Detecting Candidate Combinations of the Keywords Organ – Material –Technology in Regenerative Medicine

Hiroko Yamano, Hajime Sasaki, Ichiro Sakata
Policy Alternatives Research Institute, University of Tokyo, Japan

***Abstract*—Regenerative medicine, especially in the area related to biomaterials, is a rapidly advancing interdisciplinary field. Although there are many possibilities in applying materials and technologies in the treatment of different organs, little is quantitatively known about candidate combinations because this field requires extensive knowledge over multiple research fields.**

The purpose of this research is to detect candidate combinations of the keywords Organ – Material - Technology using a citation network and text-based co-word analysis. We first analyzed the citation network of academic papers to visualize an overview of regenerative medicine. We retrieved papers from the Web of Science using specific search queries. Using a topological-based method, papers were categorized into clusters according to the feature words.

After obtaining the high-ranked keywords of the main clusters, we divided them into three regions: Organ, Material, and Technology. By comparing the co-occurrence of words among the three regions, we detected plausible linkages within a cluster as well as existing linkages within each paper in the whole dataset. The results suggest the potential effectiveness of new combinations that have not yet been examined in detail, and can be used to predict niches and missing links in regenerative medicine.

I. INTRODUCTION

Currently, it has become more difficult for researchers to capture an overview and trends of their research interest destination due to the expansion of academic papers [1][2]. In the interdisciplinary field, the difficulty increases even more when multiplying the adjacent regions concerned. In addition to the massive amount of total information, there is a tremendous number of possible combinations of contents among related regions. This paper proposes a method of reducing the researchers' load, and of exploring the adjacent regions lying outside their specialty that would also help to inspire new research ideas in regenerative medicine.

Regenerative medicine is a highly interdisciplinary field involving engineering, life sciences, and material sciences that have been pursued for treating organs or tissues in the human body. There have been three general strategies for the creation of tissues: cells, signal molecules, and matrix [3]. The ability to repair tissue is a fundamental property of all multicellular organisms that proceeds with the interaction of cells and the extra-cellular matrix (ECM): the physico-chemical environment of cells. Although previous attempts to elaborate the complexity of the repair process have not succeeded in providing both "seed and soil" in therapy [4], recent advances in life science have revealed the importance of ECM that regulates cell behavior. In fact, there is a growing realization that, for cell therapies to succeed, it may not be sufficient to implant cells without appropriate ECM [5][6]. If humans are the model organism, cross-disciplinary collaborations and technology integrations will be essential for moving away from the study of single cells towards the medical goal [7].

Several attempts have been made to reveal the trends of regenerative medicine using citation analysis of papers [8][9][10][11][12]. Shibata [8] and Chen [9][10] proposed a method of detecting emerging research trends analyzing the whole structure of the citation network. Zhao [11] found the dominance of a few research areas in this field using all-author citation counting. He also found author selection had a greater effect on mapping results. [12]. Bibliometrics and text mining techniques have also been applied to this field [13][14]. Li [13] compared the term frequency of author keywords in each delimited period. An [14] analyzed the co-occurrence of words in medical subject headings. Both Li and An examined trends using the metadata of papers in the stem cell field, but there are difficulties in term selection. Although term frequency or experts consulting can be clues for selecting terms, it is not sufficient for targeting the proper abstraction level or scope of terms from the perspective of guiding researchers in exploring new branches. In short, these text-based approaches have not succeeded in providing more than a global overview in regenerative medicine, either.

Advanced investigations into literature-based discovery are seen in different fields. Ittipanuvat [15] examined the linkages between technology and a social issue: robotics and gerontology. He proposed a method of assisting humans in selecting terms, using semantic similarity to measure relatedness between citation network clusters of the two domains. However, as the relatedness was simply examined by comparing one value that represents each cluster pair, the distribution of terms in the cluster is not deeply considered. In addition, although a term's importance was calculated by *tf-idf*, the criteria for the final selection of terms are unclear as both the top- and bottom-listed terms were referred to alike.

In this study, we aim to detect potential candidate combinations of the keywords Organ – Material - Technology in regenerative medicine. To meet this challenge, we made a conditional model combining two values of the keywords, which are evaluated as "important, but not researched well". We applied citation network analysis and text-based co-word analysis to evaluate the frequency of the combination of keywords in academic papers. Our results can offer researchers a guide to examining adjacent regions in this field.

II. METHODOLOGY

An outline of the analysis schema in this research is shown in Fig. 1. First, we analyzed the citation network of academic papers to visualize and grasp an overview of the targeted knowledge domain, according to the methodology of previous research [1][2]. To focus on specific areas in regenerative medicine, we retrieved papers using the queries described in the next paragraph. After visualizing the clusters of the citation network, we classified the keywords of each cluster into three regions: Organ, Material, and Technology. Then, we selected a cluster by degree of share of the classification. To detect the candidate combinations in the three regions, we made a conditional model, which defines the range of values of the classified keywords. Finally, we made heatmaps to evaluate this model, visualizing the frequency of co-occurrence by the classification.

