Utterances Assessment in Chat Conversations

Mihai Dascalu, Stefan Trausan-Matu, Philippe Dessus

Utterances Assessment in Chat Conversations

Mihai Dascalu[1,2], Stefan Trausan-Matu[1,3], Philippe Dessus[4]

[1] University "Politehnica" of Bucharest,
313, Splaiul Indepentei, 060042 Bucharest, ROMANIA

[2] S.C. CCT S.R.L.,
30, Gh Bratianu, 011413 Bucharest, ROMANIA

[3] Romanian Academy Research Institute for Artificial Intelligence,
13, Calea 13 Septembrie, Bucharest, ROMANIA

[4] Grenoble University,
1251, av. Centrale, BP 47, F-38040 Grenoble CEDEX 9, FRANCE
mikedascalu@yahoo.com, stefan.trausan@cs.pub.ro, Philippe.Dessus@upmf-grenoble.fr

Abstract. With the continuous evolution of collaborative environments, the needs of automatic analyses and assessment of participants in instant messenger conferences (chat) have become essential. For these aims, on one hand, a series of factors based on natural language processing (including lexical analysis and Latent Semantic Analysis) and data-mining have been taken into consideration. On the other hand, in order to thoroughly assess participants, measures as Page's essay grading, readability and social networks analysis metrics were computed. The weights of each factor in the overall grading system are optimized using a genetic algorithm whose entries are provided by a perceptron in order to ensure numerical stability. A gold standard has been used for evaluating the system's performance.

Keywords: assessment of collaboration, analysis of discourse in conversation, social networks, LSA, Computer Supported Collaborative Learning.

1 Introduction

As a result of the ongoing evolution of the web, new collaboration tools emerged and with them the desire to thoroughly process large amounts of information automatically. From the Computer Supported Collaborative Learning's (CSCL) point of view [1], chats play an important role and have become more and more used in the effective learning process. On the other hand, manual assessment of chats is a time consuming process from the teacher's side, and therefore the need to develop applications that can aid the evaluation process has become essential. From this perspective the major improvement targeted by this paper is the development of an automatic assessment system in order to evaluate each participant in a chat environment. A series of natural language processing and social network analysis methods were used, in addition with other computed metrics for assessment.

Special issue: Natural Language Processing and its Applications.
Research in Computing Science 46, 2010, pp. 323-334

Received 24/11/09
Accepted 16/01/10
Final version 09/03/10

The system was used for CSCL chats in which teams of 4-8 students were asked to discuss, without a moderator, the benefits of online collaboration tools. Each of the students was assigned to support a collaborative technology (wikis, blogs, chats and forums), arguing both pros and cons for it. The language was English and the environment used was Concert Chat [6], which offers the possibility of explicit referencing previous utterances. From the obtained corpus, 80 chats were afterwards manually evaluated by a student from a different year for not influencing the assessment process.

The next section of this paper will present the metrics used in the evaluation process starting from the simplest, as readability or Page's factors, initially used for essay grading [3], moving to social network analysis and finally *Latent Semantic Analysis* (LSA) for a semantic approach of the marking system. The third section evaluates the system.

2 The Evaluation Process

Communication between participants in a chat is conveyed through language in a written form. Lexical, syntactic, and semantic information are the three levels used to describe the features of written utterances [2], and will be taken into account for the analysis of a participant's involvement in a chat. First, *surface metrics* are computed for all the utterances of a participant in order to determine factors like fluency, spelling, diction or utterance structure [2, 3]. All these factors are combined and a mark is obtained for each participant without taking into consideration a lexical or a semantic analysis of what they are actually discussing. At the same level *readability* ease measures are computed.

The next step is *grammatical and morphological analysis* based on spellchecking, stemming, tokenization and part of speech tagging. Eventually, a *semantic evaluation* is performed using LSA [4]. For assessing the on-topic grade of each utterance a set of predefined keywords for all corpus chats is taken into consideration.

Moreover, at the surface and at the semantic levels, metrics specific to social networks are applied for proper assessment of participants' involvement and similarities with the overall chat and predefined topics of the discussion.

