

Ю.В. Морозова, И.А. Уртамова

Томский государственный университет систем управления и радиоэлектроники, г. Томск, Россия

МЕТОДИКА АНАЛИЗА ЭЛЕКТРОННОГО УЧЕБНОГО КОНТЕНТА

Форма представления учебного контента оказывает существенное влияние на качество процесса образования. Хорошие результаты дистанционного обучения могут быть получены, когда учебный ресурс тщательно разработан, высокоинформативен, понятен, удобочитаем. В статье рассмотрены критерии анализа электронного учебного контента: информационная насыщенность; абстрактность; удобочитаемость; водность; плотность ключевых слов. Предложена оригинальная методика анализа текста электронных курсов в системах дистанционного обучения, которая позволяет создавать рекомендации для авторов учебно-методических комплексов и исключить некоторые ошибки при их составлении, которые в значительной степени влияют на качество электронного учебного контента.

Ключевые слова: дистанционное обучение, электронный учебный контент, критерии анализа электронного контента, информационная насыщенность, абстрактность, удобочитаемость, водность, плотность ключевых слов.

Электронное обучение прочно закрепилось и получило широкое применение в образовательной среде. Официально дистанционное образование существует в России с 1992 г., когда была принята Концепция создания и развития единой системы дистанционного образования для повышения доступности и качества учебных программ по всей территории страны [1].

Основой дистанционных технологий является методическое обеспечение в виде электронного учебного контента. *Электронный учебный контент* – это содержимое, которое загружается в систему дистанционного обучения, предназначенное для непосредственного восприятия пользователем с целью обучения или ориентации в учебном процессе [2].

Форма представления учебного контента оказывает существенное влияние на качество процесса образования. Хорошие результаты дистанционного обучения могут быть получены, когда учебный ресурс тщательно разработан, высокоинформативен, понятен, удобочитаем.

В настоящее время на факультете дистанционного образования (ФДО) Томского государственного университета систем управления и радиоэлектроники (ТУСУР) разработано 310 электронных курсов по техническим и гуманитарным дисциплинам, содержимое которых необходимо проанализировать. Существующие СДО Moodle, edX, A-tutor и др. не предоставляют разработчикам дистанционных учебных курсов возможность анализировать и оценивать качество учебного контента [3].

На сегодняшний день в сфере анализа учебных текстов проведено немало исследований и предложено еще больше различных критериев оценивания проведенного анализа. В первую очередь стоит выделить количественные критерии от качественных, так как ко вторым практически невозможно применить конкретные математические модели для расчета, и во многих случаях для их оценки возникает необходимость в экспертной группе и множественной проверке на потенциальных читателях.

Количественные же критерии поддаются автоматизации и в большинстве своем имеют конкретные шкалы оценивания. Но не стоит брать любой из таких критериев для анализа контента. Из-за достаточно широкого выбора многие из них могут быть плохо оптимизированы, иметь неточные метрики или вовсе не подходить для русскоязычных текстов.

Для анализа учебного контента были выбраны следующие критерии:

- информационная насыщенность;
- абстрактность;
- удобочитаемость;
- водность;
- плотность ключевых слов.

Что немаловажно, все из этих критериев достаточно хорошо исследованы, чтобы иметь большую точность оценки. Ниже каждый из критериев рассмотрен более подробно.

Информационная насыщенность. Оценивается количеством введенных в текст новых понятий, существующих в явном и скрытом виде.

Информационная насыщенность (1) рассчитывается как процентное соотношение количества встречающихся в тексте новых понятий к общему числу слов:

$$X_{\text{ин}} = (X_{\text{кс}} / X_{\text{окс}}) \cdot 100 \%, \quad (1)$$

где $X_{\text{кс}}$ – количество новых понятий; $X_{\text{окс}}$ – общее количество слов в тексте.

Количественная оценка информационной насыщенности текста может быть совершенно разная для каждого конкретного электронного курса. Но есть два основных показателя, на которые стоит опираться при ее анализе. Во-первых, информационная насыщенность должна стремиться к 100 %, т.е. текст должен быть максимально наполнен новой и полезной информацией. Но на практике такое встречается очень редко, поэтому вторым показателем является минимальный порог этой оценки, он составляет 30 %. Если оценка упала ниже минимального порога, значит, электронный курс почти не несет в себе новой информации для студента.