Taking into account the importance of the cell environment, we refined the corpus to the area related to the extra-cellular matrix, which is an advanced interdisciplinary area in regenerative medicine. The query set of prior research is designed to retain wide coverage of citation data of regenerative medicine by using multiple queries: “regenerative medicine*”, “ES cell*”, “embryonic stem cell*”, “embryo-derived stem cell*”, “ips cell*”, “pluripotent stem cell*”, “adult stem cell*”, or “somatic stem cell*” [8]. In this study, we further extended the corpus by adding “tissue engineering*” to the query set, and filtered the data by “extra-cellular matrix*” or “ECM*” (Fig. 1). We retrieved academic papers including these queries in the titles, abstracts, keywords, and keywords plus from the following citation database: the Science Citation Index (SCI-EXPANDED), the Social Sciences Citation Index (SSCI), and the Arts & Humanities Citation Index (A&HCD) compiled by Thomson Reuters.

Using citation network analysis, the network is divided into clusters by the topological clustering method [16]. The maximum connected component of the citation network in each cluster is visualized by using a large graph layout (LGL) [17]. LGL is a spring layout algorithm to visualize large networks, which was originally designed for studying biological models. The intra-cluster links were provided with the same color to seize the positional relation of each cluster. After visualizing, we analyzed the characteristics of the top six clusters that have more than 200 papers, by comparing the distribution of their keywords and keywords plus. Both of the keywords were ranked by *tf-icf*, a measure of the importance of the term, based on the method of term frequency multiplied by inverse document frequency *(tf-idf)*. The *tf-icf* weight of term *i* in cluster *s* is the term frequency multiplied by the inverse cluster frequency, given by:

$$W_{i,s} = tf_{i,s} \times icf_i, \tag{1}$$

where

$$tf_{i,s} = \frac{tC_{i,s}}{N_s} \text{ and} \tag{2}$$

$$icf_i = \log\left(\frac{M}{cf_i}\right). \tag{3}$$

Here, $tC_{i,s}$ is the number of occurrences of term *i* in cluster *s*, N_s is the total number of terms in cluster *s*, cf_i is the number of clusters containing term *i*, and M is the total number of clusters.

In this study, we compared high-ranked keywords in each cluster, and classified the top 100 keywords into three regions: Organ, Material, and Technology. The criteria for the classification are as follows. We included cell names and cell adhesion molecules in the class of Organ, whereas extra-cellular-matrix-related molecules were included in Material. Disease names were classified if they were used or related to an organ whereas liquid factors, like growth factors, were not classified because their localization in the body is unclear. We excluded general terms that represent species like “human”, “mouse”, “porcine”, “organ”, “tissue”, and

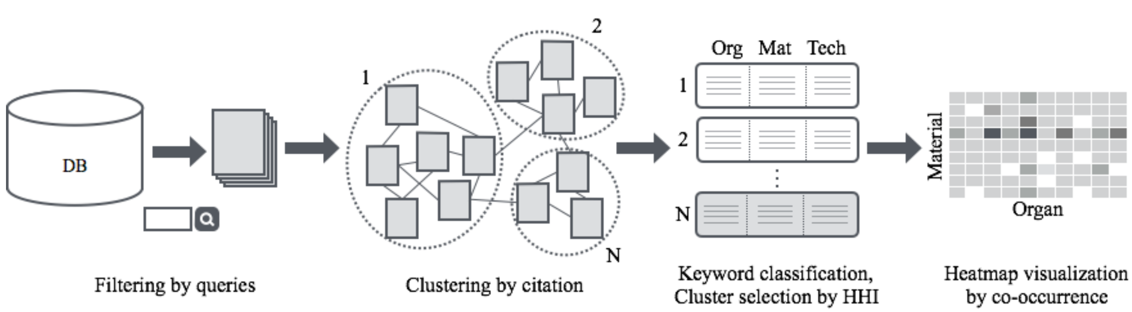

Fig. 1 Outline of methodology in this paper.

Regenerative Medicine	ES cell	iPS cell	Tissue Engineering
Extracellular Matrix			

("regenerative medicine*" or "ES cell*" or "embryonic stem cell*" or "embryo-derived stem cell*" or "ips cell*" or "pluripotent stem cell*" or "adult stem cell*" or "somatic stem cell*" or "tissue engineering*") **and** ("extracellular matrix*" or “ECM*")

Fig. 2 Query set to refine the corpus of regenerative medicine.

"cell", as well as materials such as "fiber", "substrate", "film", "sheet", "material", and "biomaterial", from the classification. We classified related abbreviations, but not adjectives and verbs unless they were used or related to the classified keywords. Keywords consisting of two or three terms were also classified.

