Document Analysis and Stability for Comprehension of University Evaluation Reports

Masaaki Ida and Susumu Shibui

***Abstract*—Document analysis for various reports assisted by computer is becoming one of the necessary techniquesinmodern information and communication age, which is expected to be an effective tool for management improvement and advancedinnovation.In document analysis, natural language processing and multivariate analysis methods based on information technologyare popular research techniques, which deepen our understanding of accumulated document data, such as evaluation reports, and also have possibility to lead new knowledge discovery. Utilizing these techniques, in this paper we examinetextual contents of evaluation reports of National University Corporation Evaluation in Japan. We operatekeyword extraction and multivariate analysis for grasping global and local features of the evaluation reports. Moreover, we consider stability of document analysis which effects to thecomprehension of evaluation reports in case of textual data fluctuation.**

***Index Terms*—Document analysis, stability, document comprehension, evaluation reports, university evaluation.**

I. INTRODUCTION

Various textual data analysis methods and visualization techniques have been developed so far to comprehend document information, such as "Morphological analysis", "Cluster analysis", and "Support vector machine". Development of these methods leads to deep our understanding of document information. In this paper, we especially focus on textual information of *Student learning outcome* in university evaluation reports.

In recent years, accountability and information disclosure of public sectors are important issues in all over the world. This is also applied to higher education institutions, so thaturgent demand for accountability and open official informationhas caused necessity for developing public higher education database[1], [2], where statistics and other information about colleges and universities could be accessed in order to clarify their accountability. Adding to the information of cost data, admissions data, and completion rates of the institution, university data should contain the data of *Student learning outcomes* to improve the *quality of education*of universities[3]. Learning outcomes is described that "Learning outcomes refer to the personal changes or benefits that follow as a result of learning. Such changes or benefits can be measured in terms of abilities or achievements" [4]. However, it is not necessarily clear how to define appropriate indicators for *measuring student learning outcomes*.

Manuscript received March 16, 2012; revised April 6, 2012.

Authors are with National Institution for Academic Degrees and University Evaluation, 1-29-1 Gakuen-nishimachi, Kodaira, Tokyo,187-8587Japan (e-mail: ida@ niad.ac.jp).

In the university evaluation, collecting objective indicators is substantially important to perform evidence-based evaluation.We have been trying to recognize indicators for measuring student learning outcomes from the documents of *peer*-reviewed *evaluation reports* of National University Corporation Evaluation in Japan [5].Aiming at appropriate analysis of document information, we must understand various features of documents and analysis methods.

Generally in document information, contents of evaluation reports, there exist various vagueness and uncertainty such as fluctuations of notations, synonyms, ambiguity of document contents in limited text volume, existence of essential or inconsequential keywords, and so on. Therefore, we always confront various difficulties in our interpretation and recognition of evaluation reports. It is considerably important problem how to comprehend the results of textual data analysis.

II. EVALUATION REPORT AND LEARNING OUTCOMES

A. Peer-Reviewed Evaluation Reports

We investigated the textual data originated from the results of university evaluation which was performed in fiscal 2008 for all national universities in Japan by National Institution for Academic Degrees and University Evaluation, NIAD-UE[5], [6]. Schematic process of the evaluation is as follows:

- Each university corporation preparesself-assessment and performance report based on two key documents, NIAD-UE's "Evaluation Guidelines" and "Guidelines for Performance Report", and submits the report to NIAD-UE.
- Their performance is examined by evaluation committeethroughout the evaluation process based on Peer-review on submitted report and site visit. Evaluators of committee read the reports and compile necessary information for evaluation, such as rates, performance, and result of questionnaire.
- Evaluation committee produces *Peer-reviewed evaluation report*, documents of the result of education and research evaluation based on the information. The evaluation report also includes four grade judgments and description of reason of the judgments, e.g., rates are good, performances are excellent, result of the questionnaire to the students is normal.

Fig. 1 shows some parts of evaluation reportsrelated to learning outcomes; "Academic Achievement" including "Judgment" and "Reason" [5].

