applied sciences

MDPI

Article

Design and Implementation of a Big Data Evaluator Recommendation System Using Deep Learning Methodology

Sukil Cha [1,2], Mun Y. Yi [3] and Sekyoung Youm [1,*]

[1] Department of Industrial and Systems Engineering, Dongguk University, Seoul 04620, Korea; chasi@dgu.ac.kr
[2] Knowledge and Information Service Team, National Research Foundation of Korea, Daejeon 34113, Korea
[3] Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea; munyi@kaist.ac.kr
* Correspondence: skyoum@dongguk.edu; Tel.: +82-2-2260-3377

Received: 9 October 2020; Accepted: 9 November 2020; Published: 11 November 2020

check for updates

Abstract: As the number of researchers in South Korea has grown, there is increasing dissatisfaction with the selection process for national research and development (R&D) projects among unsuccessful applicants. In this study, we designed a system that can recommend the best possible R&D evaluators using big data that are collected from related systems, refined, and analyzed. Our big data recommendation system compares keywords extracted from applications and from the full-text of the achievements of the evaluator candidates. Weights for different keywords are scored using the term frequency–inverse document frequency algorithm. Comparing the keywords extracted from the achievement of the evaluator candidates', a project comparison module searches, scores, and ranks these achievements similarly to the project applications. The similarity scoring module calculates the overall similarity scores for different candidates based on the project comparison module scores. To assess the performance of the evaluator candidate recommendation system, 61 applications in three Review Board (RB) research fields (system fusion, organic biochemistry, and Korean literature) were recommended as the evaluator candidates by the recommendation system in the same manner as the RB's recommendation. Our tests reveal that the evaluator candidates recommended by the Korean Review Board and those recommended by our system for 61 applications in different areas, were the same. However, our system performed the recommendation in less time with no bias and fewer personnel. The system requiresrevisions to reflect qualitative indicators, such as journal reputation, before it can entirely replace the current evaluator recommendation process.

Keywords: big data; R&D evaluator; similarity scoring; recommendation system; deep learning

1. Introduction

The platform and analysis of research on big data technology were first developed in the first decade of the new millennium, receiving increasing attention in the 2010s. The big data platform was developed through exploratory data technology, which captures significant information from diverse sources to generate predictive thorough intelligent analysis [1]. Moreover, the research, which was previously limited to information and communications technology, is currently applied to diverse and wide-ranging areas, for example, humanities, sociology, and convergence. In the public sector, the existing data accumulated by different agencies, along with real-time data that flow through various smart platforms, are also being combined and analyzed. Thus, a variety of technologcial research is underway to use big data for solutions to various problems [2].

In the research and management of national research and development (R&D) projects, considerable effort and demand have been aimed at providing improved services by tapping into big data for the entire process (planning→project selection→project agreement→implementation check and performance assessment→research cost management and calculation→performance management). However, in South Korea, as the number of researchers seeking support is increasing and overwhelming, the country's R&D budget, the possibility of research projects being tapped is decreasing, which increases dissatisfaction and doubt among unsuccessful applicants concerning the overall project selection process. The number of objections lodged annually with the National Research Foundation of Korea is about 1100, which accounts for about 1.9% of newly filed projects (=1139 out of 60,000). For the National Science Foundation in the United States, about 30 (0.08%) to 50 (0.13%) of about 38,000 annual applications result in an officially requested reexamination [3]. In assessing this, many variables, including the size of the R&D budget, must be considered. Nonetheless, a strong possibility exists that the South Korean project management agency will face many objections from its applicants.

The selection of research projects is a process where fairness is juxtaposed with professionalism. The process must consider the professionalism and specialty involved in national R&D and the required fair and equal selection of beneficiaries of the country's R&D funds. Fairness in the public sector is a factor that should take priority, and various approaches have been suggested and applied. These approaches include blind reviews and the required recusal of evaluators due to a possible conflict of interest or lack of impartiality. A blind review is designed to prevent the influence of personal connections in research, as indicated in the application, with a greater focus on the facts charted in the application than on personal skills. To this end, indicating personal information (such as name, employer, and achievements) on the application is prohibited. The required recusal of evaluators due to a possible conflict of interest or lack of impartiality is a program that keeps those off the evaluation committee who happen to be family or school relations of the principal investigator for a candidate project, share the same employer with any of the researchers participating in the candidate project, or are peer reviewers of the principal investigator in the candidate project.

The evaluator, who reviews the projects filed by researchers and has the final say on funding, represents the most crucial factor in the evaluation process. A blind review and the required recusal of evaluators because of a possible conflict of interest or lack of impartiality are procedures designed to preclude the identified risk factors lest an evaluator develops a bias in the evaluation. Once selected as a project evaluator, however, the criteria and results of the relative evaluations of the projects are completely up to the evaluator. As this is inviolable, selecting a professional and fair evaluator is crucial. The evaluator selection procedure is insulated from a single person's domination by ensuring the participation of diverse people.