After classifying the keywords, we investigated the degree of share of the three regions in each cluster. We calculated the sum of the term frequency and Herfindahl-Hirschman Index (HHI), which is typically used to investigate industrial oligopoly, defined as

$$H_{i,s} = \sum_{i=1}^{N_s} (tf_i)^2. \quad (4)$$

For further investigation, we selected a cluster with the lowest value of HHI, as the risk of counting senseless combinations was comparatively lower in the selected cluster, which will be explained later. To simplify the following calculation and evaluation, we added a unique ID to each keyword. These IDs were composed of three digits, which represent cluster, classification, and *tf-icf* ranking. For instance, the ID of the keyword "disc" is "3o5", which means that the keyword belongs to the third cluster in the whole network, is categorized in Organ, and has a *tf-icf* ranking in the cluster of 5. We counted the number of papers that have pairs of keywords from different regions in their abstracts, which we call "co-occurrence" in this study. We used the keywords from the selected cluster, and counted their co-occurrence in the paper abstracts from every cluster.

To detect possible candidate combinations of Organ - Material - Technology, we made the conditional model shown in Fig. 3. Dark red represents a high degree of *tf-icf*, and the distance between the circles represents the number of co-occurrences. The designated candidate combination of keywords is suggested by the dotted line. We assumed that keywords with high *tf-icf* tend to co-occur more, which means that many researches have already applied the keyword pair, and that those with a low *tf-icf* tend to co-occur less, which means that few researches have applied the keyword pair. In this model, we distinguished the keyword combination with a high *tf-icf* and low co-occurrence, which would be evaluated as "important, but not researched well".

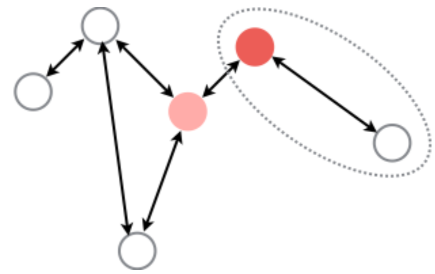

Fig. 3 Model of candidate combination of keywords

We distinguished the pairs that have conditions of fewer than 10 co-occurrences, *tf-icf* rankings of 5 or less and 60 or less in each region. The pairs whose sum of *tf-icf* ranking was lower than 30 were also included. We summed up the co-occurrence of the pairs with the same meaning in their keywords, and excluded the pairs when they exceeded the setting value.

Finally, we made heatmaps that represent co-occurrence of all combinations of the classified keywords, to evaluate this model and its results. The keywords Organ and Material were applied to columns and rows representing each *tf-icf* ranking. We compared the distribution of the co-occurrence in abstracts from every cluster with those from the selected cluster.

III. RESULTS

The number of papers on regenerative medicine published before the end of 2014 was 50,654, using the queries of the previous research. By adding the query "tissue engineering*", the number increased to 77,272. By filtering the papers with the query "extra-cellular matrix*" or "ECM*", the number decreased to 6,852. In the 6,852 papers obtained, 6,281 papers were included in the largest connected component in the citation network to be divided into 39 clusters. The number of annual publications in each cluster is shown in Fig. 4. The top four clusters are increasing rapidly, whereas the others are increasing gently.

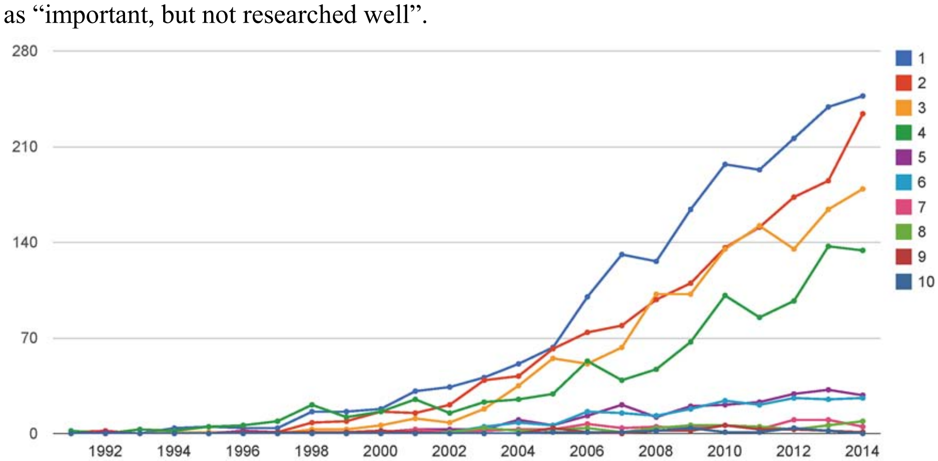

Fig. 4 Number of papers including the query set for regenerative medicine in each cluster.

