We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

8

Facet Decomposition and Discouse Analysis: Visualization of Conflict Structure

Hayeong Jeong[1], Kiyoshi Kobayashi[1],
Tsuyoshi Hatori[2] and Shiramatsu Shun[3]
[1]*Kyoto University*
[2]*Ehime University*
[3]*Nagoya Institute of Technology*
Japan

1. Introduction

Public debate has become an important process in the planning of public projects where stakeholders including citizens, experts, enterprises, and administrators meet face-to-face to discuss the desirability of a public project in its early planning stages. The roles of public debate are: to clarify the pros and cons of a project, that is, evaluate available information and proposed approaches; to promote mutual understanding of the stakeholders' different perspectives; and to find resolutions that are mutually beneficial for all stakeholders.

The stakeholders or debate participants have varying values, concerns, and expectations on the benefits and costs of a project and as such, their interests do not necessarily coincide with or complement each other. To promote their interests, debate participants strategically reiterate and legitimize their opinions. The debate facilitator must identify appropriate methods for recognizing and managing uncooperative communication in order for a debate to arrive at a fair consensus. To facilitate effective group communication such as debates, it is important to develop scientific management tools for managing debate content and conflict situations.

Advances in information technology has led to the development of various debate support systems based on corpus linguistics (Shiramatsu *et al*., 2007; Jeong *et al*. 2007) designed to assist facilitators and participants in achieving clear communication and arriving at a consensus by analyzing information from actual dialogue data. Unfortunately, these systems are yet to be fully applied to studies in public discourse. There have been few studies that examined the effectiveness of debate support systems in alleviating stakeholder interest conflict. In addition, systematic and objective means for summarizing and analyzing debate content have not been sufficiently developed. In order to effectively and efficiently manage public debate, techniques that enable deeper understanding of participant opinions and management of conflict are essential.

This study proposes a method for visualizing conflict structures in public debates by analyzing debate minutes using a corpus-based discourse analysis. In this method, the opinion of participants recorded in the debate minutes (hereafter called "primary corpus")

are decomposed into four facets based on the following defined conceptual categories—debate purpose, means, evidences, and attitudes toward the subject project. Using this facets-tagged corpus (hereafter called "secondary corpus"), this study quantitatively measures the content similarity among the recorded opinions and the level of interest dissonance among the participants. The rest of this paper is organized as follows. Section 2 describes the basic idea of the study. Section 3 explains the definition and decomposition of the facets using a support vector machine (SVM). Section 4 presents the discourse analysis using the secondary corpus and visualizes the conflict structure and debate pattern changes. The applicability of the proposed methodology is then discussed by applying it to a real case of public debate. Section 5 presents the conclusions and limitations of the study.

2. The basic idea

2.1 Related studies

Previous studies on debate support systems have focused on summarizing debate contents by extracting important topics and exploring semantic meanings using statistical indicators of term frequency and co-occurency (Horita and Kanno, 2001; Hiromi *et al*, 2006). While these techniques clarify the main topic of debate and participants' interests from a debate minute using natural language processing and text mining, they do not identify conflict structures and underlying dynamic interaction in a multiple debate process.

Nasukawa and Yi (2003) proposed a sentiment analysis approach for extracting polar opposite sentiments of subjects from a document by identifying the semantic relationship between the subject term and semantic expressions such as "good" and "bad," instead of classifying the entire document as positive or negative. By defining the terms of sentiment expression, it is possible to calculate an individual's attitude toward the subject term and determine whether participants in a discussion have different attitudes toward to subject term While this method certainly has its merits, it is unable to determine the reason for the differences in attitude and the underlying relationships in the debate. In order to enhance mutual understanding and consensus building in debates, it is necessary to clarify the conflict structure and its change throughout multiple debates, evaluate the significance of debate context on participants' sentiments, and develop facilitation techniques for managing conflict structures.

Opinion mining is a relatively new research field that includes review analysis, opinion extraction, and answering opinion questions. Opinion mining utilizes natural language processing techniques and facilitates deep semantic classification of opinions. Sophisticated techniques for deep semantic classification have not yet been developed for and applied to public debates as they require substantial linguistic knowledge and the ability to process additional non-literal information such as heuristic rules. The rules-based approaches to opinion mining also have some limitations. Somasundaran and Wiebe (2010) examined opinions in ideological online debates by developing an argument lexicon using the Multi-Perspective Question Answering (MPQA) annotations. Meanwhile, the Statement Map (Murakami *et al*., 2010) and WISDOM (Akamine *et al*., 2009) methods focus on analyzing information credibility through opinion mining techniques. These methods classify semantic relationships between subjective statements such as agreement and conflict relationships using rules-based approaches that work well with experimental data but generally are more

difficult to implement than machine learning approaches. Recognizing Textual Entailment (RTE) Challenges (Dagan, 2006; PASCAL, 2005) is a method that focuses on the semantic relationships between texts to determine whether the meaning of one text is derived from another text. Although the RTE datasets included semantic relations such as agreement and conflict, Marnefe (2008) found that opinions in real world debates were much more difficult to classify and as such, more sophisticated classification methods are required for RTE to work in the real world. Cross-document Structure Theory (CST) is another approach for identifying semantic relationships between sentences (Radev, 2000). The CSTBank Corpus built by Radev (2000) is annotated with 18 kinds of semantic relationships including agreement and conflict. As with the RTE, the CST approach does not effectively classify opinions in debates.

The rules-based approaches discussed above clarify linguistic characteristics using detailed categorizations and rules to ensure accuracy in analysis. However, they are limited in that they are difficult to implement and their classifications are too simplified to accurately classify more complicated opinions in real-world debates. This study proposes a simpler, more effective approach based on machine learning techniques for visualizing conflict structures in terms of rapid prototyping. Machine learning approaches are more effective and perform better than rules-based approaches. We then apply this approach to a series of real public debates where participants with different backgrounds argue their opinions.

2.2 Content Analysis

Debate participants' different cognitive frames and subjective interests lead to differing opinions. In order to effectively summarize the debate contents, we need to maintain the relevance between textual and contextual information using qualitative and descriptive analyses (Hashiuchi, 2003). In this study, we used the method of content analysis proposed by Stone *et al*. (1966). Content analysis refers to ". . . any research technique for making inferences by systematically and objectively identifying specified characteristics within text" (Stone *et al*., 1966). The purpose of content analysis is to examine the corpus data of dialogue (e.g., debate minutes), taking into consideration the relevance between the words and the context embedded in the dialogue. Krippendorff (1980) identified the three most important features of content analysis: it is an *unobtrusive* technique; it can assess *unstructured* material; and is *context-sensitive*. These features allow researchers to examine relatively unstructured data for meanings, symbolism, expressed information, and the role this information plays in the lives of the data sources, while at the same time acknowledging the textuality of the data, that is, recognizing that the data are read and understood by others and interpreting this data based their own contexts (Koreniusa *et al*., 2007).

