The Past, Present, and Future in Steel Research using Big Data

Hyunjoung Lee
Research Institute of Iron and Steel Technology
Yonsei Univ.
Seoul, Rep. of Korea
mktbridge@yonsei.ac.kr

Il Sohn
Materials Science and Engineering
Yonsei Univ.
Seoul, Rep. of Korea
ilsohn@yonsei.ac.kr

Abstract—This research describes the knowledge framework of steel research in the journal of Steel Research International from 1990 to 2013. 5700 keywords derived from 2430 papers were analyzed with big data analysis. Frequent keywords of steel research were found to be microstructure, finite element method, transformation induced plasticity. Through centrality analysis, keywords with a high centrality were microstructure, finite element method, mechanical properties, and continuous casting. From the cohesiveness analysis, eight cohesive subgroups were identified. The results of this study contributes valuable insights into understanding the knowledge framework of the steel research field and research trends for Steel Research International.

Keywords—big data analysis; steel research; knowledge structure; co-occurrence analysis; centrality analysis; cohesive subgroup

I. INTRODUCTION

Steel research has dramatically developed over the past several decades with the development of steel industry. Despite the increased number of research papers and researchers in the steel research field and enhanced diversity, there has been a lack of research focused on trend analysis in the field. Although countless research studies have been published in Steel Research International since 1990, there has yet to be an analysis done on the journal regarding the comprehensive academic advance and major field of research significance being carried out by global researchers.

In the present paper, the big data analysis of abstract keywords in Steel Research International has been investigated, building data on current research themes in steel for future research development efforts. This analysis can provide a link between past, present, and future research in the area of iron and steel. For example, current classification methods were accomplished through mostly insight and a broad division between physical and chemical metallurgy. However, a more detailed categorization can also be realized through quantitative data analysis, which gives a more objective and detailed research categorization resulting in a defined knowledge structure base.

This research is mainly designed to analyze the knowledge structure in steel research and identifies comprehensive research trends related to steel research.

II. THEORETICAL BACKGROUND

A. Knowledge structure analysis

Methods used to identify knowledge structures can be broadly divided in to qualitative and quantitative. The qualitative approach relies upon a small number of experts, which is a relatively simple, but is highly subjective and defined to a specific knowledge area and experience based on the experts that could partially neglect the total knowledge structure analysis.[1] To overcome the restrictions of the qualitative method, a more reliable quantitative method can be used. The quantitative method focuses on identifying certain patterns in research concepts and explaining them through mathematical and statistical methods.[2]

The quantitative way falls into two categories. The author co-citation analysis, a useful tool to identify and prove the knowledge structure of a certain research area through resorting to authors as an analytical unit, and the keyword co-occurrence analysis, compromised with the context of research papers having handled keyword combination of the same paper. The author co-citation analysis describe authors of a paper within a distinct research field and gives information on the influence of the authors determining the leading contributors to the field of research. However, there is the restriction that the research paper’s context, which is the most meaningful facet of the paper, could be ignored from the analysis.[3, 4] Additionally, the keyword co-occurrence analysis, which implements the representative keywords of the research subject in the papers, shows to be more useful for increased awareness of the knowledge structure in the research area.[6, 7] The keyword or concept that is most frequently mentioned in a certain research area can represent the field because it normally
corresponds to the terminology of the methods, concepts, materials, or equipment common to that field.

B. Social network analysis

Social Network Analysis (SNA) is a method to deduce network structures, patterns and forms and to explain the relationship between actors by quantitatively analyzing networks consisting of nodes, actors, and links.[5, 6] SNA differs from previous social science methods in that previous research was focused on an individual’s independence, not interdependency. According to the analytical approach, an actor is expressed as a node and a relationship is expressed as a link. A set of nodes and links forms a social network.[7] This study will apply SNA to investigated the knowledge structure of steel research with various analytical factors designed for this. First, the node that has the power of influence and dominance in a network can be identified through a centrality concept.[8] Second, the structural characteristic in a knowledge network can be identified by considering the cohesiveness of the subgroup.[9] With the acceleration of convergence in research fields, describing the research concepts in distinct fields have become an essential priority for researchers. The review of big articles and extracting the important research subjects within the context of the paper is the favored method of analysis. However, this method is complicated needing significant time and effort not to mention a final result that can be difficult to comprehend since different opinions exists on describing the concepts between various researchers. In this study, we were able to verify structural characteristics in steel research by implementing SNA to describe the steel research movement in the journal of Steel Research International (Described hence forth as “The Journal”).

III. METHODOLOGY

Analysis of this research has been performed through data collection, composition of keyword co-occurrence matrix, and SNA (centrality and cohesive subgroup analyses)

A. Data collection

First, the process of selecting papers was performed. 24 years of articles increasing to almost 2400 papers presented in the Journal from 1990 volume 61 to 2013 vol. 84 have been investigated. Journals typically contain technical articles, reviews, commentaries, editorial rules, and correction statements. In this study, technical articles and reviews have been classified and used, excluding others since the contents do not directly relate to the research concepts in question. Next, from this big data analysis, the keywords of the articles have been derived to discern the knowledge structure of steel research. Keywords represent the contents of relevant research papers. The list of keywords produced by the authors has been used. However, this is limited by the lack of standardization. To overcome these limitations, filtering was carried out on the extracted keywords in this study. After the filtering, almost 5700 keywords remained from 54 different countries. Visualization network of keywords is indicated in Fig. 2.