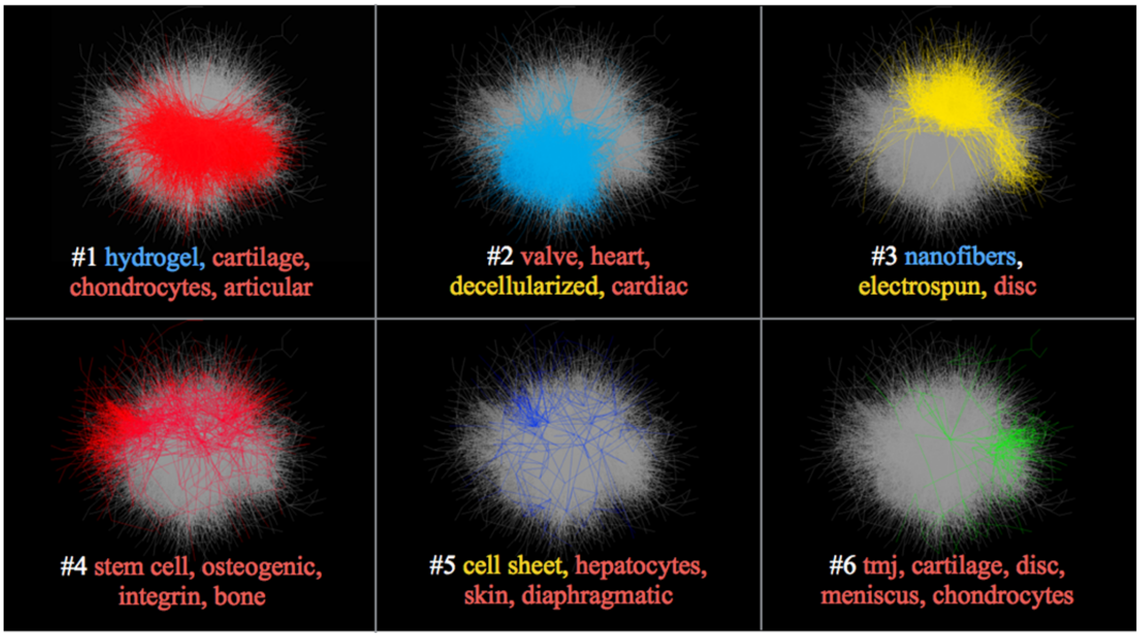

Fig. 5 Visualization of the citation network of the top six clusters in regenerative medicine.

TABLE 1 ORGAN - MATERIAL - TECHNOLOGY CLASSIFICATION OF THE KEYWORDS IN THE TOP SIX CLUSTERS

Cluster	Ranking	Organ	Ranking	Material	Ranking	Technology
1	2	cartilage	1	hydrogel	24	compression
1	3	chondrocytes	7	peptide	34	dimensional
1	4	articular	8	alginate	42	mechanical property
2	1	valve	34	elastin	3	decellularized
2	2	heart	73	gel	5	decellularization
2	4	heart valve	81	hydrogel	9	graft
3	5	disc	1	nanofibers	2	electrospun
3	15	intervertebral disc	3	nanofibrous	4	electrospinning
3	17	bone	7	nanofibrous scaffold	13	composite
4	2	stem cell	4	integrin	96	mineralization
4	3	osteogenic	20	laminin	141	dimensional
4	5	bone	52	integrins	237	bioreactor
5	4	hepatocytes	14	gelatin	2	cell sheet
5	7	skin	18	polymer	12	temperature responsive
5	8	diaphragmatic	22	pnipaam	13	detachment
6	1	tmj	17	tgf	24	hypoxia
6	2	cartilage	34	gag	66	piii
6	3	disc	35	collagen type	94	oxygen tension

Analyzing the characteristics of the top six clusters that have more than 200 papers, we realized that the keywords of the clusters can be classified into three regions: Organ, Material, and Technology. The position of the top six clusters in the whole network, and the high-ranked keywords in each cluster, are shown in Fig. 5. Keywords that represent Organ, Material, and Technology are colored red, blue, and yellow, respectively. We assumed that the deviation among the three regions would be different according to the cluster, comparing the *tf-icf* ranking of each classified keyword (Table 1).

To confirm this, we classified the top 100 keywords into the three regions, and estimated the degree of share of each cluster, by calculating the sum of the term frequency and the Herfindahl-Hirschman Index (HHI) value (Table 2). As a result, the clusters were indicated by different levels of deviation. Comparing high-ranked papers in each cluster, we noticed that papers in cluster 4 and cluster 6 were basic research in histology, whose keywords in Technology were few. Papers in cluster 2 were related to decellularization, which is a technology to isolate the extra-cellular matrix to prevent organ rejection caused by antibodies on cell surfaces. The methods used to lyse, kill cells, or break bonds and to remove cells from the matrix by physical or chemical or enzymatic treatments, whose keywords were related to Material, were few. Taking these results into account, we selected cluster 3, which had the lowest HHI value, for further investigation. As the three regions were relatively uniform in cluster 3, the risk of counting senseless combinations was lower than the other clusters. Cluster 3 consisted of 1,226 papers related to nanotechnology.