In the evaluator selection process, the person in charge organizes a pool of evaluators from among researchers who have demonstrated certain minimum achievements as listed in the research system. The roles are divided between the Review Board (RB) and the chief RB whose members are recommended from academic societies and agencies operating in the specific areas of study and are selected through the screening process by a review committee in a separate procedure. From the pool of evaluators, three times as many evaluator candidates as needed (nine are recommended if three are needed) are recommended by the RB based on the connection to the candidate project and the achievements of the evaluator candidates. The three-fold pool is divided into groups (A, B, and C) by the chief RB. Subsequently, a program manager decides which group to use first to recruit an evaluator. This procedure dictates the sequence in which the evaluator candidates are contacted if a specific group does not have sufficient researchers if the evaluator candidates do not accept the request to serve as evaluators.

The recommendation of evaluator candidates takes considerable time and effort by many people to ensure that the evaluator selection process is not overly influenced by a single person's subjective experience and preferences. However, the use of specific and objective criteria and techniques for selecting evaluators should enable the selection of these evaluators without such procedures.

In this study, we explore how to innovatively improve the process by selecting evaluators through a system (machine) instead of using several systems and procedures to minimize the influence that may be exerted when people are involved in evaluator selection. In other words, we aim to design a system that can recommend the best possible evaluators using big data that are collected from related systems and are refined and analyzed to select candidates using diverse wide-ranging data according to objective criteria to register evaluator candidates after receiving project proposals.

Thus, the system evaluates using fair and objective criteria and procedures to select excellent researchers. Through this, we aim is to create a virtual cycle in which our research results can be used continually to perform the exploratory investigation and selection of appropriate research projects.

2. Literature Review: Related Work

The basic data to be used in the big data-based evaluator recommendation system are unstructured data, such as the assignment application form and original document such as thesis. In this regard, a study related to the text mining analysis method should be preceded by targeting the extracted part in the form of information (text) from each original text file and the text information.

To analyze text, in most cases, a tokenization process (tokenize) that separates sentences into individual words is required. In the case of Korean, morpheme analysis is performed to separate sentences into individual words, and morphemes, such as nouns and verbs, are extracted. The frequency of occurrence can be determined to infer the subject of the document or attitude toward the document. In addition, the frequency of occurrence is used in various studies because it can determine the frequency of the simultaneous occurrence of words, identify the relevance or correlation between words, and express the information in a network graph [4,5].

Topic modeling is an algorithm that extracts a specific topic from numerous text data. Representative topic modeling algorithms include latent semantic analysis and latent Dirichlet allocation (LDA), and the topic can be inferred using the word2vec algorithm. Blei et al. proposed LDA, an algorithm that can check the topic of a document based on a probability technique. The LDA is a procedural probability distribution model for finding potentially meaningful topics in a large literature group. Given the probability distribution θ of the subjects in the document and the probability distribution z of the words constituting each subject, the sampling process of probabilistically selecting the subject constituting the document and the words present in the selected subject is repeated. It is a model that generates an arbitrary document [6]. The LDA estimates z and θ using parameter values of α, the prior probability distribution of topics within a given document and a predefined document, and β, which is the prior probability distribution of words within the topic. The LDA techniques are often used to analyze technologies and research trends in various fields, such as deriving a topic model using LDA techniques for "Science" journals and "Yale Law" journals, or analyzing research trends in biomedical fields [7–10].

Word2vec is an algorithm that implements the room proposed in "Efficient Estimation of Word Representations in Vector Space" written by Google researchers [11]. Word2vec learns through text documents and learns other words that appear adjacent to one word as related words in an artificial neural network, i.e., as the order of words is closer and frequently appears, words have similar vector values. It is the same as LSA in that a word is quantified as a vector value, but the difference is that it is trained using a neural network while considering the distribution of words located before and after a specific word. Learning algorithms include, continuous bag of words (CBOW) and skip gram; CBOW creates a multidimensional vector of a specific word using the surrounding words of the specific word, whereas skip- gram creates a multidimensional vector of the surrounding words based on a specific word. In most text mining techniques, all stop words are often deleted during data preprocessing, but the Word2vec technology generates vector values with up to 300-dimensional vector values for each word based on other words that occur around specific words. It can be applied and can be used in various fields such as topic modeling, machine translation, stock price prediction, or recommendation systems [12–15].

Research on the automatic evaluator recommendation system for selecting evaluators for R&D projects has been attempted in numerous ways. In "Automatic Recommendation of Panel Pool Using a Probabilistic Ontology and Research Networks," a probabilistic ontology that can automatically calculate the association between each category of a multifaceted academic field classification table was applied. Through this, the scope of the reviewer recommendation was comprehensively expanded and the ranking of reviewers reflecting their expertise was made possible. In addition, by establishing a researcher network based on joint research between researchers and using it as a rule for the exclusion of reviewers, a study was conducted to ensure fair reviewer recommendations [16].

