

Стариченко Борис Евгеньевич,

доктор педагогических наук, профессор, заведующий кафедрой информационно-коммуникационных технологий в образовании, Уральский государственный педагогический университет; 620075, г. Екатеринбург, ул. К. Либкнехта, 9; e-mail: b.e.starichenko@uspu.su.

Устинов Максим Андреевич,

студент 4-го курса Института математики, физики, информатики и технологий, Уральский государственный педагогический университет; 620075, г. Екатеринбург, ул. К. Либкнехта, 9; e-mail: m.ustinov@uspu.su.

ПРОГРАММА АВТОМАТИЗАЦИИ КОНТРОЛЯ ОФОРМЛЕНИЯ ТЕКСТОВЫХ ДОКУМЕНТОВ

КЛЮЧЕВЫЕ СЛОВА: нормоконтроль документа; автоматизация проверки; текстовые документы; оформление документов; эталоны текстовых документов; режимы проверки; работа с документами.

АННОТАЦИЯ. В статье выделяется ряд практически значимых задач, связанных с соответствием оформления текстовых документов установленным требованиям: функционирование редакций различных изданий, составление научных отчетов, обучение работе с текстовым редактором, заполнение форм документов, нормоконтроль и пр. Используемая в настоящее время «ручная» их проверка существенно замедляет подготовку документов. Несомненно актуальным является решение задачи автоматизации проверки оформления электронных текстовых документов. Информационный анализ не выявил разработок, которые обеспечивали бы контроль оформления произвольного офисного текстового документа. Обосновывается выбор языка C# в качестве инструментальной среды разработки. Приводится UML-диаграмма основной модели работы приложения. Кратко описывается основной алгоритм проверки оформления текста. Результатом разработки явилось приложение MAPDoc (Модуль Автоматизации Проверки Документов). Построенные на едином алгоритме проверки выделяются несколько различных по функциональному назначению режимов проверки. Режим «Учебный» предназначен для использования в целях обучения, когда учащимся дается задание по формированию документа единого для всех содержания и оформления. Эталон представляет собой правильно оформленный текст. В ходе проверки осуществляется последовательное сопоставление элементов форматирования каждого абзаца документа ученика с эталоном. Описанный режим может быть использован учеником в процессе тренажа или учителем в ходе контроля. В режиме «Редакторский» проверяемый документ может иметь авторское содержание, но последовательность используемых в нем стилей оформления должна соответствовать эталону. Проверка осуществляется уполномоченным лицом («редактором»). Однако перспективной и удобной представляется возможность самостоятельной проверки автором с помощью представляемой программы своей рукописи с целью обеспечения ее соответствия требованиям редакции (учреждения).

Starichenko Boris Evgenyevich,

Doctor of Pedagogy, Professor, Head of Department of Information and Communication Technologies in Education, Ural State Pedagogical University, Ekaterinburg, Russia.

Ustinov Maxim Andreevich,

Student of the 4th Year, Institute of Mathematics, Physics, Informatics and Technology of the Urals State University, Ekaterinburg, Russia.

PROGRAM OF AUTOMATION VERIFICATION OF TEXT DOCUMENTS FORMATTING

KEYWORDS: regulatory document control; automatic check; text document; document format; text document models; check mode; document processing.

ABSTRACT. The article highlights a number of practically significant tasks related to the correspondence of text documents to the established requirements: the functioning of various publication requirements, compiling of scientific reports, teaching to work with a text editing programs, filling of forms of documents, standard control and others. Currently used «manual» checks significantly slow the execution of documents. It is undoubtedly relevant to solve the problem of automated verification of electronic text documents formatting. Information analysis did not reveal developments that would ensure verification over the design of an arbitrary office text document. The choice of the C# language as a tool for development is justified. The UML diagram of the main application model is given. The article briefly describes the basic algorithm for verifying a text. The result of the development was the application MAPDoc (Module for Automating the Verification of Documents). Constructed on a single verification algorithm, there are several different modes of verification. The «Training» mode is intended to be used for the purposes of training, when students are given the task of formatting a document of uniform content and design. The standard is a right-formatted text. During the check, the formatting elements of each paragraph of the student's document are consistently compared with the standard. The described mode can be used by the student during the training process or by the teacher during the control. In the «Editorial» mode, the document being checked can have copyright content, but the sequence of the design styles used in it must correspond to the standard. Verification is carried out by an authorized person («editor»). However, perspective and convenient is the possibility of self-checking by the author with the help of the submitted program of his manuscript in order to ensure its compliance with the requirements of the editorial office.