Discourse analysis is a type of content analysis that clarifies debate content and structure by targeting language use in the real world (Schiffrin *et al.,* 2003). Discourse analysis reveals the structural and functional mechanism of language in relation to the utterances in their context. Discourse has traditionally been defined as "anything beyond the sentence." Schiffrin (1994) referred to "utterances" rather then "sentences" because utterances are *contextualized* sentences, that is, they are context- and text-bound. In discourse analysis, context refers to the information that surrounds a sentence and determines its meaning, and includes a broad range of information such as the speaker's knowledge, beliefs, facial expressions, gestures, and social circumstance, and culture.

Hatori *et al*. (2006) suggested a discourse analysis based on facet theory. In this method, utterances in a debate are decomposed into three conceptual units (hereafter called "facets")—"what" should be solved, "how" it should be solved, and "why" it should be solved, in order to effectively examine interest conflict and cognitive dissonance among debate participants. This facet decomposition task depends on how researchers analyze and interpret the utterances in a debate. However, to be able to analyze multiple debates with huge amounts of language data, improved reproducibility and computational approaches are required (Francis, 1982).

This study develops a computational method of facet decomposition and discourse analysis based on corpus linguistics. Corpus is generally defined as a collection of text or written and spoken material that highlights how a language is used. Corpus linguistics is a linguistic method that explains the meanings of words and phrases using a corpus. It is a reconstructive method for analysis of language data using a computer (Sinclair, 1991; Stubbs, 2002; Wang, 2005; McCarthy, 1998). In this study we use SVM learning to decompose debate utterances into facets. Then, using this facets-tagged corpus, we propose an approach for calculating the differences among the utterances of the debate participants and visualize their dynamic change.

2.3 Management of public debate

A public debate is essentially a third-party committee where stakeholders, including experts, public enterprises, and citizens, share their perspectives on public projects in their early planning stage. The purpose of this third-party committee is to promote mutual understanding among the debate participants and ensure transparency of the debate system by opening the discussion process to the various stakeholders and making available information on related projects. The third-party committee is not involved in the decision-making, but rather in the collection of information related to public project and the building of public trust in the decision makers (Hatori *et al*., 2008).

The third-party committee is directly entrusted by the government and the citizens and as such, must suggest solutions that are socially desirable and beneficial to the government and its citizens. As representatives of the government and its citizens, committee members must assess solutions based on their social perspective (i.e., check for adequacy) and technical and professional perspectives (i.e., check for rigor). In short, the third-party committee is expected to assess project-related information for adequacy and rigor to facilitate decision making.

Previous studies have used questionnaires and interviews to understand the perspectives of stakeholders on public projects. According to Giddens (1990), these methods are based on the segmented, asynchronous communication between the stakeholders and thus suffer from the problem of faceless commitments. These methods are limited in that questionnaire may contain the subjective perspectives of the researchers and as such both the researcher and participants readily trust the questions and answers on the questionnaire, respectively. In addition, the questionnaire may contain the assumptions and perspectives of the researchers on the projects which the respondents or citizens do not necessarily recognize. In contrast, a debate within a third-party committee realizes the richness of social communication facilitated by the direct meeting among stakeholders. Such debates ensure the effective collection of information and transparency in the decision-making process.

Debate participants are interested only in certain aspects of an issue, and as such, are motivated to steer the discussion toward their desired goals. Debate participants such as experts and engineers usually evaluate the benefits of the projects based on scientific and professional evidences highlighting rigor, while general citizens evaluate them based on their common sense and their interests, checking for adequacy. If debate participants consider only one aspect of the issue, that is, either rigor or adequacy only, the debate would not reach a consensus. The most important role of public debate involving the third-party committee is to achieve a common perspective on the project under discussion by recognizing each stakeholder's interest and point of view. To evaluate the progress of the debate toward mutual understanding and consensus, it is necessary to accumulate the debate minutes and summarize the discussions. This study proposes a method of discourse analysis that manages public debate by summarizing the debate contents and tracking the progress of the debate process toward its goal.

3. Creating the facet-tagged corpus

3.1 Facet decomposition and the Support Vector Machine (SVM)

In this study, we define utterance as the language used to establish one's position on an issue by addressing the position's pros and cons. During a debate, participants legitimize their positions with technical and scientific evidences or their common sense and self-interests. They strengthen their positions by expressing positive or negative attitudes toward another participant's utterances depending on whether those utterances coincide or contrast their positions. Based on these assumptions, we identified the four facets to be used in this study: facet A refers the positions of participants or their pros and cons on the issue; facet B refers their efforts to strengthen their positions using evidence; facet C refers the characteristics of evidences, that is, whether they pertain to either adequacy or rigor; and facet D refers to the participants' interpretation of each other's utterances, that is, their expression of positive or negative attitudes toward each other's utterances. We coded the utterances using a combination of four facets (i.e., stractable). In this study, the utterances in the debate minutes comprise the primary corpus and the utterances encoded with facet stractables comprise the secondary corpus. Utterances that do not explicitly pertain to a position are considered neutral and as such, classified under a third category for each facet. Section 3.2 presents the facet stractables in more detail.

In order to create the secondary corpus, we encoded the utterances with the facet stractables. Owing to the enormous amount of data in the primary corpus, decomposing the utterances based on facets is not easy. As such, we created the secondary corpus using a statistical facet learning model that reduces the efforts required of researcher while at the same time maintaining the logical consistency of the primary corpus. This statistical facet learning model is based on a pattern recognition technique that classifies the huge amount of primary corpus into the secondary corpus through the SVM, a type of statistical learning machine. Pattern recognition is generally defined as a method for extracting data features such as those in letters, video, and audio and determining the data categories based on the standard pattern of the features. This study decomposed the facets using the SVM model originally developed by Vapnik (see Kudo, 2002). The SVM is a binary sorter where there is only one border identifying a sample's location. Because there are three categories for each

facet, that is, $X \in \{A,B,C,D\}$, this study uses two bounding hyperplanes ${y_h}^X(U_i),(h=1,2)$. These bounding hyperplanes are defined by a high-dimensional discrimination function that determines the amount of terms included in the primary corpus and maximizes the distance between the border and each utterance. Although the discrimination function is usually represented by a nonlinear kernel function, we used a d-dimensional polynomial function that flexibly approximates the function. The SVM model used in this study is able to process large amounts of data and produces good estimation results. In this model, the utterances in the primary corpus, $U_i(i=1,\cdots,m)$, are each classified into one of the three categories of facets, $F^X(U_i) \in \{X(-1),X(0),X(1)\}$. $F^X(U_i)$ determines the corresponding relationship and is defined as:

$$F^X : U_i \mapsto \{X(-1),X(0),X(1)\} \quad (1)$$

$$(X \in \{A,B,C,D\})$$

The two types of kernel functions, ${y_1}^X(U_i)$, ${y_2}^X(U_i)$ for the facets $X \in \{A,B,C,D\}$ with m units of utterances $U_i(i=1,\cdots,m)$ which are located in the n dimensional space′ (n is the total number of terms in the primary corpus) are defined as follows:

i. $F^X(U_i) = X(1)$ if ${y_1}^X(U_i) > 0$;

ii. $F^X(U_i) = X(0)$ if ${y_1}^X(U_i) \le 0$ and ${y_2}^X(U_j) > 0$; and

iii. $F^X(U_i) = X(-1)$ if ${y_2}^X(U_i) \le 0$.