Абстрактность. Оценивается доля слов в тексте, обозначающих абстрактные смысловые объекты, т.е. такие, которые не доступны непосредственному чувственному восприятию. Предполагается, что чем больше доля абстрактных слов в учебном тексте, тем сложнее он будет усваиваться студентами.

Для оценки абстрактности изложения предлагается использовать методику Я.А. Микка. Согласно этой методике абстрактность имен существительных определяется подсчетом слов с абстрактными суффиксами. Такими суффиксами в русском языке являются: -ость, -есть, -мость, -нность, -ие, -ание, -ение, -ние, -ество, -изм и др. [4].

Основываясь на работах Я.А. Микка 1970 г., введем индекс абстрактности текста (2):

$$X_{\text{абс}} = (X_{\text{ас}} / X_{\text{окс}}) \cdot 100 \%, \quad (2)$$

где $X_{\text{ас}}$ – количество абстрактных имен существительных; $X_{\text{окс}}$ – общее количество слов в тексте.

В качестве шкалы оценивания индекса абстрактности используется метрика, разработанная Людвигом Райнером. Немецкий ученый выявил соотношение между использованием в тексте различных типов слов и понятностью этого текста [5].

Таблица 1

Шкала оценки абстрактности текста

Уровень понятности	Количество абстрактных существительных (на 100 слов текста)
Очень легко понять	До 4
Легко понять	5–8
Понятно	9–15
Трудно понять	16–20
Очень трудно понять	21 и более

Как следует из табл. 1, абстрактные слова использовать можно, но в очень небольших количествах (5–15 %) от общего количества слов в тексте. Если же доля абстрактных существительных превышает 20 %, это сигнал к их немедленному сокращению. Иначе текст однозначно окажется непонятным [5].

Удобочитаемость. Индекс удобочитаемости – мера определения сложности восприятия текста читателем. Индекс удобочитаемости может вычисляться на основе нескольких параметров: длины предложений, слов, удельного количества наиболее частотных (или редких) слов и т.д.

Для реализации поставленной задачи удобочитаемость рассчитывается с помощью индекса Ганнинга (индекс Фога, «туманность» Ганнинга).

Индекс Ганнинга позволяет определить минимальный возраст читателя, которому будет понятен анализируемый текст. Индекс туманности измеряет сложность чтения исходя из средней длины предложения и процента слов, состоящих более чем из четырех слогов. Чем выше индекс туманности, тем сложнее читать текст. Индекс Ганнинга для русскоязычных текстов рассчитывается по формуле (3):

$$\begin{aligned} \text{Индекс Ганнинга} = & \\ = & 0,4 \left[0,78 \cdot \frac{\text{кол-во слов}}{\text{кол-во предложений}} + \right. \\ & \left. + 100 \cdot \frac{\text{кол-во сложных слов}}{\text{кол-во слов}} \right], \end{aligned} \quad (3)$$

где «кол-во сложных слов» – количество слов с числом слогов больше четырех (такое число слогов определено с учетом грамматики русского языка); 0,78 – поправочный коэффициент для русского языка [6].

В табл. 2 приведена шкала оценки индекса Фога [6].

Таблица 2

Проверка удобочитаемости по Фогу

Показатель	Уровень образования
17–20	Выпускник
13–16	Студент колледжа или вуза
11–12	Учащийся старших классов школы
9–10	Учащийся средних классов школы
6–8	Учащийся начальной школы

Чем меньше значение индекса, тем большей аудитории он будет понятен.

Водность. Водность текста – это процент содержания в нем ничего не значащих, не несущих полезной информации слов (стоп-слов). Исходя из предложенного определения, индекс водности текста будет рассчитываться по следующей формуле (4):

$$X_v = (X_{cc} / X_{окс}) \cdot 100 \%, \quad (4)$$

где X_{cc} – количество стоп-слов в тексте; $X_{окс}$ – общее количество слов в тексте.

Примерами стоп-слов могут являться местоимения, союзы, сокращения и др. С полным списком стоп-слов можно ознакомиться в [7].

При расчете водности текста используют следующие метрики [8]:

- до 15 % – отсутствие «воды» в тексте;
- 15–30 % – естественное содержание «воды» в тексте;
- 31–60 % – превышенное содержание «воды» в тексте;
- более 60 % – очень высокое содержание «воды» в тексте.

Максимально допустимым показателем водности считается 60 %. Оптимальный показатель – от 15 до 30 %.

Плотность ключевых слов. Для расчета плотности ключевых слов была использована формула классической плотности.