Введение

Во многих сферах деятельности текстовый документ, создаваемый некоторым автором, должен строго соответствовать установленным правилам оформления. Примерами может служить необходимость выполнения требований редакций к оформлению рукописей, соблюдения стандартов подготовки технической документации, удовлетворения положений по оформлению научно-технических отчетов, диссертаций, учебных работ, соответствия формам документов, подаваемых в различные учреждения, и т. п.

В отечественном делопроизводстве требования оформления итогов научно-исследовательских, проектно-конструкторских, конструкторско-технологических и проектно-технологических работ (отчетов НИР) определяются ГОСТ 7.32-2017 [6]. В частности, установленным им требованиям должны соответствовать все работы научного и учебного характера, в том числе, студенческие.

В работе компаний весьма важным оказывается приведение видов документов к единому стандарту во избежание возможных проблем, связанных с разным форматированием документов внутри одной компании или между разными компаниями. Регулирование форматирования данных документов обеспечивает ГОСТ Р 7.0.97-2016, согласно которому для организационно-распорядительных документов (уставов, положений, правил, инструкций, регламентов, постановлений, распоряжений, приказов, решений, протоколов, договоров, актов, писем, справок и др.) определены правила их оформления, в том числе с применением информационных технологий [7]. Под требование указанного ГОСТа попадают все документы, упоминаемые в «ОК 011-93 Общероссийский классификатор управленческой документации» [10].

Свои требования к оформлению рукописей имеют все редакции научных изданий. В частности, весьма строгими являются правила оформления статей для журналов Scopus и Web of Science, включающие требования к представлению ссылок APA [12].

Отдельно следует остановиться на проверке документов учебного характера. В процессе освоения технологий подготовки электронных документов учащиеся различного уровня (от школьников до пенсионеров) учатся использовать инструменты форматирования. Другим примером может служить нормоконтроль выпускных работ, который в обязательном порядке осуществляется перед их представлением к защите. Многие студенты участвуют в различных научных конкурсах, проектах, в которых также необходимо оформлять результаты в

строгом соответствии с требованиями.

Во всех перечисленных и иных подобных ситуациях в учреждении, куда представляется документ, уполномоченными лицами (в учебных учреждениях — преподавателем) осуществляется проверка соответствия документа установленным требованиям оформления и в случае его отсутствия документ возвращается на доработку. Несмотря на развитие и распространение электронных форм подготовки и представления документов, контроль правильности их оформления до настоящего времени осуществляется «вручную». Это приводит к значительным затратам времени проверяющего и затягивает процесс согласования документа для его автора. При этом в работе Д. А. Романова отмечается, что в общем времени согласования документа исправление ошибок оформления занимает более 30% [13].

В то же время, если исходить из того, что требования форматирования являются формальными, следовательно, проверка на соответствие им имеет алгоритмический характер, а это в свою очередь означает, что она может быть автоматизирована. Общая идея работы подобной программной системы состоит в том, что необходимые требования к оформлению могут быть формализованы в виде эталона — некоторого документа, с которым в дальнейшем программно-нормоконтролер будет сравнивать элементы оформления проверяемого документа.

Таким образом, представляется актуальным решение проблемы программной автоматизации процесса контроля соответствия оформления текстовых документов на предмет их соответствия некоторому наперед заданному эталону оформления.

Обзор подходов к решению проблемы

Анализ информационных источников позволил выделить следующие категории программных систем проверки текстовых документов:

1. Системы проверки текстовых заимствований (плагиата) (см., например, обзоры [2; 8]).