B. Analytical procedures

The analysis was performed in the following arrangement. First, the frequency of the collected data per year was investigated. Frequency analysis is a typical technique to explain the advances arising in the research contents as time passes.[10] Next, the centrality and cohesive subgroup analysis were performed. The centrality indicators - degree, closeness, and betweenness centrality - and the keywords located in the center change depending on the definition; the keyword has diverse connected relationships with other keywords or has control or influence on other keywords.[11, 12] A keyword which is connected with many other keywords is identifiable through degree-centrality. The geodesic distance among keywords measures the closeness-centrality which is distance between keywords within the network that regards not only direct connection but also indirect connection. The betweenness-centrality can measure whether a keyword is available for the role of broker to build steel research boundaries. This is checked by the frequency of linking the shortest path among keywords. Cohesive subgroup analysis provides a recognizing of the subgroup of research in steel research. The presence of a keyword group can be considered to be a sub research area. Cohesive subgroup analysis of keywords in the network structure allows the grouping of closely connected keywords into subgroups of steel research simplifying a complex network structure.[11] In addition, the structural linking between and within subgroups of keywords can be determined.

Finally, proving the research area mapping, the properties of the subgroups of research concepts can be recognized. The research area mapping describes the inter connection of the
keywords between individual research areas and the intra connection of the keywords within research area on a two dimensional plane resulting in the understanding of development process of the research, [13, 14] Inter-connection between the subgroups is the outside relation between a keyword within a defined research group and keywords of a separate research group. Intra-connection of the subgroups manages with keywords within the delineated research group and its relation with other keywords within the same research group. The research area mapping can be recognized as a meaningful indicator because areas in the developmental stages in steel research, and areas that need further research, are revealed through this process.

IV. RESULTS AND DISCUSSION

A. Keyword occurrence analysis

Among various analytical methods targeting keywords or concepts, a general approach is to count the frequency of keyword occurrence analysis. Since frequently occurring keywords usually represent the dominant steel research area in corresponding periods, it is advantageous to identify the transition of research chronologically through changes in frequently used keywords, [10] Table I was displayed the top 10 keywords that were appeared from 1990 to 2013. And Fig. 2 was visualized the relation between the top 10 keywords. Through visualization, relationships between keywords could be identified that couldn't be identified by text formats.

During the period, keywords occurring more than 50 times are microstructure, finite element method, transformation induced plasticity, mechanical properties, continuous casting, slag, and blast furnace. It may be reflected that the 1997 Kyoto protocol and the forecasted regulation of CO2 needed the automotive and manufacturing industries to assume more closely into environmental issues and in particular the auto industry to develop new light weight automotive vehicles using TRIP. This trend extends with the improvement and extension of advanced high strength steels such as TWIP (twinning-induced plasticity). In addition analysis, microstructure co-occurred most with mechanical properties (21 times), heat treatment (8 times), hardness (6 times), and phase transformation (5 times). Finite element method was mainly researched with sheet metal forming (5 times), springback (4 times), forging (4 times).

<table>
<thead>
<tr>
<th>No.</th>
<th>Keywords</th>
<th>Freq. of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>microstructure</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>finite element method</td>
<td>79</td>
</tr>
<tr>
<td>3</td>
<td>transformation induced plasticity</td>
<td>72</td>
</tr>
<tr>
<td>4</td>
<td>mechanical properties</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>continuous casting</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>slag</td>
<td>54</td>
</tr>
<tr>
<td>7</td>
<td>blast furnace</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>kinetics</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>stainless steel</td>
<td>41</td>
</tr>
<tr>
<td>10</td>
<td>heat transfer</td>
<td>39</td>
</tr>
</tbody>
</table>

B. Centrality analysis

Centrality analysis was performed to identify the core keywords among those that have occurred over a 24-year span in Steel Research International. The analysis target 5700 keywords. In Table II, the keywords and the values of centrality are shown. Here, the results of degree, closeness, and betweenness centrality, which are thought to have meaningful value for this research, are suggested.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>microstructure</td>
<td>3.05</td>
<td>34.40</td>
<td>9.53</td>
</tr>
<tr>
<td>finite element method</td>
<td>4.72</td>
<td>30.90</td>
<td>7.71</td>
</tr>
<tr>
<td>continuous casting</td>
<td>3.87</td>
<td>32.41</td>
<td>6.43</td>
</tr>
<tr>
<td>transformation induced plasticity</td>
<td>3.50</td>
<td>31.38</td>
<td>4.67</td>
</tr>
<tr>
<td>mechanical properties</td>
<td>3.30</td>
<td>32.57</td>
<td>5.05</td>
</tr>
<tr>
<td>slag</td>
<td>3.29</td>
<td>31.87</td>
<td>5.35</td>
</tr>
<tr>
<td>blast furnace</td>
<td>2.90</td>
<td>29.44</td>
<td>4.03</td>
</tr>
<tr>
<td>kinetics</td>
<td>2.44</td>
<td>31.74</td>
<td>4.66</td>
</tr>
<tr>
<td>stainless steel</td>
<td>2.33</td>
<td>31.32</td>
<td>3.61</td>
</tr>
<tr>
<td>heat transfer</td>
<td>2.01</td>
<td>29.91</td>
<td>2.77</td>
</tr>
</tbody>
</table>