TABLE 2 SUM OF TERM FREQUENCY AND HERFINDAHL-HIRSCHMAN INDEX (HHI) VALUE IN THE TOP SIX CLUSTERS

No.	Cluster name	Sum of term frequency			HHI value
		Organ	Material	Technology	
1	Hydrogel and Cartilage	0.0300	0.0270	0.0100	0.0017
2	Valve and Decellularization	0.0535	0.0025	0.0190	0.0032
3	Nanofibers and Electrospun	0.0283	0.0182	0.0085	0.0012
4	Bone and Integrin	0.0442	0.0056	0.0005	0.0020
5	Cell sheet and Hepatocytes	0.0346	0.0094	0.0102	0.0014
6	Tmj and Cartilage	0.0692	0.0098	0.0014	0.0049

TABLE 3 IDS OF THE TOP 100 KEYWORDS IN THE NANOTECHNOLOGY CLUSTER, WHICH REPRESENT CLUSTER, CLASSIFICATION, AND RANKING OF *TF-ICF*

ID	Organ	ID	Material	ID	Technology
3o5	disc	3m1	nanofibers	3t2	electrospun
3o15	intervertebral disc	3m3	nanofibrous	3t4	electrospinning
3o17	bone	3m7	nanofibrous scaffold	3t13	composite
3o19	intervertebral	3m8	nanofiber	3t37	electrospun scaffold
3o22	nerve	3m12	pcl	3t56	electrospun nanofibers
3o27	ligament	3m14	chitosan	3t58	electrospun fiber
3o30	tendon	3m16	polymer	3t88	degradation
3o31	osteogenic	3m18	silk		
3o32	mesenchymal stem	3m39	gelatin		
3o33	mesenchymal stem cell	3m57	plla		
3o34	stem	3m60	fibroin		
3o35	stem cell	3m73	lactide		
3o38	osteoblast	3m76	nano fibrous		
3o40	nucleus pulposus	3m77	plga		
3o41	pulposus	3m83	nanofiber scaffold		
3o44	msc	3m84	lactic		
3o45	bone tissue	3m85	bioactive glass		
3o49	mesenchymal	3m89	silk fibroin		
3o50	nucleus	3m96	nanofibres		
3o53	cartilage	3m98	glass		
3o62	corneal				
3o94	osteogenic differentiation				
3o97	fibroblast				

The unique IDs of the top 100 keywords in the nanotechnology cluster are represented in Table 3. There were 23, 20, and 7 keywords classified into Organ, Material, and Technology, respectively. The number of unclassified keywords was 50, following the classification criteria described above.

The candidate combinations of keywords are shown with their ID and co-occurrence (Table 4). Ten pairs were detected within the setting range of the conditional model. However, there were pairs that had more than 200 co-occurrences whose *tf-icf* ranking was not very high.

To evaluate this model and its results, we made heatmaps that represent the co-occurrence of all combinations of the classified keywords in Organ and Material in the nanotechnology cluster (Figs. 6 and 7). The depth of blue represents the degree of co-occurrence. Cells with more than 100 co-occurrences were colored the same dark blue. Cells with no combination were colored gray. The numbers in the columns and rows represent the *tf-icf* ranking. In both heatmaps, Material 16 and Organ 34, which represent “polymer” and “stem”, are the highest among all keywords. Material 12, 57, and 77 that represent “pcl”, “plla”, and “plga” did not co-occur with any keywords in Organ, whereas every keyword in Organ co-occurred with some or all keywords in Material. Co-occurrence was not always high in the pair of high-ranked *tf-icf* in both heatmaps. In the nanotechnology cluster, high-ranked keywords in Material tended to co-occur more with the keywords in Organ generally (Fig. 7), whereas no tendency were seen in the co-occurrence of high-ranked keywords in general, in all clusters (Fig. 6).

TABLE 4 CANDIDATE COMBINATION OF KEYWORDS IN THE NANOTECHNOLOGY CLUSTER

Keyword1	Keyword2	ID1	ID2	Co-occurrence
disc	fibroin	3o5	3m60	3
ligament	nanofibrous	3o27	3m3	6
tendon	nanofibrous	3o30	3m3	5
nucleus	nanofibrous	3o50	3m3	1
tendon	electrospinning	3o30	3t4	7
nucleus	electrospinning	3o50	3t4	2
nucleus	electrospun	3o50	3t2	1
intervertebral disc	chitosan	3o15	3m14	3
intervertebral disc	electrospinning	3o15	3t4	2
intervertebral	electrospinning	3o19	3t4	2
bone	polymer	3o17	3m16	258
stem	polymer	3o34	3m16	490
stem cell	polymer	3o35	3m16	253
cartilage	polymer	3o53	3m16	203