2. Системы проверки орфографии и грамматики, используемые в целом ряде случаев, когда оказывается недостаточно встроенного орфографического контроля редактора MS Word (см., например, обзоры [3; 11]).

3. Системы проверки технической документации в отдельных производственных отраслях [1; 9].

4. Система аналитической обработки текстовой информации, обеспечивающая проверку документов при их согласовании и проведении юридической экспертизы [13]. Частью проверки является выявление некоторых погрешностей оформления документа: ошибки форматирования абзацев, не-

правильное использование шрифтов, сбой в нумерации, неправильное оформление приложений, неверные форматы дат и чисел, неправильная структура документа.

Помимо перечисленных программных систем достаточно подробно описываются приемы автоматизации обработки текстового документа при его вводе (наборе) в редактор MS Word, которые позволяют автоматически форматировать текст в соответствии с установленными настройками (шаблоном, стилями) [4; 5].

В публикациях И. А. Самойловой описана разработанная ею программа «Нормоконтроль», предназначенная для автоматизации процесса проверки дипломных работ студентов Карагандинского государственного университета имени акад. Е. А. Букетова на соответствие требований, предъявляемых к оформлению выпускных работ [14; 15]. Приложение выполнено с использованием языка VBA для MS Word и обеспечивает реализацию следующих функций: проверка документа; занесение сведений в журнал нормоконтроля (с указанием даты и времени проверки, количества ошибок); получение файла с подробным описанием ошибок. Из описания ясно, что программа работает с единственным эталоном и предполагает проверку только одного определенного типа документа.

В процессе анализа библиографических и иных информационных источников авторам не удалось выявить программных систем,

обеспечивающих контроль оформления текстовых документов по произвольному, наперед заданному эталону. Этим определяется новизна и оригинальность нашей разработки.

Методы исследования

В результате сопоставительного анализа возможных программных систем для решения задачи — Java (Apache POI, Aspose for Java), VBA, C#, предпочтение было отдано C#, благодаря ее бесплатности, среднему порогу вхождения для программирования, а также большой функциональности в работе с офисными документами [16]. Связь языка с объектами документов MS Office производится через пространство имен Microsoft.Office.Interop. Данная библиотека позволяет обратиться к любому COM-интерфейсу (объекту) документа, а также к любому его свойству, т. е. эта библиотека позволяет описать любое действие, выполняемое пользователем графически, с помощью кода.

Разрабатываемое приложение должно обеспечивать проверку оформления текстовых документов по произвольному, заранее описанному шаблону. Проверка производится путем сопоставления документа пользователя с эталоном. В качестве эталона может быть выбран (задается проверяющим) любой файл формата .doc, .docx, .docm, .dot, .dotx или .dotm. На рис. 1 представлена UML диаграмма прецедентов, описывающая все возможные взаимодействия:

