Acta Technica 62 No. 1B/2017, 391–402

Semantic analysis of English vocabulary based on random feature selection

Jufang Gong[1]

Abstract. The false statements of English fact on the Internet have seriously affected people's effective access to information, and how to determine whether the English fact statement is semantically credible becomes an urgent problem to be solved. A kind of Multi-answer English fact Statements Verification model (MFSV) based on random feature selection for the English vocabulary is proposed in this paper. In view of the characteristics of English fact statements, this model collects the English vocabulary information that is related to the English fact statements to be verified from the Internet, and measures the random feature selection corresponding to the English fact statements. At the same time, this model takes into consideration the difference in the semantic credibility of the relevant English vocabulary information, measures the semantic credibility of the relevant English vocabulary information source from the two aspects of popularity and importance, and obtains the semantic credibility ranking of the relevant English vocabulary information. According to the random feature selection and the semantic credibility ranking, the contribution of the relevant English vocabulary information to the semantic credibility verification of the corresponding English fact statements is measured, based on which the verification of the semantic credibility of the English fact statements to be judged is realized. And a series of experiments have verified the rationality and accuracy of the semantic verification of the model.

Key words. English fact statements, random feature selection, semantic verification, semantic credibility ranking.

1. Introduction

The Internet is an important source of information for people to obtain information; however, false statements of English fact on the Internet have seriously affected people's effective access to information. Therefore, how to determine whether the English fact statement is semantically credible becomes an urgent problem to be solved. On the Internet, the English fact descriptive information is mainly expressed in the form of English Fact Statement, which only expresses the true semantics of

[1]Department of Foreign Languages of Xiangsihu College, Guangxi University for Nationalities, Nanning, Guangxi, 530008, China

the information delivered, and such English fact statement is a semantically credible English fact statement; on the contrary, the English fact statement is not credible. According to the characteristics of the English fact statements, it can be seen that for any negative English fact statement, there is always a corresponding affirmative English fact statement, and the semantic credibility verification of the negative English fact statements can be realized through the semantic a credibility verification of the corresponding affirmative English fact statement. In this paper, only the credibility verification on the affirmative English fact statements is conducted.

When determining whether an English fact statement is semantically credible, the unit of doubt in the English fact statement will be specified [1], that is, the part in the English fact statement that the users need to verify. In this case, the English fact statements can be regarded as the answer to a subject. When the subject corresponding to the English fact statements has the only correct answer, the English fact statement is the only answer to the English fact statement; on the other hand, it is called the multi-answer English fact statement. For example, in the specified English fact statement "Obama is American president", "Obama" is the unit of doubt, and then the subject of the English fact statement corresponds to: "Who is American president", and the answer to the subject is unique. Therefore, the English fact statement "Obama is American" is the only answer to the English fact statement. When the unit of doubt is designated as "American president", the corresponding subject is "What is Obama", and the subject has multiple correct answers, including "Obama is a Christian", "Obama is American president" and so on. In this case, the English fact statement is a multi-answer English fact statement.

The basic idea of the information semantic credibility verification is to realize the semantic credibility verification of the information by acquiring and analyzing the English vocabulary information related to the information to be judged. Literatures [2–3] put forward the verification method for the news information. Through obtaining the relevant information of the news information to be verified from the news website, the consistency and objectivity of the content of the relevant information and the news information to be verified is taken into consideration, so as to realize the credibility verification of the news information semantics. Literature [4] puts forward an event semantics credibility verification method. Through obtaining the relevant information to be verified from the relevant website where the event occurs, the relevance of the relevant information and the event to be verified is analyzed from the three dimensions of time, space and characters. At the same time, the semantic credibility of the information source is taken into account, to achieve the verification function of the auxiliary users for the authenticity of the event. Literature [5, 6] put forward an English fact statement semantic credibility verification system Verify that is irrelevant to the direct verification domain. Through the search engine, Verify obtains the English fact relevant information to be verified, from which the English fact that can be compared is identified; then, the relevant information of the English fact that can be compared is obtained, respectively, and the rating is conducted for the comparable English facts selected from the aspects of the English vocabulary feature of the relevant information and the source features, and so on. The one with the highest rating is the semantic credible English fact statement.