In our study, we examine the document data for*bachelor degree program* (357 faculties of national universities in

Japan). Main descriptions of the student learning outcomes in the evaluation reports, are included in *evaluation viewpoint 4-2 of standard 4*: “The academic achievements, credentials, and abilities students acquired”. We examinetextual data from this part of the reports. The document data consists of judgment and the textual data describing the reason of judgment (Avg. = 132.5 characters in Japanese, SD=54.1, Range=49 to 454).

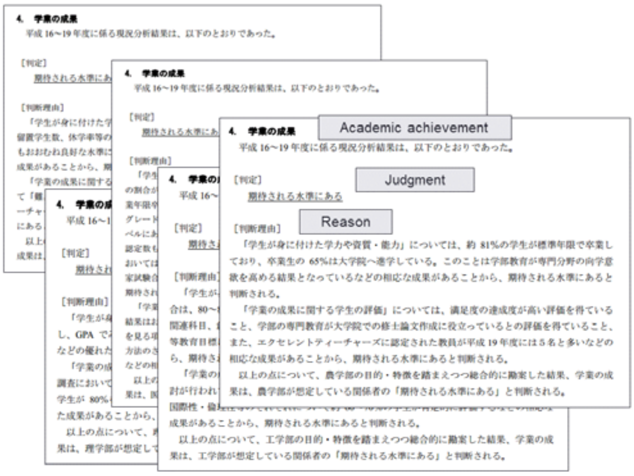

Fig. 1. Peer-reviewed evaluation report.

B. Extraction process and indicator of category

Process of our document analysis is described as follows: “*Morphological analysis*”of national language processing is adopted to decomposetextual data into set of words automatically. Extraction of important keywords or indicators from the set of wordsindicating learning outcomes was manually operated. This is owing to no clear definition for indicators of learning outcomes. Some keywords or indicators aredescribed in “Guidelines for Performance Report” and used in “University Information Database” [2],[5].

Fig. 2 shows our“working sheet” in document analysis process and what the data looks like. In each row of the sheet, data consist of “university name”, “faculty name”, “textual description of reasons for evaluation”, and “judgment”. Categories of indicators were constructed in exploratory operation by try and error. Our previous research paper [6] describesthe details of extraction process, qualitative descriptions and quantitative indicators, and relationship between qualitative indicators and rank of evaluation result.

Eventually we construct18 categories of indicators and categorical datain Table I.

III. Multivariate Analysis And Interpretation

A. Qualification Method III

In this section, for the categorical data extracted from evaluation reports, we applyone of the multivariate methods, “*Qualification method III*” [7]. Fig. 3 shows the result of this method displayed in two-dimension scatter plot. We can investigate global tendency of our data which are responded to the 18 categories.In the first axis (horizontal axis) of Fig. 3, some apparent features are recognized that many faculties of “Medicine”, “Dentistry”, and “Pharmaceutical sciences” are located on left side in this figure. Moreover, it is confirmed that same faculty names, e.g., faculty of “Engineering” or “Economics”, are gathered at some extent and make local clusters; therefore, necessity of analysis with clustered faculties is acknowledged.

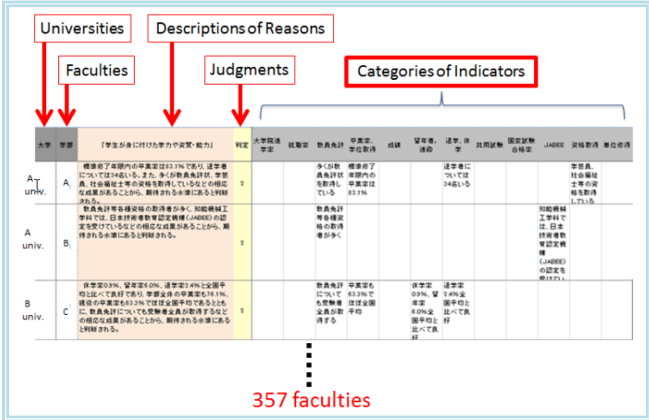

Fig.2. Working sheet used in document analysis process.