Similarly, there is an expert recommendation system using big data. When researchers start researching in a new field, they seek advice from experts or set up a research direction based on expert research papers. At this time, the existing academic search service provides information on research papers for each field, but does not provide experts in each field, so researchers must directly determine the expert based on the searched article. To improve this, they designed and implemented a search system for experts in each academic field using big data processing based on the information published in the conference. This system used big data distributed storage technology to store and manage a large amount of papers. In addition, the system uses big data distributed processing technology to identify experts and analyze information related to the experts. The distributed results are displayed through a web page when a user requests an expert search. Researchers can receive considerable help researchingby receiving recommendations from experts in the research field through the proposed system [17].

3. Designing a Big Data Evaluator Recommendation System

Although evaluator recommendation systems designed to select evaluators for filed projects have been created and operated continuously, the previous systems have focused only on showing a list of evaluator candidates. Thus, these systems play a larger role in presenting researchers in the specific areas of study and scanning their achievements as an auxiliary arm for evaluator recommendations than in finding appropriate evaluators for the projects.

The previous method that identifies researchers suitable for project proposals consists of matching researchers to the correct areas by comparing the categories of research the submitter selects in the application and the code of the area of evaluation that the candidate chooses. This method, which matches a project with an evaluator in a simple manner without causing contention (because it is the area chosen by the applicant), has limitations due to the categorization of areas.

First, because dynamic areas of research comprise real-time changes, data cannot be comprehended only through regular updates of the data. Moreover, currently, many studies take the form of convergence research and its variations cannot be contained in these classifications. Second, an applicant (or an evaluator candidate) may not select the appropriate research areas in the application. As there are diverse areas, including the area that a project proposal applies for, classification is sometimes difficult. The wide-ranging variations may not be captured in the categorization of the areas or the researcher may select an area that is different from the one indicated on the application (because, e.g., the area is more likely to be supported). Thus, after selecting a research area in high demand, a project evaluator, who is unrelated to the person's actual research topic may be selected. Third, the selection of evaluators based on the categorization of areas may tap researchers who flaunt their achievements and recognition in the areas instead of those who are most suitable for the submitted projects. Researchers who flaunt their achievements and recognition in specific areas may not be those who can best evaluate specific projects. Finally, a recommender's subjective preferences (such as a preference for specific journals or preferential treatment of patents) may be reflected in the selection of evaluator candidates.

Evaluator recommendations in the big data recommendation system, as suggested in this study, break from the previous selection method through the categorization of areas that identify pairs of the most similar project applications and evaluator candidates. The system compares keywords extracted from the applications through the big data system and the major keywords extracted from the data in the full text of the diverse achievements fo the evaluator candidates. Identifying evaluator candidates who register high levels of similarity with project applications and selecting evaluators by determining priorities based on their similarities not only enables the selection of evaluators who are the most suitable for the submitted projects, but also improves conventional practices, such as three-fold recommendations, organization of groups, and determination of priorities.

3.1. Structure of the Evaluator Recommendation System

As shown in Figure 1, the big data evaluator recommendation system starts with the collection of documents received from the project submission system. The project submission system in the form of the relational database (RDB) collects both the information related to the project submission (such as submission year, project number, principal investigator, project name, and research budget) and the original text of the submitted project. The collection includes copies of the same data in the old system to prevent the performance of the actual service provision system from being compromised and preserving the original copies (to prevent deformity due to refining).

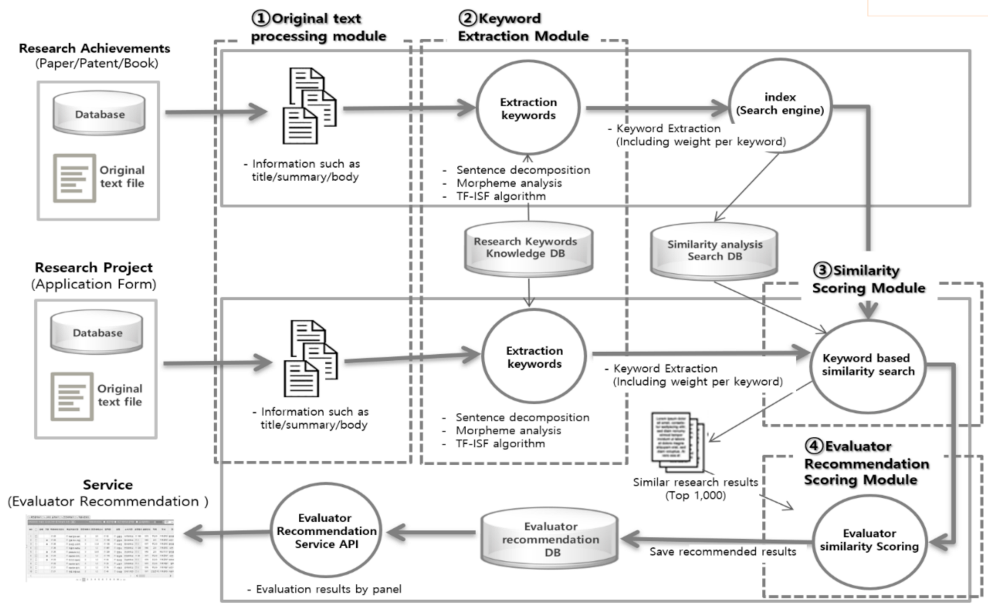

Figure 1. Original text processing module.

