

О.С. Иванов, В.А. Лохвицкий, А.С. Дудкин, С.С. Титов

КОМПЛЕКСНАЯ АВТОМАТИЧЕСКАЯ ИНТЕЛЛЕКТУАЛЬНАЯ
ОБРАБОТКА ИНТЕРНЕТ-КОНТЕНТА, ВЫВОДИМОГО НА МОНИТОР
В ХОДЕ ПОЛЬЗОВАТЕЛЬСКОЙ СЕССИИ*

Рассмотрены вопросы комплексной автоматической интеллектуальной обработки интернет-контента, выводимого на монитор в ходе пользовательской сессии. Представлены основные критерии оценки информации и ее классификации в интернете. Описаны механизмы работы алгоритмов автоматического анализа различных видов интернет-контента. Приведен пример работы программных алгоритмов в режиме ручного, или экспертного, анализа выведенного на монитор пользователя интернет-контента. Предложен подход к разработке макета программно-аппаратного комплекса, используемого для выявления и интеллектуальной обработки деструктивного мультимедийного интернет-контента.

Ключевые слова: психическое здоровье, интернет-контент, алгоритмы автоматического анализа, программное средство.

O.S. Ivanov, V.A. Lokhvitsky, A.S. Dudkin, S.S. Titov

INTEGRATED AUTOMATIC INTELLIGENT PROCESSING INTERNET
CONTENT DISPLAYED ON A MONITOR DURING USER SESSION

The issues of complex automatic intelligent processing of Internet content displayed on the monitor during a user session are considered. The main criteria for assessing information and its classification on the Internet are presented. The mechanisms of operation of algorithms for automatic analysis of various types of Internet content are described. An example of the work of a complex of these software algorithms in the manual, or expert mode, analysis of the Internet content displayed on the user's monitor is given. An approach to the development of a model of a software and hardware complex used to identify and intelligently process destructive multimedia Internet content is proposed.

Keywords: mental health, Internet content, algorithms of automatic analysis, software tool.

Вводные замечания

Мультимедийная информация, выводимая на монитор пользователя, может быть разнообразной. В общем виде практически невозможно спрогнозировать, какой именно интернет-контент в данную сессию пользователь будет просматривать.

Пользователь – акцептор интернет-контента – может целенаправленно искать какую-то нужную ему информацию или сталкиваться с некоторым инфотрафиком случайным образом. Для таких случаев представляется целесообразным предоставить в распоряжение пользователей инструмент защиты их психического (в том числе психологического) здоровья [5] от действия деструктивного интернет-контента и приемов вредоносного социального инжиниринга [18].

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-29-22064 «Модели и методы выявления и интеллектуальной обработки деструктивного мультимедийного интернет-контента».

Информационная безопасность

Например, существует мнение, что телевизионная или интернет-реклама преимущественно негативно воспринимаются зрителями [1, 20]. Прерывающая просмотр фильма или передачи реклама вносит диссонанс в мыслительный процесс и переживания зрителя основного контента. Когнитивно-аффективный диссонанс проявляется в раздражении, которое, в свою очередь, резко изменяет эмоционально-когнитивный настрой зрителя и его поведенческие реакции.

Кроме того, реклама в самом широком смысле – это совокупность рекламных технологий, применяемых для продвижения какой-либо информации. Реклама навязывает акцептору нужные транслятору (рекламодателю) мысли и эмоции, а значит, и поведение [15].

По ряду формальных признаков рекламу как сложный комплексный контент можно отнести к негативному, деструктивному, классу информационных сообщений и запретить законодательно для трансляции по публичным инфоканалам. Но поскольку в рекламной и ассоциированной с ней индустриях сходятся интересы самых разных хозяйствующих субъектов и потребителей, то для обоснования подобного заключения требуются веские аргументы и воля [22]. Аналогичным образом влияет на пользователя рекламный интернет-контент за счет встроенных или всплывающих рекламных баннеров-блоков и/или квазислучайных переходов пользователя на сторонние, не запрашиваемые им интернет-ресурсы, содержащие контент с бесполезным содержанием. Таким образом, поскольку рекламные сообщения содержат (1) провоцирующий к определенному поведению контент, (2) в котором применяются специальные технологии социального инжиниринга (3) и восприятие которого нежелательно для акцептора, то рекламу (ее большую часть) вполне корректно отнести к вирусному, то есть деструктивному, контенту.