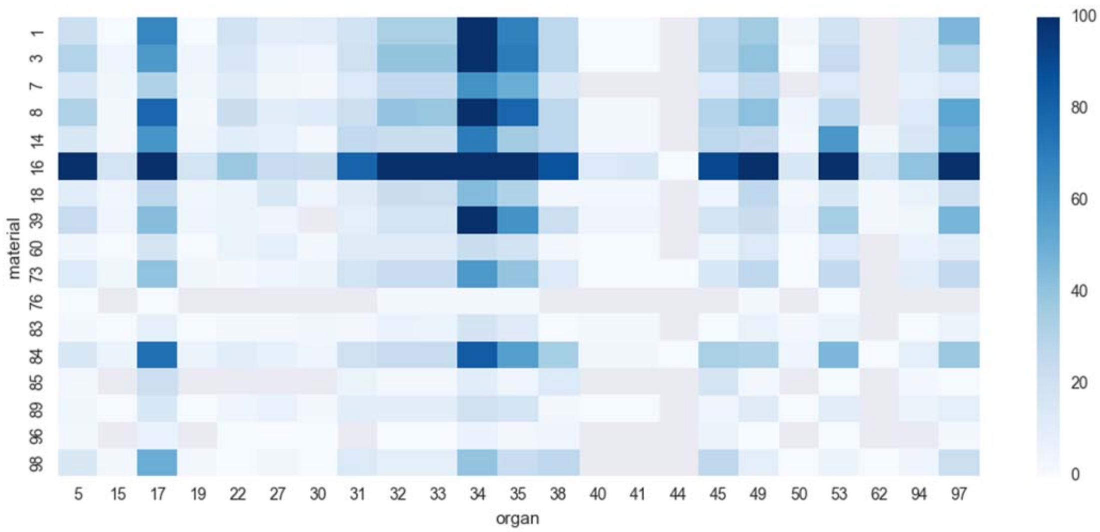

Fig. 6 Heatmap of Organ - Material co-occurrence in all clusters.

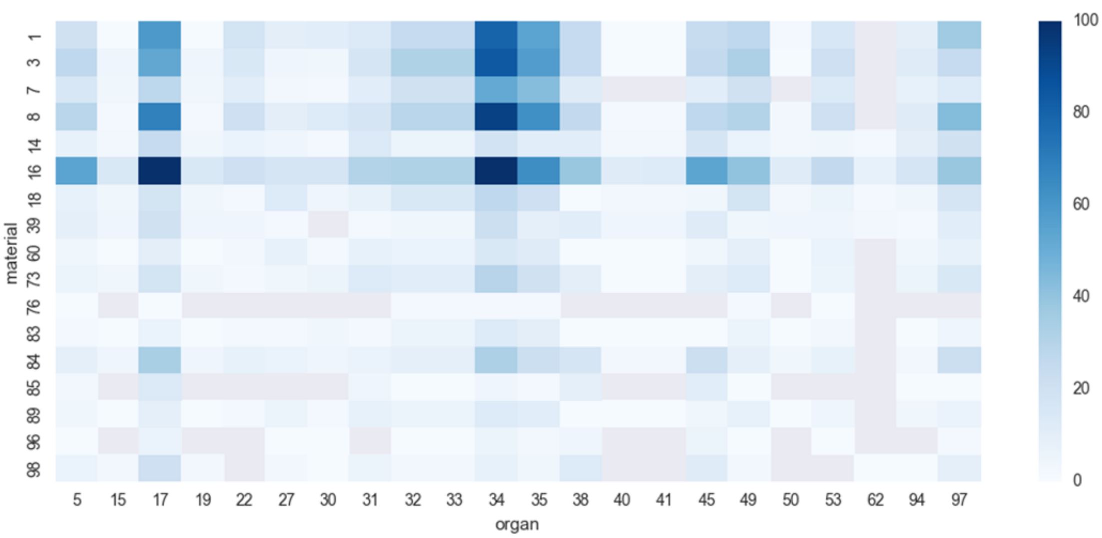

Fig. 7 Heatmap of Organ - Material co-occurrence in the nanotechnology cluster.

However, we detected a deviation similarity and outlier between the keywords in the columns and rows. For example, Organ 53 and 97, which represent "cartilage" and "fibroblast", were similar in the deviation of co-occurrence, but they had an extreme deviation in the cells corresponding to Material 8 and 14, which represent "nanofiber" and "chitosan".

IV. DISCUSSION

Previous studies have proposed methods of showing the possible academic fields of association, fusion, or technology transfer, which can be effective in decision making for R&D managers or policy makers [8][15]. On the other hand, the proposed method in this study was designed to provide a more concrete guide for researchers.

The topological positions of the top six clusters represent the relatedness of each cluster calculated by citation data (Fig.5). In cluster 3, which was the nanotechnology cluster, high-ranked keywords were distributed in the three regions, Organ, Material, and Technology.

In this study, we selected the nanotechnology cluster whose HHI value was the lowest (Table 2). The clusters with a high HHI value had a risk of having some reason for the share; for example, basic researches do not tend to refer to Technology, or methods of extracting whole ECM do not tend to refer to each Material. Therefore, we did not investigate high-HHI-value clusters, but this does not mean that the nanotechnology cluster is the only cluster to be investigated. It would be meaningful to investigate the lower-value HHI clusters, such as the "hydrogel and cartilage" cluster or the "cell sheet and hepatocytes" cluster. According to an expert hearing in the field of nanotechnology, it is reasonable to have such a low HHI value, because nanotechnology is a field that cannot be established without collaboration with researchers from various fields.