The original collected texts are processed and analyzed first through the original text processing module. The data for different projects are processed and saved as files so they can be used on the big data platform. The process also indicates where in the documents, the data are extracted from, such as the subject, summary, and body, so that different scores can be awarded to the same keywords and weighted later.

Subsequently, the keyword extracting module performs sentence breakdown, such as the analysis of morphemes, by processing the data. In this way, keywords are extracted, and the weights for different keywords are scored using the term frequency–inverse document frequency (TF-IDF) algorithm. Keywords extracted in this way are stored in the research keyword knowledge database and are accumulated as basic data that can improve the data quality by sorting synonyms and unused words, e.g., The core of the big data recommendation system is recommendation quality, so quality improvement activities should be continuously conducted. Various information are stored in the research keyword knowledge database in the form of an ontology for efficient quality improvement activities in the future.

The collection of data on the research achievements of the evaluator candidates is conducted using a method similar to that of the project application data collection. The data from the original text files, primarily in hwp, doc, and pdf format, are extracted in the same manner in the text processing module, and the keyword extraction module is indexed in the search engine for comparative analysis with the project application data. Unlike in the case of project applications, the data on the research achievements of the evaluator candidate are collected regularly, and preliminary indexing reflects changes only in the modified copies through daily placement. The system in its early phase requires the initial indexing of all researchers. As the data on all researchers cannot be complete, priority is placed on key evaluators whose problems due to omitted data can be minimized.

By comparing the keyword values extracted from project applications and the index extracted from the research achievements of the evaluator candidates, the project comparison module searches, scores, and ranks tachievements similar to the project applications. The similarity scoring module calculates the overall similarity scores for evaluator candidates based on the scores that the project comparison module creates for different achievements. These scores are used as evidential scores to determine the most appropriate evaluator candidates for evaluating the specific project application.

The original text data from the research project applications and papers, require preliminary work through the original text processing module to extract keywords. The most common formats for the research project applications and papers are hwp, doc, and pdf, which require varying text extraction modes. Doc and hwp files can be extracted through open libraries and do not involve significant errors. However, pdfs face many issues because data, such as line breaks and encoding, may be lost due to conversions from text file formats such as doc or hwp. In this study, the proof of concept phase employs a simplification mode that operates by removing all line breaks and analyzing the morphemes to process pdfs. In the actual realization of the function, errors are minimized using widely available software (in this case, Wise Tea from WISEnut Inc., Seongnam, Korea). Once through this module phase, data are converted to the same text form from the variously formatted files (hwp, doc, and pdf) and then used in the keyword extraction module.

3.1.1. Keyword Extraction Module

To analyze the extracted text, language judgment is performed first. If the Korean encoding characters comprise over 25% of the total characters in the extracted text, the document is regarded as a Korean-language document. When the Korean encoding characters make up less than 25%, it is regarded as an English-language document. In Korean literature, documents that include less than 25% of the Korean language are written primarily in Sino-Korean and in some English-language journals, Korean-language abstracts are often included; hence, the 25% ratio for the Korean language.

The next phase is pretreatment, which comprises document indexing, morpheme analysis, and lemmatizing. Document indexing is performed for keyword extraction and to enable a search later using Lucene, which is a typical method. Morpheme analysis is a process that determines and classifies not only morphemes but also diverse linguistic attribute structures, such as word stems, prefixes/suffixes, and parts of speech. Lemmatization refers to the extraction of word stems, which is a process that removes the derivative meanings of words through their morphological and lexicographical analysis and finds basic lexicographical words, that is, commonly used standard words.

The last phase is keyword extraction. As major research keywords are nouns, like technical terms, all the nouns are extracted through the abovementioned pretreatment. Then, major keywords are extracted in the TF-IDF from each document. Given that big data characteristically require processingnumerous documents, the documents are indexed in advance with Lucene to ensure prompt TF-IDF scoring.

3.1.2. Similarity Scoring Module

The similarity among applications is identified by calculating keyword similarity scores (using TF-IDF) and then analyzing the similarity between these and the words extracted from the papers (using the BM25 algorithm). By calculating the cosine similarity (or co-occurrence) from the frequency of keywords in the extracted papers shown simultaneously in one sentence, the connectivity scores that suggest if the keywords appear in similar manners are measured. The TF-IDF of each word is applied to this score (scores are regularized and multiplied) to give connectivity weight to relations between words of greater importance.

$$score(D,Q) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{f(q_i,D) \cdot (k_1+1)}{f(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)} \tag{1}$$

3.1.3. Evaluator Recommendation Scoring Module

The final scores are provided by granting weights specific to the importance of different documents following the document scores calculated through the similarity scoring module. The weights given to different types of documents are as follows: 1.5 for SCI-grade papers, 1.1 for non-SCI-grade papers, 1.2 for registered patents, 0.9 for filed patents, 0.8 for research projects, and 0.6 for other periodicals. Thus, we assume that considering the type of the achievements singled out in similar documents, achievements with greater weights register higher levels of understanding of the project application. This suggests that the evaluator candidates who have published other periodicals similar to the project applications are more suitable as evaluators of the project applications than those candidates who have written similar "SCI-grade papers."