В этой связи программное средство (ПС), осуществляющее автоматический интеллектуальный анализ медиаконтента, может стать инструментом защиты пользователя от нежелательных ему сообщений [13], в частности таких как интернет-реклама.

Критерии оценки и классификации информации

С помощью приемов и технологий социальной инженерии [19], на которых построено распространение вирусного контента [21], могут распространяться (и нередко распространяются) информационные материалы запрещенного и/или ограниченного доступа [11]. Соответственно, наличие в распоряжении пользователя инструмента, предотвращающего его контакт с нежелательной информацией в сети Интернет, будет способствовать сохранению его психического здоровья [7].

Вопрос о критериях оценки информации и ее классификации на категории многогранен. Какой бы ни была информация, по одной из классификационных систем ее условно можно разделить на следующие классы:

- 1) *тексты* (с учетом языковой группы текста, стилистики, объема и др. [4, 13]);
- 2) *звуковая* (аудиальная), классифицируемая на подклассы: устная речь (монолог, диалог, полилог, какофония и др.), естественные звуки, искусственные звуки, которые, в свою очередь, можно разделить на гармоничные, агармоничные и др. [17];
- 3) *графический контент*, представленный статической (изображения, фотографии, картины или символы) и квазистатической визуальной информацией (специальный тип изображений, представленных неподвижными пикселями, но расположенными так, что за

Комплексная автоматическая интеллектуальная обработка...

счет особенностей психофизиологии восприятия изображений человеком (у субъекта возникает иллюзия подвижности неподвижного изображения) [16], а также динамическими изображениями (короткие, зацикленные изображения в формате .gif или видеоролики разной продолжительности).

В зависимости от того, какой тип (класс) контента выводится на монитор пользователя, он (контент) может быть обработан разными алгоритмами [3].

Если это тексты, то обработка выполняется на предмет оценки, например, языка, читабельности букв (размер, шрифт, контраст «символ – фон» и др.), орфографии (правописания), наличия/отсутствия в тексте ключевых слов и фраз, определения стиля текста (проза, стихотворение, технический текст, рекламный и др.). Для автоматического анализа звуков используются алгоритмы, реализованные программными продуктами или средствами распознавания речи и оценки ее языковой принадлежности, а также алгоритмы определения гармоничных и агармоничных звуков [12].

На рисунке 1 представлена когнитивная карта, визуализирующая вышеизложенное.

