УДК 004.056 **DOI** 10.15622/ia.20.4.1

И.В. Котенко, И.Б. Саенко, А.А. Браницкий,
И.Б. Паращук, Д.А. Гайфулина

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ЦИФРОВОГО СЕТЕВОГО КОНТЕНТА ДЛЯ ЕГО ЗАЩИТЫ ОТ НЕЖЕЛАТЕЛЬНОЙ ИНФОРМАЦИИ

Котенко И.В., Саенко И.Б, Браницкий А.А., Паращук И.Б., Гайфулина Д.А. **Интеллектуальная система аналитической обработки цифрового сетевого контента для его защиты от нежелательной информации.**

Аннотация. В настоящее время Интернет и социальные сети как среда распространения цифрового сетевого контента становятся одной их важнейших угроз персональной, общественной и государственной информационной безопасности. Возникает необходимость защиты личности, общества и государства от нежелательной информации. В научно-методическом плане проблема защиты от нежелательной информации имеет крайне небольшое количество решений. Этим определяется актуальность представленных в статье результатов, направленных на разработку интеллектуальной системы аналитической обработки цифрового сетевого контента для защиты от нежелательной информации. В статье рассматриваются концептуальные основы построения такой системы, раскрывающие содержание понятия нежелательной информации и представляющие общую архитектуру системы. Приводятся модели и алгоритмы функционирования наиболее характерных компонентов системы, таких как компонент распределенного сканирования сети, компонент многоаспектной классификации сетевых информационных объектов, компонент устранения неполноты и противоречивости и компонент принятия решений. Представлены результаты реализации и экспериментальной оценки системных компонентов, которые продемонстрировали способность системы отвечать предъявляемым требованиям по полноте и точности обнаружения и противодействию нежелательной информации в условиях ее неполноты и противоречивости.

Ключевые слова: интеллектуальная система, цифровой сетевой контент, нежелательная информация, классификация, нечеткие знания, принятие решений.

1. Введение. Стремительное внедрение глобальной сети Интернет и построенных на ее основе социальных сетей в государственную, производственно-экономическую и социально-культурную сферы современного общества является мощным стимулом их дальнейшего развития. В то же время Интернет и социальные сети становятся одной их важнейших угроз персональной, общественной и государственной информационной безопасности. Возникает необходимость защиты личности, общества и государства от нежелательной информации, которая распространяется через глобальные компьютерные сети и способна нанести вред здоровью граждан или мотивировать их к противоправному поведению. В ведущих странах мира, включая Россию, защита от нежелательной сетевой информации регулируется национальным законодательством. Сайты с нежелательным контентом блокируются и заносятся в черные списки. Однако обнаружение нежелательных сайтов и

формирование черных списков осуществляется, как правило, в ручном режиме, а экспертное суждение о принадлежности информации к той или иной категории всегда является субъективным. По этой причине оно может быть недостаточно полным и/или ошибочным, что требует добавления в процесс выявления и противодействия нежелательной информации методов устранения ее неполноты и противоречивости.. Кроме того, при ручном режиме анализа Интернет-контента достаточно сложно обеспечить выполнение требований по своевременности реагирования на появление новых информационных объектов и изменение содержимого существующих ресурсов.

В научно-методическом плане проблема защиты от нежелательной информации имеет небольшое количество научно-технических решений. Несмотря на то, что за последние годы появились методики и реализации отдельных компонентов такого рода систем защиты, они или находятся на начальной стадии разработки и внедрения, или не реализуют полного спектра предполагаемых возможностей [1-3].

Целью настоящей статьи является изложение результатов исследований, посвященных построению и функционированию перспективной интеллектуальной системы, предназначенной для аналитической обработки цифрового сетевого контента в интересах защиты от нежелательной информации. Для выявления и противодействия нежелательной информации в данной системе используются методы машинного обучения и методы обработки нечетких данных. С целью удовлетворения этих требований в системе предложен и реализуется ряд специфических процессов обработки данных, которые также выделяют разработанную систему от других подобных систем. К числу этих процессов следует отнести: одновременный анализ большого количества источников данных о смысловом наполнении информационного объекта; многоуровневая классификация цифровых информационных объектов; использование методов обработки неполной и противоречивой информации; применение различных способов противодействия нежелательной информации в зависимости от целевой аудитории и некоторые другие.