The recommendation scores for different evaluator candidates are calculated by weighting and totaling the different evaluator documents with the final similarity scores that reflect the weights for the different types of documents. Specifically, in the first step that involves the submitted projects, the similarity is analyzed to arrive at a list of similar documents and then the similarity scores for different documents are rearranged by assigning specific weights to different types of documents as noted above. Then, when multiple documents by the same evaluator candidate are found among the documents, once the similarities have been rearranged, the final recommendation scores for the candidates are calculated by weighting and totaling the similarities and priorities among the documents.

[Formula for obtaining the total of weighted evaluator candidate recommendation scores] the similarity of the author's highest scoring document + ∑(similarity/priority) (2)

The recommendation score for evaluation candidates follows the method of Equation (2), and an exampleis provided as follows: The similarity of document A (9000) + the similarity of document D (5000)/document priority (4) + the similarity of document F (3600)/document priority (6) = 10,850 (points); and the similarity of document B (7000) + the similarity of document E (4000)/document priority (5) = 7800 (points). Based on the final calculated totals of the weighted candidate recommendation scores, n persons with the top recommendation scores (set differently for different projects) are recommended as evaluator candidates.

3.2. Big Data Platform Architecture

The steps to build a big data platform architecture are shown in Figure 2, and the details are as follows. ① The types of data to be collected from the legacy system are analyzed and their volumes are checked ② Collect first-time raw data into the big data platform, which involves cases such as structured→structured, structured→unstructured, and unstructured→unstructured. ③ The collected data are then structured and stored as data that cover services and data identity. ④ Collected data, which serve as the basic data for analysis, are stored in a data lake, regardless of type. ⑤ The structured data are transformed, and quality management for the raw data of the Operational Data Store (ODS) is performed. ⑥ The analysis of the structured data is performed by Sqoop. ⑦ A Hadoop data analyst performs the unstructured analysis. ⑧ To meet the objective of data storage for user services, data are stored as structured data. ⑨ Services, such as evaluator recommendations, are performed based on the results of the analysis of the raw data and big data of the legacy system.

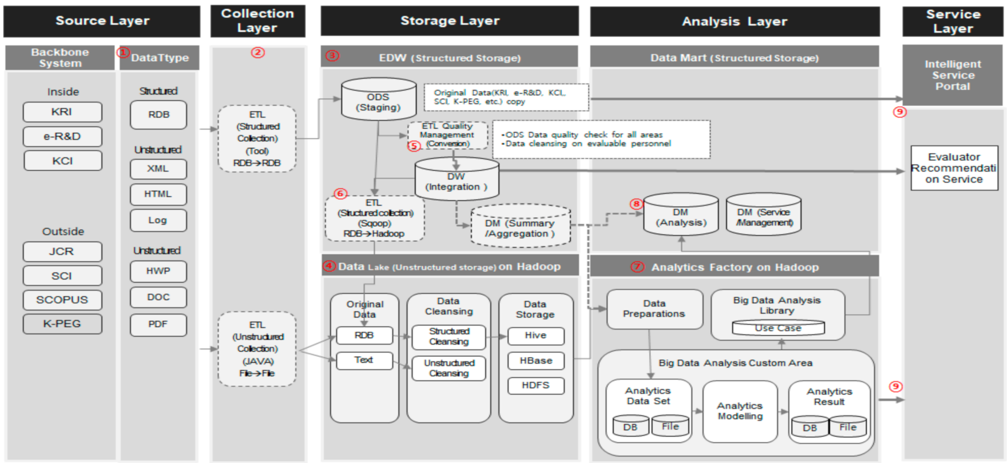

Figure 2. Big data platform architecture.

The big data platform described in this paper minimizes the time spent in collection by keeping the structured source data of the RDB management system (RDBMS) structured in the same system to reduce system load. Moreover, by managing the enterprise data warehouse (EDW) as an RDBMS, the development and promotion of data are easier. For data quality management, the RDBMS is easier to use and faster. All the source data, structured or unstructured, are stored in the data lake of the Hadoop system for analysis, and the analysis results are transferred to DataMart to ensure ease of use by service portals.

4. Results

4.1. Evaluator Candidate Recommendation System Implementation

The big data evaluator recommendation system was built in a cloud environment for future scalability and immediate response. The introduction of Hyper Converged Infrastructure (HCI) equipment, which has various advantages in constructing a new environment, was considered first. The HCI integrates storage, network, virtualization, and security, making operation management simple and convenient, and has the advantage of reducing costs. We introduced Nutanix (San Jose, CA, USA), one of the HCI equipment, installed Virtual Machine (VM) on it, and constructed the Hadoop system (Apache Software Foundation, Forest Hill, MD, USA) as one of the VMs. Hadoop system was set up with a total of 4 nodes, and it consisted of one Name Node-Master Hive Meta Store and 3 Data Nodes. Each node consisted of CPU: 8 core, memory: 40 GB, storage: 4.3 TB. There is no separate cluster configuration, and if the service is expanded in the future, new HCI equipment will be introduced to improve performance.

