packages can be used to reduce development time.

5.6. Vue.js

Vue.js was used because codes that had been processed on the server side have recently been handled in browsers, and it can be managed systematically by using a framework. Because the overall format for each category is similar and only the layout for selecting detailed options is different, it was simple to implement using Vue.js.

Figure 9 is the main screen of the system, and has been arranged for users to have quick access to the categories they use. The categories are classified into electronic devices, household appliances, and kitchen appliances. Electronic devices consist of laptops, smartphones, and tablets. Household appliances consist of a TV, air conditioner, and refrigerator. Kitchen appliances consist of an electric rice cooker, microwave, and induction.

Figure 10 shows the selection screen of the notebook category option.

After accessing the category of products that the user wants to purchase, they can easily find items by selecting specific options and searching for keywords. If you click on the notebook of the electronic device, detailed values for the manufacturer, CPU, RAM, HDD, monitor size, and price can be displayed.

Figure 11 represents a list of products searched. The system outputs a list of products according to the options selected by users. It displays the title, text, price, date and posted site of the postings, and outputs them from the latest one according to the posted dates. It provides information on buyers and sellers, as well as reliable step-by-step site information.

In the case of the service described in this paper, when the product URL is retrieved and the product is clicked, it is sent to the URL with the product. The downside of this method is that if the user who posted the product deletes the product posting, they cannot access it. If a user deletes a post even after collecting the product, dummy data are accumulated in the DB, which wastes unnecessary resources. To prevent this, we check whether the URL is valid when accessing the site, and if the URL is valid, we connect to the location where the product is located. If the URL is not valid, the data are deleted from the DB.
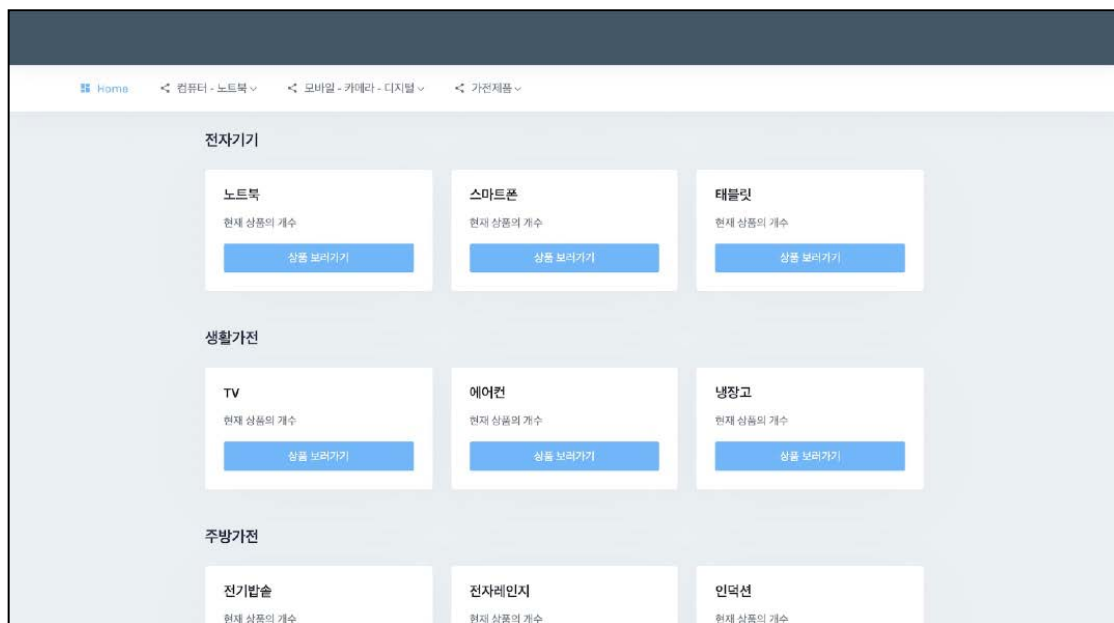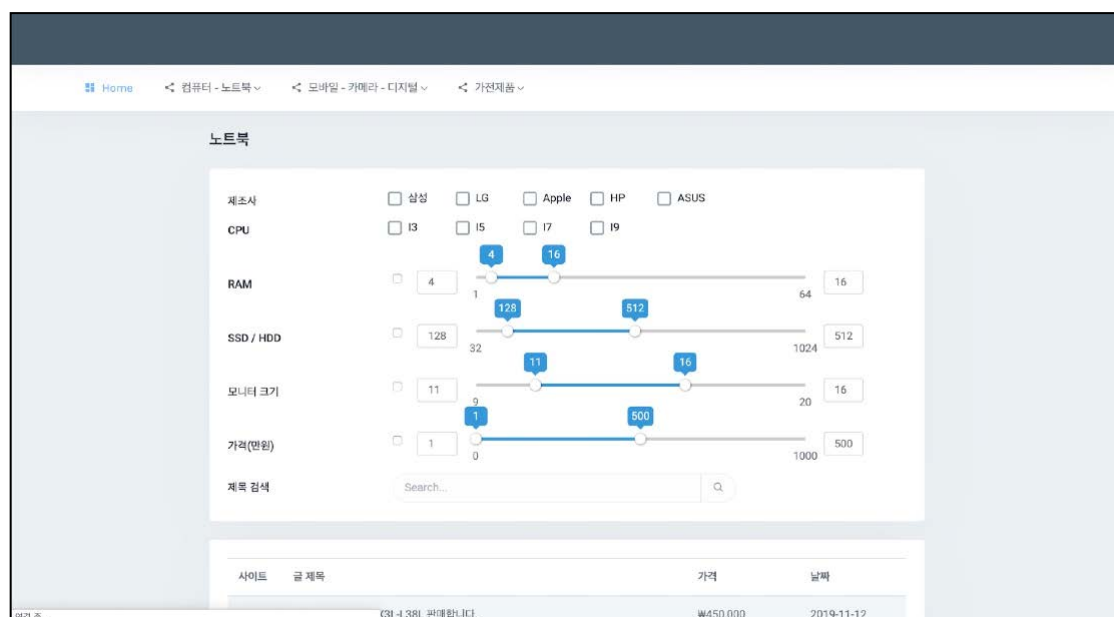
Figure 9. Main Interface of System.