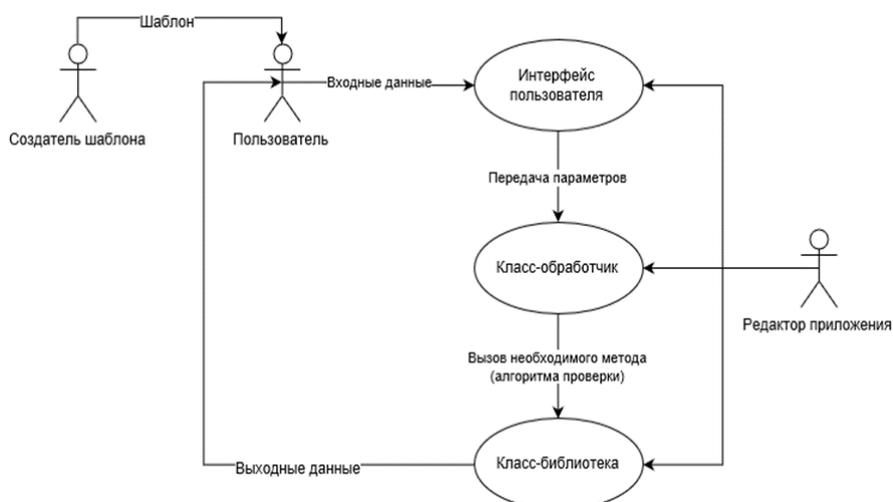


Рис. 1. UML диаграмма основной модели работы приложения

С учетом возможной расширяемости и модифицируемости характера проверки, а также особенностей текстового процессора MS Word, был построен универсальный алгоритм проверки. После старта метода запускается цикл, отвечающий за прохождение по всем абзацам проверяемого документа. Внутри данного цикла проверяются стиль всего абзаца, свойства абзаца, такие как

межстрочный интервал, отступ до и после, выравнивание и т. д. Внутри этого цикла находится также цикл, отвечающий за проверку стиля каждого слова в текущем абзаце. В свою очередь, в него включен еще один цикл, отвечающий за проверку стиля каждой буквы в текущем слове. Помимо этого, в цикл верхнего уровня входит цикл, обеспечивающий проверку свойств слов в текущем

абзаце. Проверяются такие свойства как жирный текст, курсивный текст, шрифт, размер шрифта и т. д. Данный алгоритм был построен с учетом восприятия свойств текста самим текстовым процессором и позволяет выявить все ошибки оформления документа,

описываемые алгоритмом.

Результатом разработки стало приложение MAPDoc (МАПДок — Модуль Автоматизации Проверки Документов), интерфейс которого представлен на рис. 2.

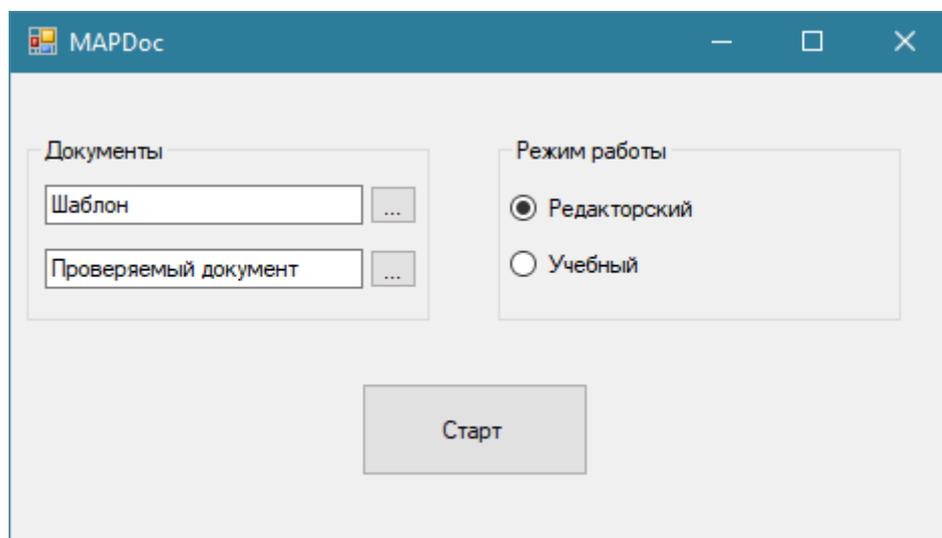


Рис. 2. Интерфейс приложения MAPDoc

Результаты и дискуссия

Хотя разработанный алгоритм носит универсальный характер, программа, созданная на его основе, предусматривает несколько различных по функциональному назначению режимов проверки.

Режим «Учебный». Режим предназначен для использования в целях обучения, когда учащимся дается задание по формированию документа единого для всех содержания и оформления. Эталон представляет собой правильно оформленный текст. В ходе проверки осуществляется последовательное сопоставление элементов форматирования каждого абзаца документа ученика с эталоном. Им однозначно определяется формат абзаца, следующего за проверяемым. Протокол проверки учебного документа содержит сведения о несоответствии элементов его форматирования эталонным.