Классическая плотность – это квадратный корень из числа, обозначающего частоту употребления слова в тексте (5). Плотность не может быть меньше 2,64. Даже если слово использовано в тексте менее 7 раз, корень квадратный в этом случае извлекается из числа 7. Именно классическая плотность взята за основу расчетов плотности ключевых слов.

В качестве общей плотности текста принимается значение плотности самого часто используемого слова.

$$Ч = \max(\sqrt{КС_i}), \quad (5)$$

где КС – количество повторений i -го ключевого слова.

Оптимальным значением считается 5–7 % [8].

Дополнительные параметры. В качестве дополнительных параметров принимаются общие стандарты ФДО, применяемые к разработке электронных учебных курсов.

Данные параметры оцениваются по следующей шкале:

- 1 – показатель соответствует заявленным требованиям;
- 0 – не соответствует заявленным требованиям.

Общая оценка дополнительных параметров рассчитывается исходя из суммы полученных показателей и оценивается от 0 до 1 соответственно шкале:

- 1 – сумма показателей больше 3;
- 0,5 – сумма показателей равна 3;
- 0 – сумма показателей меньше 3.

Ниже подробно расписаны все параметры, учитывающиеся при расчете данной оценки.

1. Соответствие глоссарию. Все термины, добавленные автором в глоссарий, должны быть хотя бы раз употреблены в тексте электронного курса. Если хотя бы один термин не соответствует данному утверждению, параметру присваивается значение 0.

2. Наличие визуального представления информации. В этом параметре учитывается присутствие в электронном курсе таких элементов, как иллюстрации и таблицы. Если курс не содержит ни одного из предложенных элементов, параметру присваивается значение 0.

3. Содержание в тексте ссылок на включенные в курс иллюстрации и таблицы. Ссылка должна быть использована в тексте хотя бы один раз и может присутствовать как в кратком виде (например, «рис. 1»), так и полном виде (например, «рисунок 1»). Если в курсе не присутствует ссылка на хотя бы одну иллюстрацию, параметр будет равен 0. Если в курсе не содержится иллюстраций и таблиц, т.е. предыдущий параметр равен 0, то данному параметру автоматически присваивается значение 0.

4. Содержание в тексте ссылок на содержащиеся в курсе пронумерованные формулы. Логика

Таблица 3

Соответствие шкал

Общая шкала	0	0,5	1
Информационная насыщенность	0–29 %	30–49 %	50–100 %
Абстрактность	21–100 %	16–20 %	0–15 %
Удобочитаемость	17–20	–	6–16
Водность	61–100 %	0–14 % 31–60 %	15–30 %
Плотность ключевых слов	0–4 %	8–100 %	5–7 %
Дополнительные параметры	0	0,5	1

вычисления данного параметра такая же, как и у предыдущего. Единственным отличием является то, что если в курсе не присутствует ни одной формулы, параметру присваивается значение 1, так как наличие формул не является обязательным.

5. Размер специальных блоков для выделения важной информации. Такой блок представляет собой пиктограмму, отделяющую содержащийся в ней текст от основного. Тем самым автор может привлечь внимание студентов к определенным частям курса. Размер таких блоков не должен превышать 1 200 символов, включая пробелы.

Математическая модель. Для получения итоговой оценки электронного учебного курса описанные выше параметры необходимо привести к единой шкале, тем самым сделав их сопоставимыми. В качестве общей шкалы была выбрана следующая:

- 1 – значение показателя является оптимальным;
- 0,5 – значение показателя лежит в пределах допустимой нормы;
- 0 – значение выходит за рамки допустимой нормы.

В табл. 3 приведено соответствие между описанными выше шкалами показателей и общей шкалой оценивания.

Итоговая оценка курса рассчитывается по формуле (6):

$$X = \sum_{i=1}^6 x_i, \quad (6)$$

где x_i – оценка i -го показателя, оцененного относительно общей шкалы.

Таким образом, мы получаем итоговую оценку, лежащую в диапазоне от 0 до 6 и соответствующую следующей шкале:

- 0–3 – низкая оценка курса. Данный показатель возможен в случаях, когда большинство

показателей выходят за рамки допустимой нормы либо часть из них лежат в пределах нормы, но не являются оптимальными;

- 3–5 – средняя оценка курса. Учитывает случаи, когда большинство показателей лежат в пределах нормы, а остальные являются оптимальными либо большинство показателей являются оптимальными, а меньшая часть может выходить за пределы нормы;

- 5–6 – высокая оценка курса. Возможна в случае, когда практически все показатели оптимальны.