TABLE I: Categories Of Indicators

	Categories of indicators	frequency
1	Graduation and Degree Awarding Requirements	179
2	State of Credit Acquisition	99
3	Teacher License	96
4	Progression to Next Grade	90
5	Acquisition of other Licenses	80
6	MD, DD, Pharmacist License	73
7	Prize and Awards	58
8	Improvement of Educational Structure	48
9	Withdrawal	47
10	Academic Scores, GPA	40
11	Graduation Thesis	37
12	Advancement to Graduate School	33
13	Published Research Papers, Research Presentation	29
14	JABEE Certification	25
15	Questionnaire to Stakeholders	18
16	CBT (Computer based testing for medicine)	18
17	TOEIC, TOEFL	10
18	Bar, CPA (certified public accountant) and Public Official Exam.	9

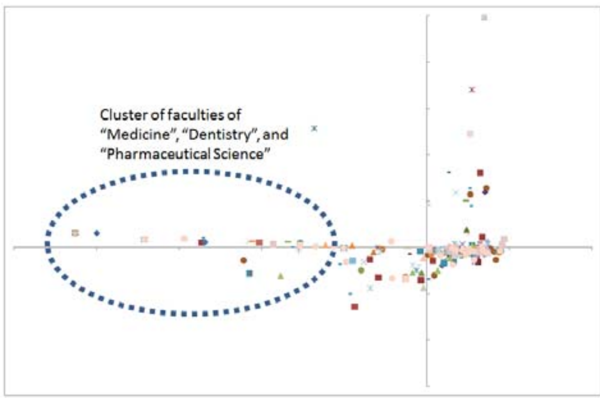

Fig.3. Analysis by qualification method III.

B. Correspondence Analysis

In order to investigate the tendency of categorical data more precisely, we classify the data into 10 major faculties as shown in Table II:

TABLE II: FACULTIES

	Faculties	frequency
1	Agriculture	75
2	Dentistry	27
3	Economics	63
4	Education	108
5	Engineering	137
6	Law	32
7	Letters	57
8	Medicine	94
9	Pharmaceutical Sciences	37
10	Science	80

In this classification process to the major faculties, we exclude some combine-named faculties such as "Science-Engineering". Therefore, total number of data reduced to 75 percent of our original data.

We applied "*Correspondence analysis*" [8] on the data, and the result is shown in Fig. 4. As expected, along the first axis (horizontal axis) of the figure, we clearly recognize faculties of "Medicine", "Dentistry", "Pharmaceutical sciences", and categories of "MD, DD, and Pharmacist License" which are related to the faculties located in left side. Moreover, we can see other features of the data in the second axis (vertical axis), that category of "Bar, CPA, and Public Official Exam" is located in upper side of the figure and "JABEE Certification" and faculty of "Engineering" are located in lower side of the figure.

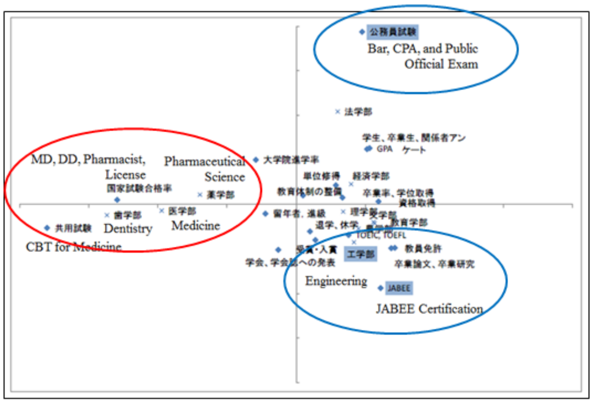

Fig.4. Correspondence analysis.

C. Hierarchical Cluster Analysis

Applying other multivariate analysis, we can investigate*proximity*feature of faculties and categories."*Hierarchical cluster analysis*(*Ward method*)"[9]derivestwo kinds of resultsdisplayedin Fig.5 and 6, which indicate hierarchical proximity. These figures show more intensive proximity tendencies, i.e., a set of faculties of "Medicine", "Dentistry", "Pharmaceutical science",a set of facultiesof "Economics" and"Law" faculties, a set of facultiesof "Education" and"Science"and so on.