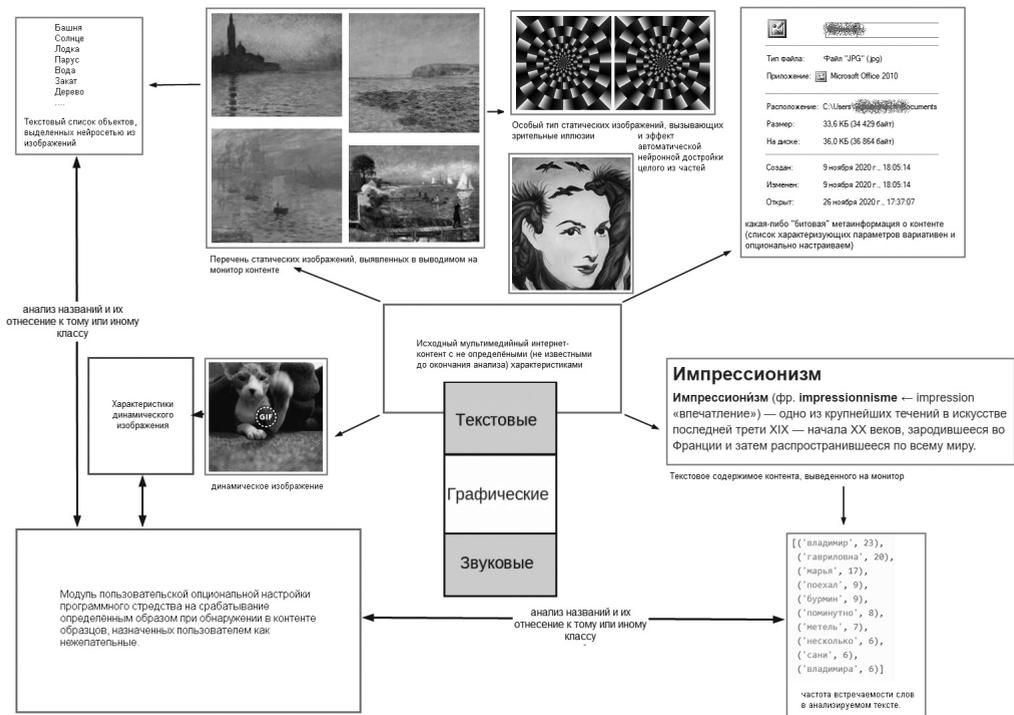


Рис. 1. Когнитивная карта архитектуры программного средства

Для автоматического анализа статических изображений используются алгоритмы: 1) оценки представленности пиксельных полей определенных диапазонов цветовых тонов; 2) определения на изображении графических примитивов; 3) оценки пропорциональности объектов и другие итерационно подключаемые/отключаемые программные модули, реализующие соответствующие целевые алгоритмы.

Информационная безопасность

Для автоматического анализа видеороликов в дополнение к алгоритмам выявления наличия и структурного анализа текстов, анализа динамики акустического поля [12], наличия графических примитивов, пиксельных полей, объектов используются алгоритмы оценки переходов от кадра к кадру по значениям всех перечисленных выше параметров.

Таким образом, для осуществления автоматического программного интеллектуально-го анализа мультимедийного интернет-контента с исходно неизвестным с точки зрения влияния на психическое здоровье потребителя содержимым целесообразно использовать различные алгоритмы обработки контента, взаимодополняющие свою информативность за счет объединения их в комплекс [10], тем более что такое объединение не исключает автономной работы каждого такого программного модуля.

На рисунке 2 представлен частный вариант результата работы комплекса описанных выше программных алгоритмов в режиме ручного, или экспертного, анализа выведенного на монитор пользователя интернет-контента.

