

Estimation of Factor Scores from Feature Values of English Question and Answer Statements

Yuya Yokoyama

Graduate School of Life and Environmental Sciences
Kyoto Prefectural University
Kyoto, Japan
y_yokoyama@mei.kpu.ac.jp

Teruhisa Hochin, Hiroki Nomiya

Faculty of Information and Human Sciences
Kyoto Institute of Technology
Kyoto, Japan
{hochin, nomiya}@kit.ac.jp

Abstract— In order to eliminate mismatches between the intentions of questioners and respondents of Question and Answer (Q&A) sites, nine factors of impressions for Japanese statements have experimentally been obtained. Nine factors have also been obtained from the impression of English Q&A statements. This paper estimates factor scores of English Q&A statements through multiple regression analysis. These are words and characters, syntactic information, and appearance percentages. It is shown that estimation accuracies of all of the nine factors are very good.

Keywords—Q&A site; factor; factor score; multiple regression analysis; feature value of statements;

I. INTRODUCTION

Recently, the number of people using Question and Answer (Q&A) sites on the Internet has been increasing. Q&A sites are online communities where users can manually post questions and answers. Thus, these sites are thought to be databases containing enormous amounts of knowledge to solve various problems. When a user posts a question, others may respond. The questioner selects the most appropriate response as the “Best Answer” (BA) and awards the respondent with a certain number of points that serve as a fee. The BA is the response statement the questioner subjectively finds most satisfying.

As the number of users of Q&A sites increases, and more questions are posted, it becomes harder for respondents to select questions that match their specialty and interests. Consequently, a question posed by a user may not be seen or answered by qualified respondents. Moreover, if an appropriate respondent is not encountered, mismatching may occur, which may cause the following problems:

- A questioner may acquire incorrect knowledge from inappropriate answers.
- Respondents may not have the necessary knowledge to properly answer the question, and thus the problem remains unsolved.
- Users may be offended by answers that contain abusive words, slanders, or statements against public order and standards of decency.

Our goal is to present questions to users who are qualified to properly answer them, thus avoiding the problems described above. The impressions of sixty statements posted on Yahoo!

Chiebukuro [1] have been evaluated [2]. By applying factor analysis to the experimental results, nine factors were obtained.

Factor scores obtained through factor analysis represent the impressions of the statements, and this is necessary for estimating the factor scores of other statements. The statements were estimated using multiple regression on the feature values assigned to the statements. Feature values include the syntactic information of the statements, such as the classes of the words in the statement (e.g., nouns or verbs), and the number of appearances (or the percentages) of each of the alphanumeric characters [3]. Moreover, word imageability, the expressions in the closing sentence, word familiarity, and word validity were also adopted as feature values. The overall estimation accuracy for all nine factors was proved to be good.

Most of the feature values, however, are dependent on Japanese. Thus, it is necessary to look into how extensive our method can be applied in other languages as well [4]. Factors that describe the impression of English Q&A statements were obtained. Experimental materials evaluated are statements actually posted to Yahoo! Answers [5]. As a result of factor analysis, it has been shown that some main factors could be obtained in English as well. It has also been shown that different type of factors could be obtained from English statements. Factor scores were calculated as well. The factor scores obtained, however, are only the thirty Q&A English statements used for the experiment. Therefore, it is required to calculate the factor scores of any English statements.

This paper tries to estimate the factor scores from feature values of English statements. The feature values include words and characters, syntactic information, and appearance percentages. The feature values extracted are then selected with the consideration of multicollinearity. Factor scores are estimated from those feature values selected. As a result of estimation, it has been shown that estimation accuracies of all of the factors are very good.

The remainder of this paper is organized as follows. In Section 2, related works are explained. Our previous works are summarized in Section 3. Estimation results of factor scores are described in Section 4. Considerations towards the results are provided in Section 5. Finally, Section 6 concludes the paper.