К числу новых результатов исследований, которые освещает статья, относятся: (1) концептуальные основы построения и функционирования предлагаемой интеллектуальной системы; (2) модели и методы работы основных компонентов системы; (3) ключевые аспекты реализации и результаты экспериментальной оценки использования системы для решения возлагаемых на нее задач. Этим определяется дальнейшая структура статьи. В разделе 2 приводятся результаты анализа релевантных работ. Раздел 3 содержит концептуальные основы, раскрывающие понятие нежелательной информации, а также общую структуру целевой

системы. Реализация и экспериментальная оценка системы обсуждаются в разделе 4. Раздел 5 содержит заключительные выводы и направления дальнейших исследований.

2. Состояние исследований. Несмотря на то, что в последние годы появились методы и реализации отдельных компонентов такого рода систем защиты, они либо находятся на начальной стадии разработки и внедрения, либо не реализуют весь спектр ожидаемых возможностей. Так, в [4-9] рассматриваются некоторые механизмы обнаружения и противодействия вредоносной информации в сетевых информационных объектах. В этих документах излагаются решения для определения надежных оценок цифрового сетевого контента. Рассмотренные в них механизмы основаны на методах классификации информации, методах интеллектуальной обработки данных и фильтрации спама. Однако эти механизмы не ориентированы на работу в условиях семантической неопределенности информационного содержания.

В работах [10-12] рассматриваются различные методы анализа социальных сетей для обнаружения и выбора мер противодействия вредоносной информации. В [10] для обнаружения вредоносной информации используются алгоритмы поиска по описанию события, идентификации пользователей различных сетей и поиска по группам пользователей. Методы количественной и качественной оценки информационных воздействий в социальных сетях, основанные на табличных и графических инструментах для представления метрик и расчета метрик, обсуждаются в [11]. В [12] рассмотрены подходы к определению демографических характеристик пользователей социальных сетей. Однако, поскольку помимо социальных сетей существуют и другие источники нежелательной информации, эти подходы нельзя считать универсальными.

На наш взгляд, в наибольшей степени для обнаружения и противодействия нежелательной информации подходят методы анализа трафика на основе классификации веб-страниц. Эти методы могут быть основаны на контент-анализе внутренних свойств веб-страниц [13]. Бинарный классификатор, основанный на выявлении групп внутренних свойств HTML-документов, используется для обучения систем классификации веб-страниц в работах [14, 15]. В [16] показано, что обучение классификаторов обнаружению и противодействию нежелательной информации может быть реализовано на основе комбинации значимых функций веб-страниц. Однако методы, представленные в [13-16], не ориентированы на анализ содержания веб-страниц, то есть веб-контента.

В ряде работ обнаружение и противодействие нежелательной информации реализуется с использованием алгоритмов классификации тем веб-контента [17, 18]. В этом случае поиск вредоносной информации осуществляется по URL-адресам. Однако этот метод уменьшает спектр характеристик нежелательной информации, которые необходимо анализировать, и, соответственно, уменьшает диапазон контрмер. Некоторое время был популярен подход, основанный на анализе ссылок в веб-контенте. Такой анализ позволял реализовать иерархическую и объединенную классификацию веб-контента [19, 20]. Для классификации использовались модели на основе метода SVM. Для классификации веб-сайтов в [21] предложен алгоритм Link Information Categorization (LIC), который основан на методе классификации kNN. Классификация страниц с помощью алгоритма kNN также была исследована в [22], где различным терминам и тегам присваиваются соответствующие веса. В [23] выполняется классификация веб-сайтов с помощью анализа существительных, извлекаемых из веб-страниц. В качестве метода классификации используется Decision Tree. В [24] рассмотрен метод, заключающийся в поиске и извлечении значимого текста из тегов с последующим применением классификатора Naïve Bayes к полученным выборкам. Такой же подход в сочетании с методами противодействия вредоносной информации упоминается в [25, 26].

Выявление и противодействие нежелательной информации в реальных условиях, то есть когда обработка и оценка свойств нежелательной информации осуществляется в условиях неполноты и неопределенности, требует использования подходов, основанных на методах, моделях и алгоритмах устранения неопределенности и неполноты. Например, обработка неопределенной информации различного типа и поддержка принятия решений обычно реализуются с использованием искусственных нейронных сетей [27-30], нечетких множеств [31, 32] и нейронечетких сетей [33]. Применение этих методов для обнаружения и противодействия нежелательной информации является довольно сложной задачей. Однако преимущества этих методов заключаются в том, что они позволяют выбирать меры противодействия вредоносной информации на основе оценки семантического содержания информационных объектов в условиях неполноты и неопределенности. Эти методы также будут рассмотрены в статье.