IV. SENSITIVITY ANALYSIS AND COMPREHENSION

A. Sensitivity ofcorrespondence analysis

As mentioned in the introduction of this paper, we should consider various vagueness and ambiguity in the document analysis. In order to cope with this difficult issue, we apply "*Sensitivity analysis*"to data fluctuation of document data. Detailed mathematical sensitivity analysis on correspondence analysis was already developed and described in reference [10].By using our mathematical method, we can calculate variation or movement of the data.

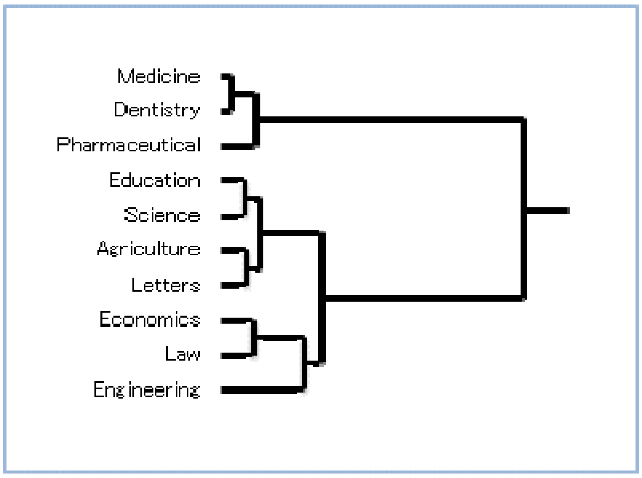

Fig.5. Cluster analysis for faculties (Ward method).

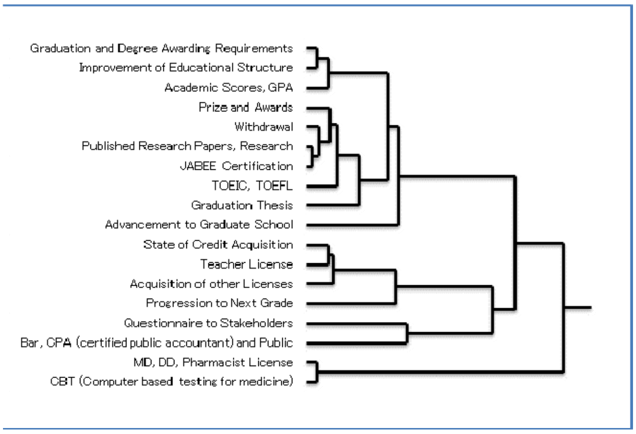

Fig.6. Cluster analysis for categories (Ward method).

Fig.7 shows the influence of data fluctuations on the number of combination of ("Faculty of Engineering", "Bar, CPA, and Public Official Exam"). Green arrows in the figure mean variations of location of "faculties". On the other hand, red arrows mean variations of location of "categories of indicators".Especially, in the right side in the figure, we can seelong red arrows and green arrows. This situation means that considerable variations by data fluctuations are estimated.

Actually,in case whenthe number of elements in combinationof ("Faculty of Engineering", "Bar, CPA, and Public Official Exam") is increased by 10 percent (+5 elements), resultant locations are shown in Fig. 8. Some elements located in upper and lower side in Fig. 4 move to centerpart of Fig. 8.

Adding to this influence of the data fluctuation, we can find other location variations, e.g., "JABEE Certification" is strongly affected and moved by the fluctuation.

Here we examine further data fluctuations. Result from fluctuation on the number of combination of("Faculty of Pharmaceutical Science", "Bar, CPA, and Public Official Exam") can be calculated by our mathematical method. Detailed variations of green and red arrows are shown in Fig. 9.

Fig. 10 shows the actual resultant location bythe fluctuations. Comparing to the original location in Fig. 4, we can see that locations of "Faculty of Pharmaceutical science" and "Bar, CPA, and Public Official Exam" are strongly effected and gathered to the center part of Fig. 10.