Описанный режим может быть использован в процессе тренажа, в ходе которого у обучаемого вырабатываются умения оформлять текст в соответствии с установленными требованиями; в этом случае проверку производит сам ученик, добиваясь полного соответствия оформления своего текста эталонному. При необходимости контрольного оценивания работ учеников проверку осуществляет преподаватель.

Режим «Редакторский». Этот режим можно также назвать «нормативным». Эталон представляет собой правильно оформленный документ с некоторой последовательностью стилей оформления. Примером может служить статья, оформленная в соот-

ветствии с требованиями редакции, или документ, подаваемый пользователем в некоторое учреждение, или текст студенческой выпускной квалификационной работы. Проверяемый документ может иметь авторское содержание, но последовательность используемых в нем стилей оформления должна соответствовать эталону. При этом в эталоне указывается, какие стили из установленного набора могут использоваться в абзаце, следующим за данным. Например, абзац, располагающийся после предыдущего со стилем «Основной текст», может иметь стили «Основной текст», «Нумерованный список», «Таблица», но не может иметь стиль «Заголовок», «Аннотация», «Автор». В протокол проверки выводятся сведения об отклонении последовательности стилей от допустимой. Проверка осуществляется уполномоченным лицом («редактором»). Однако перспективной и удобной видится возможность обеспечения доступа самого автора к программе проверки с нужным эталоном, например, на сайте редакции (учреждения) — в этом случае автор сам сможет проверять рукопись и добиваться ее соответствия требованиям редакции, и уже после этого направлять ее в редакцию.

Таким образом, использование описанного приложения MAPDoc позволяет автоматизировать проверку оформления в целом ряде значимых для практики ситуаций.

Заключение

В заключении хотелось бы обратить внимание на два основных достоинства, которые

авторы усматривают в своей разработке.

Во-первых, приложение автоматизирует рутинную работу лиц, связанных с проверкой оформления электронных рукописей. При этом исключаются ошибки оформления, которые могут быть незаметны «на глаз» — размер шрифта, межстрочный интервал, отступ, размер полей и пр.

Во-вторых, предложенное в работе решение позволяет существенным образом изменить порядок взаимодействия автора документа и учреждения, куда документ предполагается представить. Учреждение каким-то образом делает доступной программу автоматизированной проверки с нужным встроенным шаблоном. Автор, пользуясь ею, самостоятельно проверяет свой документ и при необходимости исправляет элементы оформления. Подобная схема используется,

например, при проверке на плагиат. Тем самым устраняется необходимость содержать в штате редакций специальных лиц, занятых контролем оформления рукописей. То же самое можно сказать и о нормоконтроле студенческих работ, диссертаций.

Развитие идей работы видится в создании приложения, которое смогло бы на основании заданного эталона генерировать промежуточный документ с полями ввода фрагментов текста нужных стилей, в которые автор заносит свою информацию в произвольном формате. По завершении ввода осуществляется форматирование документа в соответствии с требованиями эталона. Для автора подобное приложение заметным образом облегчило бы подготовку электронной рукописи, а редактора избавило бы от необходимости проверки ее оформления.