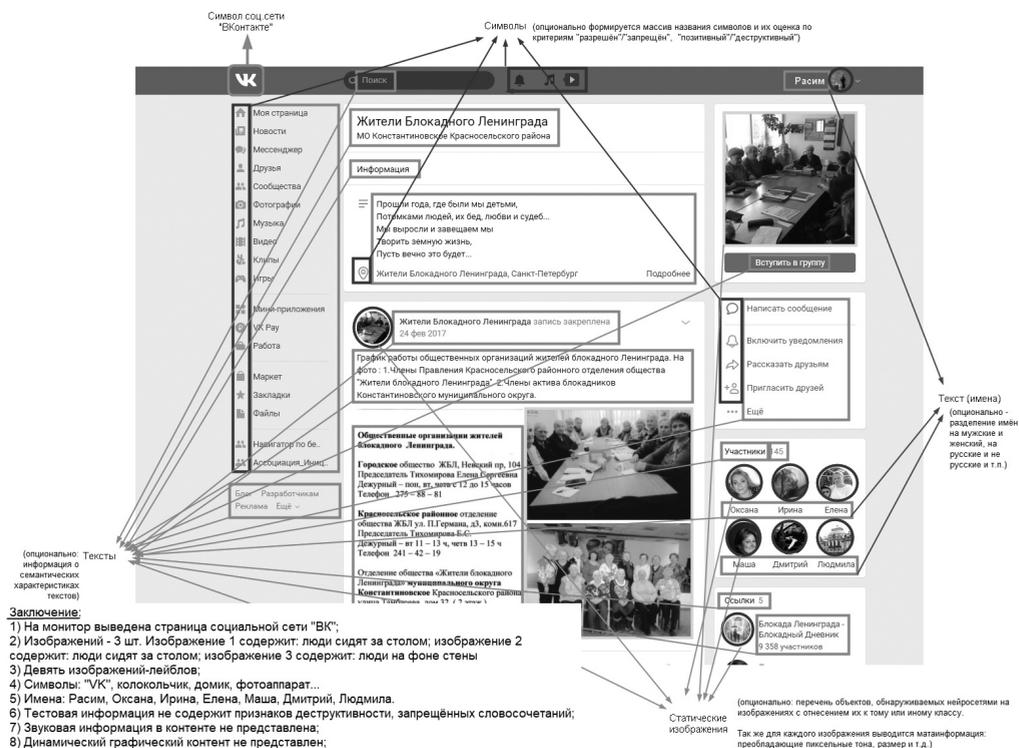


Рис. 2. Результаты комплексного автоматического интеллектуального анализа интернет-контента, выведенного на монитор пользователя

Программное средство по обработке интернет-контента

Функциональность разрабатываемого ПС по обработке мультимедийного контента аналогична описанному.

На первом шаге программным средством анализируются физические характеристики контента: яркостно-контрастные характеристики, параметры мелькания (изменения светового потока во времени), параметры акустического сопровождения изображений [12].

Комплексная автоматическая интеллектуальная обработка...

По результатам анализа физических характеристик формируется заключение о вероятном характере влияния этих (физических) характеристик контента на сенсорные системы человека: «допустимое», «настороженное», «недопустимое», «неопределенное» [7].

На втором шаге ПС выделяет на изображении блоки (области) с качественно разным содержанием: тексты, графика – символы, изображения, «гифки», видеоролики. Далее каждый тип блоков обрабатывается специализированными релевантными программными модулями (алгоритмами). В частности, области, содержащие текст, анализируются программными модулями, анализирующими параметры текстов [9]. В ходе такого анализа в тексте выделяются слова и (или) фразы, которые, в свою очередь, анализируются на предмет наличия в них негативного смысла; определяется стиль изложения, язык текста и др.

Области, содержащие статичные изображения, обрабатываются программными модулями, называемыми нейросетями. Результатом работы подобных нейросетей является текстовый перечень выявленных объектов. Этот список становится доступным для анализа другими программными модулями на предмет обнаружения/необнаружения в нем объектов, относимых, в том числе, к негативной (деструктивной) группе [2].

Области с динамической графикой обрабатываются соответствующим данному типу контента образом [14].

В приведенном описании функциональности создаваемого ПС указано, что результатами интеллектуального анализа, выведенного на монитор пользователя мультимедийного интернет-контента, являются списки с названиями объектов, а также, где это возможно, характеристики объектов с точки зрения их влияния на психическое здоровье акцептора информации.

В работе [7] авторы подробно рассмотрели, что разделение мультимедийного интернет-контента на «нейтральный», «позитивный» и «деструктивный» является сложной задачей, не имеющей общего решения. Любое решение будет частным, контекстно зависимым, то есть отнесение контента к деструктивному зависит от того, кто и для каких целей воспринимает анализируемый контент.

В разрабатываемом ПС предусмотрен блок пользовательских опциональных настроек, в базу шаблонов которого пользователь может заранее внести итерационный перечень слов-названий и прочих характеристик – числовых значений, символов и др., являющихся образцами-идентификаторами, с которыми автоматически выполняется сравнение названий объектов. Соответственно, если сличение и идентификация произошли, то ПС сигнализирует пользователю опционально указанным способом о наличии в просматриваемом контенте признаков деструктивной (обозначенной пользователем как таковая) информации, а также указывает, на какой именно «объект» или «группу объектов» произошло срабатывание. С этого момента пользователь имеет возможность задать «поведение» ПС при обнаружении им подобных признаков в другой «порции» контента.

Ввиду огромного разнообразия (скорее всего, принципиально неисчислимого) вариантов подаваемой на монитор пользователя информации архитектура ПС реализуется на основе принципа открытости – принятого в WWW-сообществе соглашения, суть которого формулируется следующим образом: кто угодно может сказать что угодно и о чем угодно. Применение данного принципа в архитектуре ПС дает возможность обеспечить принципиальную итерационность программных модулей (мультиагентный подход [24])

Информационная безопасность

без потери достигнутой функциональности; также обеспечивается возможность произвольно менять количество самих модулей в составе ПС без потери его работоспособности. Данный принцип реализован в технологиях Semantic Web [25]; варианты его практической реализации представлены программными средствами онтологического инжиниринга [8, 23].