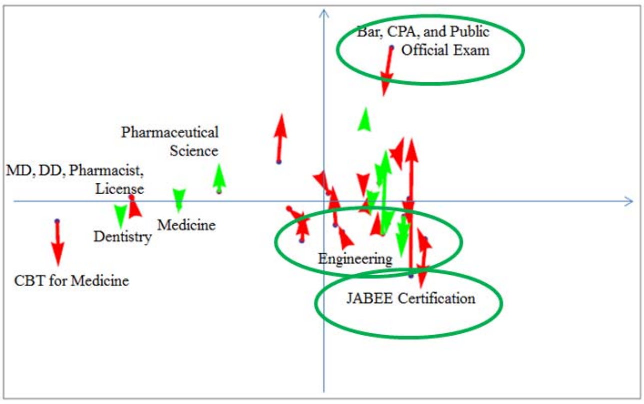

Fig. 7. Sensitivity analysis for data fluctuation: ("faculty of engineering", "Bar, CPA, and public official exam").

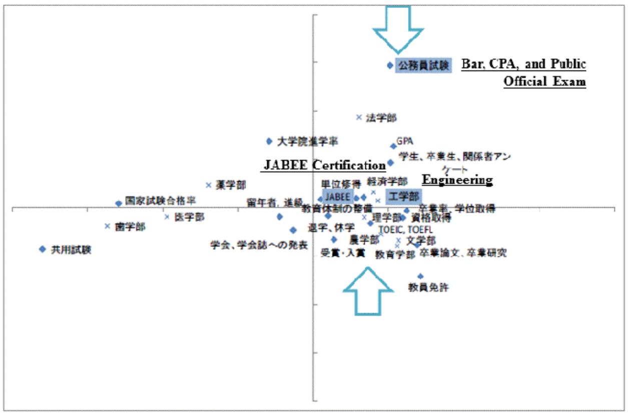

Fig. 8. Influence of fluctuation: ("faculty of engineering", "Bar, CPA, and public official exam").

From the strategic point of view, this means that if departments of "Pharmaceutical science"increase keywords related to "Bar, CPA, and Public Official Exam", then comparing to Fig. 4, the location of "Bar, CPA, and Public Official Exam"and"Pharmaceutical science"would connecteach other more strongly.

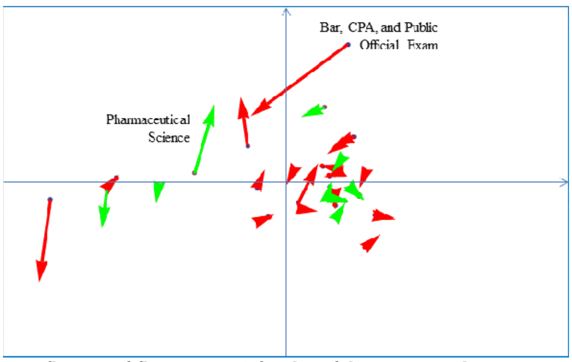

Fig. 9. Influence of fluctuation: ("faculty ofpharmaceutical science", "Bar, CPA, and public official exam").

Similarly, as shown in Fig. 11 and Fig. 12, departments of "Medical"and"Dental" science have similar characteristics for the key words related to "Bar, CPA, and Public Official Exam".We can recognize the movements related to the data fluctuation.

As seen in this section, for document data fluctuations, we should consider sensitivity for correspondence analysis and its influence to interpretation of evaluation reports.

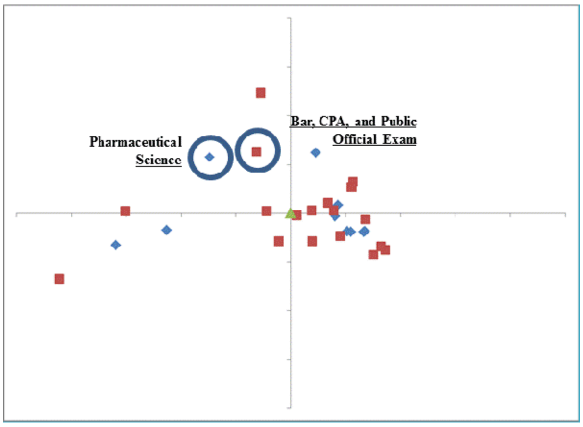

Fig. 10. Result of fluctuation: ("faculty ofpharmaceutical science", "Bar, CPA, and public official exam").