In addition, we define the facet vector $\mathbf{G}_i$ for an utterance U_i using 0–1 variables ($\delta_i^y \in \{0,1\}$) in order to represent the four facets $X \in \{A,B,C,D\}$ corresponding to each of the categories $y \in \{X(-1),X(0),X(1)\}$:

$$\begin{aligned}\mathbf{G}_i = \{&\delta_i^{A(-1)},\delta_i^{A(0)},\delta_i^{A(1)},\delta_i^{B(-1)},\delta_i^{B(0)},\delta_i^{B(1)},\\ &\delta_i^{C(-1)},\delta_i^{C(0)},\delta_i^{C(1)},\delta_i^{D(-1)},\delta_i^{D(0)},\delta_i^{D(1)}\}\end{aligned} \quad (2)$$

The set of facet vector $\mathbf{G}_i$ for the all utterances $U_i(i=1,\cdots,m)$ is the secondary corpus.

3.2 Case study

This paper's case study examines the debates of the Yodo-River committee which was established in 2001 in order to obtain advice for the planning and policy handling of a river improvement project related to the building of a dam and gather the opinions of the representatives of affected citizens and public organizations. The Yodo-river committee meetings consists of a general meeting, four regional meetings, five theme meetings, five working group meetings, and three meetings of the sub-working group. A total of 400 meetings have been held since 2001. The debate minutes are available from the committee website and are downloadable as PDF files. The minutes record the names of speakers and their utterances chronologically.

From these meetings, we selected 14 debates between citizens and experts or between administrators (i.e., river managers) and experts. Participants were classified into five

groups according to their roles: facilitator; expert; citizen with a "con" opinion; citizen with a "pro" opinion; and administrator. The minutes from the 14 debates consisted of 8,831 utterances. A unit of utterance is a single sentence ended by punctuation mark.

3.3 Outline of the facet decomposition

Four researchers from construction consulting companies with experience in mobilizing public involvement made a training set of facet decomposition. To maintain the objectivity of the training set, two researchers conducted facet decomposition on the same utterances. We completed the training set the corresponding facet decomposition. Table 1 shows the framework of the four facets with the facet elements and the typical language use.

An utterance is coded as facet A(1) if it describes the pros of the project. Statements about the advantages of the project are examples of A(1). Meanwhile, disapproving opinions are classified as facet A(-1) and include statements that mention the disadvantages of the project and suggest new alternatives. Utterances that include evidences are coded as facet B(1) and those without evidences as B(-1). Utterances with rigorous or adequate evidences are coded as C(1) or C(-1), respectively. Facet C(1) includes statements that cite evidences from studies or experiments while facet C(-1) includes statements with evidences based on individual experiences or feelings. Finally, facet D(1) represents positive attitudes toward the project, signaling affirmation or belief in the project. Facet D(-1) represents negative attitudes toward the project. Conflicts can be identified by comparing the difference between the responses in facet A and those in facet D. Facet B determines the plausibility of arguments while facet C sheds light on the participants' reasoning systems. If an utterance cannot be easily coded using the above facets, it is classified under the neutral category, $X(0)$, $X \in \{A, B, C, D\}$.

Below are two examples of how utterances are encoded using the facets framework:

Example 1: The dam construction is the only way to ensure the safety of the town's residents given the limited public finances. → A (1) B (1) C (-1) D (-1)
Example 2: There have been many regional development projects that constructed dams but only a few of them succeeded. → A (-1) B (1) C (1) D (-1)

The first utterance example is decomposed into facet A(1)B(1)C(-1)D(-1). The expression "is the only way" indicates that the speaker agrees to the dam construction, hence, A (1), but does not have a positive attitude toward it, hence, D(-1). The expression "given the limited public finances" is an evidence and is therefore classified as B(1); however, this evidence is also based on the speaker's anxiety and concern, and as such, is classified as C(-1). Meanwhile, the second utterance example expresses pessimism regarding the dam construction, hence, A(-1). The phrase "only a few of them have succeeded" expresses the speaker's reason for his negative view and thus is denoted as B(1), C(1), and D(-1).

As mentioned earlier, researchers created and completed the training utterance set of facet decomposition. Of the total 8,831 utterances, 34% were decomposed as facet A, 65% as facet B, 59% as facet C, and 52% as facet D. Over half of the utterances are selected as training utterence excepting facet A. The training utterance set determines the statistical facet

	Definition	Elements		
		$X(1)$	$X(-1)$	$X(0)$
Facet A	Refers to the position of the participants (i.e., whether they have taken a "pro" or "con" position on the project.	Pros	Cons	Neutral utterances including greetings, questions, and other utterances during the progress of the debate. These utterances are categorized as X(0) and not X(1) or X(-1).
Facet B	Refers to whether participants provided evidences to strengthen their position and help others understand their perspectives	With evidence, including scientific, mathematical, and psychological evidence	Without Evidence	
Facet C	Refers to the characteristics of the evidences (i.e., whether they are rigorous or adequate)	Rigor of evidence focusing on statistical facets, scientific facts, existing facts, and engagement	Adequacy of evidence focusing on psychological facts (e.g., anxiety, trust, impression), sensitizing concepts (values, norms, belief), community spirit, ethical facts, fairness, rightfulness, etc.	
Facet D	Refers to the participants' bargaining attitudes to the project	Positive attitudes such as agreement, esteem, encouragement, compromise, etc.	Negative attitudes such as avoidance, constraint, doubt, distrust, disregard, ignorance, disruption, etc.	

Table 1. Framework of Facet Decomposition.

decomposition using the SVM. We then investigate the accuracy of the SVM decomposition by calculating the matching rate between the results done by the researchers and that by the SVM.

The accuracy of the SVM decomposition was estimated by using kernel-fold cross validation wherein we divide the 8,831 utterances into K subgroups and use K-1 subgroups as the training set (i.e., the utterances facet-encoded by researcher) while the remaining subgroup is decomposed into facets using the SVM. In this study, $K = 10$. By changing the subgroup of the training set to the other subgroup, the statistical facet decomposition for all the utterances using the SVM can be implemented a total of K times. The dummy variable δ_i^y indicates the categorized elements $y \in \{X(-1), X(0), X(1)\}$ of each facet $X \in \{A, B, C, D\}$ of utterances $U_i (i = 1, \cdots, m)$. α refers to the matching frequency of the results of the facet decomposition from both the researchers and the SVM where $\delta_i^y = 1$. β refers to the frequency that training set was $\delta_i^y = 1$ but was categorized by the SVM as $\delta_i^y = 0$. γ refers to the frequency that training set was $\delta_i^y = 0$ but was identified by the SVM

as $\delta_i^y = 1$. Finally, λ refers to the frequency that both the researchers and the SVM decomposed the utterance as $\delta_i^y = 0$. The accuracy of the SVM model was determined by the F-value calculated as follows:

$$A = \frac{\alpha + \lambda}{\alpha + \beta + \gamma + \lambda} \tag{3a}$$

$$F = \frac{2P \times R}{P + R} \tag{3b}$$

where $P = \frac{\beta}{\alpha + \beta}$, and $R = \frac{\alpha}{\alpha + \gamma}$.