3. Концептуальные основы интеллектуальной системы аналитической обработки цифрового сетевого контента. В настоящем разделе рассматриваются понятие нежелательной информации и общая архитектура предлагаемой системы.

Нежелательная информация воспринимается зачастую как элемент информационного воздействия. Информационный эффект *R* от информационного воздействия трактуется как основной поражающий фактор информационной войны. Он представляет собой воздействие информационным потоком на информационную систему как на объект атаки. Объектом атаки может являться отдельный человек, коллектив людей (некоторая организация) и даже государство в целом. Естественно, что эффект может быть как отрицательный, так и положительный. Однако воздействие с положительным эффектом мы рассматривать не будем. Поэтому будем считать, что цель такого воздействия заключается в достижении негативных структурных и/или функциональных изменений системы за счет приема и обработки этой информации. Формально информационный эффект определяется следующим образом:

$$R = IE\ (IO), \tag{1}$$

где *IE* — функция, определяющая некоторое информационное воздействие, *IO* — информационный объект, *R* — результат воздействия.

Информационный объект (ИО) *IO* есть логически цельный блок информации, представленный в определенной фиксированной форме, который создан и используется в информационной деятельности человека. Формально связь этого понятия с другими понятиями представляется следующим образом: $IO \in I$, то есть ИО является элементом множества всей анализируемой информации *I*.

Использование понятия ИО позволяет предложить другой вариант определения нежелательной информации, основывающийся на анализе информационных признаков ИО. Обозначим всю информация в сети Интернет как *Int*. Положим, что множество *Int* содержит опасную информацию *RI* (*Risky Information*) и безопасную информацию *SI* (*Safe Information*). Между этими понятиями справедливо следующее равенство:

$$Int = RI \cup SI\ . \tag{2}$$

Нежелательная информация (Inappropriate Information, II) есть отдельный ИО и/или совокупность объектов в сети Интернет, содержащих признаки, попадающие под категории ненужности, негодности. Наиболее ярким примером здесь являются ИО, фильтруемые системой

родительского контроля. Кроме того, будем относить к категории нежелательной информации также сомнительную и вредоносную информацию, упоминания которой иногда встречаются в литературе.

Сомнительная информация (*Dubious Information, DI*) есть отдельный ИО и/или совокупность объектов в сети Интернет, содержащие признаки, попадающие под категории опасности. Например, фишинговый сайт, недоверенный ресурс, ресурс с низкой репутацией. Объект, содержащий ложную (фейковую) информацию или дезинформацию, также относится к данному типу.

Вредоносная информация (Harmful information, HI) есть отдельный ИО и/или совокупность объектов в сети Интернет, содержащие признаки, по которым информация запрещена к распространению. Например, под эту категорию попадает информация, включенная в федеральный список экстремистских материалов, конфиденциальная информация, персональные данные и т.д.

Используя введенные обозначения для различных типов информации, можно сформировать между ними следующие соотношения:

$$II \subseteq RI\ ,\ \left(DI \cup HI\right) \subseteq II\ . \qquad (3)$$

Следует отметить, что в общем случае пересечение множеств *DI* и *HI* не является пустым множеством, то есть один и тот же ИО может быть отнесен как к *DI*, так и к *HI*.

Общая архитектура системы. Предлагаемая интеллектуальная система аналитической обработки цифрового сетевого контента (ИСАОЦСК) имеет общую архитектуру, показанную на рисунке 1. Архитектура содержит три уровня:

1) сбора и предварительной обработки данных о безопасности сетевых ИО;

2) оценивания смыслового содержания ИО;

3) выработки мер противодействия выявленной в ИО нежелательной информации.

Исходными данными для такой системы являются информационные объекты сети Интернет и социальных сетей. Результаты, полученные с помощью ИСАОЦСК, используются пользователями (администраторами безопасности), отвечающими за защиту от нежелательной информации. Потребителями результатов функционирования ИСАОЦСК являются регуляторы телекоммуникационного сектора.