2.1 Surface Analysis

In order to perform a detailed surface analysis two categories of factors are taken into consideration at a lexical level: Page's essay grading proxes and readability. Page's idea was that computers could be used to automatically evaluate and grade student essays as effective as any human teacher using only simple measures – statistically and easily detectable attributes [5]. The main purpose was to prove that computers could grade as well, but with *less effort and time*, therefore enabling teachers to *assign more writing*. So the goal was to improve the student's capabilities by practice, having at hand the statistical capabilities of computers for writing analysis.

In order to perform a statistical analysis, Page correlated two concepts: *proxes* (computer approximations of interest) with human *trins* (intrinsic variables – human

measures used for evaluation). The overall results were remarkable – a correlation of 0.71 using only simple measures which proved that computer programs could predict grades quite reliably - at least the grades given by the computer correlated with the human judges as well as the humans had correlated with each other.

Starting for Page's metrics [5] for automatically grading essays, and taking into consideration Slotnick's method [5] to group them correspondingly to their intrinsic values, the following factors and values were identified in order to evaluate each participant only at the surface level:

Table 1. Categories taken into consideration and corresponding proxes

Number	Quality	Characteristic Proxes
1.	Fluency	Number of total characters, number of total words, number of different words, mean number of characters per utterance, number of utterances, number of sentences (different, because in an utterance multiple sentences can be identified)
2.	Spelling	Misspelled words, but in order to obtain a positive approach (the greater the percentage, the better) the percentage of correctly written words is used
3.	Diction	Mean and standard deviation of word length
4.	Utterance Structure	Number of utterances, mean utterance length in words, mean utterance length in characters

All the above proxes determine the average consistency of utterances. Although simple, all these factors play an important role in discovering the most important person in a chat, in other words to measure his activity. In addition, quantity is also important in its part of analyzing each participant's utterances.

Each factor has the same weight in the corresponding quality and the overall grade is obtained by using the arithmetic mean on all predefined values. All these factors, except misspelled words, are converted into percentages in order to scale them and to obtain a relative mark for all participants.

The second factor taken into account is readability. It can be defined as *reading ease* of a particular text, especially as it results from one's writing style. This factor is very important because extensive research in this field show that easy-reading text (and in our case chats and utterances) has a great impact on comprehension, retention, reading speed, and reading persistence.

Because readability implies the interaction between a participant and the collaborative environment, several features from the *reader's point of view* are essential: prior knowledge, personal skills and traits (for example intelligence), interest, and motivation.

In the currently evaluated chats, the first factor (prior knowledge) can be considered approximately the same for all students because all come from the same educational environment and share a common background. On the other hand, the remaining features vary greatly from one student to another and the last two ones are greatly reflected in their implication in the chat.

Therefore two key aspects must be taken into consideration: *involvement* and *competency*, both evaluated from the social network's point of view and with a semantic approach which will be detailed further in this paper.

Starting from Jacques Barzun's quote –"Simple English is no person's native tongue"– it is very difficult to write for a class of readers other than one's own, therefore readability plays an important role in understanding a chat. Although in a chat environment some words are omitted and syntax is usually simplified, readability still offers a good perspective of one's current level of knowledge/understanding or attitude in some cases, but all the information obtained from readability measures must be correlated with other factors.

Readability is commonly used unconsciously, based on the insight of other chat participants, but for its evaluation a readability formula is used, which is calibrated against a more labor-intensive readability survey and which matches the overall text with the expected reading level of the audience [4]. These formulas estimate the reading skill required to read the utterances in a chat and evaluate the overall complexity of the words used, therefore providing the means to target an audience.

Three formulas were computed. *The Flesch Reading Ease Readability Formula* (http://www.readabilityformulas.com/flesch-reading-ease-readability-formula.php) is one of the oldest and most accurate readability formulas, providing a simple approach to assess the grade-level of a chat participant and the difficulty of reading the current text. This score rates all utterances of a user on a 100 point scale. The higher the score, the easier it is to read, not necessarily understand the text. A score of 60 to 70 is considered to be optimal.

$$RE = 206.835 - (1.015 * ASL) - (84.6 * ASW) \quad (1)$$

RE is the **R**eadability **E**ase, **ASL** is the **A**verage **S**entence **L**ength (the number of words divided by the number of sentences) and **ASW** is the **A**verage number of **S**yllables per **W**ord (he number of syllables divided by the number of words).