ЛИТЕРАТУРА

- 1С:Автоматизированная проверка конфигураций [Электронный ресурс]. — Режим доступа: <http://v8.1c.ru/acc/> (дата обращения: 10.06.2018).
- 6 лучших сервисов для проверки на плагиат онлайн [Электронный ресурс]. — Режим доступа: <http://geek-nose.com/6-luchshix-servisov-dlya-proverki-na-plagiat-onlajn/> (дата обращения: 10.06.2018).
- 7 лучших онлайн сервисов для проверки правописания [Электронный ресурс]. — Режим доступа: <http://geek-nose.com/7-luchshix-onlajn-servisov-dlya-proverki-pravopisaniya/> (дата обращения: 10.06.2018).
- Автоматизация обработки текстового документа [Электронный ресурс]. — Режим доступа: <http://files.lbz.ru/authors/informatika/9/5-1-6.pdf> (дата обращения: 10.06.2018).
- Автоматизация обработки текстового документа [Электронный ресурс]. — Режим доступа: eos.ibi.spb.ru/umk/11_15/5/5_R3_T6.html (дата обращения: 10.06.2018).
- ГОСТ 7.32-2017 СИБИД. Отчет о научно-исследовательской работе. Структура и правила оформления [Электронный ресурс]. — Режим доступа: <http://docs.cntd.ru/document/1200157208/> (дата обращения: 10.06.2018).
- ГОСТ Р 7.0.97-2016 Система стандартов по информации, библиотечному и издательскому делу (СИБИД). Организационно-распорядительная документация. Требования к оформлению документов [Электронный ресурс]. — Режим доступа: <http://docs.cntd.ru/document/1200142871> (дата обращения: 20.04.2018).
- Какие бывают системы проверки уникальности текста [Электронный ресурс]. — Режим доступа: <http://antiplagiatu.net/news/sistemy-proverki-unikalnosti-teksta> (дата обращения: 10.05.2018).
- Калачев Я. Б. Автоматизированный контроль качества текстов проектной документации на предприятиях топливно-энергетического комплекса [Электронный ресурс]. — Режим доступа: [http://www.eltech.ru/assets/files/nauka/dissertacii/2015/Kalachev/Avtoreferat_1.4.5-\(1\).pdf](http://www.eltech.ru/assets/files/nauka/dissertacii/2015/Kalachev/Avtoreferat_1.4.5-(1).pdf) (дата обращения: 10.05.2018).
- ОК 011-93. «Общероссийский классификатор управленческой документации» [Электронный ресурс]: утв. Постановлением Госстандарта России от 30.12.1993 № 299 (ред. от 21.12.2017). — Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_43631/ (дата обращения: 10.06.2018).
- Поваляев Е. Системы проверки орфографии [Электронный ресурс]. — Режим доступа: <https://compress.ru/article.aspx?id=9511> (дата обращения: 10.06.2018).
- Правила оформления статей для журналов Scopus и WoS [Электронный ресурс]. — Режим доступа: http://www.sci-cooperation.ru/partners/Instructions_for_authors (дата обращения: 10.06.2018).
- Романов Д. А. Системы аналитической обработки текстовой информации [Электронный ресурс]. — Режим доступа: <http://preferentum.ru/images/present.pdf> (дата обращения: 30.05.2018).
- Самойлова И. А. Сложность реализации автоматизированной проверки дипломных работ на соответствие техническим нормам // Молодой ученый. — 2016. — № 15. — С. 149–152.
- Самойлова И. А. Некоторые особенности поэтапного алгоритма программы для проверки дипломных работ на нормоконтроль [Электронный ресурс] // Молодой ученый. — 2016. — № 16. — С. 51–53. — Режим доступа: <https://moluch.ru/archive/120/33282/> (дата обращения: 10.06.2018).
- Язык программирования C# [Электронный ресурс]. — Режим доступа: <https://vadimstroganov.com/yazyki-programirovaniya/c-sharp> (дата обращения: 18.02.2018).

REFERENCES

- 1S:Avtomatizirovannaya proverka konfiguratsiy [Elektronnyy resurs]. — Rezhim dostupa: <http://v8.1c.ru/acc/> (data obrashcheniya: 10.06.2018).
- 6 luchshikh servisov dlya proverki na plagiat onlayn [Elektronnyy resurs]. — Rezhim dostupa: <http://geek-nose.com/6-luchshix-servisov-dlya-proverki-na-plagiat-onlajn/> (data obrashcheniya: 10.06.2018).
- 7 luchshikh onlayn servisov dlya proverki pravopisaniya [Elektronnyy resurs]. — Rezhim dostupa: <http://geek-nose.com/7-luchshix-onlajn-servisov-dlya-proverki-pravopisaniya/> (data obrashcheniya: 10.06.2018).