Figure 10. Notebook Category Option.

Figure 11. Product List after Filtering.

6. Service Comparison

　　This paper implemented the system and concretely described each specific matter so far. Section 6 compares the service developed in the paper with other trading services for second-hand goods at home and abroad. Table 1 below compares the characteristics of the system with those of other similar services.

Table 1. Compare Service.

	BungaeJangTer	JoongoNara	DangnMarket	Amazon	Ebay	System
Independent Trading System	O	O	X	O	O	X
Comparison of Trade Prices for used Products	X	X	X	O	X	Partially Supported
Registration of Used Products	O	O	O	O	O	Scheduled to be developed
Specific Options Search	X	X	X	Partially supported	O	O
Multi-platform Search	X	X	X	X	X	O

　　Most websites for second-hand product transactions do not support the Multi-platform Search service. Due to the nature of second-hand product transactions, there is no specific format, as they are managed by each user and consumer posting. Therefore, different products would be analyzed as the same product, because there is no particular format for data analysis due to the presence of various titles and texts. In this regard, they do not support the Multi-platform Search as it is difficult and less accurate in classifying products. The system developed in this paper can now classify up to 30 categories, mainly electronic products.

　　The websites do not provide price records or price forecast services because of problems with product classification. Recording and forecasting price requires not only the classifying of products, but also further classifying the products in terms of the exact name and model number of products. The system can type a small number of goods in one product classification.

7. Conclusions

In this paper, I developed the crawler system needed to develop an integrated transaction system for second-hand goods through Internet e-commerce transactions, defined morphological analyzers, and described the service that users can employ in the web environment by using the system developed in the paper.

The big data market is growing by more than 10% every year, with a steady increase in Internet users, and the quantity of data produced directly by the users is expanding as well. In line with this trend, the government and businesses also require the forecasting of customer needs and the analysis of the data. However, data collection, which should precede analysis in the field of big data, is more important than analysis. In this regard, the crawling API required for this process is expected to be useful to other researchers.

Figure 12 shows a list of second-hand goods that can be classified in the system developed in this paper. The system currently categorizes them as a product whose name is clearly identified, such as via model names.