The number of keywords has dramatically increased over the years, indicating that various keywords have emerged in steel research and cross over research between keywords has been performed. Based on this, it can be contended that steel research has combined and become more detailed. The degree-centrality was identified to be topmost for keywords of microstructure, finite element method, continuous casting, transformation induced plasticity, mechanical properties showing that the broad field of physical metallurgy to be pronounced than chemical metallurgy. The closeness-centrality was described to be topmost for the keywords in the order of microstructure, mechanical properties, continuous casting, slag, kinetics, stainless steel and shows the research topic that can speedily spread knowledge to the whole steel research network. The betweenness-centrality, which can best bridge the
C. Cohesive subgroup analysis

Cohesive analysis was performed to have keywords within a specific research group to have significantly more links than cross linking between research groups in order to apparently identify independent research areas that can be grouped. For cohesive subgroup analysis, keyword to keyword pairing that occurred less than five times within a particular group, which was 94.2%, were deleted from the cohesive subgroup analysis since connectivity of 5 or less is usually not described as an independent subgroup of research fields. Through this analysis, eight cohesive groups (subordinate research areas) were discovered. Table III shows each group with corresponding keywords.

In the final step, identified subordinate research fields were placed on the research area mapping depending on both the maturity and the independence of the research areas. The mappings revealed the structure of subordinate research areas and calculated the connecting of keywords in intra and inter groups. The research areas mapping, which aid in revealing research trends in steel research, can be categorized into four different types.[14] The division of subordinate research areas through mapping is shown in Table IV. The intra connection identifies how densely populated and related the keywords are within the group and the inter connection is the link between the in-group and out-group keywords.

V. CONCLUSION

The purpose of this research is to provide research area mapping for future research through the identification of knowledge structure in steel research. For this, occurrence analysis and SNA were implemented on keywords published in Steel Research International from 1990 to 2013.[15]

This research was conducted in four steps to reveal the knowledge structure in steel research. The occurrence frequency keyword was identified. High frequency keywords not only represent steel research but also aid in identifying the transition process of steel research when the data is analyzed chronologically. Based on this analysis, keywords representing steel research for the 24-year period are microstructure, finite element method, and transformation induced plasticity. Next, central keywords in steel research were describes by degree, closeness, and betweenness centrality analysis. Keywords showed high correlated between centrality indicators and the frequency of occurrence. For the research period, keywords showing high degree centrality are microstructure, finite element method, and continuous casting. The third step identified subordinate research area in steel research through cohesive subgroup analysis in SNA. Through the result of this analysis was found eight research areas. In the final step, determined subordinate research areas were placed on the research map depending on both the maturity and the independence of research areas.

In this study, several methodologies - occurrence analysis, centrality and cohesive subgroup analyses from SNA - were implemented to identify research trends, subordinated research areas and core keywords in steel research. Also, through research areas mapping, dynamic knowledge structure in steel research had been described through classification of the research areas. Additionally, this research will provide the direction to steel researchers for determining future research area.

This research tried to thoroughly consider how steel research has developed. In the analysis of journal big data, much has been learned, but also restrictions in this type of arrangement must be addressed. First, this work was restricted to the journal of Steel Research International. However, considering other journals are present and that steel research is now interdisciplinary, the present study does not involve all of steel research currently being studied. Thus, an extension of other journals of steel research in the research area of metallurgy and also interdisciplinary journals should be addressed to organize into the current data. Second, although the research areas mapping is based on the quantitative method, opinions of experts from subordinated fields in steel research are necessary to analyze qualitative meanings. Third, the keywords were picked from the listed keywords within the article. However, there have been cases, where the keywords do not sufficiently represent the main objective of the paper and the results. For a more trustworthy analysis, keywords within the article must be pre-filtered within any journal and meaningful efforts within the publishing community are underway to improve the selection of keywords within articles. Furthermore, to enhance the analysis and provide keywords are representative of the article, a system of correct keyword
selection for authors should be established during the submission and review process.

ACKNOWLEDGEMENTS

The material for this paper is currently in-press for Steel Research International[15] and details of the analysis is more comprehensive. The present work has been supported by the project from the Ministry of Trade, Industry, and Economy (MOTIE).

REFERENCES