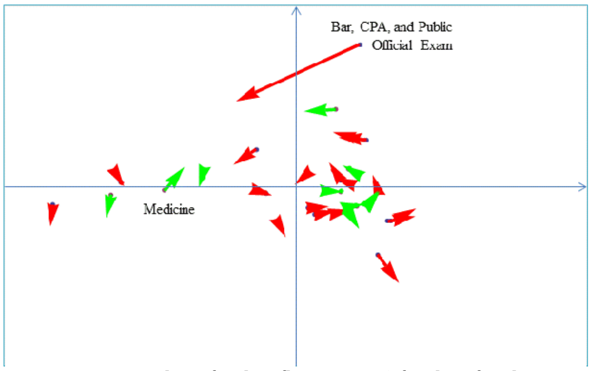

Fig. 11. Sensitivity analysis for data fluctuation: ("faculty ofmedicine", "Bar, CPA, and public official exam").

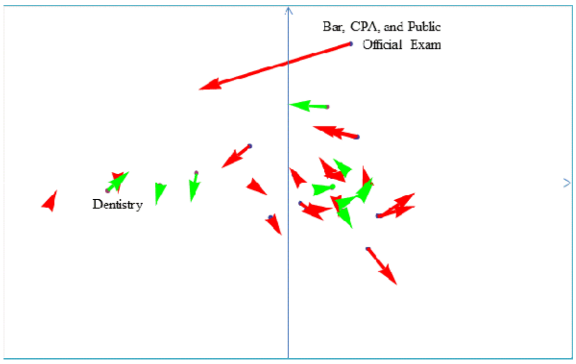

Fig. 12. Sensitivity analysis for data fluctuation: ("faculty of dentistry", "Bar, CPA, and public official exam").

B. Various Cluster Analysis methods and interpretation

In the previous sub-section, we examine direction and quantity of variations caused by data fluctuation. This sub-sectionexamines other kind of variations depending on the methods of data analysis themselves.

In Fig. 5 and 6, we examined the results of cluster analysis by"Ward method (root square distance between the gravity center and each member)". Here, we investigate other kind of cluster analysis usingalternative norms or distances to evaluate similarity between clusters [9], and their interpretations.

1) Cluster analysis for faculties

We examine three cluster analysis methods: "Group average method (root mean square distance between all pairs of data within two different clusters)","Nearest neighbor method (minimum distance)", "Furthest neighbor method (maximum distance)" for faculties as shown in Fig. 13, 14, and 15 respectively. We can see some expected strong connections among faculties in the figure, so that we cannot find any special different features comparing to the result of Fig. 5 of "Ward method". Therefore, in this case robustinterpretations on proximity for the faculties are recognized.

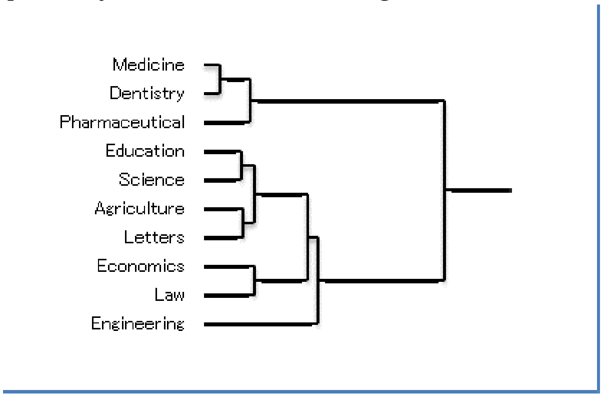

Fig. 13. Cluster analysis for faculties (Group average method).