Выводы

Таким образом, в данной работе описан и сформулирован реализуемый авторами подход к разработке макета программно-аппаратного комплекса, используемого для выявления и интеллектуальной обработки деструктивного мультимедийного интернет-контента.

Литература

1. *Ахметшина А.А.* Современная коммерческая реклама как идеологический аппарат государства. Социологический ракурс // Вестник Чувашского университета. 2009. № 4. С. 198–203.
2. *Белим С.В., Ларионов С.Б.* Алгоритм сегментации изображения с помощью искусственной нейронной сети без использования других изображений // Радиостроение. 2017. № 3. С. 43–53.
3. *Бождай А.С., Тимонин А.Ю.* Исследование процесса анализа текстовых и мультимедиа данных социального профиля из открытых источников информации // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 2. С. 19–28.
4. *Брицков Р.А.* Ранжирование информации на основе оценок и поведения пользователей // Т-Сотт – Телекоммуникации и Транспорт. 2016. Т. 10, № 1. С. 62–66.
5. *Воронина А.В.* Проблема психического здоровья и благополучия человека: обзор концепций и опыт структурно-уровневого анализа // Сибирский психологический журнал. 2005. № 21. С. 142–147.
6. *Губанова А.Ю.* Классификация электронного контента сайтов для детей: социологический анализ // Вестник РГГУ. Серия «Философия. Социология. Искусствоведение». 2015. № 7. С. 139–143.
7. *Иванов О.С., Пилькевич С.В., Гнидко К.О., Лохвицкий В.А. и др.* Обоснование критериев оценки влияния интернет-контента на психологическое здоровье потребителя информации // Социология. 2020. № 5. С. 275–282.
8. *Итинсон К.С.* Web 1.0, Web 2.0, Web 3.0: этапы развития веб-технологий и их влияние на образование // Карельский научный журнал. 2020. Т. 9, № 1 (30). С. 19–21.
9. *Кулабухов Р.О.* Оценка фоносемантического программного обеспечения для психолингвистического анализа текста с целью выявления эмоционального окраса текста // Евразийский научный журнал. 2017. № 3. С. 206–208.
10. *Курако Е.А., Орлов В.А.* Методы динамического обновления программных комплексов для систем с каскадной структурой // Проблемы управления. 2017. № 5. С. 52–60.
11. *Ловцов Д.А., Федичев А.В.* Архитектура национального классификатора правовых режимов информации ограниченного доступа // Правовая информатика. 2017. № 2. С. 35–54.
12. *Миргородская Ю.В., Бернавская М.В., Стаценко Л.Г., Чусов А.А. и др.* Объектный анализ акустического поля, создаваемого различными источниками звука в произвольном помещении // Известия Южного федерального университета. Технические науки. 2018. № 6. С. 85–94.
13. *Новиков А.А., Стрекалов И.Э., Лопатин Д.В.* Динамическая фильтрация web-контента // Гаудеамус. 2013. № 2. С. 159–163.

Комплексная автоматическая интеллектуальная обработка...