Figure 12 is a diagram of the second-hand product classification list. While most of the second-hand products sold in the market are clothing items, their model and product names are uncertain. Therefore, their accuracy is insufficient when classifying products. Thus, a future data classification algorithm will need to be researched and developed so as to improve product classification accuracy.
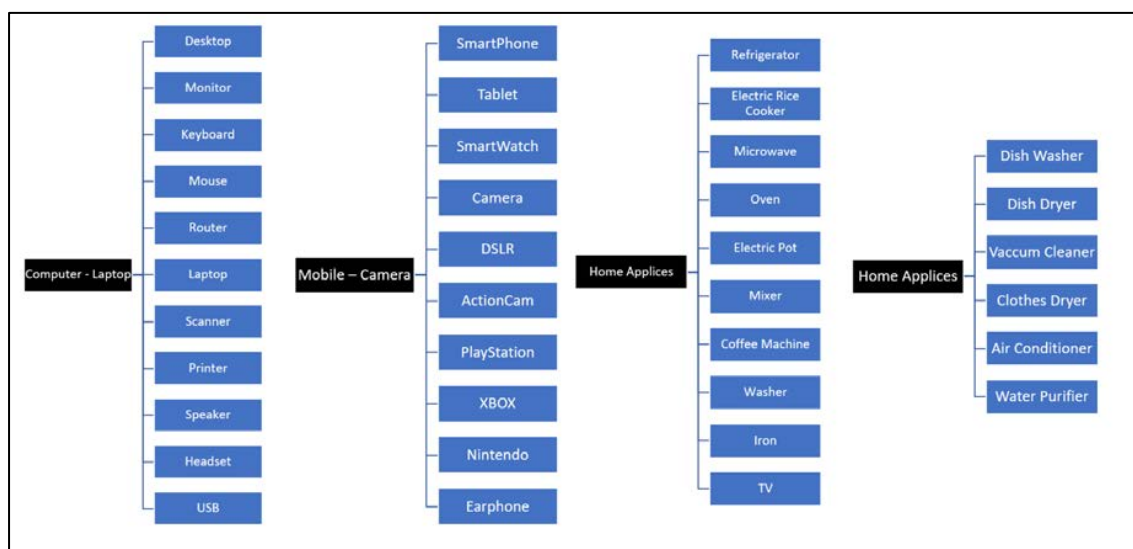

Figure 12. Second-Hand Product Classification List.

In this paper, I found the problems caused when a Korean morpheme analyzer is used for the trading bulletin of second-hand goods, and suggested a solution to resolve the issues. More than 20 million data on second-hand products were analyzed to research the morphological analyzers and solve the problems.

These 20'million data were collected using a crawler bot that implemented the product data of second-hand country services, as presented in Section 3, and included all garbage data for data analysis and pre-processing.

As a result of the use of the analyzers, accurate analysis was done for the words registered in the existing analysis dictionary. However, the analysis rate for newly coined words or sentences that include proper nouns decreased. To solve this problem, I have made the system label the products to get the desired results when processing the morphological analysis. However, the method used in this paper also has the problem of continuing to add product names whenever additional products are

posted. Therefore, I will further research and develop a method for morpheme analysis to improve the accuracy of the system in the future.