Анализ. При анализе текста учебного контента блок с оценками критериев имеет цветовое оформление (рис. 1). Слева располагается общая оценка курса. Справа можно увидеть шесть блоков, размеченных разными цветами. Цвет блока зависит от оценки критерия, приведенной к общей шкале:

- 0 – красный;
- 0,5 – желтый;
- 1 – зеленый.

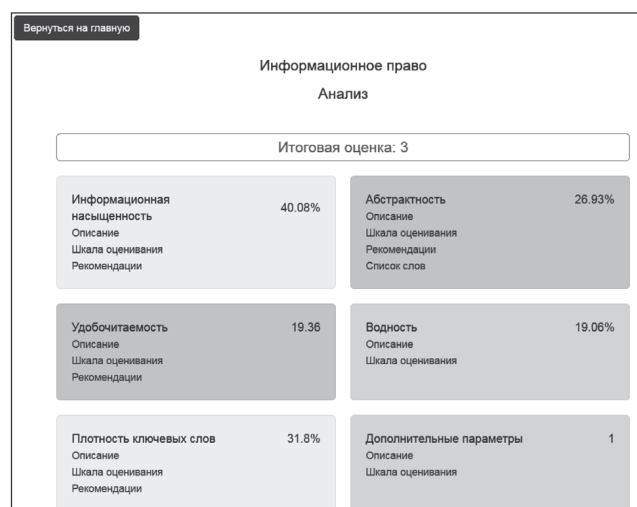


Рис. 1. Окно режима «Анализ»

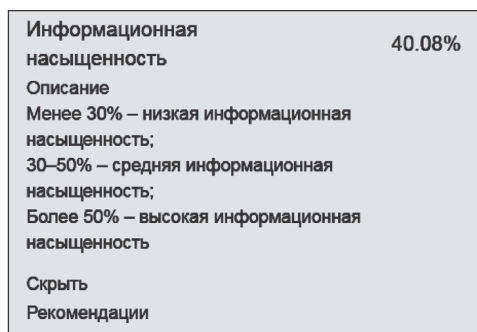


Рис. 2. Шкала оценивания критерия «Информационная насыщенность»

На начальном экране пользователь может увидеть только название критерия и его исходную оценку. У каждого блока с критерием есть 2–4 ссылки. При нажатии на них разворачивается соответствующая информация о данном критерии (рис. 2).

Ссылки имеют следующие категории:

1. Описание – краткая информация о критерии, включающая формулу расчета и входящие в нее показатели.

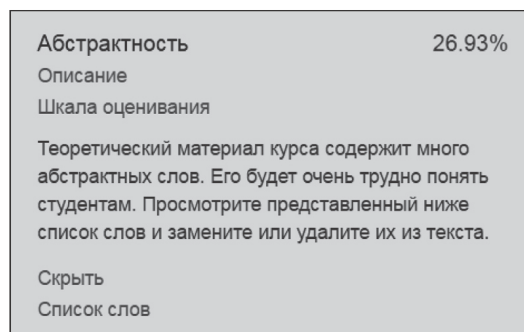


Рис. 3. Рекомендации к критерию «Абстрактность»

2. Шкала оценивания – информация об исходных метриках критерия. Данная информация дана пользователю для того, чтобы он наглядно представил, в каких диапазонах может лежать рассматриваемый критерий.

3. Рекомендации. Выводятся для критериев с оценками «0» и «0,5», кроме критерия «Дополнительные параметры» (рис. 3).

Анализ проводился на электронных курсах, разработанных на ФДО. Ниже рассмотрены резуль-