На первом уровне архитектуры ИСАОЦСК располагаются распределенные сканеры сетевых ИО. Их задача заключается в сборе ИО,

формировании облака тегов (меток, ярлыков, хештегов, ключевых слов) и приоритезации ИО.

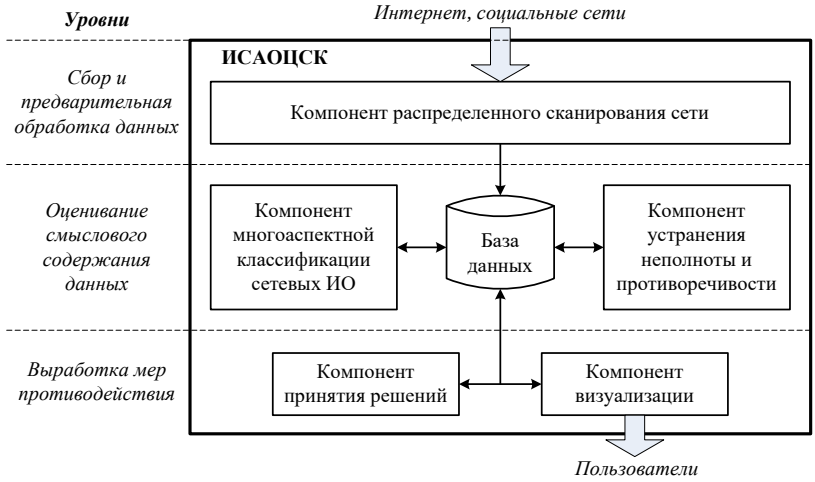

Рис. 1. Общая архитектура ИСАОЦСК

На втором уровне находятся база данных веб-контента, содержащая всю собираемую и обрабатываемую информацию об ИО, а также компонент многоаспектной классификации сетевых ИО и компонент устранения неполноты и противоречивости результатов классификации. На этом уровне исходные данные для классификаторов, сформированные с помощью распределенных сканеров, подвергаются дополнительной обработке с целью устранения неоднозначности (нечеткости) и недостоверности (недостаточности, неполноты).

На третьем уровне располагаются компонент принятий решений, осуществляющий выбор мер противодействия выявленной нежелательной информации, и компонент визуализации результатов работы системы.

Обобщенный алгоритм функционирования ИСАОЦСК можно описать следующим образом.

Шаг 1. Сбор данных о сайтах, потенциально содержащих нежелательную информацию (с помощью компонента распределенного сканирования). Помещение их на хранение в базу данных.

Шаг 2. Выявление и классификация нежелательной информации (компонент многоаспектной классификации). Если классификация про-

шла с высокой точностью, то переход на шаг 4. Иначе — принятие решения о необходимости устранения неполноты и противоречивости собранных данных.

Шаг 3. Функционирование компонента устранения неполноты и противоречивости собранных данных. Переход на шаг 2.

Шаг 4. Выработка и выбор мер противодействия нежелательной информации (компонент принятия решений).

Шаг 5. Визуальное оформление промежуточных и окончательных решений ИСАОЦСК (компонент визуализации).

Рассмотрим решения по построению и функционированию выше указанных компонентов ИСАОЦСК. Компонент визуализации, в силу его специфики, рассматривать не будем.

4. Решения по построению и функционированию компонентов системы. Главная особенность сетевого ИО (СИО), отличающая его от обычного электронного документа, состоит в наличии сложной иерархической структуры. Самым распространенным примером СИО является веб-страница, которая представляет собой набор текстовых файлов, размеченных на языке HTML. Использование HTML позволяет форматировать текст, различать в нём функциональные элементы, создавать гиперссылки и вставлять в отображаемую страницу изображения, звукозаписи и другие мультимедийные элементы. Содержимое веб-страницы называется контентом.

Основная задача компонента распределенного сканирования сети заключается в сборе информации о веб-страницах и предварительной обработке контента. Предлагается использовать в ИСАОЦСК комплекс распределенных интеллектуальных сканеров (КРИС), выполняющих параллельный анализ СИО. Подобный подход подразумевает гибкое масштабирование. Каждый сканер располагается на отдельном хосте и самостоятельно выполняет операции сбора и предобработки сетевого контента.

Исходными данными для распределенного сканирования сети является известное конечное множество X, которое включает в себя n сетевых адресов веб-страниц URL:

$$X = (URL_1, URL_2, \ldots, URL_n). \quad (4)$$

Данное множество распределяется между интеллектуальными сканерами. Каждый сканер загружает результаты сбора и предобработки представленного ему множества объектов в локальное временное хранилище. Далее информация агрегируется в общую базу данных веб-контента.