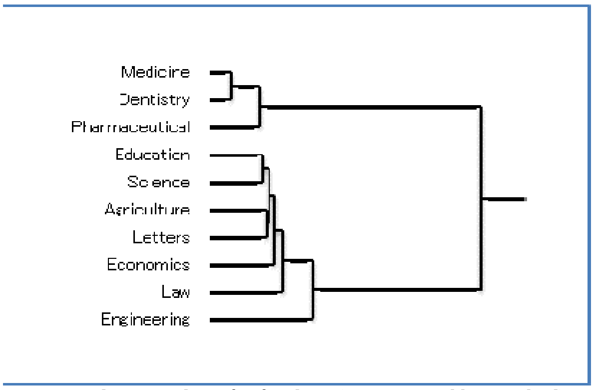

Fig. 14. Cluster analysis for faculties (Nearest neighbor method).

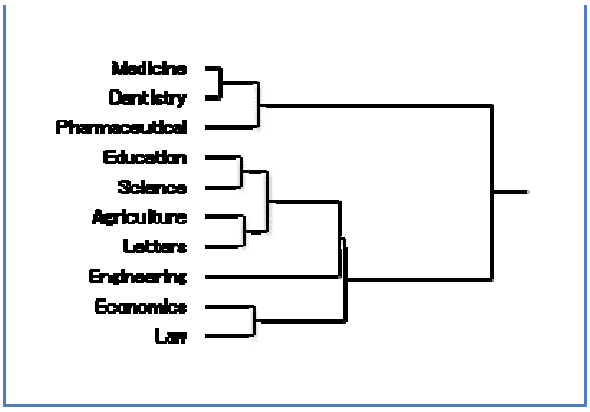

Fig. 15. Cluster analysis for faculties (Furthest neighbor method).

2) Cluster analysis for categories of indicators

In the same way, we examinefor categories of indicatorsby three cluster analysis methods: "Group average method","Nearest neighbor method", "Furthest neighbor method" as shown in Fig. 16, 17, and 18 respectively.

Comparingthese results, we find some chunks or sets of categories. For example, the set of{"Graduation and Degree Awarding Requirements", "Improvement of Educational Structure"} means that these two categories of indicators are closely related and have co-occurrence. Similarly, the set of {"Withdrawal", "Published Research Papers, Research Presentation", "and JABEE Certification"} means that indicators related to severe performance evaluation are strongly connected. Adding to these categories, "TOEIC, TOEFL" and "Prize and Awards" are related to the set.

On the other side, we can see some isolated categories of indicators.For example,"Graduation Thesis", "Bar, CPA and Public Official Exam.", "Questionnaire to Stakeholders", and the set of {"MD, DD, Pharmacist License", "CBT"} are relatively isolated to other categories.

As we examined, we can find some local robustness or local relationsin these figures, and global tendency or separate location of clusters.

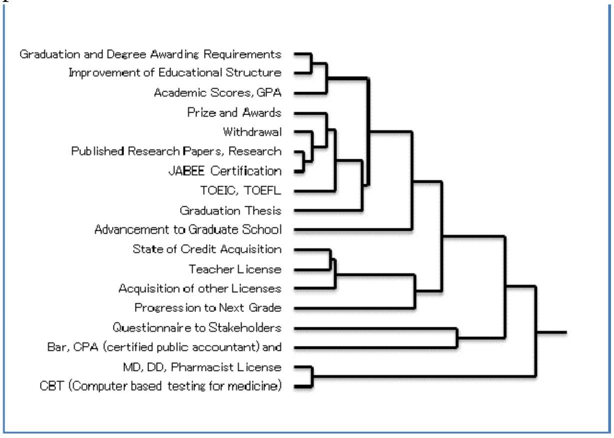

Fig. 16. Cluster analysis for categories (Group average method).