3.4 Results

Table 2 presents the results of the facet decomposition. The total number of facet decomposition performed by the researchers exceeded that of the utterances as each utterance can be classified into multiple facet elements. 35% of the utterances expressed the pros of the dam project; 33% mentioned the cons percent of the project; 31% of the utterances had evidences while 24% did not; 21% had rigorous evidence while 41% had adequate evidence; 12% expressed a positive attitude toward the project and 55% expressed a negative attitude. Over 50% of the utterances were negative responses to the project that had rigorous evidence. This result does not mean, however, that over half of the debate participants had a negative opinion of the project because the utterances were only the opinions of those who spoke. Meanwhile, in the facet decomposition resulting from the SVM, 4% expressed the pros of the project and 1% expressed the cons; 26% had evidences and 20% did not; 10% presented rigorous evidence and 24% presented adequate evidence; 5% expresses positive attitudes toward the project and 34% negative attitudes. The accuracy of the SVM decomposition was calculated using Equation 3b which resulted in an F-measure between 0.48 and 0.66. In general, the accuracy level should be between 0.6 and 0.7. There are at least two possible reasons for this low level of accuracy. First, many of the utterances were decomposed into the third category, neutral (X(0)). Second, the accuracy of the decomposition for facet A was particularly low because most participants imply rather than express the pros and cons of the project and as such, their utterances cannot be easily decomposed into the appropriate facets. The final facet classification lies on the researcher. To ensure the objectivity of the researchers' decomposition, a feedback system is used to review and revise the facet decomposition based on the discourse patterns. Subsection 4.2 presents the analysis of the discourse patterns.

The less than ideal accuracy level of the SVM decomposition is not uncommon, as only few classification techniques utilizing natural language processing for discourse data have achieved sufficient accuracy. As such, the results of the SVM decomposition in this study are still useful for determining facet decomposition. The final facet decomposition was improved by adjusting the two classifications by the SVM which supported and facilitated the researchers' tasks.

The SVM decomposition provides useful information that helps researchers in their final analysis and significantly reduces their efforts for confirming the secondary corpus. In

addition, the SVM helps monitor and improve the performance of rule-based approaches for opinion mining (e.g., Dagan, 2006; Radev, 2000). The SVM method is facilitates the comparison of the results of two approaches using disagreements to track changes in the discourse rules. Future studies should test the applicability of the SVM method to rules-based approaches.

Total 8,831 utterances		**Facet A** Cons/Pros		**Facet B** With/Without Evidences		**Facet C** Rigor/adequacy		**Facet D** Positive/Negative	
		#	%	#	%	#	%	#	%
Researchers	○	6130	35	5447	31	3715	21	2088	12
	●	5839	33	4214	24	7252	41	9716	55
	×	5693	32	8045	45	6739	38	5902	33
SVM	○	313	4	2292	26	904	10	452	5
	●	49	1	1793	20	2142	24	2959	34
	×	8469	95	4746	54	5785	66	5420	61
Final Decomposition	○	3532	40	2914	33	2031	23	1236	14
	●	3267	37	2296	26	3533	40	4946	56
	×	2782	23	3621	41	3267	37	2649	30
Accuracy		48		66		55		50	
F-value		35		54		44		34	

Note: The symbols ○, ●, and × indicate categories X(1), X(-1), and neutral, respectively.

Table 2. Facet Decomposition Results.

4. Discourse analysis of the public debate

4.1 Similarity of utterance meaning

Using the secondary corpus (i.e., the facet-encoded utterance developed in the previous section), we conducted the discourse analysis in order to examine the interest conflict among the participants and visualize the discourse patterns during the debate process. Using non-metric multi-dimensional scaling (MDS), we distributed the participants on a two-dimensional space based on the similarity of their utterance meanings (Kruskal, 1964a and 1964b; Qian, 2004). Among the Z participants $p_i \in (p_1,...,p_z)$, the dissimilarity of interest $dsim(p_i, p_j)$ between two participants p_i and p_j is represented as the distance $dis(p_i, p_j)$ in a two-dimensional space. If the interest dissimilarity between the participants is $dsim(p_i, p_j) < dsim(p_i, p_k)$, then the distance is $dis(p_i, p_j) < dis(p_i, p_k)$.

For a participant p_k, the relative frequency of dummy variables $\delta_j^y (y = A(-1), A(0), A(1), \cdots, D(-1), D(0), D(1))$ is defined, making up the facet vector from the secondary corpus as illustrated below:

$$W_y^{pk} = \frac{\sum_{j \in \Omega pk} \delta_j^y}{\sum_{j \in \Omega pk} \sum_{y=A(-1)}^{D(1)} \delta_j^y} \tag{4}$$

Here, Ωp_k indicates a set of facet vector. Facet matrix S is defined as:

$$\mathbf{S} = \begin{bmatrix} \mathbf{F}_{p1} \\ \vdots \\ \mathbf{F}_{pz} \end{bmatrix} = \begin{bmatrix} W_{A(-1)}^{p_1} & \cdots & W_{D(1)}^{p_1} \\ \vdots & \ddots & \vdots \\ W_{A(-1)}^{p_z} & \cdots & W_{D(1)}^{p_z} \end{bmatrix} \tag{5}$$

The similarity between the facet vectors of two participants is calculated based on the cosine angle distance:

$$sim(p_i, p_j) = \cos(\mathbf{F}_{pi}\mathbf{F}_{pj}) = \frac{\mathbf{F}_{pi}\mathbf{F}_{pj}}{|\mathbf{F}_{pi}| \cdot |\mathbf{F}_{pj}|}$$

$$= \frac{\sum_{y=A(-1)}^{D(1)} W_y^{p_i} W_y^{p_j}}{\sqrt{\sum_{y=A(-1)}^{D(1)} (W_y^{p_i})^2} \sqrt{\sum_{y=A(-1)}^{D(1)} (W_y^{p_j})^2}} \tag{6}$$

We can determine the dissimilarity of the facet vectors through the inverse of the cosine function:

$$dsim(p_i, p_j) = \cos^{-1}\{sim_i(p_i, p_j)\} \tag{7}$$

All distances $dis(p_i, p_j)$ are arranged in a correlation matrix D:

$$\mathbf{D} = \begin{pmatrix} dis(p_1, p_1) & \cdots & dis(p_1, p_z) \\ \vdots & \ddots & \vdots \\ dis(p_z, p_1) & \cdots & dis(p_z, p_z) \end{pmatrix} \tag{8}$$

MDS reproduces the distance $dis(p_i, p_j)$ on a two-dimensional space by minimizing the sum of the squares of the distance between $dis(p_i, p_j)$ and the dissimilarities $dsim(p_i, p_j)$. Using stress, which represents the degree of incompatibility between the dissimilarity and the distance, an accurate coordinate value is produced:

$$Stress = \sqrt{\frac{\sum_{i=1}^{z-1}\sum_{j>i}^{z}\{dis(p_i, p_j) - dsim(p_i, p_j)\}^2}{\sum_{i=1}^{z-1}\sum_{j>i}^{z}\{dis(p_i, p_j) - \overline{dis}\}^2}} \tag{9}$$

$\overline{dis}$ denotes the average of the distance between two participants:

$$\overline{dis} = \sqrt{\frac{\sum_{i=1}^{z-1}\sum_{j>i}^{z}\{dis(p_i, p_j)\}^2}{{}_zC_2}} \tag{10}$$

By changing the dimension value to a lower value, the optimal arrangement for the coordinate value in a two-dimensional space can be defined. Consequently, the semantic similarity of all individuals is illustrated as the distances in a two-dimensional space.