14. *Петров А.А.* Новейшие инструменты четвертой промышленной революции и цифровые механизмы контроля и управления обществом // *Кронос*. 2020. № 8. С. 24–34.
15. *Простотина Ю.В.* Институт рекламы в формировании моделей потребительского поведения: гендерный аспект // *Гуманитарные, социально-экономические и общественные науки*. 2019. № 1. С. 45–50.
16. *Рабичев И.Э., Котов А.В., Амирханян А.А.* Иллюзия «сцинтилляции»: вероятные механизмы ее формирования // *Теоретическая и экспериментальная психология*. 2018. Т. 11, № 1. С. 28–34.
17. *Сиднева Т.Б.* Шум и музыка: логика взаимопревращений // *Известия Российского государственного педагогического университета им. А.И. Герцена*. 2012. № 146. С. 25–33.
18. *Сырецкий Г.А.* Сквозные цифровые технологии и прорывные технологии кибер-безопасности в контексте системного инжиниринга // *Интерэкспо Гео-Сибирь*. 2018. № 7. С. 254–260.
19. *Тепляков С.П., Тимохович А.С.* Социальная инженерия. Анализ и методы защиты // *Academy*. 2018. № 7. С. 26–27.
20. *Терских М.В.* Отношение аудитории к провокационной рекламе: критерии оценки коммуникативной эффективности // *Неофилология*. 2020. Т. 6, № 21. С. 201–212.
21. *Тетерина Е.А., Бородина Е.С.* Эффективность вирусной рекламы в социальных медиа // *Наука. Общество. Государство*. 2019. Т. 7, № 3 (27). С. 173–179.
22. *Тимчак К.Н., Осипов Е.М.* Эффективность российской социальной рекламы: критерии и проблемы оценки // *Социология*. 2019. № 6. С. 126–131.
23. *Elezaj O., Yayilgan S.Y., Ahmed J., (...), Brichfeld B., Haubold C.* Crime Intelligence from Social Media Using CISMO // *Advances in Intelligent Systems and Computing*. 2021. No. 1183. Pp. 441–460.
24. *Masloboev A.V., Langhans M.* A Multi-Agent System for Management Information Support of Regional Innovations // *Научно-технический вестник информационных технологий, механики и оптики*. 2018. Т. 18, № 4. С. 630–638.
25. *Reformat M.Z.* Semantic Web: Graphs, Imprecision and Knowledge Generation // *Studies in Fuzziness and Soft Computing*. 2021. No. 394. Pp. 271–283.

Literatura

1. *Akhmetshina A.A.* Sovremennaya kommercheskaya reklama kak ideologicheskij apparat gosudarstva. Sotsiologicheskij rakurs // *Vestnik Chuvashskogo universiteta*. 2009. № 4. S. 198–203.
2. *Belim S.V., Larionov S.B.* Algoritm segmentatsii izobrazheniya s pomoshch'yu iskusstvennoj nejronnoj seti bez ispol'zovaniya drugikh izobrazhenij // *Radiostroenie*. 2017. № 3. S. 43–53.
3. *Bozhdaj A.S., Timonin A.Yu.* Issledovanie protsessa analiza tekstovykh i mul'timedia dannykh sotsial'nogo profilya iz otkrytykh istochnikov informatsii // *Izvestiya vysshikh uchebnykh zavedenij. Povolzhskij region. Tekhnicheskie nauki*. 2017. № 2. S. 19–28.
4. *Britsov R.A.* Ranzhirovanie informatsii na osnove otsenok i povedeniya pol'zovatelej // *T-Comm – Telekommunikatsii i Transport*. 2016. Т. 10, № 1. S. 62–66.
5. *Voronina A.V.* Problema psikhicheskogo zdorov'ya i blagopoluchiya cheloveka: obzor kontseptsij i opyt strukturno-urovneвого analiza // *Sibirskij psikhologicheskij zhurnal*. 2005. № 21. S. 142–147.
6. *Gubanova A.Yu.* Klassifikatsiya elektronного kontenta sajtoв dlya detej: sotsiologicheskij analiz // *Vestnik RGGU. Seriya «Filosofiya. Sotsiologiya. Iskusstvovedenie»*. 2015. № 7. S. 139–143.