Архитектура компонента распределенного сканирования сети представлена на рисунке 2.

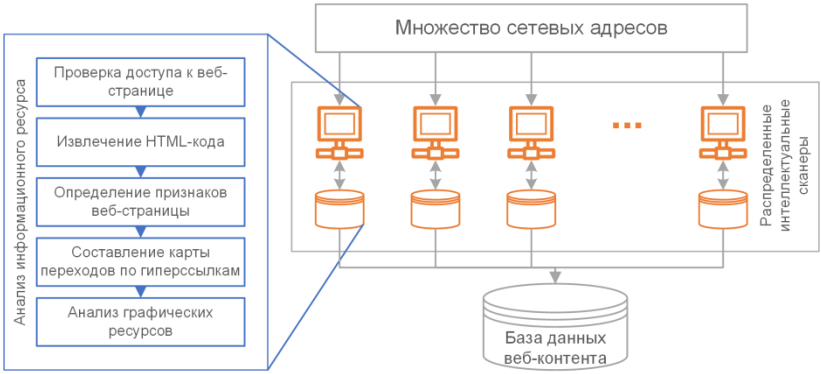

Рис. 2. Компонент распределенного сканирования сети

Для каждого СИО проверяется доступность контента по указанному адресу и присваивается соответствующий статус. При наличии доступа к веб-странице производится загрузка СИО в виде HTML-кода с указанием времени загрузки и его графических ресурсов (изображений, логотипов, элементов фона и т.п.). Формируется множество X', которое содержит HTML-коды для m доступных веб-страниц:

$$X' = (<URL_1, HTML_1>, \ldots, <URL_m, HTML_m>). \quad (5)$$

В случае отсутствия доступа к веб-странице фиксируется код и текст полученной ошибки с указанием времени попытки подключения. Причиной отсутствия доступа может быть устаревание, удаление или некорректный адрес СИО. Также код ошибки может указывать на проблемы с подключением у домена веб-страницы или у сетевого сканера. В этом случае адрес веб-страницы помещается в очередь на повторную проверку.

HTML-код веб-страницы используется для извлечения признаков виде конечного множества X'' размерностью m:

$$X'' = (<URL_1, a_{11},\ldots,a_{k1}>, \ldots, <URL_m, a_{1m},\ldots,a_{km}>), \quad (6)$$

где a_{ij} — i-й признак ($i = 1, \ldots, k$) j-ой веб-страницы ($j = 1, \ldots, m$).

В качестве признаков веб-страниц предлагается использовать:

– тип контента (текст, изображение, видео, аудио и т.д.);

– размер контента;

– текстовое содержимое;

– язык текстового содержимого;

– длину текстового содержимого;

– количество гиперссылок на внутренние СИО (принадлежащие к тому же домену);

– количество гиперссылок на внешние СИО;

– количество графических ресурсов.

Текстовое содержимое СИО представляется в виде описания блоков текста с их частотной характеристикой. Информация о взаимосвязи нескольких СИО отображается в виде карты переходов. Для каждого i-го корневого СИО определяется следующее множество:

$$H_i = (<h_{i1}, d_{i1}, l_{i1}>, \ldots, <h_{is}, d_{is}, l_{is}>), \quad (7)$$

где h_{ij} ($j = 1, \ldots, s$) — сетевой адрес j-го стороннего СИО, извлекаемого из гиперссылок на веб-странице i-го корневого СИО; d_{ij} — глубина взаимосвязи двух СИО, обозначающая число переходов от корневого СИО до текущего; l_{ij} — показатель локальности, определяющий, связан корневой СИО с внутренним ресурсом (принадлежащим тому же домену) или внешним (принадлежащим стороннему домену).

В процессе сбора данных о СИО происходит выгрузка и последующий анализ изображений, содержащихся в HTML-коде и в таблице стилей. Стили хранятся в отдельном CSS-файле, который может быть использован для любых информационных ресурсов одного домена. Предлагается выделять следующие признаки графических ресурсов:

– сетевой адрес изображения;

– сетевой адрес домена, на котором хранится изображение;

– наименование;

– графический формат файла (JPEG, JPG, PNG, GIF и т.д.);

– цветовую модель (RGB, RGBA, CMYK);

– ширину изображения в пикселях;

– длину изображения в пикселях;

– список основных цветов изображения (центроидов цветовых кластеров) и соответствующий им процент принадлежащих пикселей в формате {«код цвета» : «процент пикселей»};

– текст на изображении (если он присутствует) в формате {«код языка» : «текст»}.