The portal service operation, which provides services to users, is composed of a Linux OS, JEUS Web Application Server (TmaxSoft, Seongnam, Korea), and WebtoB Web Server (TmaxSoft, Seongnam, Korea) for web services. The basic development environment maintained the J2EE environment (Java, JSP, EJB, JDBC, XML, etc.) for interworking with the existing system, the communication was done in OpenAPI format, and the data were done in Java Script Object Notation (JSON) method. For data storage, HBase (Apache Software Foundation, Forest Hill, MD, USA) and MongoDB (MongoDB Inc., New York, NY, USA) were used for distributed processing and storage. Spark platform (Apache Software Foundation, Forest Hill, MD, USA) was used for big data analysis and processing.

The big data evaluator recommendation system planned an ontology that could store and infer various information in the knowledge database to improve recommendation quality, but the ontology could not be implemented in the actual implementation stage. In this study, we saved the knowledge database in the form of a keyword-attribute-value triple in the RDB. Through this data base, the direction of the ontology composition was examined, and the main keywords were stored here to be used as basic data for improving the quality of future evaluator recommendations. Figure 3 shows the output of the big data evaluation recommendation system. ① The "Big Data Recommendation Level" indicates the recommendation for the final evaluator candidates that have undergone pretreatment and analysis according to the abovementioned big data platform architecture. ② The major keywords of the evaluator candidate recommenders are shown in one single view. ③ The numbers of connected projects are identified, indicating how many projects of each evaluator candidate recommender relates to the specific panel. ④ The shown sequence maps the evaluator candidate recommenders and related projects.

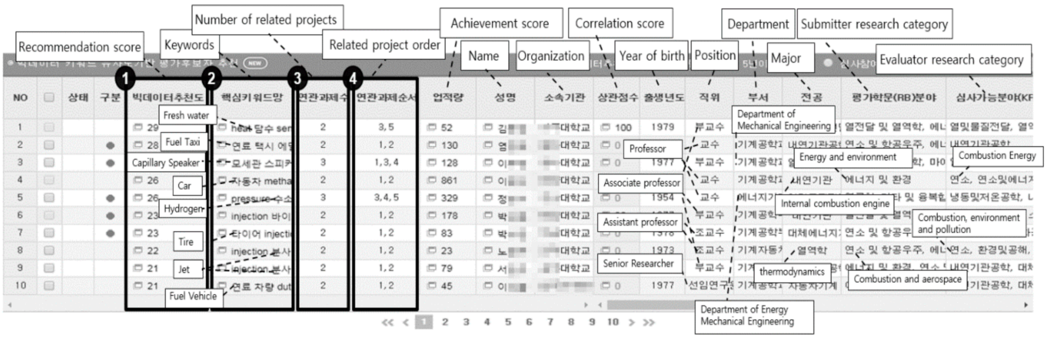

Figure 3. The output of the big data valuation recommendation system.

Figure 4 shows the output of the similar achievements and scores of the evaluator candidate recommendations. ① Documents with achievements similar to a submitted project by document type are shown. In ②, the big data recommendations for different projects are shown. This enables the users to see the grounds for determining which of the submitted projects on the panel support the recommendation. ③ The data on similar achievements are shown, which indicate the similar achievements of the evaluator candidate with respect to the submitted project. In ④, the grounds of recommendation are shown by displaying the list and summary of similar achievements.

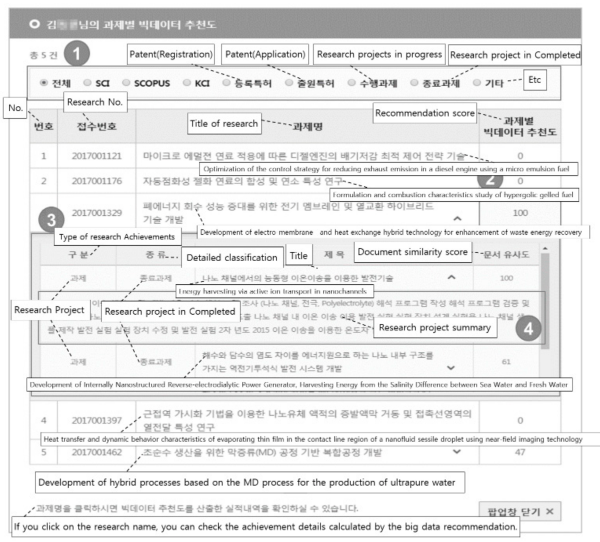

Figure 4. The outputs on the similar achievements and score of evaluator candidate recommendation.

4.2. Testing the Evaluator Candidate Recommendation System

To assess the performance of the evaluator candidate recommendation system, we conducted experiments in three RB research fields (system convergence, organic biochemistry, and Korean literature). In comparing the scores of the evaluator candidates recommended by the RB and those recommended by our system, for 61 applications in different areas (21 in system convergence, 20 in organic biochemistry, and 20 in Korean literature), our system recommended the same evaluator candidates, matching the numbers of the RB-recommended evaluator candidates (747 persons in system convergence, 413 persons in organic biochemistry, and 779 persons in Korean literature).