Figure 1 shows the distribution of the participants in this study (49 professionals, 27 administrators [i.e., river managers], and 5 citizens) on a two-dimensional space based on the similarity of the facet vectors derived from 14 debates minutes (i.e., the primary corpus). The symbols ◆, ✳, and □ denote citizens, experts, and administrators, respectively. As shown in Figure 1, the participant groups usually have similar interests. For instance, citizens are located on the leftmost part of the graph. Meanwhile the administrators are distributed on the top portion of the space, while, the experts are positioned on the bottom of the space, indicating that these two groups have interest conflict.

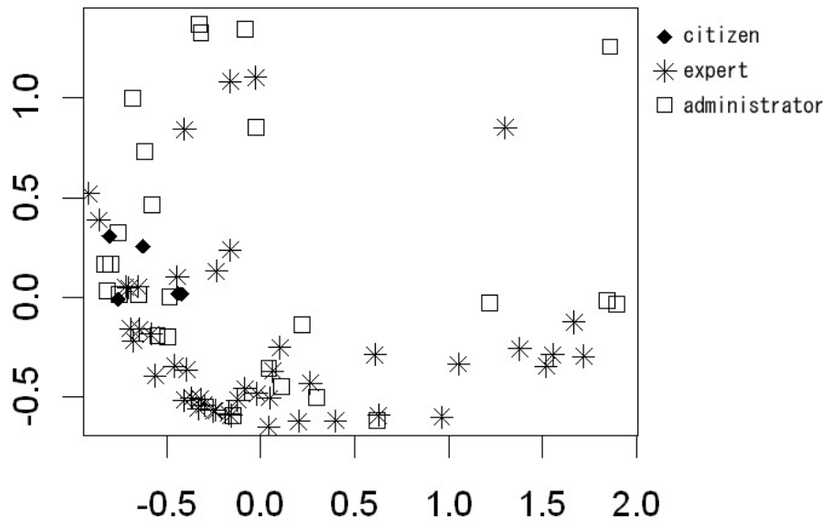


Fig. 1. Similarity of Interest Structures of Debate Participants.

Figure 2 illustrates the five types of participants who expressed their utterances: the facilitators (⊠); experts (✳); citizens with a "pro" opinion (◆); citizens with a "con" opinion (◇), and administrators (□). As shown in the figure, the distance between facilitator and expert is wider than that between facilitator and citizen or administrator, which means that the interest structure of facilitators is similar to that of experts, but substantially differs from that of citizens and administrators. In this case, it is important to examine the change in the interest structure during the debate rather than the different interest structures themselves. It is also important to remember that different interest structures are not necessarily a negative thing. On the contrary, it will help groups obtain and understand the diverse perspectives of the various stakeholders on a project. In the following subsection, we will analyze the change in interest structure during the debate process using Figure 2 once again.

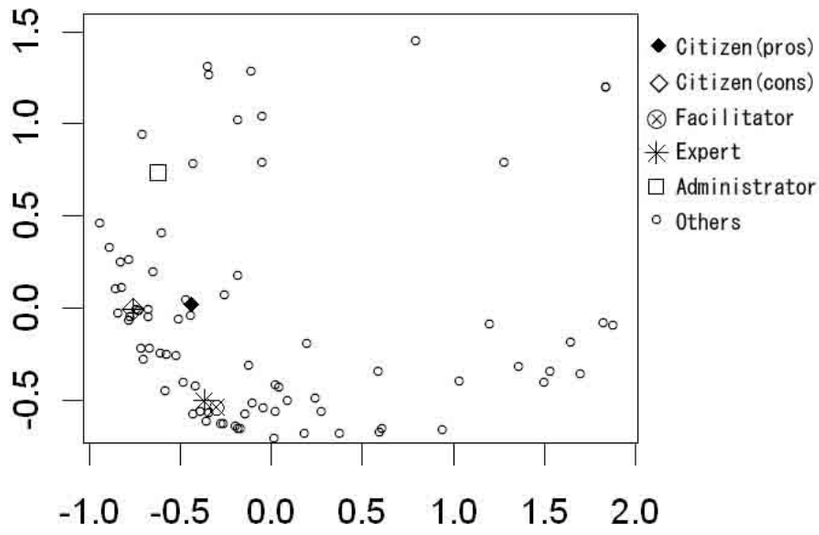


Fig. 2. The Five Types of Participants.

4.2 The change of discourse pattern

We created a time-sequence of the primary corpus and calculated the cumulative frequency of the facet in order to analyze the change in the discourse patterns of the five types of participants. Figures 3 to 7 illustrate the discourse change for the different types of participants. The horizontal axis on the graphs represents the number of time-sequence of utterances while the vertical axis represents the cumulative frequency of the elements of each facet. The line will go up if the frequency of category $X(1)$ of facet X increases. The line will go down if the frequency of category $X(-1)$ of facet X increases. The line will remain constant if the frequency of $X(1)$ matches that of $X(-1)$.

Figure 3 illustrates the temporal variation of the cumulative frequency of the facets of the facilitator. The facilitator expresses a total of 849 utterances in 8 out of the 14 debates. The results of the facet decomposition for facilitators show that the facilitators do not take a "pro" or "con" position on the project; expressed utterances without evidence; when they do provide evidence, they presented rigorous evidence during the debate in general and used adequate evidence on certain issues; and expressed more negative attitudes rather than positive attitudes throughout the debates.

In contrast, the cumulative frequency of the facets of the other types of participants exhibits a stable pattern (Figures 4 to 7). Experts had a total of 444 utterances in 6 out of the 14 debates; administrators expressed 151 utterances in one debate; citizens (pros) expressed 143 utterances in one debate; and citizens (cons) had 228 utterances in one debate. Except for the facilitators and experts, the participants were involved in only one debate; as such, their opportunities for expressing their perspectives on the project was limited, only managing to gain mutual understanding and knowledge during the debate. The results also show that administrators and experts have a tendency to take a neutral position on many facets while citizens have a tendency to repeat same categories of facets, that is, the pros or cons.