Следует отметить, что на текущем этапе разработки ИСАОЦСК в состав признаков текстовых и графических ресурсов пока еще не входят признаки, характеризующие интерпретацию ИО. В результате возможны ложноположительные ошибки, при которых к категории нежелательной информации могут быть отнесены, например, сайты с наборами данных, содержащими примеры сетевых атак. Охват признаков интерпретации ИО рассматривается как направление дальнейших исследований.

Таким образом, компонент распределенного сканирования сети реализует следующие функции: (1) обнаружение и загрузка веб-контента для заданного множества сетевых адресов; (2) структурная категоризация СИО; (3) вычисление частотных характеристик для структурных элементов веб-страниц; (4) построение карт переходов СИО.

Компонент многоаспектной классификации сетевых ИО включает в свой состав четыре модуля:

1) модуль фильтрации содержимого СИО;
2) модуль извлечения признаков из СИО;
3) модуль предобработки признаков СИО и построения обучающих выборок;
4) модуль классификации СИО.

Предназначение модуля фильтрации содержимого СИО заключается в удалении знаков препинания, а также тех единиц речи, представленных в виде отдельных слов и словосочетаний, которые не влияют на смысловое наполнение текста. К таким словам относятся местоимения, предлоги, цифры, артикли, модальные глаголы и союзы, а также те слова, которые являются свойственными одновременно для нескольких классов и используемыми в различных контекстах (например, «теперь», «тогда», «только»). Кроме этого, из рассмотрения исключаются данные, помещенные в секции комментариев, поскольку они не видны конечному пользователю.

Модуль извлечения признаков из СИО построен на основе DOM (Document Object Model) парсинга, что позволяет быстро и удобно извлекать из исходной html-страницы текст, заключенный в искомый тег. В данном модуле реализована поддержка вычисления параметров html-страницы с использованием трех типов исходных данных: структуры документа, текстового содержимого, а также URL-строки. В таблице 1 перечислены наименования html-тегов, частоты встречаемости которых формируют параметры вектора признаков в соответствии с типом исходных данных «Структура документа».

При формировании параметров, соответствующих типу исходных данных «Текстовое содержимое», применялся подход на основе

«мешка слов». Исходный документ $T = \{w_1, \ldots, w_N\}$ представляется последовательностью слов. Он преобразуется в список пар $L=\{(w_1, Q(w_1, T)), \ldots, (w_M, Q(w_M, T))\} = \{(w_1, q_1), \ldots, (w_M, q_M)\}$, где элементы $\{w_1, \ldots, w_M\} = D \subset T$ являются уникальными словами ($M \leq N$) (множество D называется словарем), а элементы $\{q_1 = Q(w_1, T), \ldots, q_M = Q(w_M, T)\}$ отражают абсолютные частоты появления соответствующих слов в документе T (Q — функция, возвращающая число вхождений слова, представленного первым аргументом, в документ, представленный вторым аргументом).

Таблица 1. Описание html-тегов, извлекаемых из сетевого ИО

№	html-тег	Описание html-тега
1	<b>	Жирное выделение текста
2	<dt>	Создание элемента в списке определений
3		Разбиение документа на фрагменты
4	<h1>	Задание заголовка первого уровня
5	<h2>	Задание заголовка второго уровня
6	<h3>	Задание заголовка третьего уровня
7	<h4>	Задание заголовка четвертого уровня
8	<h5>	Задание заголовка пятого уровня
9	<h6>	Задание заголовка шестого уровня
10	<link>	Задание связи с внешним ресурсом
11	<a>	Создание ссылки
12	<form>	Задание формы
13		Создание элемента маркированного или нумерованного списка
14	<i>	Курсивное выделение текста
15	<p>	Выделение абзаца

Оценка семантической схожести двух документов сводится к вычислению количества слов, которые одновременно встречаются в этих документах. Чем меньшее количество общих слов содержится в обоих документах, тем ниже их уровень семантической схожести. Следует отметить, что с увеличением объема документа модель «мешка слов» становится особенно чувствительной к сравнению документов. В этом случае для уменьшения временных затрат рассматривается множество $L' = \{(w, q) \mid (w, q) \in L \wedge q > h \geq 0\}$ вместо L, что позволяет игнорировать редко встречающиеся слова, частота вхождения которых в документ не превосходит заранее заданного числового порога h.