The RB recommends three times the number of recommended evaluators, but the number of recommended evaluators differs for each field because the recommendation of similar candidates is requested for verification of the system. A total of 747 candidates were recommended for 61 tasks, so about 32 candidates per task were recommended. In the case of organic biochemistry, 413 similar researchers were recommended and relatively small numbers compared to other fields were suggested as similar raters. The similarity was measured by applying the similarity Equation (1) to all recommended candidates. Table 1 shows the average value of the similarity of each candidate for evaluation.

Table 1. The results of the recommendation from the Review Board (RB) and developed system.

RB Areas	Average RB Recommendation	Average Machine Recommendation	Difference (Times)
System convergence	0.01864	0.45478	24.39
Organic biochemistry	0.06571	0.60992	9.28
Korean literature	0.02114	0.14178	6.70
Average	0.03516	0.40216	11.4

Table 1 presents the results of the recommendations from the RB and our system. The results show the average scores of the researchers recommended by the RB and the system. The similarity score was measured according to the method described in the evaluator recommendation scoring module. The quality may vary by recommending more numbers than the method of recommending three times the number needed, which is the standard of the actual evaluator recommendation procedure. However, the average scores for the researchers recommended by the system are about 11 times higher than those of the RB-recommended researchers, which is quite a substantial difference. However, the results consider the similarity among the text in the database, and the difference does not mean that the conventional RB recommendations are not valid.

We considered various methods to review the validity of the big data evaluator recommendation system. It was difficult for people to individually check the results of applying various weights after extracting about 120 keywords using big data. Thus, after having the RB use the system on a trial basis, a satisfaction survey (5-point scale: very dissatisfied, somewhat dissatisfied, neutral, somewhat satisfied, and very satisfied) was conducted. The satisfaction survey found that 84.9% of 119 respondents said they were somewhat satisfied or very satisfied.

The system requires revisions in many respects before it can entirely replace the current evaluator recommendation process. For example, it must reflect various qualitative indicators such as journal reputation and the distinction of the first author. Moreover, in natural sciences, the more significant SCI databases are currently unable to collect original texts other than metadata. Despite these partial revisions, the fundamental quality improvement requires more diverse data.

5. Conclusions

This study presented the design for an evaluator recommendation system based on a big data platform. Finding a researcher who is suitable for a designated research project with limited content using the top-down method is not often compared; thus, human work is appropriated in terms of expertise and efficiency. However, to evaluate each project targeting applicants for all academic fields using a bottom-up method, researchers with expertise in the relevant content must be identified, which is difficult. The National Research Foundation of Korea evaluates about 60,000 projects per year. There have been attempts to solve this problem, but the reason the possibility of a soluation has increased at this point is due to the big data environment.

Previously, only limited information was collected and managed due to the limitations of H/W and cost issues. Therefore, there was a limit to finding a solution through data matching. Due to the characteristics of the big data platform architecture, reliability increases as more data are accumulated. The National Research Foundation of Korea, the source of the original data in this study, is at the point at which 10 years have passed since the data were accumulated by building a new large-scale system. Since sufficient data for big data analysis were secured, this study could also be attempted.

Instead of using the academic field, which has limited information on the subject applicant, the main keywords were extracted through the content analysis of the original text of the assignment application form, and the evaluation candidates also extracted keywords through the analysis of their thesis, patents, etc., in place of the academic field. In addition, it tried to recommend the best candidates through matching between these two extracted keywords. Through this method, it wis meaningful to establish an objective evaluation candidate recommendation system based on data, instead of a recommendation based on individual preferences and subjective experience in the performance of the individual in charge of recommending evaluator candidates.

However, it is difficult to replace the existing evaluation task by simply building a data-based system. Toimprove work through the system, securing the user's reliability must first be prioritized. It is necessary to visualize what algorithm the system operates with and what evidential data result in the outcome of being chosen as an evaluator candidate. Due to the nature of big data, the presentation of algorithms and evidential data should be considered in various ways such as word cloud, score calculation details, and verification of the highest score performance.

In addition, it is also necessary to prepare other measures to check the recommended results other than measuring the similarity between the growers suggested through this study. This could be the actual commission rate of the recommended evaluator, the evaluation satisfaction survey or other methods.

Second, the system must be enhanced for quality improvement. The characteristics of each discipline, qualitative index, and algorithm improvement can be considered. The characteristics of each discipline may be given different weights under the assumption that the importance of each field is different, even if the keywords are extracted identically. For example, when the keyword COVID-19 is derived, it must be assessed differently in the fields of Korean literature and organic biochemistry. By assigning keyword weights for each field in this way, the quality of the chosen evaluators can be improved.

Next, the qualitative index is the consideration of the author classification, JCR, and K-PEG. When thesis results that are similar to the subject of the project applicants research are extracted for evaluation concerning the evaluator candidates, their adequacy scores should be measured differently fregarding the written paper for the lead author and co-authors. In addition, in the case of a paper or patent with excellent quality, a little more weight may be given in the evaluation of the adequacy of the evaluator candidate.