II. RELATED WORKS

There have been a number of prior works investigating Q&A sites; estimating BAs [6-8], introducing users to answer

statements [9, 10], and inspecting the quality or tendency of answer statements [12-14], etc.

Several works have tackled the estimation of BAs. Blooma *et al.* used both five textual and five non-textual features to predict the BAs [6]. It was found that textual features influenced the quality of the answers more than the non-textual ones did. The analogical reasoning approach [7] has found the BA by using links between questions and answers contained in the previous knowledge base. They used three textual features, seven statistical ones, and five user interactions. Nishihara *et al.* have proposed a way to detect the BA to a question [8]. They obtained the BA to a question by noticing the affinity between the closing expressions of questioners and respondents, and the clustering combinations of questions and BAs.

Several works have proposed introducing users to answers. Jurczyk *et al.* used link analysis to detect users who were authorities for specific question categories [9]. They considered three points for analysis; the difference between positive/negative evaluation for answers rated by other users, the proportion of answer statements chosen as BAs, and the rate value when chosen as a BA by the questioner. As a result of analysis, good precision was obtained for some categories, not for all the categories. Riahi *et al.* have investigated a way to provide appropriate experts with a newly posted question [10]. Profiles are constructed on the basis of their answering history, and several measures are used. In some parts of the dataset, their proposed model showed better performance than others at recommending new questions to experts. This is, however, the result for StackOverflow [11], Q&A site specialized in programmers. Thus, this result is for the limited categories, not for all the categories.

Several works have inspected the quality or tendency of answer statements. Agichtein *et al.* used the content and usage features of Q&A statements to assess their quality [12]. Of the twenty major features determining the quality of a question, eleven were related to web information and nine were obtained through questions. Liu *et al.* have analyzed the question types so that alternatives to the BAs can be found out [13]. Then they have proposed applying automatic summarization techniques to shortened answers or BAs. Their experimental results have shown that question-type focused summarizations lead to the drastic improvements of answer quality. Raban focused on self-presentation expressions between questioners and respondents related to value of information [14]. The examinations conducted on Google Answers indicated that implicit cues in textual online communication can make Q&A sites successful and sustainable.

III. PREVIOUS WORKS

A. Factors of Statements

An experiment was conducted to evaluate impressions of answers. There were forty-one evaluators, and they evaluated the style or content of statements and assigned labels from a group of fifty words [2]. Twelve sets of questions and answers were evaluated, and these included three from each of four major categories: Auction, PC, Love, and Political/Social Problems; the categories were chosen from those actually

TABLE I. JAPANESE FACTORS AND CORRESPONDING IMPRESSION WORDS.

Factors	Impression words			
1st (Accuracy)	Persuasive	Fluent	Important	Appropriate
	Wonderful	Refreshing	Skillful	
	Fulfilling	Beautiful	Favorable	
	Courteous	Real	Accurate	
2nd (Displeasure)	Uncomfortable	Resentful	Thoughtless	Disillusioning
	Fearful	Amazing	Regrettable	Unjust
3rd (Creativity)	Creative	Unexpected	Special	
	Original	Marvelous		
4th (Ease)	Easy	Clear	Difficult	
5th (Persistence)	Minute	Persistent	Long	
6th (Ambiguity)	Ambiguous	Insufficient		
7th (Moving)	Warm-hearted	Impressive		
8th (Effort)	Touching			
9th (Hotness)	Hot	Powerful		

posted at Yahoo! Chiebukuro in 2005 [1]. Factor analysis was applied to the experimental results, and nine factors were obtained and shown in Table 1. The factors indicate the nature of statements, as explained by the various impression words assigned to that statement; they were named *accuracy*, *displeasure*, *creativity*, *ease*, *persistence*, *ambiguity*, *moving*, *effort*, and *hotness*.