The experimental results showed the deviation of keywords in Organ, Material, and Technology among the clusters in regenerative medicine. Under the setting criteria, we showed the candidate combinations of keywords with high *tf-icf* and low co-occurrence (Table 4).

For example, we detected the combination of “disc” and “fibroin” whose co-occurrence count was only 3, even if the importance of “disc” is the highest in the nanotechnology cluster. Among the three papers that referred to both “disc” and “fibroin”, Bhattacharjee’s paper [18] had the highest citation number of 15. He developed a designed scaffold of silk fibroin with chondroitin sulphate (CS), resembling the annulus fibrosus of intervertebral discs, in 2012. Human chondrocytes cultured over the scaffold uniformly followed the silk fiber alignment and deposited ECM maintaining their re-differentiation. Although the outcome of the combined effect of cell/matrix alignment and chondrogenic support reveals the availability of the combination of “disc” and “fibroin”, few researches have experimented with this combination. Therefore, there is a possibility of a niche or missing link lying between “disc” and “fibroin”.

The same can be applied for other combinations such as “ligament” and “nanofibrous”, or “tendon” and “nanofibrous”, as their co-occurrence was 6 and 5 count, even if the *tf-icf* ranking of “nanofibrous” is third. We noticed adjectival, plural, or compound synonyms, as “nanofibrous” is similar to “nanofibers”, “nanofiber”, “nanofibrous scaffold”, “nano fibrous”, “nanofiber scaffold”, and “nanofibres”. One solution is simply to combine the values of these synonyms. However, the distinction between noun and adjective or other minor differences of inflection also has certain meanings for the researchers in this field. For further investigation, the abstraction level of the keywords should be modified to an adequate range according to the purpose. If the abstraction level is high, the outcome will be global; if the abstraction level is low, the outcome will be detailed.

There is the possibility of another classification or sub-classification into regions such as “liquid factor”, “substrate property”, and “structure of scaffold”, instead of Organ - Material - Technology. To examine these, it is necessary to expand the corpus, as keywords corresponding to these regions were rare in the top 100 keywords. We limited the classification range to the top 100 keywords this time, while there are 134,328 keywords in the nanotechnology cluster. To expand the corpus of keywords, the classification process is expected to be automated, like using corresponding dictionaries. However, compared to the existing Organ or tissue expression dataset, it is difficult to find a proper dataset of Material or Technology, as these are rapidly expanding categories. Therefore, in using such dictionaries or datasets, we should be aware of the limitation of their range. In addition, if we expand the corpus, the possibility of noise arising, as well as the desired outcome suited to a specific region, will increase. There is a trade-off in selecting the threshold in each case.

Using heatmaps, we visualized the co-occurrence of all combinations of the keywords Organ and Material in the nanotechnology cluster (Figs. 6 and 7). Contrary to our expectation, co-occurrence was not always high in the pair of high-ranked *tf-icf*, and there were many candidate combinations within the setting range of the conditional model. It is an inevitable result, as *tf-icf* reflects the inverse cluster frequency, which is a characteristic that other clusters do not have. Different deviation would be seen if we take *tf* or *tf-idf* on behalf of *tf-icf*. The deviation similarity and outlier between the keywords in the same region we detected this time would be useful, as there is potential for development into a recommendation function. For example, "mesenchymal" and "cartilage" that correspond to the lane of Organ 49 and 53, are similar in the deviation of the co-occurrence of the keywords in Material. However, as for the co-occurrence of "nanofiber" and "chitosan" of Material 8 and 14, the pattern was reversed between the lane of "mesenchymal" and "cartilage". Therefore, we can recommend the combinations of "mesenchymal" - "chitosan" and "cartilage"-"nanofiber" as missing links to be connected.

The idea of using co-occurrence of context words to capture the characteristics of words is implemented as vector space semantic representations, like word2vec [19] or Glove [20]. As with the semantic representations, the candidate combinations of this study have no golden standard. To verify the results, we should set some task to see the performance. For further study, the time-series variations of co-occurrence would be the key for the task to predict the stage of the combinations.

From the expert hearing in the nanotechnology field, the results of this study are highly useful for researchers who are beginning to explore new fields, as a first step to understanding the pyramid of the targeted academic region. It was suggested that the composition of the clusters in the high-co-occurrence group would also be meaningful. Another suggestion from the expert hearing was that, as silk and chitosan are the general materials that have conventionally been used in nanotechnology, there is a possibility that the research phase has already passed for these materials, which resulted in the low co-occurrence. However, even if these materials have been generally used, there still remains the possibility that the proposed combination has not yet been the subject of experiment.

V. CONCLUSION

In this research, we proposed a method of detecting candidate combinations of keywords in regenerative medicine, using a citation network and text-based co-word analysis. We first analyzed the clusters in the citation network of academic papers to visualize an overview of regenerative medicine. After obtaining high-ranked keywords of the main clusters, we divided them into three regions: Organ, Material, and Technology. By comparing the co-occurrence and *tf-icf* among the three regions, we detected plausible linkages within the keywords in the selected cluster as well as the existing linkages within each paper in the whole dataset. The results suggest the potential effectiveness of new combinations that have not yet been examined in detail, and can be used to predict niches and missing links in regenerative medicine.

REFERENCES

[1] I. Sakata, “Knowledge Structuring Tools for Decision Support Service : An Overview of Citation-Based Approach,” in *Global Perspectives on Service Science: Japan*, S.K.Kwan, Y.Sawatani, and J.C.Spohrer, Eds. Springer, 2015.

[2] Y. Kajikawa, J. Ohno, Y. Takeda, K. Matsushima, and H. Komiyama, “Creating an academic landscape of sustainability science: An analysis of the citation network,” *Sustain. Sci.*, vol. 2, no. 2, pp. 221–231, 2007.

[3] R. Langer and J. P. Vacanti, “Tissue Engineering,” *Sci. 14 May 1993*, vol. 260 VN -, no. 5110.

[4] G. C. Gurtner, S. Werner, Y. Barrandon, and M. T. Longaker, “Wound repair and regeneration,” vol. 453, no. May, 2008.

[5] F. M. Watt and W. T. S. Huck, “Role of the extracellular matrix in regulating stem cell fate,” *Nat. Rev. Mol. Cell Biol.*, vol. 14 VN - r, no. 8, pp. 467–473, 2013.

[6] C. Bonnans, J. Chou, and Z. Werb, “Remodelling the extracellular matrix in development and disease I E r,” *Nat. Publ. Gr.*, vol. 15, no. 12, pp. 786–801, 2014.

[7] K. R. Chien, “Regenerative medicine and human models of human disease.,” *Nature*, vol. 453, no. 7193, pp. 302–305, 2008.

[8] N. Shibata, Y. Kajikawa, Y. Takeda, I. Sakata, and K. Matsushima, “Detecting emerging research fronts in regenerative medicine by the citation network analysis of scienti fi c publications,” *Technol. Forecast. Soc. Chang.*, vol. 78, no. 2, pp. 274–282, 2011.

[9] C. Chen, Z. Hu, S. Liu, and H. Tseng, “Emerging trends in regenerative medicine : a scientometric analysis in CiteSpace,” no. November 2011, pp. 593–608, 2012.

[10] C. Chen, R. Dubin, and M. C. Kim, “Emerging trends and new developments in regenerative medicine: a scientometric update (2000 – 2014),” *Expert Opin. Biol. Ther.*, vol. 2598, no. January, pp. 1–23, 2014.

[11] D. Zhao and A. Strotmann, “Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field,” *Scientometrics*, vol. 87 VN - r, no. 1, pp. 115–131, 2010.

[12] D. Zhao and A. Strotmann, “Counting first, last, or all authors in citation analysis: A comprehensive comparison in the highly collaborative stem cell research field,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 62 VN - r, no. 4, pp. 654–676, 2011.

[13] L. L. Li, G. Ding, N. Feng, M. H. Wang, and Y. S. Ho, “Global stem cell research trend: Bibliometric analysis as a tool for mapping of trends from 1991 to 2006,” *Scientometrics*, vol. 80, no. 1, pp. 39–58, 2009.

[14] X. Y. An and Q. Q. Wu, “Co-word analysis of the trends in stem cells field based on subject heading weighting,” *Scientometrics*, vol. 88 VN - r, no. 1, pp. 133–144, 2011.

[15] V. Ittipanuvat, K. Fujita, I. Sakata, and Y. Kajikawa, “Finding linkage between technology and social issue: A Literature Based Discovery approach,” *J. Eng. Technol. Manag. - JET-M*, vol. 32, pp. 160–184, 2014.

[16] M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” *Phys. Rev. E*, vol. 69, no. 6, p. 066133, 2004.

[17] A. T. Adai, S. V Date, S. Wieland, and E. M. Marcotte, “LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks,” *J. Mol. Biol.*, vol. 340, pp. 179–190, 2004.

[18] M. Bhattacharjee, S. Miot, A. Gorecka, K. Singha, M. Loparic, S. Dickinson, A. Das, N. S. Bhavesh, A. R. Ray, I. Martin, and S. Ghosh, “Oriented lamellar silk fibrous scaffolds to drive cartilage matrix orientation: Towards annulus fibrosus tissue engineering,” *Acta Biomater.*, vol. 8, no. 9, pp. 3313–3325, 2012.

[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” *Nips*, pp. 1–9, 2013.

[20] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation,” *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.*, pp. 1532–1543, 2014.