Regarding algorithm improvement, it is expected that the quality of the candidate recommendation can be improved if the learning-based machine learning algorithm, which currently is actively applied and studied, is applied instead of the statisticallybased algorithm used in this study.

However, the cost reduction effect must be considered in the process of recommending evaluator candidates. Naturally, cost savings occur concerning replacing the existing evaluator candidate recommendation process through the RB with the system. Currently, it is at the stage of reviewing the possibility of the system, so measurement is difficult, but if the system is actually applied and compared with the existing process and is measured, it is possible to confirm the details of the cost reduction.

The system can also be considered in terms of user satisfaction. The goal is to conduct an assessment that is satisfactory to the applicant. If the system is applied, an evaluation is performed, and a satisfaction survey is conducted on the project applicant, the limitations and improvements of the system can be determined.

Although it may be difficult to collectively enhance the above-listed improvements, it is expected that the big data-based evaluator recommendation system implemented in this study can be gradually improved if changes are applied in stages.

Author Contributions: S.C. was responsible for methodology, analysis, and writing—original draft preparation, M.Y.Y. was responsible for the main algorithm and conceptualization and S.Y. was responsible for methodology, review, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Dongguk University Research Fund 2020 (S-2020-G0001-00050).

Acknowledgments: This work was supported by the Dongguk University Research Fund 2020 (S-2020-G0001-00050)

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

References

1. Blei, D.M. Probabilistic Topic Models. *Commun. ACM* **2012**, *55*, 77–84. [CrossRef]
2. Kim, H.J.; Ra, J.H.; Jeon, W.R.; Kim, P. A Study on Construction of Platform Using Spectrum Big Dat. *Smart Media J.* **2020**, *9*, 99–109.
3. Park, C.; Choi, Y.; Whang, J.; Lee, S.; Lee, J.; Jun, S.; Ahn, W.; Lee, Y.; Cho, Y.; Park, W.; et al. *Plan of Evaluation System for Operation of Research Project*; NRF Issue Report; NRF: Yucheondang, Korean, 2015; Volume 9, pp. 6–10.
4. Jeong, C.W.; Kim, J.J. Analysis of Foresight Keywords in Construction Using Complexity Network Method. *J. Korean Digit. Archit. Inter. Assoc.* **2012**, *12*, 15–23.
5. Kim, H.J.; Song, M. A Study on the Research Trends in Domestic/International Information Science Articles by Co-Word Analysis. *J. Korean Soc. Inf. Manag.* **2014**, *31*, 99–118.
6. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
7. Byeongmun, J.; Taehwan, K.; Jin, L.; Jungsun, K. Twitter Topic Extraction and Topic Category Decision Using LDA Model. *Korea Inst. Inf. Sci. Eng.* **2015**, *12*, 787–788.
8. Park, J.H.; Song, M. A Study on the Research Trends in Library & Information Science in Korea Using Topic Modeling. *J. Korean Soc. Inf. Manag.* **2013**, *30*, 7–32.
9. Kim, K.S.; Jung, H.G.; Lee, H.J.; Lee, H.J. Collaborative Filtering Using Topic Models for Rating Based Recommender Systems. Available online: http://www.koreascience.or.kr/article/CFKO201229149570751.page (accessed on 8 October 2020).
10. Chae, S.H.; Lim, J.I.; Kang, J. A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce. *J. Intell. Inf. Syst.* **2015**, *21*, 53–77.
11. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
12. Park, D.H.; Choi, W.S.; Kim, H.J.; Lee, S.L. Web document classification based on hangeul morpheme and keyword analyses. *KIPS Trans. PartD* **2012**, *19*, 263–270. [CrossRef]
13. Kang, S.; Seo, J. An Iterative Approach to Graph-Based Word Sense Disambiguation Using Word2vec. *Korean J. Cogn. Sci.* **2016**, *27*, 43–60.
14. Kim, S.K.; Cho, H.J.; Kang, J.Y. The status of using text mining in academic research and analysis methods. *J. Inf. Technol. Archit.* **2016**, *13*, 317–329.
15. Park, J.H.; Park, E.J.; Jo, D.J. Automated Text Analysis of North Korean New Year Addresses, 1946–2015. *Korean Political Sci. Rev.* **2015**, *49*, 27–61.

16. Lee, J.Y.; Lee, J.Y.; Kang, I.S.; Shin, S.K.; Jung, H.M. Automatic Recommendation of Panel Pool using a Probabilistic Ontology and Researcher Networks. *J. Korean Soc. Inf. Manag.* **2007**, *24*, 43–65. [CrossRef]
17. Choi, D.; Kim, M.; Kim, D.; Lee, S.; Han, J.; Seo, I.; Lim, J.; Bok, K.; Yoo, J. Design and Implementation of an Expert Search System Using Academic Data in Big Data Processing Platforms. *J. Korea Contents Assoc.* **2017**, *17*, 100–114.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.