Среди ограничений, присущих данной модели, следует отметить невозможность учета контекста, в котором может использоваться то или иное слово. Для устранения этого недостатка прибегают к построе-

нию матрицы семантических связей [34]. С этой целью разработан алгоритм вычисления семантической схожести наиболее употребительных слов на основе модели word2vec [35] (рис. 3).

В данном алгоритме для каждого класса S_i формируется документ R, объединяющий все документы этого класса. Для документа R вычисляется функция *find_common_words* (R, d), которая извлекает d слов, наиболее часто встречающихся внутри R. Результат применения данной функции обозначается как W. После этого внутри каждого документа класса S_i находятся c наиболее употребительных слов, для которых вычисляется семантическая схожесть с каждым словом из набора W. С этой целью применяется функция *word2vec*. Полученные результаты записываются в виде компонентов вектора признаков, соответствующих типу исходных данных «Текстовое содержимое».

Входные данные: $\Omega = \{S_1, S_2, S_3, ...\} = \{arts, business, computers, ...\}$ — классы документов
$\Phi = \left\{\{T_{ij}\}_{j=1}^{N_i}\right\}_{i=1}^{\#\Omega}$ — набор документов с разделением по классам
$c = 3$ — количество наиболее употребительных слов внутри документа
$d = 5$ — количество наиболее употребительных слов внутри совокупности документов, принадлежащих одному и тому же классу

Выходные данные: $\Psi = \left\{\{F_{ij}\}_{j=1}^{N_i}\right\}_{i=1}^{\#\Omega}$ — набор признаков документов

для каждого $i \in \{1, \ldots, \#\Omega\}$ **выполнять**
  $R \coloneqq \varnothing$
  для каждого $j \in \{1, \ldots, N_i\}$ **выполнять**
    $R \coloneqq R \cup T_{ij}$
  $W \coloneqq find\_common\_words(R, d)$
  для каждого $j \in \{1, \ldots, N_i\}$ **выполнять**
    $F_{ij} \coloneqq \varnothing$
    для каждого $v \in find\_common\_words(T_{ij}, c)$ **выполнять**
      для каждого $w \in W$ **выполнять**
        $F_{ij} \coloneqq F_{ij} \cup \{word2vec(v, w)\}$

Рис. 3. Алгоритм вычисления семантической схожести наиболее употребительных слов

Если в качестве исходных данных используются URL-строки, то тогда для формирования признаков СИО проверяется признак вхождения в эти строки десяти наиболее употребительных слов, характерных для каждой категории.

В модуле предобработки признаков СИО и построения обучающих выборок реализована поддержка минимаксной нормализации и разбиения исходной выборки на тестовую и обучающую части.

В функционировании модуля классификации СИО выделяются два режима: режим обучения и режим анализа. В первом режиме выполняется настройка классификаторов путем итеративного предъявления на их вход последовательностей обучающих векторов и последующей корректировки внутренних параметров классификаторов. Во втором режиме осуществляется выделение класса анализируемого ИО, включая характер и степень вредоносности сетевого контента.

Компонент устранения неполноты и противоречивости результатов классификации СИО предназначен для устранения неопределенности (нечеткости, неполноты и противоречивости) оценки СИО. Такая оценка почти во всех случаях осуществляется в условиях неопределенности исходных данных — измеряемых, моделируемых или наблюдаемых в шумах атрибутов СИО (текстовых, графических, числовых, булевых, ординальных, номинальных и т.д.). Полагается, что основным источником такой неопределенности является «Текстовое содержимое» веб-страниц. При этом доминирующими видами неопределенности являются неоднозначность (нечеткость, противоречивость) и недостаточность (неполнота) исходных данных.

Неопределенность оценки СИО вызвана нестационарностью поступления информации, нечеткостью, неполнотой и противоречивостью идентификации признаков такой информации, динамикой функционирования системы защиты и воздействиями дестабилизирующих (зачастую антагонистических) факторов внешней среды, а также неопределенностью целей и несогласованностью задач обнаружения и противодействия нежелательной информации.