The *Gunning's Fog Index* (or FOG) *Readability Formula* (http://www.readabilityformulas.com/gunning-fog-readability-formula.php) is based on Robert Gunning's opinion that newspapers and business documents were full of "fog" and unnecessary complexity. The index indicates the number of years of formal education a reader of average intelligence would need to understand the text on the first reading. A drawback of the Fog Index is that not all multi-syllabic words are difficult, but for computational issues, the consideration that all words above 2 syllables are complex is used.

$$FOG = (ASL + PHW) * 0.4 \quad (2)$$

ASL is the **A**verage **S**entence **L**ength (the number of words divided by the number of sentences) and **PHW** is the **P**ercentage of **H**ard **W**ords (in current implementation words with more than 2 syllables and not containing a dash)

The Flesch Grade Level Readability Formula (http://www.readabilityformulas.com/flesch-grade-level-readability-formula.php) rates utterances on U.S. grade school level. So a score of 8.0 means that the document can be understood by an eighth grader. This score makes it easier to judge the readability level of various texts in order to assign them to students. Also, a document

whose score is between 7.0 and 8.0 is considered to be optimal, since it will be highly readable.

$$FKRA = (0{,}39 * ASL) + (11{,}8 * ASW) - 15{,}59 \quad (3)$$

FKRA is the **F**lesch-**K**incaid **R**eading **A**ge, **ASL** is the **A**verage **S**entence **L**ength (the number of words divided by the number of sentences) and **ASW** is the **A**verage number of **S**yllable per **W**ord (the number of syllables divided by the number of words)

For each given chat, the system performs and evaluates all the 3 previous formulas and provides to the user detailed information for each participant. Also relative correlations between these factors and the manual annotation grades are computed in order to evaluate their relevance related to the overall grading process.

2.2 Social Networks Analysis

In addition to quantity and quality measures computed starting from the utterances, social factors are also taken into account in our approach. Consequently, a graph is generated from the chat transcript in concordance with the utterances exchanged by the participants. Nodes are participants in a collaborative environment and ties are generated based on explicit links (obtained from the explicit referencing facility of the chat environment used [6], which enables participants to manually add links during the conversation for marking subsequent utterances derived from a specific one).

From the point of view of social networks, *various metrics* are computed in order to determine the most competitive participant in chat: degree (indegree, outdegree), centrality (closeness centrality, graph centrality, eigen–values) and user ranking similar to the well known *Google Page Rank Algorithm* [7]. These metrics are applied first on the effective number of interchanged utterances between participants providing a quantitative approach; Second, the metrics are applied to the sum of utterance marks based on a semantic evaluation of each utterance; the evaluation process will be discussed in section 2.5 and, based on the results obtained for each utterance, a new graph is built on which all social metrics are applied. This provides the basis for a qualitative evaluation of the chat.

All the identified metrics used in the social network analysis are *relative* in the sense they provide markings relevant only compared with other participants in the same chat, not with those from other chats. This is the main reason why all factors are scaled between all the participants, giving each participant a weighted percentage from the overall performance of all participants.

2.3 LSA and the Corresponding Learning Process

Latent Semantic Analysis is a technique based on the *vector-space based model* [10, 14]. It is used for analyzing relationships between a set of documents and terms contained within by projecting them in sets of concepts related to those documents [9, 10]. LSA starts from a *term-document array* which describes the occurrence of each term in all the corpus documents. LSA transforms the occurrence matrix into a

relation between terms and concepts, and a relation between those concepts and the corresponding documents. Thus, the terms and the documents are now indirectly related through concepts [10, 13]. This transformation is obtained by a singular-value decomposition of the matrix and a reduction of its dimensionality.

Our system uses words from a chat corpus. The first step in the learning process, after spell–checking, is stop words elimination (very frequent and irrelevant words like "the", "a", "an", "to", etc.) from each utterance. The next step is POS tagging and, in case of verbs, these are stemmed in order to decrease the number of corresponding forms identified in chats by keeping track of only the verb's stem (the meaning of all forms is actually the same, but in LSA only one form is learnt). All other words are left in their identified forms, adding corresponding tagging because same words, but with different POS tags have other contextual senses, and therefore semantic neighbors [11].

Once the term-document matrix is populated, *Tf-Idf* (term frequency - inverse document frequency [13]) is computed. The final steps are the singular value decomposition (SVD) and the projection of the array in order to reduce its dimensions. According to [12], the optimal empiric value for *k* is 300, a value used in current experiments at which multiple sources concord.

Another important aspect in the LSA learning process is segmentation which is the process of dividing chats taking into consideration units with similar meaning and high internal cohesion. In the current implementation, the chat is divided between participants because of the considered unity and cohesion between utterances from the same participant. These documents are afterwards divided into segments using fixed non-overlapping windows. In this case contiguous segments are less effective because of intertwined themes present in chats and these aspects will be dealt with in future improvements of the marking system.

LSA is used for evaluating the proximity between two words by the *cosine measure*:

$$Sim(word_1, word_2) = \frac{\sum_{i=1}^{k} word_{1,i}, word_{2,i}}{\sqrt{\sum_{i=1}^{k} word_{1,i}^2} \times \sqrt{\sum_{i=1}^{k} word_{2,i}^2}} \cdot \quad (4)$$

Similarities between utterances and similarities of utterances related with the entire document are used in order to assess the importance of each utterance compared with the entire chat or with a predefined set of keywords referenced as a new document:

$$Vector(utterance) = \sum_{i=1} (1 + \log(no\_occurence(word_i)) * vector(word_i) \cdot \quad (5)$$

$$Sim(utterance_1, utterance_2) = Sim\big(Vector(utterance_1), Vector(utterance_2)\big). \quad (6)$$

2.4 The Utterance and Participants' Evaluation Process

2.4.1 The Utterance Marking Process

The first aspect that needs to be taken care of is building the graph of utterances which highlights the correlations between utterances on the basis of explicit references.

In order to evaluate each sentence, after finishing the morphological and lexical analysis three steps are processed:

1. Evaluate each utterance individually taking into consideration several features: the effective length of initial utterance; the number of occurrences of all keywords which remain after eliminating stop words, spell-checking and stemming; the level at which the current utterance is situated in the overall thread (similar to a Breadth-First search in the utterance space/threads based only on explicit links); the branching factor corresponding with the actual number of derived utterances from current one; the correlation / similarity with the overall chat; the correlation / similaritude with a set of predefined set of topics of discussion.

This mark combines the quantitative approach (the length of the sentence starting from the assumption that a piece of information should be more valuable if transmitted in multiple messages, linked together, and expressed in more words, not only to impress, but also meaningful in the context) with a qualitative one (the use of LSA and keywords).

In the process of evaluating each utterance, the semantic value is evaluated with the help of likelihood between the terms used in the current utterance (those after preliminary processing) and the whole document, respectively those from a list of predefined topics of discussion.

The formulas used for evaluating each utterance are:

$$mark_{empiric} = \left(\frac{length(initial\_utterance)}{10} + \frac{9}{10} \times \sum_{word}^{remaining} mark(word) \right) \times . \\ \times emphasis \tag{7}$$

$$mark(word) = length(word) * (1 + \log(no\_occurences)) \ . \tag{8}$$

$$emphasis = (1 + \log(level)) \times (1 + \log(branching\_factor) \times \\ \times Sim(utterance, whole\_document) \times \\ \times Sim(utterance, predefined\_keywords) \tag{9}$$

2. Emphasize Utterance Marks. Each thread obtained by chaining utterances based upon explicit links has a global maximum around which all utterance marks are increased correspondingly with a Gaussian distribution:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \text{ where:} \quad (10)$$

$$\sigma = \frac{\max(id\_utter\_thread) - \min(id\_utter\_thread)}{2}; \quad (11)$$

$$\mu = id\_utterance\_with\_highest\_mark\ . \quad (12)$$

Therefore each utterance mark is multiplied by a factor of 1 + *p*(*currrent_utterance*).

3. *Determine the final grade for each utterance in the current thread*

Based upon the empiric mark, the final mark of the utterance is obtained for each utterance in its corresponding thread:

$$mark_{final} = mark_{final}(prev\_utter) + coefficient \times mark_{empiric}\ , \quad (13)$$

where the coefficient is determined from the type of the current utterance and the one to which it is tied to.

For the coefficient determination, identification of speech acts plays an important role: verbs, punctuation signs and certain keywords are inspected. Starting from a set of predefined types of speech acts, the coefficients are obtained from a predefined matrix. These predefined values were determined after analyzing and estimating the impact of the current utterance considering only the previous one in the thread (similar to a Markov process). The grade of a discussion thread may be raised or lowered by each utterance. Therefore, depending on the type of an utterance and the identified speech acts, the final mark might have a positive or negative value.

2.4.2 Participant Grading

The in-degree, out-degree, closeness and graph centrality, eigen–values and rank factors are applied on the matrix with the number of interchanged utterances between participants and the matrix which takes into consideration the empiric mark of an utterance instead of the default value of 1. Therefore, in the second approach quality, not quantity is important (an element [*i*, *j*] equals the sum of $mark_{empiric}$ for each utterance from participant *i* to participant *j*), providing a deeper analysis of chats using a social network's approach based on a semantic utterance evaluation.

Each of the analysis factors (applied on both matrixes) is converted to a percentage (current grade/sum of all grades for each factor, except the case of eigen centrality where the conversion is made automatically by multiplying with 100 the corresponding eigen–value in absolute value). The final grade takes into consideration all these factors (including those from the surface analysis) and their corresponding weights:

$$final\_grade_i = \sum_k weight_k \times percentage_{k,i}\ , \quad (24)$$

where *k* is a factor used in the final evaluation of the participant *i* and the weight of each factor is read from a configuration file.

After all measures are computed and using the grades from human evaluators, the Pearson correlation for each factor is determined, providing the means to assess the importance and the relevance compared with the manual grades taken as reference.

General information about the chat – for example overall grade correlation, absolute and relative correctness – are also determined and displayed by the system.

2.5 Optimizing each Metric's Grade

The scope of the designed algorithm is to determine the optimal weights for each given factor in order to have the highest correlation with the manual annotator grades. A *series of constraints* had to be applied. First, *minimal/maximum values* for each weight are considered. For example, a minimum of 2% in order to take into consideration at least a small part of each factor, and maximum 40% in order to give all factors a chance and not simply obtain a solution with all factors 0% besides the one with the best overall correlation – 100%. Second, *the Sum of all factors* must be 100%**.** Third**,** obtain *maximum mean correlation* for all chats in the corpus.

In this case, the system has two components. A *perceptron* is used for obtaining fast solutions as inputs for the genetic algorithm. The main advantages for using this kind of network are the capacity to learn and adapt from examples, the fast convergence, the numerical stability; search in the weight space for optimal solution; duality and correlation between inputs and weights.

Secondly, a *genetic algorithm* is used for fine-tuning the solutions given by the neural network, also keeping in mind the predefined constraints. This algorithm operates over a population of chromosomes which represent potential solutions. Each generation represents and approximation of the solution - the determination of optimal weights in order to assure the best overall correlation, not the best distance between automatic grades and annotator ones. Correlation is expressed as an arithmetic mean of all correlations per chat because of the differences between evaluator styles.

The scope of this algorithm is to **maximize** the **overall correlation**, and specific characteristics of the implemented algorithm are:

- ***Initialization***: 2/3 of initial population obtained via Neural Networks (perceptron), the rest is randomly generated in order to avoid local;
- ***Fixed number of*** 100 chromosomes per population;
- ***Fitness*** - overall correlation of all chats from the corpus evaluated as a mean of all individual correlations;
- ***Selection*** – roulette based or elitist selection - the higher the fitness, the greater the possibility a participant is selected for crossover;
- ***Correction*** – a necessary operator in order to assure that the initial constraint are satisfied: if above or below minim/maximum values, reinitialize weight starting from threshold and adding a random quantity to it; if overall sum of percentages different from 100% adjust randomly weights with steps of 1/precision;
- ***Crossover*** - is based on *Real Intermediate Recombination* which has the highest dispersion of newly generated weights - select a random alpha for

each factor between [-0,25; 1,25]; the relative distance between 2 chromosomes selected for crossover must be at least 20% in order to apply the operator over them;

- Use ***CHC optimization***, with a little modification - generate N children and retain 20% of the best newly generated chromosomes; 20% of best parents are kept in the new generation and the rest is made of the best remaining individuals;
- ***Multiple populations*** that exchange best individuals - add after 10 generations the best individual to a common list and replace the worst individual with a randomly selected one from the list;
- After reaching ***convergence*** of a population (consecutively 20% of the maximum number of generations have the same best individual), reinitialize population = keep best 10% of existing individuals, obtain 30% via neural networks, and generate the remaining randomly;

The solution for determining the optimal weights combines the two approaches in order to obtain benefits from both – numerical stable solutions from neural networks and the flexibility of genetic algorithms in adjusting these partial solutions.

3 System Evaluation

The initial running configuration used by the system was: 10% for Page's Grading, 5% for social networks factors applied on the number of interchanged utterances, and 10% for the semantic social network factors applied on utterance marks. The overall results obtained with these weights are: *Relative correctness*: 77.44%, *Absolute correctness*: 70.07%, Correlation: 0.514.

Relative correctness and absolute correctness represent absolute/relative distances in a one-dimensional space, where the annotator's grade and the one obtained automatically using the Ch.A.M.P. system are taken into consideration for the given corpus. Eventually, the final results (as arithmetic means for each of the 3 individual measures determined per chat) are also displayed.

The results after multiple runs of the weight optimization system (all with 4 concurrent populations) show that most importance in the manual evaluation process is given to the following factors:

Table 2. Results after multiple runs of the weight optimization system, with regards to factors with a corresponding percentage $\geq 10\%$

Percentage	Factor
20-25%	Page's Grading methods - so only surface analysis factors
10-15%	Indegree from the social network's point of view, applied on number of interchanged utterances
30-40%	Outdegree also determined by the number of outgoing utterances – somehow a participant's gregariousness measure
$\approx$ 10%	Semantic graph centrality – the only measure with a higher importance applied which relies on utterance marks

All remaining factor are evaluated below 5%, therefore don't have high importance in the final grading process. The overall results, with regards to correlation optimization, obtained after running the genetic algorithm are: *Relative correctness*: ≈ 46.83%, *Absolute correctness*: ≈ 45.70%, *Correlation*: ≈ 0.594.

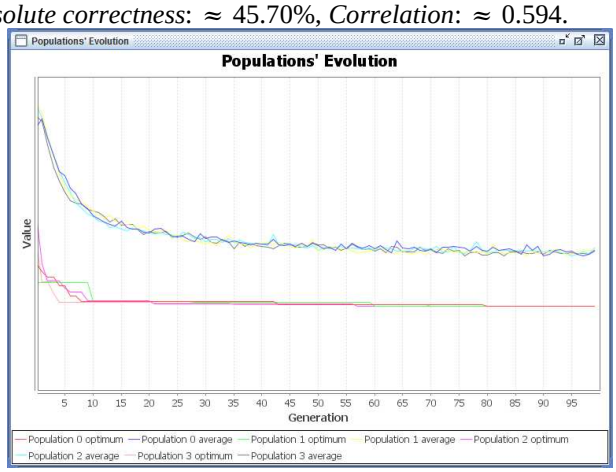

Fig. 1. Convergence to an optimal solution using 4 populations with the visualization of optimum/average chromosomes

The spikes from each population's average fitness are determined by newly inserted individuals or by the population reinitialization. After the first 10 iterations important improvements can be observed, whereas after 30 generations the optimum chromosomes of each population stagnate. Only population reinitializations and chromosome interchanges provide minor improvements in the current solution.

Our results entail several conclusions: The human grading process uses a predominantly quantitative approach; Uncorrelated evaluations and different styles/principles used by different human annotators are the main causes for lowering the overall correlation and correctness; The improvement of correlation was in the detriment of absolute/relative correctness; Convergence of the genetic algorithm can be considered after 30 generations.

4 Conclusions

The results obtained from our system allow us to conclude that the evaluation of a participant's overall contribution in a chat environment can be achieved. Also we strongly believe that with further tuning of the weights, better LSA learning and increased number of social network factors (including those applied to the entire network) will increase performance and reliability of the results obtained. Moreover, the subjective factor in manual evaluation is also present and influences the overall correctness.

In present, evaluations and tuning of the assessment system are performed in the LTfLL project, in which the work presented in the paper is one of the modules for feedback generation [16].

Acknowledgements

The research presented in this paper was partially performed under the FP7 EU STREP project LTfLL. We would like to thank all the students of University "Politehnica" of Bucharest, Computer Science Department, who participated in our experiments and provided the inputs for generating our golden standard.

References

1. Stahl, G.: Group Cognition: Computer Support for Building Collaborative Knowledge. MIT Press (2006)
2. Anderson, J. R.: Cognitive psychology and its implications, New York, Freeman (1985)
3. Page, E. B. Paulus, D. H.: Analysis of essays by computer. Predicting Overall Quality, U.S. Department of Health, Education and Welfare (1968)
4. http://www.streetdirectory.com/travel_guide/15672/writing/all_about_readability_formulas__and_why_writers_need_to_use_them.html
5. Wresch, W.: The Imminence of Grading Essays by Computer--25 Years Later. Computers and Composition 10(2), 45-58, retrieved from http:// computersandcomposition.osu.edu/archives/v10/10_2_html/10_2_5_Wresch.html (1993)
6. Holmer, T., Kienle, A. & Wessner, M.: Explicit Referencing in Learning Chats: Needs and Acceptance. In: Nejdl, W., Tochtermann, K., (eds.): Innovative Approaches for Learning and Knowledge Sharing- ECTEL, LNCS, 4227, Springer, pp. 170–184 (2006)
7. Dascălu, M., Chioaşcă, E.-V. Trăuşan-Matu, S.: ASAP – An Advanced System for assessing chat participants. In: D. Dochev, M. Pistore, and P. Traverso (Eds.): AIMSA 2008, LNAI 5253, Springer, pp. 58–68 (2008)
8. Bakhtin, M. M.: Problems of Dostoevsky's poetics (Edited and translated by Caryl Emerson). Minneapolis: University of Minnesota Press (1993)
9. http://lsa.colorado.edu/
10. Landauer, K. Th., Foltz, W. P., Laham, D.: An Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284 (1998)
11. Wiemer-Hastings, P., Zipitria, I.: Rules for syntax, vectors for semantics. In: proceeding of the 23rd Annual Conference of the Cognitive Science Society (2001).
12. Lemaire, B.: Limites de la lemmatisation pour l'extraction de significations. JADT 2008: 9es Journées internationales d'Analyse statistique des Données Textuelles (2008).
13. Manning, C., Schütze, H.: Foundations of statistical Natural Language Processing. MIT Press, Cambridge (Mass.) (1999)
14. Miller, T.: Latent semantic analysis and the construction of coherent extracts. In: Nicolov, N. Botcheva, K., Angelova, G. and Mitkov, R., (eds.), Recent Advances in Natural Language Processing III. John Benjamins, Amsterdam/Philadelphia, pp. 277–286 (2004)
15. Fernandez, S., Velazquez, P., Mandin, S.: Les systèmes de résumé automatique sont-ils vraiment des mauvais élèves?. In: JADT 2008: 9es Journées internationales d'Analyse Statistique des Données Textuelles (2008)
16. Stefan Trausan-Matu, Traian Rebedea, A Polyphonic Model and System for Inter-animation Analysis in Chat Conversations with Multiple Participants, in A. Gelbukh (Ed.): CICLing 2010, LNCS 6008, Springer, 2010, pp. 354–363