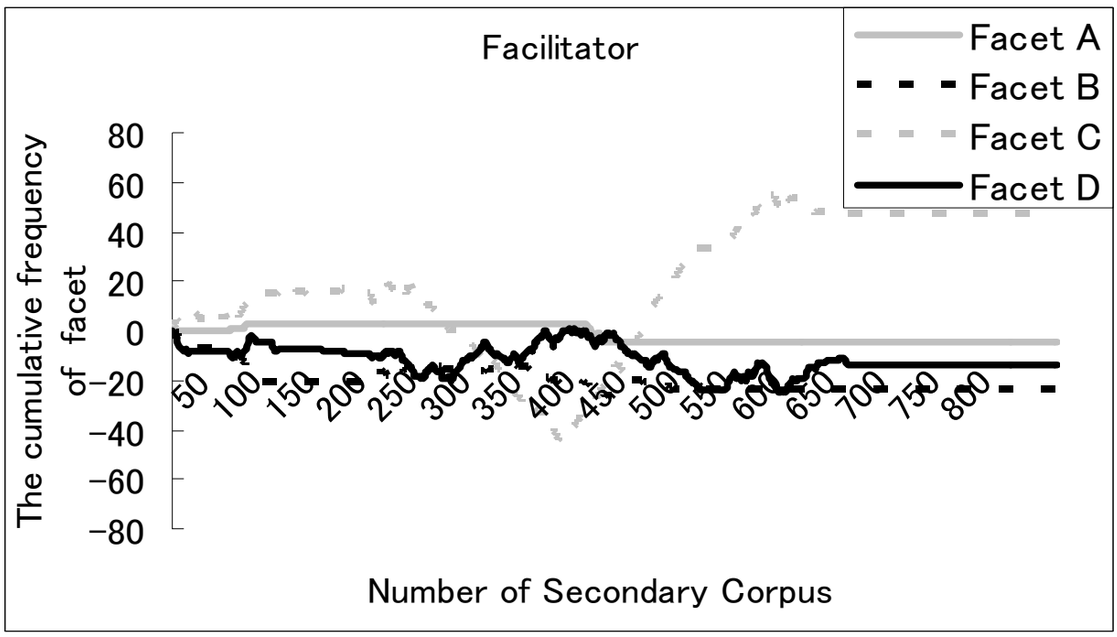


Fig. 3. The Discourse Pattern Change of the Facilitator.

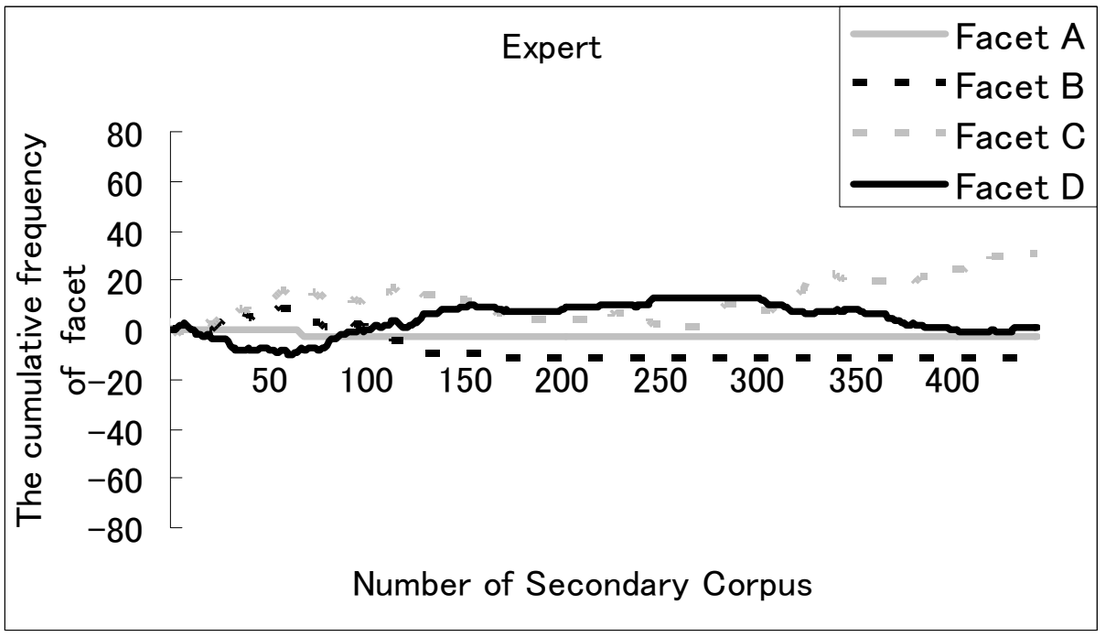


Fig. 4. The Discourse of Pattern Change of the Experts.

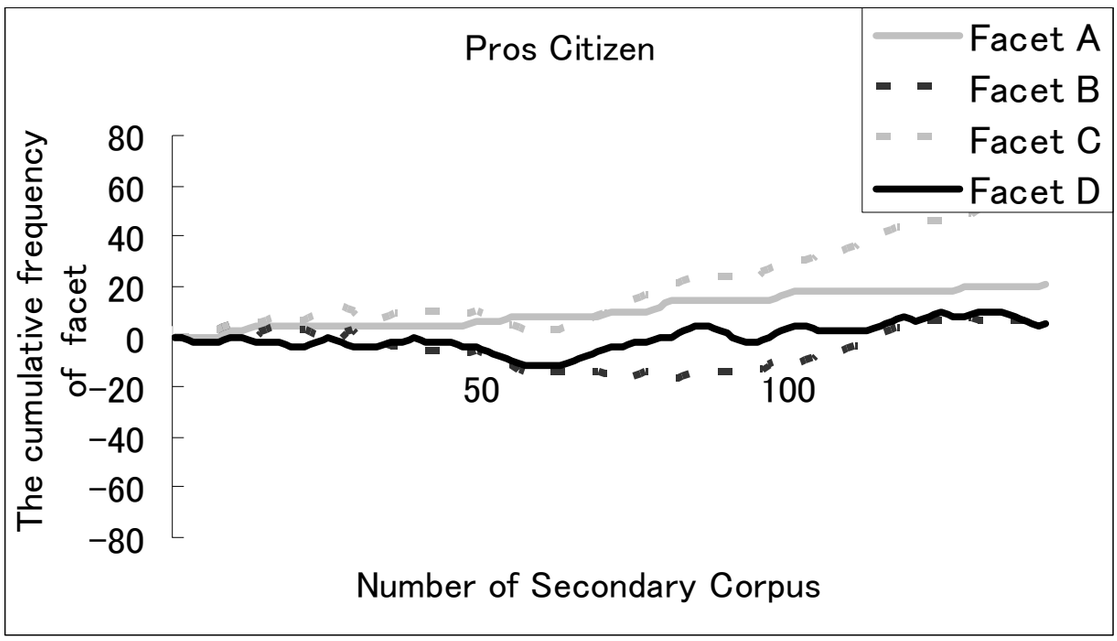


Fig. 5. The Discourse Pattern Change of Citizens (Pros).