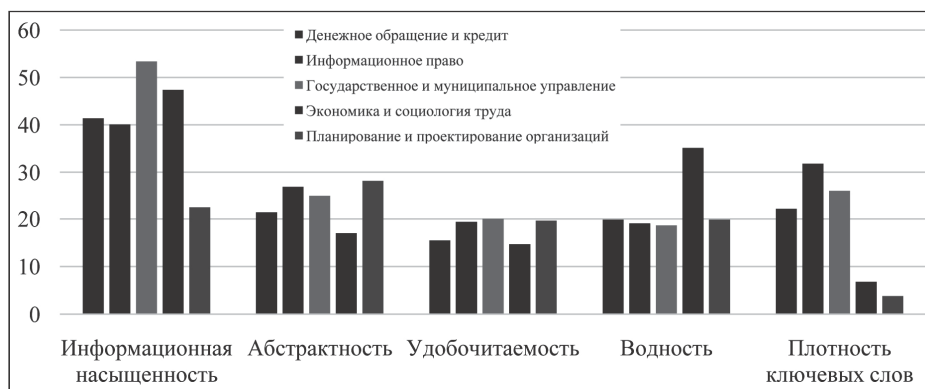


Рис. 4. Сравнение оценок критериев для социально-экономических дисциплин

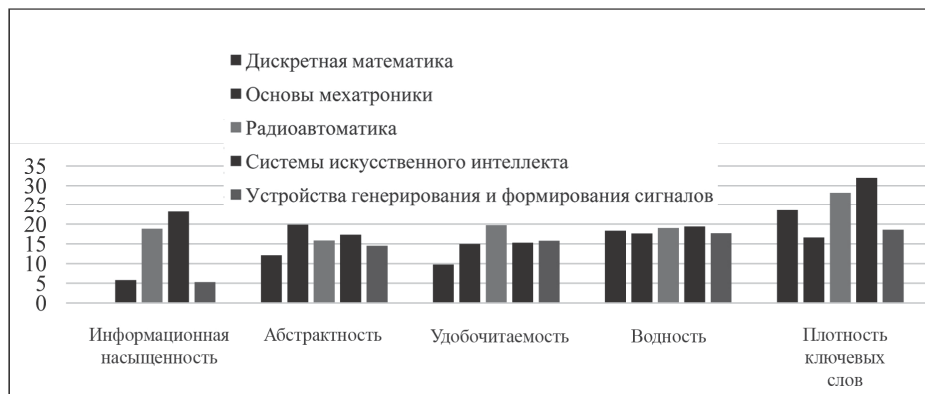


Рис. 5. Сравнение оценок критериев для технических дисциплин

таты анализа 10 курсов (5 – по техническим дисциплинам, 5 – по социально-экономическим).

Анализируя итоговые оценки по каждому виду курсов, можно отметить, что социально-экономические дисциплины уступают техническим, показавшим лучшие результаты при тестировании.

На рис. 4–5 представлено сравнение оценок критериев для каждого вида дисциплин.

Предложенная методика анализа учебного контента позволяет создавать рекомендации для авторов учебно-методических комплексов и исключить некоторые ошибки при их составлении, которые в значительной степени влияют на освоение студентом электронного учебного контента:

- перегруженность текста научными терминами;
- частое использование отвлеченных понятий;
- несоответствие содержания пособия заявленному курсу;
- использование большого количества громоздких предложений;
- недостаточность визуального представления информации.

ЛИТЕРАТУРА

1. Концепция создания и развитие системы дистанционного образования в России. – М.: Госкомвуз, 1995.
2. Семеновских Т.В. Методика электронного обучения / Т.В. Семеновских, С.Ф. Шляпина; ред. В.И. Загвязинский. – Тюмень, 2015. – 56 с.
3. Готская И.Б. Аналитическая записка «Выбор системы дистанционного обучения» / И.Б. Готская, В.М. Жучков, А.В. Кораблев / РГПУ им. А.И. Герцена [Электронный ресурс]. – Режим доступа: <https://ra-kurs.spb.ru/2/0/3/1/?id=13> (дата обращения: 21.06.2017).
4. Микк Я.А. Оптимизация сложности учебного текста: в помощь авторам и редакторам / Я.А. Микк. – М.: Просвещение, 1981. – 119 с.
5. Колесниченко А.В. Практическая журналистика: учеб. пособие / А.В. Колесниченко. – М.: Изд-во Моск. ун-та, 2008.
6. Булгакова Н.В. Технологии обработки текстовой информации средствами Microsoft Word: метод. рекомендации к выполнению практических заданий и лабораторных работ / Н.В. Булгакова, А.А. Чиркина. – Витебск : ВГУ им. П.М. Машерова, 2016. – 43 с.
7. Список стоп-слов [Электронный ресурс]. – Режим доступа: <https://contentmonster.ru/empty/list> (дата обращения: 02.05.2017).
8. Деменко А.В., Рыбанов А.А. Разработка информационной системы, осуществляющей оценку качества текстового веб-контента // Материалы VII Международной студенческой электронной научной конференции «Студенческий научный форум» [Электронный ресурс]. – Режим доступа:

www.scienceforum.ru/2015/812/9666 (дата обращения: 02.05.2017).

Morozova Y.V., Urtamova I.A.

Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia

METHODOLOGY OF ANALYSIS OF ELECTRONIC EDUCATIONAL CONTENT

Keywords: distance learning, e-learning content, criteria for analyzing e-content, information saturation, abstractness, readability, water content, keyword density.

Currently, 310 electronic courses on technical and humanitarian disciplines have been developed at the Faculty of Distance Learning (FDL) of the Tomsk State University of Control Systems and Radioelectronics (TUSUR), the basis of which is methodical provision in the form of educational content. The form of presentation of educational content has a significant impact on the quality of the education process. Good results of distance learning can be obtained when the training resource is carefully designed, highly informative, understandable, and readable.

To date, there are many studies on the analysis of electronic content in various areas, including domestic ones. In most of them qualitative analysis criteria have been singled out, which are difficult to imagine as a universal means of evaluating texts, not to mention the automation of calculating such criteria. Quantitative same criteria are amenable to automation and most of them have specific scales of evaluation.

The article considers the following criteria for the analysis of electronic learning content: information saturation; abstractness; readability; water content; keyword density.

Information saturation is estimated by the number of new concepts introduced in the text that exist in an explicit and hidden form. The abstractness of the text is determined by the fraction of words in the text denoting abstract semantic objects, that is, those that are not accessible to immediate sensory perception. The readability index determines the minimum age of the reader, who will understand the text being analyzed. The water content of a text is the percentage of content in it of meaningless words that do not contain useful information (stop words). The density of the keyword is the square root of the

number indicating the frequency of word usage in the text. The density of the most frequently used word is taken as the general density of the text.

The main importance is that all of these criteria are sufficiently well researched to have a high accuracy of estimation.

As additional parameters, the general FDL standards applied to the development of e-learning courses are adopted: compliance of the glossary; presence of visual representation of information; the content in the text of the links to the included illustrations, tables and formulas; the size of special blocks for highlighting important information.

To obtain the final evaluation of the electronic training course, the parameters described above are reduced to a single scale, which makes them comparable.

The article also offers an original method for analyzing the text of e-courses in distance learning systems, which allows creating recommendations for authors of educational and methodological complexes and eliminating some mistakes in their compilation that significantly affect the quality of electronic learning content.

REFERENCES

1. *Koncepcija sozdaniya i razvitiya sistemy distancionnogo obrazovaniya v Rossii*. – M.: Goskomvuz, 1995.
2. *Semenovskiy T.V. Metodika elektronnoy obucheniya* / T.V. Semenovskiy, S.F. Shljapina; red. V.I. Zagvazinskiy. – Tjumen', 2015. – 56 s.
3. *Gotskaja I.B. Analiticheskaya zapiska «Vybor sistemy distancionnogo obucheniya»* / I.B. Gotskaja, V.M. Zhuchkov, A.V. Korablev / RGPU im. A.I. Gercena [Elektronnyy resurs]. – Rezhim dostupa: <https://ra-kurs.spb.ru/2/0/3/1/?id=13> (data obrashheniya: 21.06.2017).
4. *Mikk Ja.A. Optimizatsiya slozhnosti uchebnogo teksta: v pomoshh' avtoram i redaktoram* / Ja.A. Mikk. – M.: Prosveshchenie, 1981. – 119 s.
5. *Kolesnichenko A.V. Prakticheskaya zhurnalistika: ucheb. posobie* / A.V. Kolesnichenko. – M.: Izd-vo Mosk. un-ta, 2008.
6. *Bulgakova N.V. Tehnologii obrabotki tekstovoy informatsii sredstvami Microsoft Word: metod. rekomendatsii k vypolneniyu prakticheskikh zadaniy i laboratornykh rabot* / N.V. Bulgakova, A.A. Chirkina. – Vitebsk : VGU im. P.M. Masherova, 2016. – 43 s.
7. *Spisok stop-slov* [Elektronnyy resurs]. – Rezhim dostupa: <https://contentmonster.ru/empty/list> (data obrashheniya: 02.05.2017).
8. *Demenko A.V., Rybanov A.A. Razrabotka informatsionnoy sistemy, osushchestvlyajushhej ocenku kachestva tekstovogo veb-kontenta* // Materialy VII Mezhdunarodnoy studencheskoj elektronnoy nauchnoy konferentsii «Studencheskiy nauchnyy forum» [Elektronnyy resurs]. – Rezhim dostupa: www.scienceforum.ru/2015/812/9666 (data obrashheniya: 02.05.2017).