B. Estimation of Factor Scores

1) Feature Values of Statements

The factor scores were obtained for only the sixty statements used in the experiment. To be able to estimate the factor scores of other statements, multiple regression analysis was applied to their feature values [3]. Overall, seventy-seven feature values were adopted; these are summarized, as follows:

- Syntactic information: the number and length of statements, number or percentage of word classes (e.g., nouns and verbs), and punctuation (e.g., exclamation and question marks).
- Word imageability: a subjective characteristic that evaluates to what degree imagination is aroused by the words.
- Expression in the closing sentence: fundamental Japanese words; zo, da, yo, ne, ka, na, shi, desu, masu, tai, and nai.

TABLE II. ENGLISH FACTORS AND CORRESPONDING IMPRESSION WORDS.

Factors	Impression words		
1st (Accuracy)	Accurate	Real	Clear
	Appropriate	Important	Fulfilling
	Dull	Insufficient	
	Ambiguous	Thoughtless	
2nd (Evaluation)	Beautiful	Marvelous	Skillful
	Wonderful	Minute/Detailed	Persistent
	Fluent	Sharp	Warm-hearted
3rd (Disappointment)	Resentful	Regrettable	Shocked
	Fearful	Hot/Intense	
4th (Discomfort)	Unjust	Uncomfortable	
5th (Novelty)	Special	Unexpected	Refreshing
6th (Potency)	Persuasive	Powerful	Impressive
7th (Difficulty)	Difficult		
8th (Politeness)	Courteous	Fun	
9th (Nostalgia)	Dear/Nostalgic		

- Word familiarity: an index representing the familiarity of an evaluator for a word.
- Notation validity: an index indicating how valid a word is in the given context.

2) Estimation Result

Multiple regression analysis was performed on the sixty questions and answers employed in the impression evaluation experiment. Multiple correlation coefficients (MCCs), which show the goodness of the estimation, were above 0.9 for all nine factors [3]. Therefore, it was shown that the estimation accuracies of all of the factors are very good.

C. Obtaining Factors in English

Most of the feature values in our study are dependent on Japanese. Therefore, it is required to see how extensive our study could be applied in other languages [4]. As a first step, factors that describe impression of English statements were obtained in the similar fashion as explained in Section 3-A.

In order to directly compare our English method with Japanese one, most of the impression words used are the same ones shown in Table 1. With using those fifty impression words, impression evaluation experiment was conducted for 4 foreign subjects (3 males: Mexican, Vietnamese, and French, and 1 female: French). The experiment was conducted on desktop PC. Experiment materials were six sets of Q&A English statements from Yahoo! Answers (three each from “Computer & Internet” and “Auction ebay.”) These materials were extracted from sourceforge [15], where data on Yahoo! Answers are open to public so that research on Community Question Answering will be facilitated. Similar as the previous experiment depicted in Section 3-A, each set consists of one question and four answer statements including the “BA.”

Factor analysis was applied to the experimental results. As a result of the analysis, nine factors were obtained and shown in Table 2. Both similarity and difference were confirmed between the Japanese factors and English ones. As for the similarity, regardless of languages, some main factors could be

TABLE IV. FEATURE VALUES ADOPTED.

Word & Char	Base form			Appearance percentage		
	CC	NN	TO	CC(%)	NN(%)	TO(%)
Word	CC	NN	TO	CC(%)	NN(%)	TO(%)
Char	CD	NP	UH	CD(%)	NP(%)	UH(%)
Word/SENT	DT	PDT	VB	DT(%)	PDT(%)	VB(%)
Char/SENT	EX	POS	VD	EX(%)	POS(%)	VD(%)
	FW	PP	VH	FW(%)	PP(%)	VH(%)
	IN	RB	VV	IN(%)	RB(%)	VV(%)
	JJ	RP	WH	JJ(%)	RP(%)	WH(%)
	LS	SENT	GESY	LS(%)	SENT(%)	GESY(%)
	MD	SYM	CUSY	MD(%)	SYM(%)	CUSY(%)

obtained in common. Especially, a factor named “Accuracy” appears as the 1st factor in both Japanese and English. In addition, similar kinds of factors could be obtained in both languages. On the other hand, factors could be obtained as a different form. A Japanese factor named “Displeasure” was obtained as subdivided two English factors named “Disappointment” and “Discomfort.” Moreover, some factors could be obtained in one language while they failed to in another language. For example, the 5th and 8th Japanese factors named “Persistence” and “Effort” were not obtained in English. On the other hand, the 8th and 9th English factors named “Politeness” and “Nostalgia” did not appear in Japanese.

IV. ESTIMATION OF FACTOR SCORES FROM ENGLISH FEATURE VALUES OF STATEMENTS

A. Approach

Similar to the obtainment of factors, factor scores are estimated from English feature values of statements, in the similar fashion as explained in Section 3-B. Most of the feature values are extracted through morphological analysis applied to the statements of experimental materials used in Section 3-C. Feature values are then selected with the consideration of multicollinearity. Factor scores are estimated from those feature values selected through multiple regression analysis. As a result, estimation accuracies of all the factors are very good.

B. Explanatory Variables

1) Words and characters

In considering feature values of statements, the number of word/character of each statement is extracted. Each variable is denoted as “Word” and “Char.” Moreover, the ratio of word/character to sentence is taken into consideration as well. Each variable is denoted as “Word/SENT” and “Char/SENT.”

2) Syntactic information of statements

Besides the feature values above, syntactic information is extracted through morphological analysis applied to the statements of experimental materials used in Section 3-C. TreeTagger [16] is installed to run morphological analysis. Morphological analysis breaks down a set of sentences into each word, its Part-of-Speech (POS) tag, and basic form. Descriptions of POS tags are explained in Appendix [17].

Some words are classified as more profoundly subdivided tags as follows [17]:

TABLE III. BASE TAGS AND THEIR SUBDIVIDED TAGS.

Base	Subdivided	Description	Base	Subdivided	Description
IN	IN/that	Complementizer	VH	VHD	Past form of VH
JJ	JJR	Comparative of JJ		VHG	Gerund/participle of VH
	JJS	Superlative of JJ		VHN	Past participle of VH
NN	NNS	Plural of NN		VHZ	Pres, 3rd p. sing of VH
NP	NPS	Plural of NP		VHP	Pres non-3rd p. of VH
PP	PPS	Possessive pronoun	VV	VVD	Past form of VV
RB	RBR	Comparative of RB		VVG	Gerund/participle of VV
	RBS	Superlative of RB		VVN	Past participle of VV
VB	VBD	Past form of VB		VVZ	Pres, 3rd p. sing of VV
	VBG	Gerund/participle of VB	VVP	Pres non-3rd p. of VV	
	VBN	Past participle of VB	WH	WDT	Wh-determiner
	VBZ	Pres, 3rd p. sing of VB		WP	Wh-pronoun
VBP	Pres non-3rd p. of VB	WPS		Possessive wh-pronoun	
VD	VDD	Past form of VD	WRB	Wh-adverb	
	VDG	Gerund/participle of VD			
	VDN	Past participle of VD			
	VDZ	Pres, 3rd p. sing of VD			
	VDP	Pres non-3rd p. of VD			

- One type of prepositions or subordinate conjunctions (IN): complementizer
- Comparative/Superlative form of adjective (JJ)/adverb (RB)
- Plural form of noun (NN)/proper noun (NP)
- Different form of pronoun (PP)
- Different tense of verbs (VB, VD, VH, VV)
- Wh-words (WH)

In those cases, the more deeply subdivided tag is equally regarded as its base tag. Each base tag and its subdivided tags are shown in Table 3. Base (Subdivided, respectively) tags are given in the column entitled “Base.” (“Subdivided.”)

3) Appearance percentage

For the words given base POS tags, the appearance of each Q&A statement is considered. Taking an example of the words given a “CC” tag, the ratio of their appearance to all the words in a statement is considered and denoted as “CC(%)” The appearance percentages of the rest of base forms are also considered in the same fashion as “CC(%)”

4) Feature values adopted

Through the procedures 1) to 3), fifty-eight explanatory variables are initially extracted. These are shown in Table 4. Each explanatory variable is classified as “Word & Char,” “Base form,” and “Appearance percentage.”

In applying multiple regression analysis, it is required that explanatory variables have no correlation each other, and the following conditions must be considered.

1. Selection of the explanatory variable whose correlation coefficient to a dependent variable is high.
2. Exclusion of one of two criterion variables whose correlation coefficient is high.

If the condition 1) is not satisfied, partial regression coefficients cannot be properly calculated. This situation is called “multicollinearity.”

In order to avoid multicollinearity, either of the explanatory variables, whose correlation coefficients are high, is excluded from explanatory variables. As a result, the number of the

TABLE V. FEATURE VALUES ADOPTED.

Eg	Variable	Eg	Variable	Eg	Variable
Eg1	Word	Eg11	CC(%)	Eg19	PP(%)
Eg2	Word/SENT	Eg12	EX(%)	Eg20	RB(%)
Eg3	CD	Eg13	IN(%)	Eg21	TO(%)
Eg4	FW	Eg14	JJ(%)	Eg22	VH(%)
Eg5	MD	Eg15	MD(%)	Eg23	VV(%)
Eg6	PDT	Eg16	NN(%)	Eg24	WH(%)
Eg7	RP	Eg17	NP(%)	Eg25	GESY(%)
Eg8	UH	Eg18	POS(%)	Eg26	CUSY(%)
Eg9	VD				
Eg10	CUSY				

TABLE VI. EVs WITH HIGHER SPRCS.

1st (Accuracy)		2nd (Evaluation)		3rd (Disappointment)	
EV	SPRC	EV	SPRC	EV	SPRC
Eg5	1.56	Eg10	2.24	Eg1	3.72
Eg16	1.31	Eg5	1.23	Eg26	1.98
Eg10	1.24	Eg20	-1.14	Eg15	1.93
Eg6	1.01	Eg26	-2.03	Eg20	1.22
Eg12	1.00	Eg1	-3.16	Eg13	1.19
Eg13	-1.13			Eg10	-2.04
Eg26	-1.14			Eg5	-2.12
Eg1	-3.76				

4th (Discomfort)		5th (Novelty)		6th (Potency)	
EV	SPRC	EV	SPRC	EV	SPRC
Eg5	1.33	Eg24	1.13	Eg5	3.90
Eg26	1.22	Eg25	1.04	Eg25	2.02
Eg9	1.14	Eg1	-1.85	Eg16	1.92
Eg13	-1.18			Eg9	1.66
Eg14	-1.32			Eg24	1.56
Eg23	-1.36			Eg12	1.51
				Eg6	1.43
				Eg22	1.10
				Eg3	1.09
				Eg18	-1.17
				Eg14	-1.62
				Eg23	-1.71
				Eg15	-2.41
				Eg13	-2.55
				Eg1	-4.30

7th (Difficulty)		8th (Politeness)		9th (Nostalgia)	
EV	SPRC	EV	SPRC	EV	SPRC
Eg1	2.31	Eg5	2.10	Eg1	4.96
Eg15	1.01	Eg10	1.45	Eg15	1.28
Eg5	-1.14	Eg13	-1.12	Eg16	-1.35
		Eg23	-1.29	Eg9	-1.37
		Eg26	-1.92	Eg3	-1.39
		Eg15	-2.25	Eg6	-1.82
		Eg1	-2.64	Eg12	-2.12
				Eg5	-2.37

TABLE VII. MULTIPLE CORRELATION COEFFICIENTS.

Factor	MCC
1st (Accuracy)	0.905
2nd (Evaluation)	0.999
3rd (Disappointment)	0.953
4th (Discomfort)	0.971
5th (Novelty)	0.936
6th (Potency)	0.954
7th (Difficulty)	0.990
8th (Politeness)	0.919
9th (Nostalgia)	0.992

explanatory variables is twenty-six. Those feature values are shaded in Table 4. They are summarized and shown in Table 5. Each explanatory variable is denoted as Eg1, Eg2, ..., Eg26.

C. Estimation Result

Through multiple regression analysis, factor scores of nine factors are estimated from the twenty-six feature values shown in Table 5. Estimation results of the nine factors are not shown due to limited space. Explanatory variables (EVs) whose absolute values of standardized partial regression coefficient (SPRCs) are over 1.0 are shown in Table 6. MCCs of nine factors are shown in Table 7. They are above 0.9 for all nine factors. Therefore, it is shown that the estimation accuracies of all of the factors are very good.

V. CONSIDERATIONS

In order to explain each factor, EVs are focused on whose absolute values of the SPRCs are above 1.0.

- 1st factor (accuracy): Eg5, Eg16, and Eg10 have particularly larger positive SPRC values. Therefore, more modals, currency symbols, and higher percent of nouns could contribute to higher factor scores for accuracy. On the other hand, Eg13, Eg26 and Eg1 have strong negative SPRC values. Thus, smaller percentage of prepositions and currency symbols and fewer words could lead to higher factor scores of accuracy.
- 2nd factor (evaluation): Eg10 and Eg5 have large positive SPRC values. Thus, more currency symbols and modals could result in higher factor scores for evaluation. Meanwhile, Eg20 and Eg1 have higher negative SPRC values. Thus, smaller percentage of adverbs and fewer words could contribute to higher factor scores of evaluation.
- 3rd factor (disappointment): Eg1, Eg26, and Eg15 have particularly large positive SPRC values. Therefore, more words, higher percentage of currency symbols and modals could result in higher factor scores for disappointment. On the other hand, Eg5 and Eg10 have strong negative SPRC values. Thus, fewer currency symbols and modals could lead to higher factor scores of disappointment.
- 4th factor (discomfort): Eg5, Eg26, and Eg9 have stronger positive SPRC values. Thus, more modals and do-verbs, and higher percent of current symbols could lead to higher factor scores for discomfort. Meanwhile, Eg13, Eg14 and Eg23 have higher negative SPRC values. Therefore, smaller percentage of prepositions, adjectives, and general verbs could result in higher factor scores of discomfort.
- 5th factor (novelty): Eg24 and Eg25 have larger positive SPRC values. Therefore, higher percentage of wh-words and general symbols could lead to higher factor scores for novelty. On the other hand, Eg1 has a strong negative SPRC value. Thus, fewer words could lead to higher factor scores of novelty.
- 6th factor (potency): Eg5, Eg26, and Eg16 have particularly large positive SPRC values. Thus, more modals and higher currency symbols and nouns could contribute to higher factor scores for potency. Meanwhile, Eg15, Eg13 and Eg1 have stronger negative SPRC values. Therefore, smaller percentage of general symbols, prepositions, and fewer words could result in higher factor scores of potency.

- 7th factor (difficulty): Eg1 and Eg15 have larger positive SPRC values. Thus, more words and higher percentage of modals could result in higher factor scores for difficulty. On the other hand, Eg5 has a stronger negative SPRC value. Therefore, fewer modals could contribute to higher factor scores of difficulty.
- 8th factor (politeness): Eg5 and Eg10 have higher positive SPRC values. Therefore, more modals and currency symbols could lead to higher factor scores for politeness. Meanwhile, Eg26, Eg15 and Eg1 have stronger negative SPRC values. Thus, smaller percentage of currency symbols and modals, and fewer words could contribute to higher factor scores of politeness.
- 9th factor (nostalgia): Eg1 and Eg15 have larger positive SPRC values. Therefore more words and higher percent of modals could result in higher factor scores for nostalgia. On the other hand, Eg6, Eg12 and Eg5 have particularly stronger negative SPRC values. Thus, fewer predeterminers and modals, and smaller percentage of existential there could contribute to higher factor scores of nostalgia.

Compared with the method dependent on Japanese, feature values extracted through morphological analysis were sufficient to estimate factor scores. This could have resulted from the smaller size of English samples than that of Japanese ones. Thus, additional experiment with more subjects and samples needs to be conducted. If more samples could be gained through another experiment, it might be necessary to consider feature values other than those based on morphological analysis.

VI. CONCLUSIONS

In this paper, factor scores were estimated from feature values of English statements. Feature values were extracted through morphological analysis applied to the experimental materials. These feature values included words and characters, syntactic information, and appearance percentages. The feature values extracted were selected with the consideration of multicollinearity. As a result of estimation, it was shown that estimation accuracies of all of the factors are very good. Each factor was tried to be explained through major feature values of each factor.

For future work, an additional experiment with more subjects and samples is required. As all the foreign subjects were non-native English, an experiment with native English subjects must be conducted. It is also required to use the estimated scores to estimate the BAs. In order to detect appropriate answerers, the characteristics of users must be revealed and used for investigation.

ACKNOWLEDGMENT

This research was partially supported by the Ministry of Education, Science, Sports, and Culture, Grant-in-Aid for Scientific Research (B), 25280110, 2013-2016, and by the Japan Society for the Promotion of Science, Grant Number 26008587, and 2015-2016.

REFERENCES

- [1] Yahoo! Chiebukuro (URL, in Japanese), <http://chiebukuro.yahoo.co.jp/>, 2016.
- [2] Yokoyama, Y., Hochin, T., Nomiya, H. and Satoh, T.: Obtaining Factors Describing Impression of Questions and Answers and Estimation of their Scores from Feature Values of Statements, Software and Network Engineering, pp.1-13, Springer, 2012.
- [3] Yokoyama, Y., Hochin, T. and Nomiya, H.: Using Feature Values of Statements to Improve the Estimation Accuracy of Factor Scores of Impressions of Question and Answer Statements, International Journal of Affective Engineering, Vol. 13 (2014) No. 1 Special Issue on ISAE 2013, pp.19-26, 2014.
- [4] Yokoyama, Y., Hochin, T. and Nomiya, H.: Factors Describing Impression of English Question and Answer Statements, Proc. of 11th Spring Conference of Japan Society of Kansei Engineering, G12-3, 2016.
- [5] Yahoo! Answers (URL), <https://answers.yahoo.com>, 2016.
- [6] Blooma, M.J. and Chua, A.Y.K. and Goh, D.H.L.: A Predictive Framework for Retrieving the Best Answer, Proc. of 2008 ACM Symposium on Applied Computing (SAC08), pp.1107-1111, 2008.
- [7] Wang, X.J., Tu, X., Feng, D. and Zhang, L.: Ranking Community Answers by Modeling Question-Answer Relationships via Analogical Reasoning, Proc. of 32nd Int'l ACM SIGIR Conf., pp.179-186, 2009.
- [8] Nishihara, Y., Matsumura, N. and Yachida, M.: Understanding of Writing Style Patterns between Q&A in Knowledge Sharing Community, the 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 1H2-7, 2008.
- [9] Jurczyk, P. and Agichtein, E.: Discovering Authorities in Question Answer Communities by Using Link Analysis, Proc. of 16th ACM Conference on Information and Knowledge Management (CIKM), pp. 919-922, 2007.
- [10] Riahi, F., Zolaktaf, Z. Shafiei, M. and Milios, E.: Finding Expert users in Community Question Answering, Proc. of the 21st International Conference Companion on World Wide Web (WWW12), pp.791-798, 2012.
- [11] Stack Overflow (URL), <http://stackoverflow.com>, 2016.
- [12] Agichtein, E., Carlos, C., Donato, D., Gionis, A., and Gilad, M.: Finding High-Quality Content in Social Media, Proc. of the Int'l Conf. on Web Search and Web Data Mining (WSDM08), pp.183-194, 2008.
- [13] Liu, Y., Li, S., Cao, Y., Lin, C., Han, D., and Yu, Y.: Understanding and Summarizing Answers in Community-Based Question Answering Services, Proc. of the 22nd International Conference on Computational Linguistics, pp.497-504, 2008.
- [14] Raban, D. R.: Self-Presentation and the Value of Information in Q&A Websites, Journal of the American Society for Information Science and Technology Volume 60, Issue 12, pp.2465-2473, 2009.
- [15] sourceforge, Yahoo! Answers Datasets, Summary (URL), <http://sourceforge.net/projects/yahoodataset/>, 2016.
- [16] TreeTagger - a language independent part-of-speech tagger (URL), <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>, 2016.
- [17] Tree Tagger Tag Set (58 tags) (URL), <https://courses.washington.edu/hypertext/csar-v02/penntable.html>, 2016.

APPENDIX

POS tags available through TreeTagger [16] are shown in Table 8. What each POS Tag means and its example is summarized in the column entitled "Explanation" and "Example," respectively [17].

TABLE VIII. POS TAGS, EXPLANATIONS, AND EXAMPLES.

POS Tag	Explanation	Example	POS Tag	Explanation	Example
CC	coordinating conjunction	and, but, or, &	VB	verb be, base form	be
CD	cardinal number	1, three	VBD	verb be, past	was/were
DT	determiner	the	VBG	verb be, gerund/participle	being
EX	existential there	there is	VBN	verb be, past participle	been
FW	foreign word	d'œuvre	VBZ	verb be, pres, 3rd p. sing	is
IN	preposition/subord. conj.	in,of,like,after,whether	VBP	verb be, pres non-3rd p.	am/lare
IN/that	complementizer	that	VD	verb do, base form	do
JJ	adjective	green	VDD	verb do, past	did
JJR	adjective, comparative	greener	VDG	verb do gerund/participle	doing
JJS	adjective, superlative	greenest	VDN	verb do, past participle	done
LS	list marker	(1),	VDZ	verb do, pres, 3rd per.sing	does
MD	modal	could, will	VDP	verb do, pres, non-3rd per.	do
NN	noun, singular or mass	table	VH	verb have, base form	have
NNS	noun plural	tables	VHD	verb have, past	had
NP	proper noun, singular	John	VHG	verb have, gerund/participle	having
NPS	proper noun, plural	Vikings	VHN	verb have, past participle	had
PDT	predeterminer	both the boys	VHZ	verb have, pres 3rd per.sing	has
POS	possessive ending	friend's	VHP	verb have, pres non-3rd per.	have
PP	personal pronoun	I, he, it	VV	verb, base form	take
PP\$	possessive pronoun	my, his	VVD	verb, past tense	took
RB	adverb	however, usually, here, not	VVG	verb, gerund/participle	taking
RBR	adverb, comparative	better	VVN	verb, past participle	taken
RBS	adverb, superlative	best	VVP	verb, present, non-3rd p.	take
RP	particle	give up	VVZ	verb, present 3d p. sing.	takes
SENT	end punctuation	?, !, .	WDT	wh-determiner	which
SYM	symbol	@, +, *, ^, , =	WP	wh-pronoun	who, what
TO	to	to go, to him	WP\$	possessive wh-pronoun	whose
UH	interjection	uhhuhuhuh	WRB	wh-adverb	where, when
			:	general joiner	;, -, --
			\$	currency symbol	\$, £