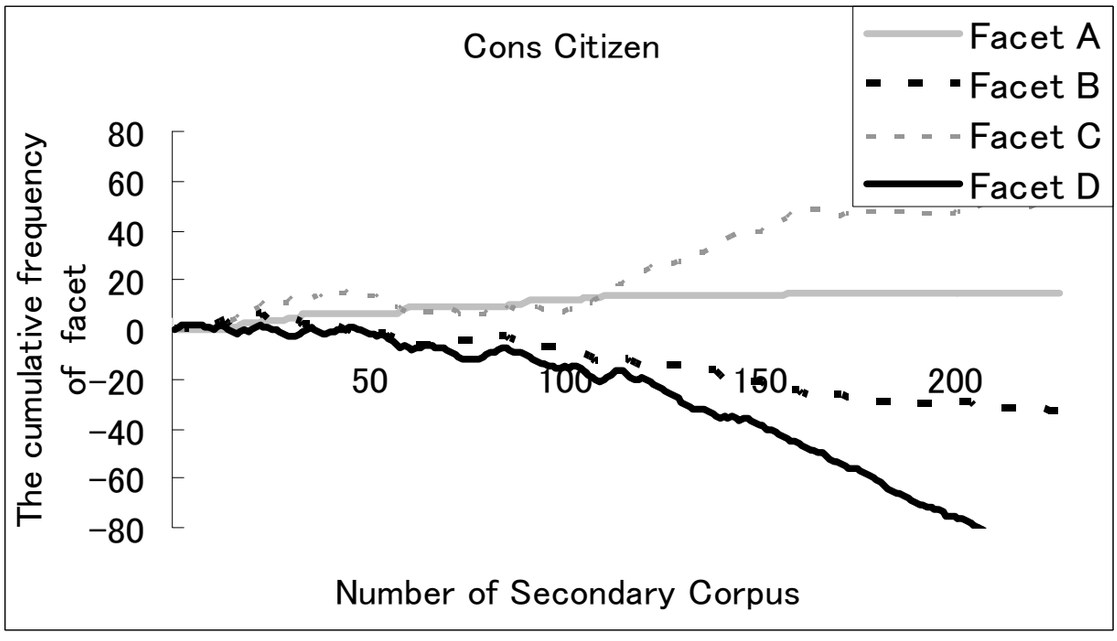


Fig. 6. The Discourse Pattern Change of Citizens (Cons).

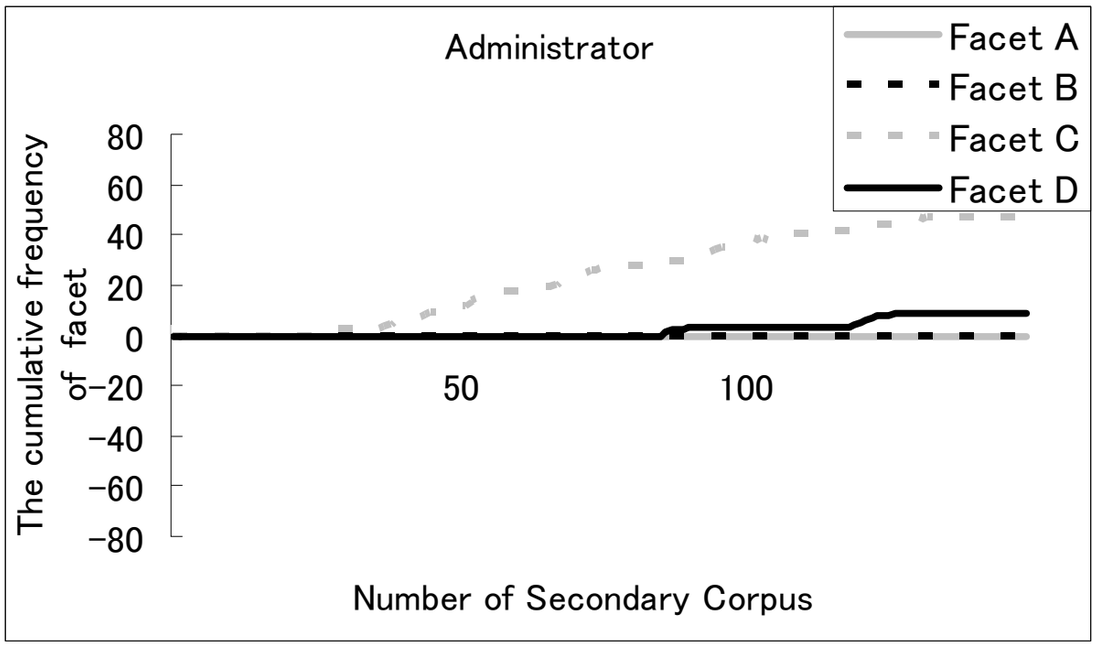


Fig. 7. The Discourse Pattern Change of the Administrators.

The cumulative frequency of facet is a simple and useful indicator for understanding the change in discourse and consequently the change in the interest structures. By examining the changes in the discourse patterns, we can uncover the possible causes of the interest conflicts.

In sum, this study's proposed method of discourse analysis enables us to identify debate content and structure and consequently the conflict structure and dynamic change throughout the debate process. This new approach overcomes the limitations previous existing approaches.

5. Conclusions

This study proposed a new method for examining the interest conflict of participants and discourse pattern changes in the process of debate using discourse analysis. We applied this method to a series of debates on a real public project during which we uncovered interest conflict among certain types of participants.

Results of our analysis suggest that during public debates, it is important to identify and adopt facilitation techniques that help identify discourse patterns, which in turn uncovers the cause of interest conflicts. This will help the debate participants examine the different perspectives of stakeholders and arrive at a consensus.

While the proposed discourse analysis method in this study helps manage and support the debate progress, it is not without its limitations. First, the validity of the proposed method relies mainly on the facet decomposition framework. Further empirical analysis is needed to confirm this validity. Second, the accuracy of the facet decomposition needs to be improved. Facet decomposition resulting from the use of the SVM has low accuracy levels for highly

context-dependent utterances. Further studies may address this problem using a feedback system that reviews the results of discourse pattern changes. Third, utterances are dependent on the debate context and the characteristics of the participants. As such, it is necessary to improve the method by taking into account latent variables such as social context and unobservable individual characteristics. Fourth, it is easy to lose diverse contextual information in the process of facet decomposition. It is therefore necessary to build a database that includes other contextual information such as the right to speak (e.g., utterance turn or length of utterance) or the social relationship between participants. Finally, further normative research of public debate is needed. This involves developing normative rules of public debate and evaluating models of desirability of public debate. By overcoming these limitations, facilitation techniques can be vastly improved.

6. References

Akamine, S., Kawahara, D., Kato, Y., Nakagawa, T., Inui, K., Kurohashi, S., and Kidawara, Y. (2009). WISDOM: A web information credibility analysis system, *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*, pp. 1-4.

Dagan, I., Glickman, O., and Magnini, B. (2006). The PASCAL recognising textual entailment challenge, *Machine learning challenges. Lecture notes in computer science*, Vol. 3944, Springer, pp. 177-190.

Ericsson, K.A. and Simon, H.A. (1993). *Protocol analysis: verbal reports as data*, MIT Press.Francis, N. (1982). Problems of assembling and computerizing large corpora, In: Johansson, S. (ed.): *Computer corpora in English language research*, Norwegian Computing Centre for the Humanities, Bergen, pp. 7-24.