References

1. Gordon, R.J. Does the new economy measure up to the great inventions of the past? *J. Econ. Perspect.* **2000**, *14*, 49–74. [CrossRef]
2. Carlsson, B. The Digital Economy: What Is New and What Is Not? *Struct. Chang. Econ. Dyn.* **2004**, *15*, 245–264. [CrossRef]
3. Armagan, R. Yeni Ekonomi ve Turkiye. *Suleyman Demirel Universitesi IIBF Dergisi* **2000**, *5*, 139–153.
4. Akyazı, H.; Kalça, A. Yeni Ekonomi ve İktisat Bilimi. *Liberal Düşünce Dergisi* **2003**, *29*, 221–242.
5. BARIŞIK, S.; Yirmibeşcik, O. Turkiye'de Yeni Ekonomi'nin Olusum Surecini Hızlandırmaya Yonelik Uyum Cabaları. *ZKU Sosyal Bilimler Dergisi* **2006**, *2*, 39–62.
6. Viskari, S.; Pekka, S.; Marko, T. *Implementation of Open Innovation Paradigm, Cases: Cisco Systems, Dupont, IBM, Intel, Lucent, P&G, Philips and Sun Microsystems*; Lappeenranta University of Technology Research Report 189; Lappeenranta University of Technology: Lappeenranta, Finland, 2007.
7. Conboy, K.; Mikalef, P.; Dennehy, D.; Krogstie, J. Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *Eur. J. Oper. Res.* **2020**, *281*, 656–672. [CrossRef]
8. Mikalef, P.; Boura, M.; Lekakos, G.; Krogstie, J. Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *Br. J. Manag.* **2019**, *30*, 272–298. [CrossRef]
9. Taylor, T. Thinking about a new economy. *Public Interest* **2001**, *24*, 3–19.
10. Addo-Tenkorang, R.; Helo, P.T. Big data applications in operations/supplychain management: A literature review. *Comput. Ind. Eng.* **2016**, *101*, 528–543. [CrossRef]
11. Huang, B.; Jin, L.; Lu, Z.; Yan, M.; Wu, J.; Hung, P.C.; Tang, Q. RDMA-driven MongoDB: An approach of RDMA enhanced NoSQL paradigm for large-Scale data processing. *Inf. Sci.* **2019**, *502*, 376–393. [CrossRef]
12. Schäffer, E.; Mayr, A.; Fuchs, J.; Sjarov, M.; Vorndran, J.; Franke, J. Microservice-based architecture for engineering tools enabling a collaborative multi-user configuration of robot-based automation solutions. *Procedia CIRP* **2019**, *86*, 86–91. [CrossRef]
13. Fabian, K.; Philipp, B. Return of the JS: Towards a Node.js-Based Software Architecture for Combined CMS/CRM Applications. *Procedia Comput. Sci.* **2018**, *141*, 454–459.
14. Boran, F.E.; Genç, S.; Kurt, M.; Akay, D. A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. *Expert Syst. Appl.* **2009**, *36*, 11363–11368. [CrossRef]
15. Zingla, M.A.; Chiraz, L.; Slimani, Y. Short Query Expansion for Microblog Retrieval. *Procedia Comput. Sci.* **2016**, *96*, 225–234. [CrossRef]
16. Chang, E.; Dillon, T.; Gardner, W.; Talevski, A.; Rajugan, R.; Kapnoullas, T. A virtual logistics network and an E-hub as a competitive approach for small to medium size companies. In *International Conference Human Society@ Internet*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 265–271.
17. Jones, R. The C programming language. *Data Process.* **1985**, *27*, 35–38. [CrossRef]
18. Chen, K.; Kou, G.; Shang, J.; Chen, Y. Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electron. Commer. Res. Appl.* **2015**, *14*, 58–74. [CrossRef]
19. Chen, X.; Hua, L. Research on e-commerce logistics system informationization in chain. *Procedia Soc. Behav. Sci.* **2013**, *96*, 838–843.
20. Choi, T.M.; Wallace, S.W.; Wang, Y. Big data analytics in operations management. *Prod. Oper. Manag.* **2018**, *27*, 1868–1883. [CrossRef]
21. Barker, T.J.; Zabinsky, Z.B. A multicriteria decision making model for reverse logistics using analytical hierarchy process. *Omega* **2011**, *39*, 558–573. [CrossRef]

22. Zheng, X. The analytics and applications on supporting big data framework in wireless surveillance networks. *Int. J. Soc. Humanist Comput.* **2017**, *2*, 141–149. [CrossRef]

23. Chen, Y.S.; Lin, C.K.; Lin, C.Y.; Chuang, H.M.; Wang, L.C. Electronic commerce marketing-based social networks in evaluating competitive advantages using SORM. *Int. J. Soc. Humanist Comput.* **2017**, *2*, 261–277. [CrossRef]

24. Stai, E.; Karyotis, V.; Katsinis, G.; Tsiropoulou, E.E.; Papavassiliou, S. A Hyperbolic Big Data Analytics Framework within Complex and Social Networks. *Big Data Complex Soc. Netw.* **2016**, *4*, 75–88.

25. Stai, E.; Karyotis, V.; Papavassiliou, S. Exploiting socio-physical network interactions via a utility-based framework for resource management in mobile social networks. *IEEE Wirel. Commun.* **2014**, *21*, 10–17. [CrossRef]

26. Pouli, V.; Kafetzoglou, S.; Tsiropoulou, E.E.; Dimitriou, A.; Papavassiliou, S.; Vasiliki, P. Personalized multimedia content retrieval through relevance feedback techniques for enhanced user experience. In Proceedings of the 2015 13th International Conference on Telecommunications (ConTEL), Graz, Austria, 13–15 July 2015.