Информационная безопасность

7. *Ivanov O.S., Pil'kevich S.V., Gnidko K.O., Lokhvitskij V.A. i dr.* Obosnovanie kriteriev otsenki vliyaniya internet-kontenta na psikhologicheskoe zdorov'e potrebitelya informatsii // *Sotsiologiya*. 2020. № 5. S. 275–282.
8. *Itinson K.S.* Web 1.0, Web 2.0, Web 3.0: etapy razvitiya veb-tekhnologij i ikh vliyanie na obrazovanie // *Karel'skij nauchnyj zhurnal*. 2020. T. 9, № 1 (30). S. 19–21.
9. *Kulabukhov R.O.* Otsenka fonosemanticheskogo programmnoho obespecheniya dlya psikholingvisticheskogo analiza teksta s tsel'yu vyyavleniya emotsional'nogo okrasa teksta // *Evrazijskij nauchnyj zhurnal*. 2017. № 3. S. 206–208.
10. *Kurako E.A., Orlov V.L.* Metody dinamicheskogo obnovleniya programmnykh kompleksov dlya sistem s kaskadnoj strukturoj // *Problemy upravleniya*. 2017. № 5. S. 52–60.
11. *Lovtsov D.A., Fedichev A.V.* Arkhitektura natsional'nogo klassifikatora pravovykh rezhimov informatsii ogranichennoho dostupa // *Pravovaya informatika*. 2017. № 2. S. 35–54.
12. *Mirgorodskaya Yu.V., Bernavskaya M.V., Statsenko L.G., Chusov A.A. i dr.* Ob'ektnyj analiz akusticheskogo polya, sozdavaemogo razlichnymi istochnikami zvuka v proizvol'nom pomeshchenii // *Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki*. 2018. № 6. S. 85–94.
13. *Novikov A.A., Strekalov I.E., Lopatin D.V.* Dinamicheskaya fil'tratsiya web-kontenta // *Gau-deamus*. 2013. № 2. S. 159–163.
14. *Petrov A.A.* Novejshie instrumenty chetvertoj promyshlennoj revolyutsii i tsifrovye mekhanizmy kontrolya i upravleniya obshchestvom // *Kronos*. 2020. № 8. S. 24–34.
15. *Prostotina Yu.V.* Institut reklamy v formirovanii modelej potrebitel'skogo povedeniya: gendernyj aspekt // *Gumanitarnye, sotsial'no-ekonomicheskie i obshchestvennye nauki*. 2019. № 1. S. 45–50.
16. *Rabichev I.E., Kotov A.V., Amirkhanyan A.A.* Illyuziya «sintillyatsii»: veroyatnye mekhanizmy ee formirovaniya // *Teoreticheskaya i eksperimental'naya psikhologiya*. 2018. T. 11, № 1. S. 28–34.
17. *Sidneva T.B.* Shum i muzyka: logika vzaimoprevrashchenij // *Izvestiya Rossijskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena*. 2012. № 146. S. 25–33.
18. *Syretskij G.A.* Skvoznye tsifrovye tekhnologii i proryvnye tekhnologii kiber-bezopasnosti v kontekste sistemnogo inzhiniringa // *Interekspo Geo-Sibir'*. 2018. № 7. S. 254–260.
19. *Teplyakov S.P., Timokhovich A.S.* Sotsial'naya inzheneriya. Analiz i metody zashchity // *Academy*. 2018. № 7. S. 26–27.
20. *Terskikh M.V.* Otnoshenie auditorii k provokatsionnoj reklame: kriterii otsenki kommunikativnoj effektivnosti // *Neofilologiya*. 2020. T. 6, № 21. S. 201–212.
21. *Teterina E.A., Borodina E.S.* Effektivnost' virusnoj reklamy v sotsial'nykh media // *Nauka. Obshchestvo. Gosudarstvo*. 2019. T. 7, № 3 (27). S. 173–179.
22. *Timchak K.N., Osipov E.M.* Effektivnost' rossijskoj sotsial'noj reklamy: kriterii i problemy otsenki // *Sotsiologiya*. 2019. № 6. S. 126–131.
23. *Elezaj O., Yayilgan S.Y., Ahmed J., (...), Brichfeld B., Haubold C.* Crime Intelligence from Social Media Using CISMO // *Advances in Intelligent Systems and Computing*. 2021. No. 1183. Pp. 441–460.
24. *Masloboev A.V., Langhans M.* A Multi-Agent System for Management Information Support of Regional Innovations // *Nauchno-tekhnicheskij vestnik informatsionnykh tekhnologij, mekhaniki i optiki*. 2018. T. 18, № 4. S. 630–638.
25. *Reformat M.Z.* Semantic Web: Graphs, Imprecision and Knowledge Generation // *Studies in Fuzziness and Soft Computing*. 2021. No. 394. Pp. 271–283.