Giddens, A. (1990). *The consequences of modernity*. Cambridge: Polity Press.

Horita, M. and Kanno, Y. (2001). Development an information based CRANES for Public involvement Management, *Journal of Japan Society of Civil Engineers, VI*, Vol. 686, No. 52, pp. 109-120 (in Japanese).

Hashiuchi, T. (2003). *Disukōsu: Danwa no orinasu sekai*, Tokyo: Kuroshio Publishers (in Japanese).

Hatori, T., Kawayoke, T., Kobayashi, K., Natsume, T., and Fujisaki, E. (2006). Protocol analysis of a public debate using facet theory, *Journal of Japan Society of Civil Engineers*, Vol. 23, No.1, pp. 91-102 (in Japanese).

Hatori, T., Jeong, H., and Kobayashi, K. (2008). The open public debate and the impacts upon trust formation, *Journal of Japan Society of Civil Engineers D*, Vol. 64, pp. 148-169.

Hiromi, W., Tomonari, M., Atsuhiro, T., and Jun, A. (2006). Topic-oriented term extraction and term clustering for query focusing, *Transactions of Information Processing Society of Japan database,* Vol. 47, No. SIG19 (TOD32), pp. 72-85.

Jeong, H., Hatori, T., and Kobayashi, K.(2007). Discourse analysis of public debates: a corpus-based approach, *IEEE, SMC.*

Koizumi, T. (2001). *Nyumon goyoron kenkyu: riron to ouyou*, Kenkyusya (in Japanese).

Koreniusa, T., Laurikkalaa, J., and Juhola, M. (2007). On principal component analysis, cosine and Euclidean measures in information retrieval, *Information Sciences*, Vol. 177 pp. 4893-4905.

Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*, Sage Publication, Inc.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrica*, Vol. 29, pp. 1-29.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method, *Psychometrica*, Vol. 29, pp. 115-129.

Kudo, T. and Matsumoto,Y. (2002). Chunking with Support Vector Machines, *Journal of Natural Language Processing*, Vol. 9, pp. 3-22.

Lupia, A. and McCubbins, M.D. (1998). The democratic dilemma; can citizens learn what they need to know?, Cambridge University Press.

McCarthy, M.(1998). *Spoken language and applied linguistics*, Cambridge University Press.

Murakami, K., Nichols, E., Mizuno, J., Watanabe, Y., Masuda, S., Goto, H., Ohki, M., Sao, C., Matsuyoshi, S., Inui, K., and Matsumoto, Y. (2010). Statement map: reducing web information credibility noise through opinion classification. *Proceedings of the 4th Workshop on Analytics for Noisy Unstructured Text Data*, pp. 59-66.

Nasukawa, T. and Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing, Second International Conference on Knowledge Capture.

PASCAL (2005). Recognising textual entailment challenge, http://www.pascal-network.org/Challenges/RTE/.

Ong, B. S. (2005). *Towards automatic music structural analysis: identifying characteristic within —song excerpts in popular music*, Universitat Pompeu Fabra Barcelona.

Qian, G., Sural, S., Gu, Y., and Pramanik, S. (2004). Similarity between Euclidean and cosine angle distance for nearest neighbor queries, *Proceedings of The 19th Annual ACM Symposium on Applied Computing*.

Radev, D.R. (2000). Common theory of information fusion from multiple text sources step one: Cross-document structure. *Proceedings of The 1st SIGdial workshop on Discourse and dialogue*, pp. 74-83.

Schiffrin, D. (1994). *Approaches to discourse*, Blackwell Publishers.

Schiffrin, D., Tannen, D., and Hamilton, H. *(Eds.)* (2003). *The handbook of discourse analysis*, Blackwell Publishers.

Shiramatsu, S., Komatani, K., Hasida, K., Ogata, T., and Okuno, H.G. (2007). A game-theoretic model of referential coherence and its statistical verification based on large Japanese and English corpora, *Journal of Natural Language Processing*, Vol. 14(5), pp. 199-239 (in Japanese).

Sinclair, J. (1991). *Corpus, concordance, collocation*, Oxford University Press.

Sinclair, J. (1996). The search for units of meaning, *Textus*, Vol. 9, pp. 75-106.

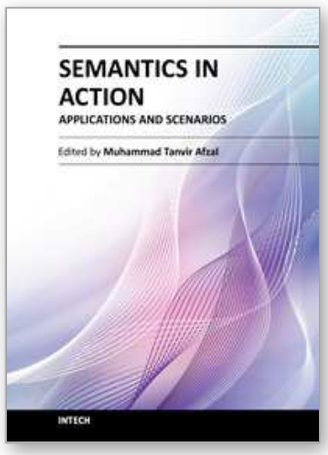
Somasundaran, S and Wiebe, J. (2010). Recognizing stances in ideological on-line debate, *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text,* pp. 116-124.

Stone, P.J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. G. (1966). *The general inquirer: a computer approach to content analysis*, MIT Press.

Stubbs, M. (2002). *Words and phrases: corpus studies of lexical semantics*, Blackwell Publishers.

Takubo, Y., Nishiyama, Y., Mitoh, H., Kameyama, M., and Katagiri, Y. (1999). *Danwa to bunmyaku*, Tokyo, Japan: Iwanami Publishers (in Japanese).

Wang, S.-P. (2005). Corpus-based approaches and discourse analysis in relation to reduplication and repetition, *Journal of Pragmatics*, Vol. 37, pp. 505-540.



Semantics in Action - Applications and Scenarios
Edited by Dr. Muhammad Tanvir Afzal

ISBN 978-953-51-0536-7
Hard cover, 266 pages
Publisher InTech
Published online 25, April, 2012
Published in print edition April, 2012

The current book is a combination of number of great ideas, applications, case studies, and practical systems in the domain of Semantics. The book has been divided into two volumes. The current one is the second volume which highlights the state-of-the-art application areas in the domain of Semantics. This volume has been divided into four sections and ten chapters. The sections include: 1) Software Engineering, 2) Applications: Semantic Cache, E-Health, Sport Video Browsing, and Power Grids, 3) Visualization, and 4) Natural Language Disambiguation. Authors across the World have contributed to debate on state-of-the-art systems, theories, models, applications areas, case studies in the domain of Semantics. Furthermore, authors have proposed new approaches to solve real life problems ranging from e-Health to power grids, video browsing to program semantics, semantic cache systems to natural language disambiguation, and public debate to software engineering.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hayeong Jeong, Kiyoshi Kobayashi, Tsuyoshi Hatori and Shiramatsu Shun (2012). Facet Decomposition and Discouse Analysis: Visualization of Conflict Structure, Semantics in Action - Applications and Scenarios, Dr. Muhammad Tanvir Afzal (Ed.), ISBN: 978-953-51-0536-7, InTech, Available from: http://www.intechopen.com/books/semantics-in-action-applications-and-scenarios/facet-decomposition-and-discouse-analysis-visualization-of-conflict-structure

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen