Turkish Online Journal of Distance Education-TOJDE October 2014 ISSN 1302-6488 Volume: 15 Number: 4 Article 19

EDUCATIONAL INFORMATION QUANTIZATION FOR IMPROVING CONTENT QUALITY IN LEARNING MANAGEMENT SYSTEMS

Alexander Aleksandrovich RYBANOV,
PhD in Technical Sciences, Associate Professor,
Informatics and programming techniques Department,
Volzhskii Polytechnic Institute,
Branch of the Volgograd State Technical University,
Volzhskii, RUSSIA

ABSTRACT

The article offers the educational information quantization method for improving content quality in Learning Management Systems. The paper considers questions concerning analysis of quality of quantized presentation of educational information, based on quantitative text parameters: average frequencies of parts of speech, used in the text; formal text readability indexes; lexical and syntactic text variety factors. The process of obtaining quantitative parameter values is focused on use of the phpMorphy morphological analysis library.

Keywords: **E-learning, distance learning, web-based e-learning system, educational text quantization, educational content, formal readability index, quantitative characteristics.**

INTRODUCTION

Development of educational content preparation tools lags behind Learning Management Systems (LMS). Success of LMS, in its turn, depends on quality and effective organization of educational content. The current LMSs, such as Moodle, Ilias, Claroline, Atutor, etc., do not allow developers of e-learning courses to assess educational content quality. At the same time, educational content assessment is aimed at determining advantages and disadvantages of educational information and at making the decision on possibility and optimum conditions of its use in e-learning. One of the directions in solving the problem of assessing educational content quality in LMSs is quantitative linguistics methods.

PROBLEM STATEMENT

Quality and effective organization of educational content influence directly the following LMS parameters (A. A.Rybanov, 2011):

- ***Educational content mastering factor*** **(K) is a ratio of the educational content mastered by LMS users during a certain time unit to the content provided to the users during this time unit:**

$$K = \frac{I_\theta}{I_\alpha}.$$

Here I_θ is the mastered content; I_α is the provided content. If the same content has been mastered by the users during various times, the factor K should be divided by the time t. To measure I_α and I_θ, comparative analysis of the user thesaurus and the educational content thesaurus can be used (A.A.Rybanov, 2013).

- ***Educational content mastering speed*** **or a ratio of the mastering factor to mastering time:**

$$K_i = \frac{t_i}{t_{avg}}.$$

Here K_i is the relative learning time factor; t_i is the time spent by i-th LMS user for mastering a certain educational content; t_{avg} is the average time spent for mastering a certain educational content by a group of LMS users.

- ***Educational content mastering retention*** **shows the level of LMS user's knowledge and skills after some time after e-learning course completion:**

$$\alpha_m = \frac{I_m}{I_\alpha}.$$

Here I_α is the provided content; I_m is the educational content retained and effectively used by the user after some time t.

LMS educational content development includes development of content preparation technologies, such as educational information quantization (V.S.Avanesov, 2012). K, K_i, and α_m factors depend on, inter alia, educational information quantization quality. An important problem is forming a quantitative criteria system for assessing educational information quantization quality.

MATHEMATICAL FORMULATION

Concept of Educational Information Quantization Process

Quantization **is dividing of educational information into different purpose (information, training, controlling, and managing) elementary fragments (educational units, steps, frames) that facilitates mastering the sense contained in each educational information fragment. Volume of the text information contained in these fragments must be limited.**

Quantization process is a transformation:

$$T' = f(T).$$

Here $T = (T_i \mid i = \overline{1,n})$ is the educational information intended for quantization, T_i is logically complete fragment of the educational information T; $T' = (T_i' \mid i = \overline{1,n})$ is quantized presentation of the educational information, T_i' is an educational information quantum associated with the fragment T_i.

System educational information quantization principle **assumes taking into account the following regularities:**

- **Large volume educational information is remembered hardly;**
- **Educational information, which is presented compactly and according to a certain system, is perceived better;**
- **Emphasizing sense units in the educational information promotes effective memorizing.**

Taking into account that the educational information quantum T_i' must contain the most informative part of the fragment T_i, requirements to the educational information quantum can be formulated as follows:

- **Educational information quantum T_i' must have a lower redundancy and a higher entropy than T_i;**
- **Educational information quantum T_i' must be smaller by volume than the corresponding educational information fragment T_i: $|T_i'| \le |T_i|$.**

The process of constructing quantum **T_i' for the educational information fragment T_i by the teacher consists of the following stages:**

- **Preparation stage (reading and comprehension of the educational information fragment T_i);**
- **Analytical stage (highlighting of the main semantic units (sentences, words, phrases), construction of the quantum T_i' structure for the educational information fragment T_i);**
- **The stage of constructing quantum T_i' for the educational information fragment T_i (the units highlighted earlier are placed in the common secondary text according to the quantum T_i' structure).**

Semantic units of the quantum T_i' for the educational information fragment T_i can be:

- **γ_1 : the full (without changes) key sentence of the initial text T_i ;**
- **γ_2 : the paraphrased key sentence of the initial text T_i ;**
- **γ_3 : the sentence constructed of the key words and phrases of the initial text T_i ;**
- **γ_4 : the sentence generalizing several sentences of the initial text T_i .**

Quality quantized educational texts ensure understanding of the educational material by most students, because division of the material into parts reduce noticeably the volume of directly perceived information and the volume of senses in each fragment, thereby improving understandability of senses of the entire educational text.

Besides, work with test items for such texts ensures mastering of each text's content.

Quantitative Characteristics of Educational Information

Quantitative linguistics **is one of the applied linguistics' areas in which language is studied by means of statistical methods (Keith Johnson, 2008).**

Advantage of quantitative text studying methods is their accuracy and unambiguity of the results. Calculation of quantitative text characteristics is necessary for solving the following tasks:

- **Determining style and genre characteristics of the texts with the purpose of their subsequent classification (J.Tuldava, 2004);**
- **Examination of text samples with the purpose of establishing authorship (J.Grieve, 2007);**
- **Speciality language teaching (V.V.Ageev, V.M.Sergevnina, E.I.Yakovleva, 2011).**

One of the content preparation technology problems is forming the system of quantitative criteria for assessing educational information quantization quality. Quantitative text characteristics can form a basis of this criteria system.

O. A. Wiio suggested using quantitative characteristics for assessing the complexity factor (O. A. Wiio, 1968), the more adjectives and adverbs in the text, the higher is the text complexity. Verb is the liveliest part of speech. Frequent using of verbs in conjugation forms results in easier remembering and understanding of the sentences. In such sentences, related words are close to each other and their relations are perceived easily. Verbs promote text understanding (R.Flesh, 1946).

The problem of automated determining of quantitative text characteristic values is important.

Software realization of the automated determining of some quantitative text characteristics is possible on the basis of the PHP based phpMorphy morphological analysis library (http://phpmorphy.sourceforge.net).

The phpMorphy library supports processing of texts in Russian, English, and German. The library is aimed at solving the following tasks:

- **Lemmatization (obtaining normal word form);**
- **Obtaining all word forms;**
- **Obtaining semigrammatical information on the word (part of speech, case, conjugation, etc.);**
- **Changing the word form according to the set grammatical characteristics;**
- **Changing the word form according to the set pattern.**

Among great number of quantitative text characteristics, let us consider the following ones:

- **Quantitative characteristics of used parts of speech;**
- **Quantitative text readability characteristics;**
- **Quantitative text variety characteristics.**

By means of the phpMorphy library, the following low-level quantitative text characteristics calculated on the basis of average frequencies of parts of speech used in the text can be determined:

- ***Analyticity index* is a ratio of the function word quantity to the total word quantity in the text;**
- ***Verb index* is a ratio of verb quantity to the total word quantity in the text;**
- ***Substantive index* is a ratio of noun quantity to the total word quantity in the text;**
- ***Adjective index* is a ratio of adjective quantity to the total word quantity in the text;**
- ***Pronoun index* is a ratio of pronoun quantity to the total word quantity in the text;**
- ***Autosemanticity index* is a ratio of meaningful word quantity to the total word quantity in the text;**
- ***Unmomentous word index* is a ratio of unmomentous word quantity to the total word quantity in the text;**
- ***Nominal lexicon index* is a ratio of the total noun and adjective quantity to the total word quantity in the text.**

Part of speech designations in the phpMorphy library are presented in Table: 1.

Table: 1
Part of speech designations in the phpMorphy library

Constant	Description
PMY_RP_NOUN	Noun
PMY_RP_ADJ_FULL	Adjective
PMY_RP_ADJ_SHORT	Short adjective
PMY_RP_INFINITIVE	Infinitive
PMY_RP_VERB	Verb in the personal form
PMY_RP_ADVERB_PARTICIPLE	Adverbial participle
PMY_RP_PARTICIPLE	Participle
PMY_RP_PARTICIPLE_SHORT	Short participle
PMY_RP_NUMERAL	Numeral
PMY_RP_NUMERAL_P	Ordinal numeral
PMY_RP_PRONOUN	Pronoun-noun
PMY_RP_PRONOUN_PREDK	Pronoun-predicative
PMY_RP_PRONOUN_P	Pronominal adjective
PMY_RP_ADV	Adverb
PMY_RP_PREDK	Predicative
PMY_RP_PREP	Preposition
PMY_RP_CONJ	Conjunction
PMY_RP_INTERJ	Interjection
PMY_RP_PARTICLE	Particle
PMY_RP_INP	Parenthesis
PMY_RP_PHRASE	Phraseological unit

Low-level quantitative text characteristics can be expressed through the part of speech designations in the phpMorphy library as follows (COUNT_WORDS is the total word quantity in the text):

- **Analyticity index :**
  ***Analyticity_index* = (PMY_RP_PREP + PMY_RP_CONJ +**
  + PMY_RP_PARTICLE) / COUNT_WORDS.
- **Verb index :**
  ***Verb_index* = (PMY_RP_INFINITIVE + PMY_RP_VERB +**
  + PMY_RP_ADVERB_PARTICIPLE + PMY_RP_PARTICIPLE +
  + PMY_RP_PARTICIPLE_SHORT) / COUNT_WORDS.
- **Substantive index:**
  ***Substantive_index* = PMY_RP_NOUN / COUNT_WORDS.**
- **Adjective index :**
  ***Adjective_index* = (PMY_RP_ADJ_FULL +**
  + PMY_RP_ADJ_SHORT) / COUNT_WORDS.
- **Pronoun index :**
  ***Pronoun_index* = (PMY_RP_PRONOUN + PMY_RP_PRONOUN_PREDK +**
  + PMY_RP_PRONOUN_P) / COUNT_WORDS.

- **Autosemanticity index :**
  ***Autosemanticity_index* = 1 - *Unmomentous_words_index*.**
- **Unmomentous word index:**
  ***Unmomentous_word_index* = ((PMY_RP_PREP + PMY_RP_CONJ +**
  + PMY_RP_PARTICLE) + (PMY_RP_PRONOUN +
  PMY_RP_PRONOUN_PREDK + PMY_RP_PRONOUN_P)) / COUNT_WORDS.
- ***Nominal lexicon index :***
  ***Nominal_lexicon_index* = (PMY_RP_NOUN + PMY_RP_ADJ_FULL +**
  + PMY_RP_ADJ_SHORT) / COUNT_WORDS.

Among quantitative text readability characteristics, the following characteristics can be highlighted: average word length in syllables and average sentence length in words. These characteristics are statistical parameters, which are used in the formulas for assessing readability and are necessary for calculating the formal readability index. These parameters can be easily expressed quantitatively and are suitable for automatic assessment.

Quantitative text variety characteristics are described by the lexical and syntactic variety factors. Since factor in not an absolute, but a relative value (within a certain value range), compared texts' lengths can be neglected within certain limits. Researching of the internal educational text "dynamics" in relation to comparing the factors in different parts of the text and their ratios to the general factor for the entire text is of theoretical interest as well.

The *lexical variety factor* is a ratio of lexeme quantity to the total word quantity in the text:

$$K_{lex} = \frac{L}{W}, \quad \textbf{(1)}$$

Here K_{lex} is the lexical variety factor; L is lexeme (word form) quantity in the text; W is the total word (the units between blanks) quantity in the text. The higher the K_{lex} value, the higher is the lexical variety of the text.

The *syntactic variety factor* is a ratio of the total sentence quantity to the total word quantity in the text:

$$K_{syn} = 1 - \frac{S}{W}, \quad \textbf{(2)}$$

Here K_{syn} is syntactic variety factor; S is sentence quantity; W is word quantity in the text.

The higher the K_{syn} value, the wordier are the sentences in the text in general, and, therefore, the higher the possibility of the variety of syntactic relations between words in a separate sentence.

Measuring Quantitative Characteristics of the Educational Information

When processing a text automatically, there can be a situation when the part of speech determining function returns several values for one word form. For example, for the word 'PROGRAM', the *getPartOfSpeech* function in the phpMorphy library returns the following array with part of speech values:

```
var_dump ($morphy-> getPartOfSpeech (' PROGRAMM'));
// array ('NOUN', 'ADJECTIVE', 'VERB')
```

Therefore the value of each quantitative text characteristic must be described by its calculation error value.

Let us set the following designations for the process of automatic calculation of word quantity in the text T, relating to the part of speech k:

- η_k **is quantity of single-value determinations of the part of speech** k;
- μ_k **is quantity of multiple-value determinations of the part of speech** k.
- θ_k **is word quantity in the part of speech** k **in the text** T.

Part of speech probability distribution in the text T is unknown. Therefore, *according to the Laplace's principle of insufficient reason,* in automatic recognizing of the parts of speech, there are no reasons to consider them to be different.
According to the principle of insufficient reason, let us assume that

$$\eta_k + \Delta_k \le \theta_k \le \eta_k + \mu_k - \Delta_k .$$

From there, let us assume that

$$\theta_k = \eta_k + \mu_k / 2 .$$

Then the absolute error Δ_k in automatic determining of the part of speech k:

$$\Delta_k = \mu_k / 2 .$$

And the relative error δ_k in automatic determining of the part of speech k:

$$\delta_k = \frac{\Delta_k}{\theta_k} \cdot 100\ \% = \frac{\mu_k}{2 \cdot \eta_k + \mu_k} \cdot 100\ \% .$$

On the basis of the values Δ_k and δ_k, let us calculate the errors for automatic determining the value of the quantitative characteristic β for the text T:

- **Absolute error** Δ_β**:** $\Delta_\beta = \frac{1}{2 \cdot W} \sum_{i \in P} \mu_i$,
- **Relative error** δ_β**:** $\delta_\beta = \frac{\sum_{i \in P} \mu_i}{2 \cdot \sum_{i \in P} \eta_i + \sum_{i \in P} \mu_i}$.

Here P is a set of parts of speech, used in calculating the quantitative characteristic β. For example, adjective index errors are calculated as follows:

$$\delta_{\text{Adjective\_index}} = \frac{\mu_{\text{PMY\_RP\_ADJ\_FULL}} + \mu_{\text{PMY\_RP\_ADJ\_SHORT}}}{2 \cdot W},$$

$$\delta_{\text{Adjective\_index}} = \frac{\mu_{\text{PMY\_RP\_ADJ\_FULL}} + \mu_{\text{PMY\_RP\_ADJ\_SHORT}}}{2 \cdot (\eta_{\text{PMY\_RP\_ADJ\_FULL}} + \eta_{\text{PMY\_RP\_ADJ\_SHORT}}) + \mu_{\text{PMY\_RP\_ADJ\_FULL}} + \mu_{\text{PMY\_RP\_ADJ\_SHORT}}}.$$

Formal Readability Index for Educational Information

The works by G.Hargis (2000), W.H.DuBay (2004), R.H.Hall and Hanna P. (2004) define the following element groups influencing readability: content, style, format, features of organization. It is necessary to distinguish between formal text readability (S. Cepni, M. Gokdere, M. Kucuk, 2002) $R_{\text{form}}(I)$, which is a function of parameters of the educational content I itself only, and individual text readability $R_{\text{ind}}(I,u)$, which depends both on characteristics of the educational content I and on properties of the reader u.

For quantitative formal readability assessment, it is possible to use the indexes offered in the works by J. Tuldava (1975) and R. Flesh (1974). J. Tuldava 's index is calculated according to the formula:

$$R(\bar{i}, \bar{j}) = \bar{i} \cdot \lg \bar{j}, \quad \textbf{(3)}$$

Here $R(\bar{i}, \bar{j})$ is formal readability index (Figure: 1), $\bar{i}$ is average word length in syllables, $\bar{j}$ is average sentence length in words. The formula (3) is developed on the basis of the regularity observed in various languages. Therefore J. Tuldava 's formula is intended for analyzing texts in different languages. The lower the value $R(\bar{i}, \bar{j})$, the better is the text perception.

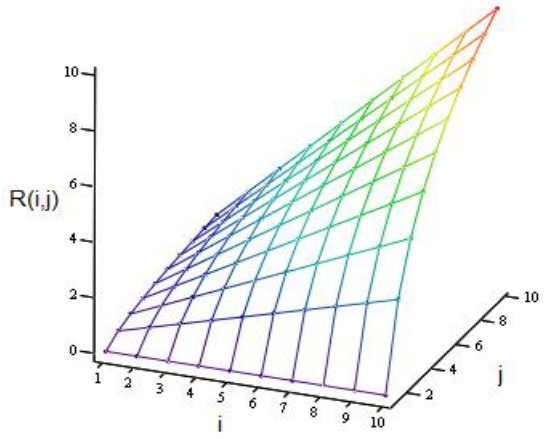

Figure: 1
The kind of the function $R(\bar{i}, \bar{j})$

R.Flesh's index is calculated according to the formula:

$$Fr(\bar{i},\bar{j}) = 206.835 - a_1 \cdot \bar{j} - a_2 \cdot \bar{i}, \quad (4)$$

Here a_1, a_2 are the language dependent factors (for English, $a_1 = 1.015$, $a_2 = 84.6$; for Russian, $a_1 = 1.3$, $a_2 = 60.1$). Correspondence between R.Flesh's index values and the linguistic variables "Readability level" and "Educational level" is shown in Table: 2.

Table: 2
Linguistic variables "Readability level" and "Educational level"
for the R.Flesh's index $Fr(\bar{i},\bar{j})$

R.Flesh's index $Fr(\bar{i},\bar{j})$	Readability level	Educational level
90–100	Very high	5th grade
80–90	High	6th grade
70–80	Above the average	7th grade
60–70	Average	8th - 9th grades
50–60	Below the average	10th - 12th grades
30–50	Low	College
0–30	Very low	Graduate

RESULTS AND DISCUSSION

A.P.Chekhov's story "The White Forehead Puppy" has been used as an experimental material. Analysis of the educational information quantization quality has been carried out on the basis of the two story's presentations: T - the initial (original) text, and T' - quantized text.

The initial text T has been divided into seven logically complete fragments $T = (T_i \mid i = \overline{1,7})$, each of which was quantized.

The obtained quantized text is also a set of seven logically complete fragments $T' = (T_i' \mid i = \overline{1,7})$; here T_i' is a quantized text fragment obtained as a result of fragment T_i quantization.

Results of automatic part of speech recognition carried out with the use of the phpMorphy library in the initial T and quantized T' texts are presented in Table: 3 and Table: 4.

Table: 3
Results of part of speech recognition in the initial text T

Part of speech k	η_k	μ_k	θ_k	Δ_k	δ_k
PMY_RP_NOUN	429	138	498.0	69.0	13.855
PMY_RP_ADJ_FULL	102	40	122.0	20.0	16.393
PMY_RP_ADJ_SHORT	5	66	38.0	33.0	86.842
PMY_RP_INFINITIVE	36	5	38.5	2.5.0	6.494
PMY_RP_VERB	285	59	314.5	29.5	9.380
PMY_RP_ADVERB_PARTICIPLE	37	4	39.0	2.0	5.128
PMY_RP_PARTICIPLE	15	4	17.0	2.0	11.765
PMY_RP_PARTICIPLE_SHORT	3	2	4.0	1.0	25
PMY_RP_NUMERAL	9	9	13.5	4.5	33.333
PMY_RP_NUMERAL_P	0	3	1.5	1.5	100
PMY_RP_PRONOUN	97	98	146.0	49.0	33.562
PMY_RP_PRONOUN_PREDK	0	0	0.0	-	-
PMY_RP_PRONOUN_P	28	69	62.5	34.5	55.200
PMY_RP_ADV	44	219	153.5	109.5	71.336
PMY_RP_PREDK	0	32	16.0	16.0	100
PMY_RP_PREP	203	35	220.5	17.5	7.937
PMY_RP_CONJ	1	254	128.0	127.0	99.219
PMY_RP_INTERJ	0	170	85.0	85.0	100
PMY_RP_PARTICLE	28	120	88.0	60.0	68.182
PMY_RP_INP	0	4	2.0	2.0	100
PMY_RP_PHRASE	0	4	2.0	2.0	100

Table: 4
Results of part of speech recognition in the quantized text T'

Part of speech k	η_k	μ_k	θ_k	Δ_k	δ_k
PMY_RP_NOUN	216	61	246.5	30.5	12.373
PMY_RP_ADJ_FULL	46	16	54.0	8.0	14.815
PMY_RP_ADJ_SHORT	3	33	19.5	16.5	84.615
PMY_RP_INFINITIVE	20	2	21.0	1.0	4.762
PMY_RP_VERB	140	32	156	16.0	10.256
PMY_RP_ADVERB_PARTICIPLE	14	1	14.5	0.5	3.448
PMY_RP_PARTICIPLE	8	2	9.0	1.0	11.111
PMY_RP_PARTICIPLE_SHORT	0	0	0.0	-	-
PMY_RP_NUMERAL	5	4	7.0	2.0	28.571
PMY_RP_NUMERAL_P	0	1	0.5	0.5	100
PMY_RP_PRONOUN	43	54	70.0	27.0	38.571
PMY_RP_PRONOUN_PREDK	0	0	0.0	-	-
PMY_RP_PRONOUN_P	16	37	34.5	18.5	53.623
PMY_RP_ADV	29	102	80.0	51.0	63.750
PMY_RP_PREDK	0	14	7.0	7.0	100
PMY_RP_PREP	103	22	114.0	11.0	9.649
PMY_RP_CONJ	0	126	63.0	63.0	100
PMY_RP_INTERJ	0	83	41.5	41.5	100
PMY_RP_PARTICLE	15	64	47.0	32.0	68.085
PMY_RP_INP	0	2	1.0	1.0	100
PMY_RP_PHRASE	0	1	0.5	0.5	100

Percentagewise, discrepancy in word distributions among parts of speech for the initial (Table: 3) and the quantized (Table: 4) texts are insignificant (Figure: 2).

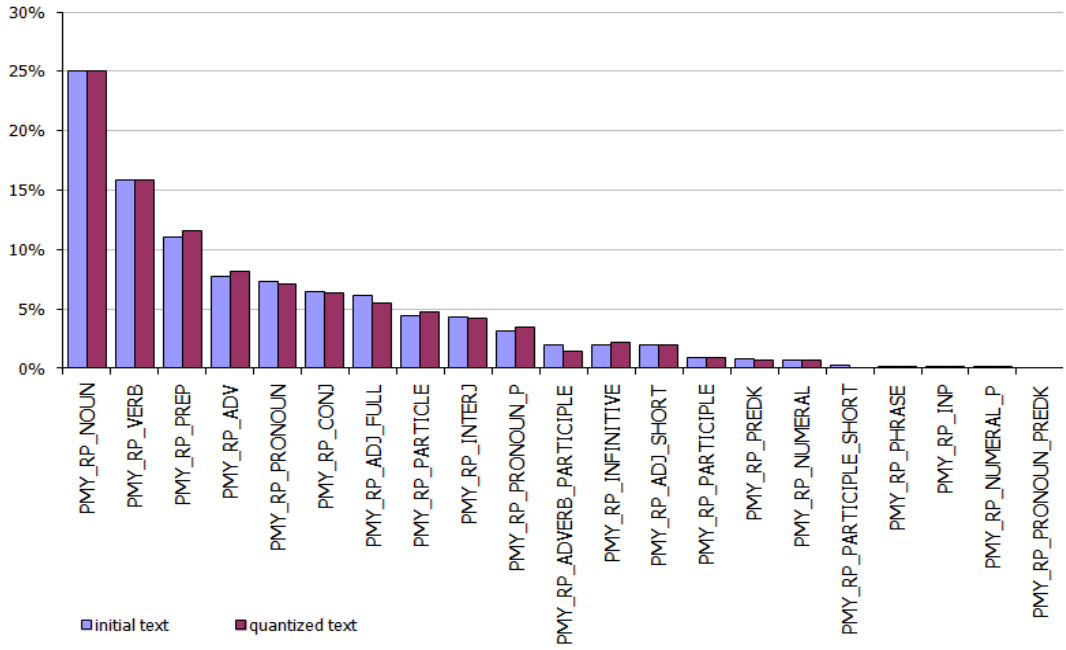

Figure: 2
Comparative analysis of relative word distributions among parts of speech for the initial and the quantized texts

Values and errors of calculating quantitative part of speech characteristics for the initial and the quantized texts are presented in Table: 5.

Table: 5
Values and errors of calculating quantitative part of speech characteristics

Quantitative Characteristic β	Initial text T			Quantized text T'		
	Value	Δ_β	δ_β	Value	Δ_β	δ_β
Analyticity index	.229	.107	46.849	.236	.112	47.321
Verb index	.216	.019	8.959	.211	.019	9.227
Substantive index	.261	.036	13.855	.259	.032	12.373
Adjective index	.084	.028	33.125	.077	.026	33.333
Pronoun index	.109	.044	40.048	.110	.048	43.541
Autosemanticity index	.662	.151	22.809	.654	.160	24.312
Unmomentous word index	.338	.151	44.651	.346	.159	46.119
Nominal lexicon index	.345	.064	18.541	.377	.058	17.188

Discrepancy between corresponding quantitative characteristics for the initial and the quantized texts are insignificant. Errors Δ_k, δ_k, Δ_β, δ_β can be used for comparative analysis of automatic text processing software programs regarding their accuracy in determining parts of speech and quantitative characteristics. Quantization results in compression of the initial text sentences by means of the following methods: exception (γ_3), replacement (γ_2), and merging (γ_4). Thus quantitative readability characteristics of the initial and the quantized texts as well as of their fragments, presented in Table: 6 and Table: 7, testify reduction of the average sentence length in words in the quantized text. Exception is only the quantized text fragments 2 and 3.

Table: 6
Quantitative readability characteristics of the initial and the quantized texts

Quantitative characteristic	Initial text T	Quantized text T'
Average word length in syllables	2.052	2.023
Average sentence length in words	14.264	12.614

Table: 7
Quantitative readability characteristics of the initial and the quantized text fragments

Text fragment	Average word length in syllables		Average sentence length in words	
	Initial text T	Quantized text T'	Initial text T	Quantized text T'
No. 1	2.095	2.153	23.695	11.800
No. 2	1.990	1.924	15.923	16.957
No. 3	2.137	2.111	13.9	31.500
No. 4	2.047	1.925	21.25	16.000
No. 5	2.147	1.963	17.875	12.000
No. 6	2.056	1.850	15.765	15.000
No. 7	2.025	2.074	9.429	9.240

Quantitative variety characteristics of the initial and the quantized texts are presented in Table: 8 and Table: 9. Changes of the factors K_{lex} and K_{syn} as a result of quantization procedure are also connected with using the exception (γ_3), replacement (γ_2), and merging (γ_4) methods.

Table: 8
Quantitative variety characteristics of the initial and the quantized texts

Quantitative characteristic	Initial text T	Quantized text T'
K_{lex}	.306	.355
K_{syn}	.944	.940

Table: 9
Quantitative variety characteristics of the initial and the quantized text fragments

Text fragment	K_{lex}		K_{syn}	
	Initial text T	Quantized text T'	Initial text T	Quantized text T'
No. 1	.688	.831	.958	.915
No. 2	.628	.703	.937	.941
No. 3	.712	.730	.928	.968
No. 4	.612	.738	.953	.938
No. 5	.706	.731	.944	.917
No. 6	.590	.700	.937	.933
No. 7	.583	.636	.894	.892

Lexical variety characterizes information saturation of the text. Reduction of the wordform repetition degree is characteristic of the quantized text, in comparison with the initial text. Therefore the lexical variety factor for the quantized text is a little higher than for the initial text (Figure: 3).

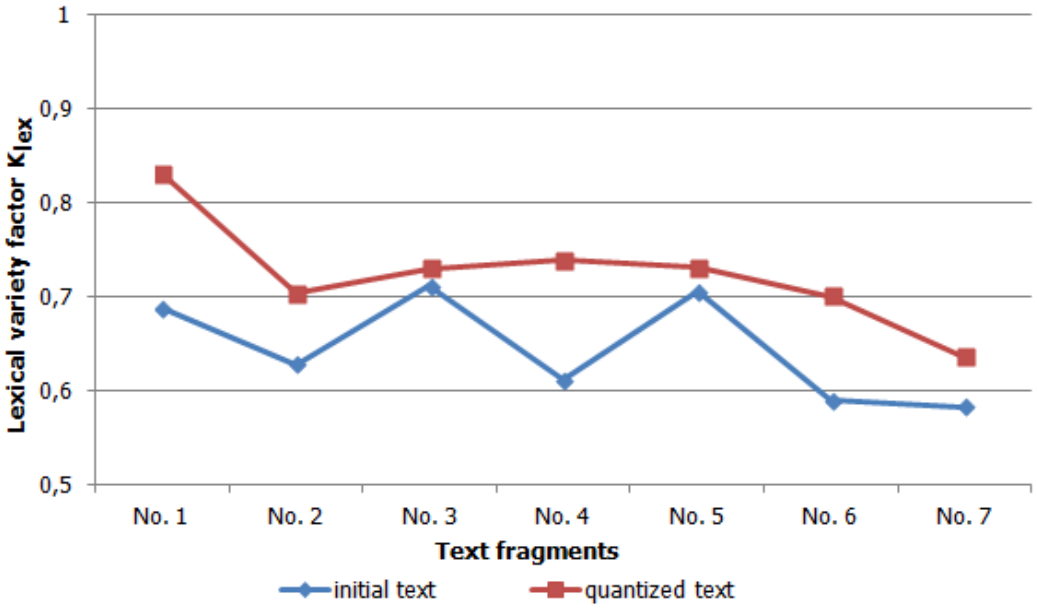

Figure: 3
Comparative analysis of the lexical variety factor for text fragments

Syntactic variety shows itself in using various syntactic means: quantization reduces the syntactic variety factor. In Figure: 4, syntactic variety factor for the quantized text fragments 2 and 3 is higher than for the initial text that indicates necessity of requantization of these fragments.

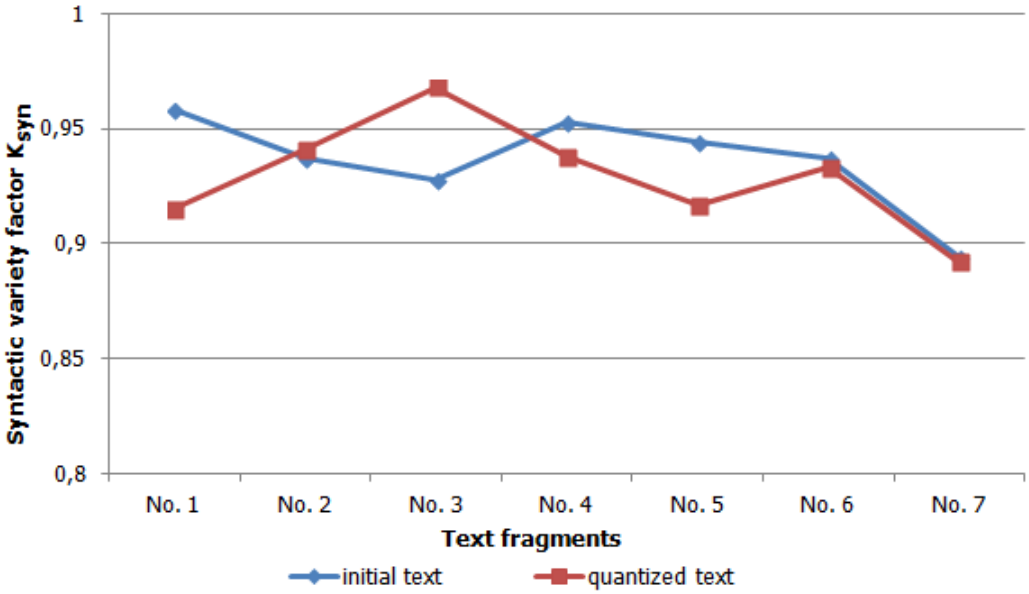

Figure: 4
Comparative analysis of the syntactic variety factor for text fragments

Let us analyze changes in formal readability of the quantized text in comparison with the initial text. Table: 10 shows the formal readability indexes $R(\bar{i},\bar{j})$ and $Fr(\bar{i},\bar{j})$ for corresponding fragments of the initial and the quantized texts.

Table: 10
Indexes $R(\bar{i},\bar{j})$ and $Fr(\bar{i},\bar{j})$ for the initial and the quantized text fragments

Text fragment	$R(\bar{i},\bar{j})$		$Fr(\bar{i},\bar{j})$	
	Initial text T	Quantized text T'	Initial text T	Quantized text T'
No. 1	2.878	2.307	50.199	62.127
No. 2	2.392	2.360	66.516	69.305
No. 3	2.442	3.163	60.350	39.007
No. 4	2.717	2.318	56.182	70.342
No. 5	2.688	2.118	54.572	73.261
No. 6	2.462	2.176	62.777	76.150
No. 7	1.973	2.002	72.860	70.200

The formal readability index $R(\bar{i},\bar{j})$ for the quantized text is equal to 2.227, and for the initial text it is equal to 2.368 that testifies better presentation of the quantized text.

At the same time, comparative analysis of the indexes $R(\bar{i},\bar{j})$ for the initial and the quantized text fragments (Figure: 5) indicates that the quantized text fragments 3 and 7 require further improvement.

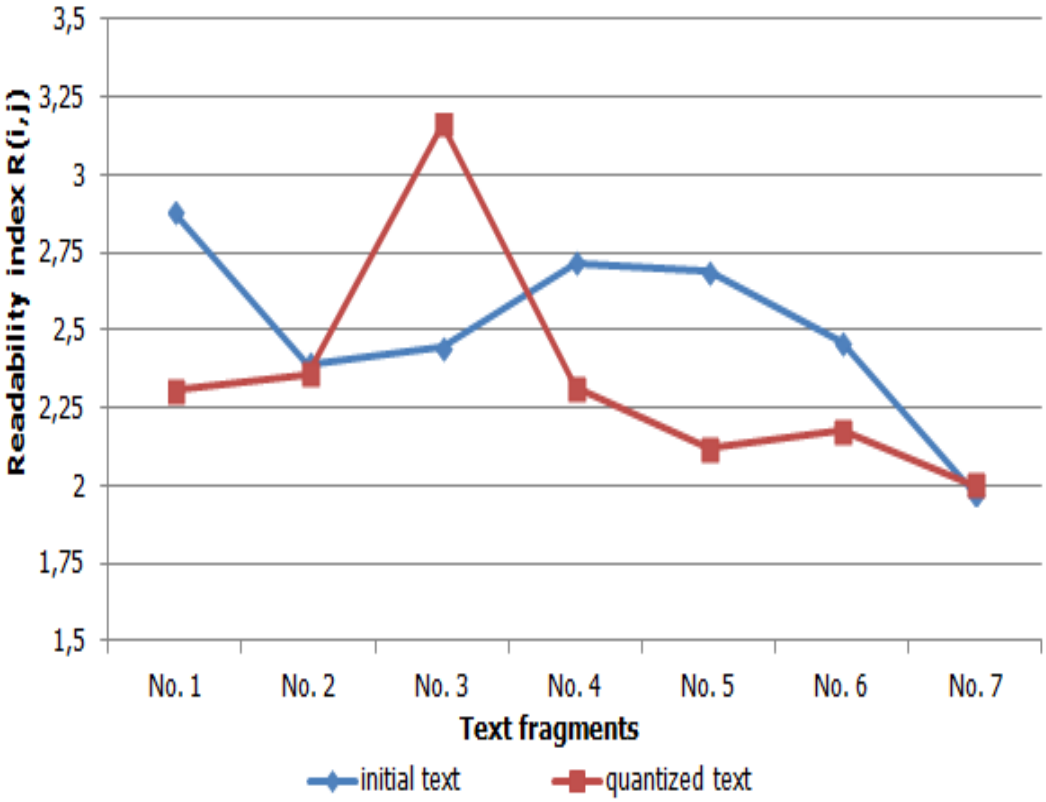

Figure: 5
Comparative analysis of the readability index $R(\bar{i},\bar{j})$ for text fragments

A similar situation is observed for the Flesh's index as well: for the quantized text, the index $Fr(\bar{i},\bar{j})$ is equal to 68.855; for the initial text, the index $Fr(\bar{i},\bar{j})$ is equal to 64.966 that also testify better presentation of the quantized text.

At the same time, comparative analysis of the indexes $Fr(\bar{i},\bar{j})$ for the initial and the quantized text fragments (Figure: 6) indicates that the quantized text fragments No.3 and No.7 require further transformation.

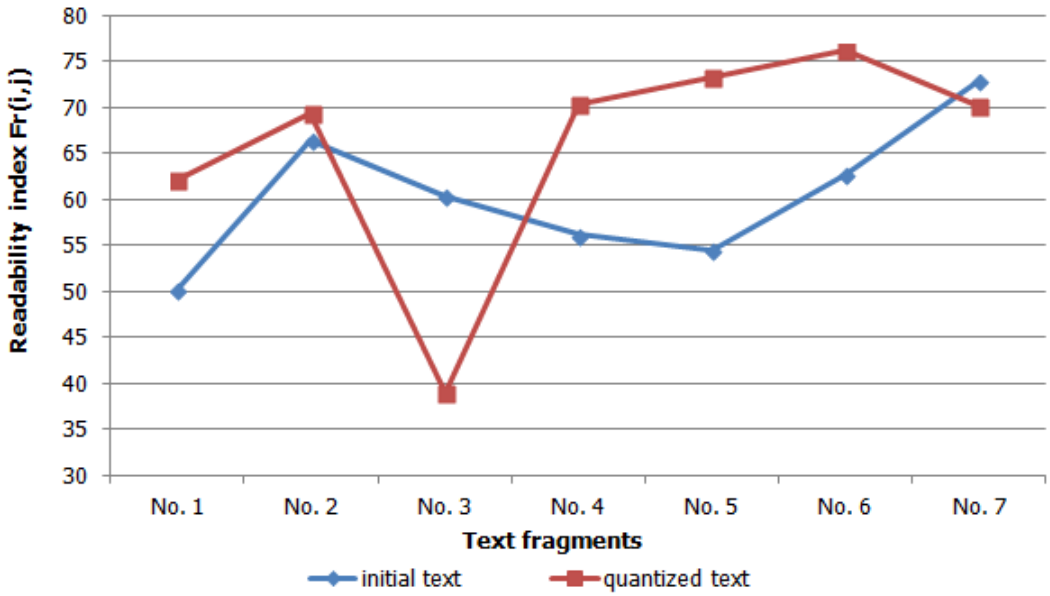

Figure: 6
Comparative analysis of the readability index $Fr(\bar{i},\bar{j})$ for text fragments

Thus the syntactic variety factor and the formal readability index for the quantized text fragments 3 and 7 show that these fragments require requantization of the educational information.

The experiment results allow to draw the following conclusions:

- **Values of the formal readability indexes and for the quantized text are better than for the initial text that testifies their better perception by the reader.**
- **Comparative analysis of formal readability index values and syntactic variety factor values for corresponding initial and quantized text fragments allows determining quantized text fragments, which require requantization of the educational information.**

CONCLUSION

The considered approach allows taking into account formal characteristics for assessing educational text quantization quality. The procedure for obtaining metrics and the method for analyzing educational text quantization quality, offered in the article, can be used for preparing educational content for LMS.

The offered system of quantitative educational content characteristics (Formulas: 1-4) is suitable for weakly structured texts. This criteria system is unsuitable for formulas, tables, graphic and multimedia objects.

Taking into account that these objects, as a rule, are not quantizable, the quantitative characteristics system (Formulas: 1-4) can be successfully used as a part of automated educational content preparation systems

BIODATA and COTACT ADDRESSES of AUTHOR

Alexander ALEKSANDROVICH RYBANOV, PhD in Technical Sciences, Associate Professor. *He's head* **of the Informatics and Programming Techniques Department, Volzhskii Polytechnic Institute, Branch of the Volgograd State Technical University. He has more than 15 years teaching and administration experience in education. His research interests are technological pedagogical content knowledge, pedagogical measurements, web-based distance education, e-learning, communication and information technologies. He's a member of the editorial board of the Vestnik Magistratury Journal published by the Scientific Publishing Centre "Colloquium" (Russia) and** *member of the Russian Academy of Natural History.*

Alexander Alexandrovich RYBANOV
PhD in Technical Sciences, Associate Professor
Informatics and programming techniques Department,
Volzhskii Polytechnic Institute,
Branch of the Volgograd State Technical University,
404121, 42a Engelsa Street, Volzhskii, Volgograd region, RUSSIA
Phone: +7 (88443) 41-22-62
URL: www.volpi.ru
Email: vit@volpi.ru

REFERENCES

Ageev, V. V., Sergevnina, V. M., Yakovleva, E. I. (2011). Means for optimization of linguodidactics for the correspondence mode of study with reduced period of training. ***Bulletin of the Nizhny Novgorod University N. I. Lobachevsky, No. 3-1,*** **37-43.**

Avanesov, V. S. (2012). Application of tasks in a test form and quantized educational texts in new educational technologie. ***Pedagogical measurements, 2,*** **75-91.**

Cepni, S., Gokdere, M., Kucuk, M. (2002). Adaptation of the readability formulas into the Turkish science textbooks. ***Energy Education Science and Technology, 10(1),*** **49-58.**

DuBay, W. H. (2004). ***The principles of readability.*** **Retrieved February 12, 2008, from http://www.nald.ca/fulltext/readab/readab.pdf**

Flesh, R. (1946). ***The Art of Plain Talk*****. New York: Haper and Brothers Publisher, 210 p.**

Flesh, R. (1974). ***The Art of Readability Writing*****, New York: Harper and Row.**

Grieve, J. (2007). Quantitative authorship attribution: An evaluation of techniques. ***Literary and Linguistic Computing, vol. 22 (3),*** **251-270**

Hall, R. H., Hanna, P. (2004). The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. ***Behaviour and Information Technology, vol. 23 (3),*** **183-195**

Hargis, G. (2000). Readability and computer documentation. ***ACM Journal of Computer Documentation, 24(3),*** **122-131.**

Johnson, K. (2008). ***Quantitative methods in linguistics*****. Malden: Blackwell.**

Rybanov, A. A. (2011). Measurement of quality for texts of electronic tutorials. ***School technologies*****,** ***6*****, 172-174.**

Rybanov, A. A. (2013). Educatee's thesaurus as an object of measuring learned material of the distance learning course. ***Turkish Online Journal of Distance Education*****,** ***Vol. 14, No. 4*****, 12-25.**

Tuldava, J. (1975). About Measurement of Text Difficulties. ***In: Proc. Of Tartu State University*****, 102–120**

Tuldava, J. (2004). The development of statistical stylistics (a survey). ***Journal of Quantitative Linguistics, vol. 11 (1-2),*** **141-151**

Wiio, O. A. (1968). Readability. Compression and Readership. ***Acta Universitatis Tamperensis, vol. 22 (A)*****, p. 161.**