In view of the defects of the aforementioned research work, this paper puts forward a kind of new multi-answer English fact statement verification (MFSV) that is irrelevant to the new domain. The model obtains the relevant English vocabulary information of the English fact statements to be verified through the search engine. In the process of verification of the English fact statements, the model takes into account the supporting relationship between the relevant English vocabulary information and the corresponding English fact statement, as well as the difference in the semantic credibility of the relevant English vocabulary information, thus making up for the defect of the first category of verification method [7]. In addition, the process of using this model to conduct English fact statement verification, it is not necessary to specify the unit of doubt of the English fact statements, or look for and analyze the English fact statement that can be compared. Therefore, this model is also applicable to the multi-answer English fact statement semantic credibility verification, so as to make up for the limitation of the second category of the verification [8–9] in the multi-answer English fact statement verification.

2. Multi-answer English fact statements verification model (MFSV)

English fact statements semantic credibility verification model MFSV is shown in Fig. 1. The model consists of four modules: The relevant English vocabulary information acquisition; random feature selection measurement; English vocabulary information semantic credibility ranking; and English facts statements semantic credibility verification. The input of the model is a statement of the English fact to be verified, and the output is the result of the semantic credibility verification of the English fact statement.

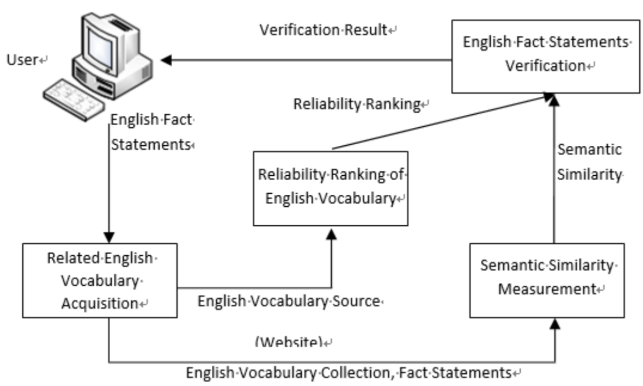

Fig. 1. Multi-answer English fact semantic credibility verification model MFSV

The related English vocabulary information acquisition module obtains the related English vocabulary information of the English fact statement to be verified. As the model is to make verification based on the English fact statement that is irrelevant to the domain, the acquisition of the relevant English vocabulary information is completed by the search engine. The input of the module is the English fact statement to be verified. And the outcome is the relevant English vocabulary information collection corresponding to the statement. The random feature selection measurement module is used to calculate the random feature selection between the relevant English vocabulary information and the corresponding English fact statement. Firstly, the sentences that make sense for the English fact statement verification is extracted from the English vocabulary information; secondly, the similarity between the sentences extracted and the corresponding English fact statement is measured, so that the English vocabulary information and the random feature selection of the corresponding English fact statement is obtained. The input of this module is the relevant English vocabulary information and the corresponding English fact statement. The output is the English vocabulary information and the random feature selection of the corresponding statement. The semantic credibility ranking module realizes the semantic credibility ranking of the relevant English vocabulary information. In this module, through obtaining the Page rank corresponding to the relevant English vocabulary information source (website) and its position in the Alex ranking, the importance ranking and the popularity ranking of the source of the relevant information is realized. From the combination of these two ranking, the relevant English vocabulary information semantic credibility ranking can be obtained. The input of this module is the relevant English vocabulary information source (website), and the output is the credibility ranking of the English vocabulary information semantic. The English fact statement semantic credibility verification module has realized the semantic credibility verification of the English fact statements. In this module, according to the similarity of the English vocabulary information and the corresponding English fact statement, as well as the English vocabulary information semantics credibility ranking, the contribution of other English vocabulary information to the corresponding English fact statement verification is measured. Combined with the measurement on the contribution made by the English vocabulary information, the semantic credibility verification of the English fact statement is realized. The input of this module is the random feature selection and the semantic credibility ranking of the relevant English vocabulary information. And the output is the verification result.

2.1. Random feature selection measurement

The similarity between the relevant English vocabulary information and the corresponding English fact statements is the basis for the measurement of the supporting relationship between the relevant English vocabulary information and the corresponding English fact statement. And this section has described the calculation process of the similarity of the relevant English vocabulary information with the corresponding English fact statements.

The random feature selection of the English vocabulary information r_i and fs is the similarity of the sentence of st_i and fs. This paper has extended the sentence similarity calculation method based on the semantics and word order [10], so as to realize the similarity measurement of st_i and fs. The traditional sentence similarity calculation method based on the semantics and word order does not take into account the influence of the adjacent word of the influence of the adjacent word to the target word on the acquisition of the matching word when generating the semantic vector and work order vector through searching for the matching word of the target word.

In the sentence similarity calculation method described in this paper, it is considered that the semantics of the words in a sentence will be affected by the adjacent words. Therefore, in the process of acquiring the optimal matching word for the target word, the adjacent word to the target word is taken as an important factor for measurement. And the optimal matching word acquisition algorithm is put forward. And on this basis, the corresponding semantic vector and word order vector is obtained. At the same time, in the process of the similarity calculation, the case of the English vocabulary information negative corresponding English fact statement is taken into consideration.

The similarity between the words is the basis of generating the semantic vectors and word order vectors. The formula to calculate the similarity between the words is shown in equation (1). Equation (1) calculates the similarity between the calculated words w_1 and w_2. Symbols l and h represent the shortest distance between $w_1$and w_2 in the Wordnet, and depth of the common category that both w_1 and w_2 belong to in the Wordnet, respectively. When $\alpha = 0.2$, $\beta = 0.45$, the similarity between the words can be measured properly through equation

$$S_w\left(w_1, w_2\right) = \left\{ \begin{array}{ll} \mathrm{e}^{-at} \cdot \frac{\mathrm{e}^{\beta h} - \mathrm{e}^{-\beta h}}{\mathrm{e}^{\beta h} + \mathrm{e}^{-\beta h}}\,, & w_1 \neq w_2\,, \\ 1, & w_1 = w_2\,. \end{array} \right\} \tag{1}$$

In equation (1), when $w_1 = w_2$, it is considered that the relevance is 1; in addition, as in the actual situation, WordNet cannot completely cover all the words that appear in the information, when w_1 or w_2 is not covered by the WordNet, there is $S_w\left(w_1, w_2\right) = 0$.

Assuming that s_1 is the sentence st_i that is extracted from r_i, and s_2 is the corresponding English fact statement fs of r_i. The next section describes the similarity calculation process taking the calculation of the similarity of s_1 and s_2, for example.

Conduct the semantic vector correlation calculation. Through the generation of the corresponding semantic vector of the sentence s_1 and s_2, the cosine similarity between the semantic vectors is calculated, so as to realize the calculation of the semantic vector correlation. Assuming that the word set after the suspension word is removed from s_1 and s_2 is $W_1 = \{w_{11}, w_{12}, \cdots w_{1n}\}$, $W_2 = \{w_{21}, w_{22}, \cdots w_{2n}\}$, $W = W_1 \bigcup W_2$ and $W = \{w_1, w_2, \cdots w_k\}$, respectively. Assuming the corresponding semantic vector to s_1 is $V_1 = \{v_{11}, v_{12}, \cdots v_{1k}\}$, the process to obtain the component v_{1i} is as follows:

(1) Suppose $w_i \in W$. If $w_i \in W_1$, then $v_{1i} = 1$.

(2) Let $w_i \in W$. If $w_i \notin W_1$, search for the optimal match word w_{bm} for w_i (target word) in s_1. If w_{bm} exists, $v_{1i} = S_w\ (w_i, w_{\mathrm{bm}})$. Otherwise, $v_{1i} = 0$ is the process to obtain the optimal match word w_{bm}.

Similarly to the method of obtaining V_1, the semantic vector V_2 corresponding to s_2 can be obtained. The semantic vector correlation of s_1 and s_2 can be obtained by calculating the cosine similarity of V_1 and V_2, as shown in the formula

$$S_s\left(s_1, s_2\right) = \frac{V_1 \bullet V_2}{\|V_1\| \bullet \|V_2\|}. \tag{2}$$

Calculate the word order vector relevance. Through generating the word order vector corresponding to the sentence, according to equation (3), and the word order vector similarity between the sentences is calculated. O_1 and O_2 in Equation (3) represent the word order vector of s_1 and s_2, respectively. The process to generate the word order vector $O_1 = \{o_{11}, o_{12}, \cdots o_{1k}\}$ is as follows:

(1) Let $w_i \in W_1$. If $w_i \in W_1$, o_{11} is the position of w_i in s_1.

(2) Let $w_i \in W_1$. If $w_i \notin W_1$, according to Algorithm 2, search for the optimal match word $w_{\rm bm}$ to w_i in s_1. If $w_{\rm bm}$ is present, o_{1i} is the position of $w_{\rm bm}$ in s_1. Otherwise, $o_{1i} = 0$. In the process of obtaining the word order vector, the optimal value of the parameter ζ involved in Algorithm 2 is 0.4.

$$S_{\rm re}\left(s_1, s_2\right) = 1 - \frac{\|O_1 - O_2\|}{\|O_1 + O_2\|}. \tag{3}$$

Conduct the random feature selection calculation. According the semantic vector correlation and word order vector correlation, the random feature selection of s_1 and s_2 can be calculated by equation (4). Since the sentences st_i and s_2 are the English fact statements fs (corresponding to r_i extracted from r_i, s_1 and s_2) are expressed using st_i and fs in equation (4), respectively. And the optimal value of the parameter θ in equation (4) is 0.85.

$$S\left(st_i, fs\right) = \begin{cases} \theta S_s\left(st_i, fs\right) + (1-\theta) S\left(st_i, fs\right) \\ \left(r_i \text{ does not have a negative tendency for } fs\right), \\ -\left(\theta S_s\left(st_i, fs\right) + (1-\theta) S\left(st_i, fs\right)\right) \\ \left(r_i \text{ has a negative tendency for } fs\right). \end{cases} \tag{4}$$

Whether r_i has the negative tendency to fs is verified according to whether the process to obtain st_i involves the negative grammatical dependency relationship representation, and whether the negative adverb in c_i, such as hardly, rarely, few, seldom and so on. As can be known from Stanford Parser, the sentence corresponding grammatical dependency between words can represent the negative tendency that is clearly existent in the sentence. For example, when a negative word not appears in a sentence, the corresponding grammatical dependency relationship is neg. And when the negative conjunction, such as rather than, and so on, occurs, it is reflected in the grammar dependency as conj_negcc, and so on. Therefore, it is possible to determine whether there is negative tendency in the sentence through the grammatical dependency relationship. In addition, when the negative adverbs, hardly, rarely, few, seldom and so on appear in the sentence, it is also the basis to verify whether there is negative tendency in the sentence. Therefore, whether r_i has a negative tendency to fs can be determined according to the following rules:

(1) Whether the grammatical dependency relationships when st_i is extracted from c_i includes the grammatical dependency relationship that represents the negative, such as neg and conj_negcc, etc.

(2) Whether c_i includes negative adverbs, such as hardly, rarely, few, seldom, scarcely, never, little, etc. If one of these aforementioned rules is met, it is considered that the corresponding English vocabulary information r_i has the negative tendency to fs. As this paper is to conduct the semantic verification for the affirmative English fact statement, in the aforementioned negative tendency verification process, the case that fs is the negative English fact statement is not considered.

2.2. English fact statement semantic credibility verification

According to the similarity of the relevant information to the corresponding statement, the threshold value k is introduced, and the relevant English vocabulary information is divided into three categories:

(1) For the supportive English vocabulary information of the corresponding statement, its set is expressed with R_{pos}, if $S(r_i, fs) \geq k$, $r_i \in R_{\text{pos}}$.

(2) For the objection to the corresponding statement English vocabulary information, and its set is expressed with R_{neg}, if $|S(r_i, fs)| \geq k$, and $S(r_i, fs) < 0$, then $r_i \in R_{\text{neg}}$.

(3) For the neutral English vocabulary information, its set is expressed with R_{neu}, if $|S(r_i, fs) < k|$, $r_i \in R_{\text{neu}}$. The contribution of the English vocabulary information to the English fact statement to be determined is determined by the corresponding semantic relevance and the ranking of the English vocabulary information in the semantic credibility ranking. The contribution of r_i to the semantic credibility verification of fs is $S(r_i, fs)/\text{Crank}$. Symbols Δ_{pos}, $\Delta_{\text{neg}}$and Δ_{neu} represent that the optimal value of the summary k of the contribution of the English vocabulary information in R_{pos}, R_{neg} and $R_{\text{neu}}$to the English fact statements is determined by the experiment.

The basic verification method suggests that if the contribution of the English vocabulary information supporting the statement is much greater than the contribution of the information against the statement in the relevant English vocabulary information corresponding to the English fact statement fs, the semantics of fs is credible and vice versa, fs is the English fact statement with unconfirmed semantic credibility. Through introducing the threshold value δ, according to the sum of Δ_{pos} and Δ_{neg} the sum and the relationship with the size of δ, the semantic credibility verification of fs can be achieved. Input the relevant English vocabulary information set R to the English fact statement fs to be verified, the semantic credibility ranking corresponding to R is Crank, and the similarity of the English vocabulary information r_i and the corresponding English fact statement fs is $S(r_i, fs)$, and the threshold value is δ; the output of the algorithm is the result of the verification on fs. Firstly, the relevant English vocabulary information $r_i = (i = 1, \cdots, n)$ is processed one by one, and Δ_{pos},Δ_{neg} are calculated; then the sum of Δ_{pos} and Δ_{neg}

is calculated. If its value is greater than or equal to the threshold value δ (the value of δ is determined by the experiment), then fs is the semantic credible English fact statement (return Ture); on the contrary, fs is the English facts statement with unconfirmed semantic credibility.

3. Experiments

In this paper, there is not public recognized data set existing in the English fact statement semantic verification research field. Therefore, the English fact statements are obtained from TREC2007 to constitute experimental dataset. 30 semantic credible unique answers to the English fact statements and 20 semantics credible multi-answer English fact statements are randomly selected from TREC2007 to constitute the dataset semantics credible English fact statement part. As in the TREC2007, any semantic credible English fact has a corresponding English fact statement that is close to it and without credible semantics, 50 corresponding non-semantically credible English fact statements are selected as the part of the non semantic credible English fact statements in the dataset. In this paper, two related English vocabulary information acquisition methods are provided as follows: To use the English fact statement as the search engine to query and access to the relevant English vocabulary information, which is referred to as FQ method for short; to use the keyword collection of the English fact statement as the search engine to query and access to the relevant English vocabulary information, which is referred to as the KQ method for short. For any English fact statement, through yahoo boss 2.0, two methods of FQ, KQ are used to obtain the first 150 search results (relevant English vocabulary information). In addition, 13 graduate students with long-term Internet experience annotate the English vocabulary information according to the relationship between the obtained English vocabulary information and the corresponding English fact statement. According to the annotation, the English vocabulary information can be divided into three categories, including supporting corresponding statements, objection to the corresponding statements and irrelevant to the corresponding statements.

The rationality and accuracy of the MFSV, English semantic credibility verification model are verified by carrying out a series of experiments.

(1) Through the experimental analysis, in the different English vocabulary information acquisition methods (FQ and KQ), the distribution of the English vocabulary information with the semantics containing the corresponding English fact statement.

(2) The value of the threshold k affects the classification of the relevant English vocabulary information, thus affecting the accuracy of the English fact statement verification, and the optimal value is obtained through the experiment.

(3) Analyze the ranking distribution of the English vocabulary information semantic credibility in different relevant English vocabulary access modes.

(4) Analyze the influence of English vocabulary information quantity n, threshold value δ, semantic credibility ranking and English vocabulary information acquisition mode on the accuracy of basic verification method.

(5) The influence of n, semantic credibility ranking and the relevant English vocabulary information acquisition mode on the accuracy of the SVM verification

method.

(6) Analyze the difference between the basic verification method and the SVM verification method in determining the accuracy.

The experiment was conducted on the platform of Intel Core 2Quad2, 66 GHz processor and 2 GB memory Windows 7.

3.1. Semantic credibility ranking of English vocabulary information

This experiment analyzes the distribution of the semantic credibility ranking in two acquisitions modes FQ and KQ. Figures 2 and 3 show in FQ and KQ mode, when the number of the English vocabulary information is 150 ($n = 150$), the distribution of CBrank, CBGrank, CFrank and CFGrank, respectively. And the horizontal coordinates represent the position of the English vocabulary information in the English vocabulary information set. The vertical coordinates represent the semantic credibility ranking of the English vocabulary information in the corresponding position.

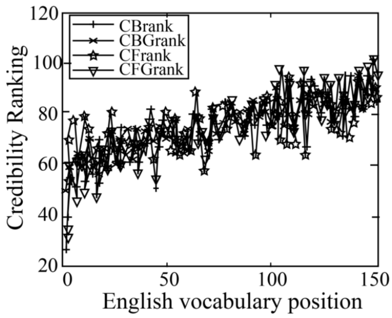

Fig. 2. Semantic credibility ranking in FQ mode

It can be seen from Fig. 2, that the semantic credibility ranking is not significantly related to the position of the English vocabulary information in the English vocabulary information set. The semantic credibility of the English vocabulary information in the English vocabulary information set is not always lower than the semantic credibility of the top ranking English vocabulary information. As CBGrank and CFrrank take the Alexa ranking interval into consideration, CBGrank and CFGrank have the characteristics of large span compared with CBrank and CFrank. As can be seen from Fig. 3, in the KQ mode, the semantic credibility ranking shows the similar trend to Fig. 2; compared with the FQ mode, the relevant English vocabulary information semantic credibility distribution in the FQ mode is more concentrated (the span in the FQ and KQ under CFGrank is $26 \sim 102$ and $24 \sim 96$, respectively), and the reason is the influence of the semantic information contained in the search when the search engine returns the search result in the FQ method compared with the KQ mode.

3.1.1. Verification case analysis The validity of the method described in this paper is illustrated taking the English fact statement "English is the primary lan-

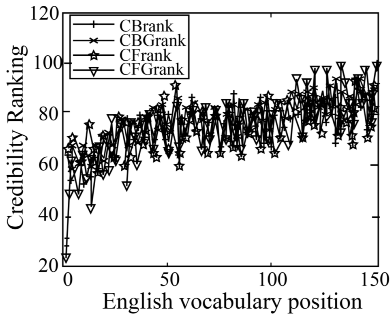

Fig. 3. Distribution of the semantic credibility ranking in the KQ mode

guage of the Philippines" as an example. As "English is the primary language of the Philippines" is a multi-answer English fact statement (when "English" is specified as the unit of doubt, the statement is multi-answer English fact statement). In order to show that the method described in this paper is still valid for the multi-answer English fact statements, another English fact statement "Filipino is the primary language of the Philippines" is verified at the same time; to illustrate the method in this paper can also correctly verify the non-semantically credible statement, the third English fact statement "Chinese is the primary language of the Philippines" is verified. Using the SVM verification method described in this paper to verify the above-mentioned English facts, and the verification result is that the first second English fact statements are true, and the third English fact statement is not true. This case illustrates the validity of the method proposed in this paper in the verification of the English fact statements.

4. Conclusion

This paper proposes a multi-answer English fact statements verification model MFSV that is irrelevant to the domains selected based on random features. This model realizes the semantic credibility verification of the English fact to be verified by acquiring and analyzing the relevant English vocabulary information corresponding to the English fact statement. In the process of semantic credibility verification of the English fact statements, the random feature selection between the relevant English vocabulary information and the English fact statements, as well as the semantic credibility of the relevant English vocabulary information is taken into account. And the relevant English vocabulary information semantic credibility and other factors are considered as well, based on which the contribution of the relevant English vocabulary information to the English fact statement semantic credibility verification, and realize the English fact semantic credibility verification. The verification model does not require specifying the English fact statement unit of doubt, which makes the verification model applicable for the unique answer to the English fact and multi-answer English fact semantic verification. As the English fact statements do not include the emotion and degree description, for the negative English fact statement semantic credibility verification, its semantic credibility verification can be conducted through the corresponding affirmative English fact statements, so as to help make the semantic credibility verification for the negative English fact statement. Therefore, in this paper, verification is conducted mainly on the affirmative English fact statements. The quality of the related English vocabulary information is the prerequisite for the accurate determination of the English facts. When the English facts to be verified are relatively complicated, the corresponding information quality is relatively low, and it is difficult to make the correct verification to such English facts. Therefore, in the future work, it is expected that the English fact statements decomposition, rewriting and other technologies should be adopted to access to the relevant high-quality English vocabulary information, so that the semantic credibility verification for the English fact statements is more accurate.

References

[1] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, A. Hussain: *Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach.* Cognitive Computation *7* (2015), No. 4, 487–499.

[2] Y. W. Hung, Y. H. Chiu, Y. C. Jou, W. H. Chen, K. S. Cheng: *Bed posture classification based on artificial neural network using fuzzy c-means and latent semantic analysis.* Journal of the Chinese Institute of Engineers *38* (2015), No. 4, 415–425.

[3] P. Munro, S. Siyambalapitiya: *Improved word comprehension in Global aphasia using a modified semantic feature analysis treatment.* Clinical Linguistics & Phonetics *31* (2017), No. 2, 119–136.

[4] J. U. Heu, I. Qasim, D. H. Lee: *FoDoSu: Multi-document summarization exploiting semantic analysis based on social Folksonomy.* Information Processing & Management *51* (2015), No. 1, 212–225.

[5] J. Wang, J. Peng, O. Liu: *A classification approach for less popular webpages based on latent semantic analysis and rough set model.* Expert Systems with Applications *42* (2015), No. 1, 642–648.

[6] M. Gambhir, V. Gupta: *Recent automatic text summarization techniques: a survey.* Artificial Intelligence Review *47* (2017), No. 1, 1–66.

[7] T. Ruette, K. Ehret, B. Szmrecsanyi: *A lectometric analysis of aggregated lexical variation in written Standard English with Semantic Vector Space models.* International Journal of Corpus Linguistics *21* (2016), No. 1, 48–79.

[8] L. van Maastricht, E. Krahmer, M. Swerts: *Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa.* Language Learning *66* (2016), No. 1, 124–158.

[9] G. Dias, R. Moraliyski, J. Cordeiro, A. Doucet: *Automatic discovery of word semantic relations using paraphrase alignment and distributional lexical semantics analysis.* Natural Language Engineering *16* (2010), No. 4, 439–467.

[10] M. Granroth-Wilding, M. Steedman: *A robust parser-interpreter for jazz chord sequences.* Journal of New Music Research *43*, (2014), No. 4, 355–374.

Received July 12, 2017