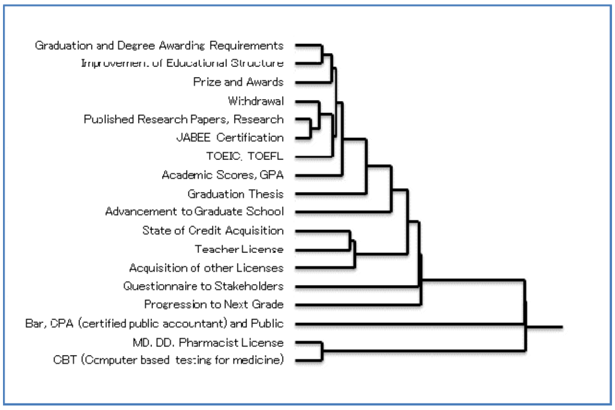

Fig. 17. Cluster analysis for categories (Nearest neighbor method).

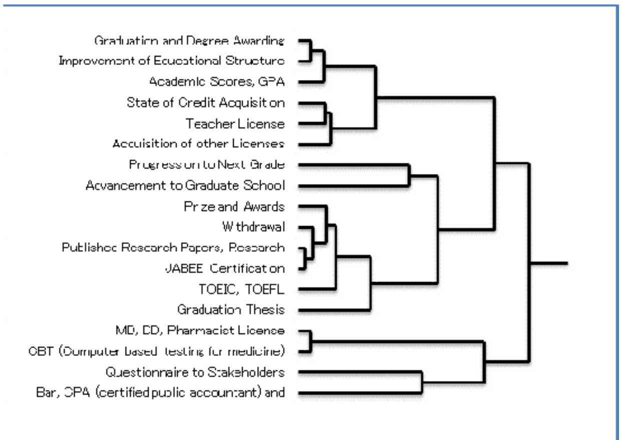

Fig. 18. Cluster analysis for categories (Furthest neighbor method).

V. Conclusion

In this paper, we considered the document information of student learning outcomes in evaluation reports of National University Corporation Evaluation in Japan. We conducted our research on information extraction and text analysis for grasping global and local features of the report. Moreover, we considered stability of interpretation for documents in case of data fluctuation.We should note that we acknowledgedthe necessity of integrated document analysisbased on sensitivity analysis and various analytical methods and the effects to document comprehension, which will deepen our understanding on accumulated document data and also have possibility to lead new knowledge discovery.

REFERENCES

[1] J. C. Burke,“The new accountability for public higher education: from regulation to results,”*Research on Academic Degrees and University Evaluation*, National Institution for Academic Degrees and University Evaluation, vol. 3, pp. 66-85, 2003.

[2] National Institution for Academic Degrees and University Evaluation,“University information database,” Available: (Japanese) http://www.niad.ac.jp/n_hyouka/jouhou/database/

[3] U.S. department of education, “The secretary of education’s commission on the future of higher education. A test of leadership: charting the future of US. higher education,” U.S. department of education: Washington D.C., 2006.

[4] OECD,“Assessment of higher education learning outcomes,”available: http://www.oecd.org/dataoecd/3/13/42803845.pdf

[5] National Institution for Academic Degrees and University Evaluation,“Evaluation reports of national university corporation evaluation in Japan,” (Japanese)
Available: http://www.niad.ac.jp/n_hyouka/kokuritsu/hyoukakekka/

[6] S. Shibui, S. H. Kim, T. Hayashi, and M. Ida,“Investigation of the indicators of student learning outcomes by means ofuniversity evaluation reports of the National University Corporation Evaluation,”*Research on National Institution for Academic Degrees and University Evaluation*, National Institution for Academic Degrees and University Evaluation, No.13, pp.1-19, 2012, (Japanese)Available: http://www.niad.ac.jp/n_shuppan/gakujutsushi/mgzn13/

[7] C. Hayashi, “On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view,”*Annals of the Institute of Statistical Mathematics*, vol. 3, pp. 69-98, 1951.

[8] J. P. Benzecri, *Correspondence Analysis Handbook*, Dekker, New York, 1994.

[9] JARS, “Clustering.” [Online]. Available: http://www.jars1974.net/pdf/12_Chapter11.pdf

[10] M. Ida, “Consideration on sensitivity for correspondence analysis and curriculum comparison,”*Integrated Uncertainty Management and Applications, Advances in Intelligent and Soft Computing 68*, Springer, pp. 547-558, 2010.