4. Avtomatizatsiya obrabotki tekstovogo dokumenta [Elektronnyy resurs]. — Rezhim dostupa: <http://files.lbz.ru/authors/informatika/9/5-1-6.pdf>. (data obrashcheniya: 10.06.2018).
5. Avtomatizatsiya obrabotki tekstovogo dokumenta [Elektronnyy resurs]. — Rezhim dostupa: eos.ibi.spb.ru/umk/11_15/5/5_R3_T6.html. (data obrashcheniya: 10.06.2018).
6. GOST 7.32-2017 SIBID. Otchet o nauchno-issledovatel'skoy rabote. Struktura i pravila oformleniya [Elektronnyy resurs]. — Rezhim dostupa: <http://docs.cntd.ru/document/1200157208/> (data obrashcheniya: 10.06.2018).
7. GOST R 7.0.97-2016 Sistema standartov po informatsii, biblioteknomu i izdatel'skomu delu (SIBID). Organizatsionno-rasporyaditel'naya dokumentatsiya. Trebovaniya k oformleniyu dokumentov [Elektronnyy resurs]. — Rezhim dostupa: <http://docs.cntd.ru/document/1200142871> (data obrashcheniya: 20.04.2018).
8. Kakie byvayut sistemy proverki unikal'nosti teksta [Elektronnyy resurs]. — Rezhim dostupa: <http://antiplagiatu.net/news/sistemy-proverki-unikalnosti-teksta> (data obrashcheniya: 10.05.2018).
9. Kalachev Ya. B. Avtomatizirovannyy kontrol' kachestva tekstov proektnoy dokumentatsii na predpriyatiyakh toplivno-energeticheskogo kompleksa [Elektronnyy resurs]. — Rezhim dostupa: [http://www.eltech.ru/assets/files/nauka/dissertacii/2015/Kalachev/Avtoreferat_1.4.5-\(1\).pdf](http://www.eltech.ru/assets/files/nauka/dissertacii/2015/Kalachev/Avtoreferat_1.4.5-(1).pdf) (data obrashcheniya: 10.05.2018).
10. OK 011-93. «Obshcherossiyskiy klassifikator upravlencheskoy dokumentatsii» [Elektronnyy resurs] : utv. Postanovleniem Gosstandarta Rossii ot 30.12.1993 № 299 (red. ot 21.12.2017). — Rezhim dostupa: http://www.consultant.ru/document/cons_doc_LAW_43631/ (data obrashcheniya: 10.06.2018).
11. Povalyaev E. Sistemy proverki orfografii [Elektronnyy resurs]. — Rezhim dostupa: <https://compress.ru/article.aspx?id=9511> (data obrashcheniya: 10.06.2018).
12. Pravila oformleniya statey dlya zhurnalov Scopus i WoS [Elektronnyy resurs]. — Rezhim dostupa: http://www.sci-cooperation.ru/partners/Instructions_for_authors (data obrashcheniya: 10.06.2018).
13. Romanov D. A. Sistemy analiticheskoy obrabotki tekstovoy informatsii [Elektronnyy resurs]. — Rezhim dostupa: <http://preferentum.ru/images/present.pdf> (data obrashcheniya: 30.05.2018).
14. Samoylova I. A. Slozhnost' realizatsii avtomatizirovannoy proverki diplomnykh rabot na sootvetstvie tekhnicheskim normam // Molodoy uchenyy. — 2016. — № 15. — S. 149–152.
15. Samoylova I. A. Nekotorye osobennosti poetapnogo algoritma programmy dlya proverki diplomnykh rabot na normokontrol' [Elektronnyy resurs] // Molodoy uchenyy. — 2016. — № 16. — S. 51–53. — Rezhim dostupa: <https://moluch.ru/archive/120/33282/> (data obrashcheniya: 10.06.2018).
16. Yazyk programmirovaniya C# [Elektronnyy resurs]. — Rezhim dostupa: <https://vadimstroganov.com/yazyki-programmirovaniya/c-sharp> (data obrashcheniya: 18.02.2018).