Общий алгоритм функционирования компонента устранения неполноты и противоречивости результатов классификации СИО опирается на модели и механизмы устранения неопределенности оценки признаков нежелательной информации (в интересах ее обнаружения и противодействия ей). Он использует методы обработки нечетких, неполных и противоречивых знаний и включает два ключевых этапа.

В основе *первого этапа*, ориентированного на устранение неопределенности классификации СИО на основе нечетких множеств [31, 32], лежит механизм поддержки принятия решения по включению (либо не включению) нечетко заданных признаков информации, циркулирующей в цифровом сетевом контенте, во множество признаков нежелательной информации. Иными словами, если в СИО наличие,

объем и номенклатура (уровень опасности) признаков какой либо сомнительной информации превышает допустимый порог (α-уровень функции принадлежности), то эта информация оценивается и классифицируется как нежелательная, потенциально вредоносная. При этом субъективная мера уверенности, с которой данная информация принадлежит нечеткому множеству признаков нежелательной информации, задается функциями принадлежности. При этом для объединения нескольких субъективных мер уверенности (мнений нескольких экспертов) используются математические операции дополнения, объединения, пересечения нечетких множеств и операция дизъюнктивного суммирования нечетких множеств.

Потребность устранять нечеткость признаков анализируемой информации возникает на фоне того, что данные признаки фактически определены, сформулированы, но их значения заданы нечетко. Эти значения поступают из множества разнородных источников и могут неоднозначно, с помощью нечетких высказываний (лингвистических термов типа «много», «сильно» и т.п.), указывать, например, на меру уверенности, с которой конкретный контент анализируемой веб-страницы принадлежит (либо не принадлежит) множеству признаков нежелательной информации.

В основе *второго этапа*, ориентированного на устранение неопределенности оценки и категоризации на основе искусственной нейронной сети (ИНС), лежит нейросетевая модель [28, 30] поиска и прогнозирования неполно и противоречиво заданных признаков анализируемой информации. Эта модель позволяет осуществлять поиск взаимосвязей между признаками и обоснованно включать (либо не включать) неполно и противоречиво заданные признаки во множество признаков нежелательной информации. Иными словами, если есть хотя бы один признак, гарантированно включаемый в состав множества признаков нежелательной информации, то можно построить такой вектор входных признаков, который учитывает неполные и противоречивые взаимосвязи всех признаков (по мнению экспертов). Тогда с помощью ИНС можно получить выходной вектор признаков с коэффициентами, характеризующими их вес (уровень опасности) и, в свою очередь, оценить и классифицировать эту информацию как нежелательную.

Итоговым результатом работы компонента устранения неполноты и противоречивости является окончательно сформулированная система признаков СИО, однозначно определяющая принадлежность (либо не принадлежность) конкретной информации к нежелательной, с учетом устранения неопределенности в рамках моделей и алгоритмов обработки нечетких, неполных и противоречивых знаний. При этом

предложенный подход позволяет работать с обоими видами неопределенности раздельно, что сокращает объем производимых вычислений и приводит к увеличению быстродействия ИСАОЦСК для защиты от нежелательной информации.

Компонент принятия решений по противодействию нежелательной информации использует в своей работе теорию принятия решений, включая методы многокритериальной оптимизации. На вход компонента поступают: (1) нежелательные СИО; (2) доступные контрмеры. Основными этапами работы компонента являются: (1) создание моделей СИО, информационной системы, противодействия и процесса противодействия; (2) выбор контрмер. На выходе компонента формируется набор выбранных мер противодействия.

Информационной системой, в которой реализовано противодействие, является Интернет. Модель информационной системы задается следующим образом: $IS = (IO, IC)$, где *IO* — сетевые ИО, *IC* — связывающие их коммуникационные средства.

Модель СИО определяется следующим образом:

$$IO = <size, role, hltype, type, state, ioaud, saud>, \quad (8)$$

где *size* — размер СИО, может иметь значения из множества {*sm*, *mi*, *la*}, *sm* — «малый», *mi* — «средний», *la* — «большой»;

role — роль СИО, может иметь значения из множества {*s*, *r*, *u*}, *s* — «отправитель», *r* — «получатель», *u* — «пользователь»;

hltype — абстрактный тип СИО, принимает значение *h*, если СИО является нежелательным, и *n* — в противном случае;

type — детальный тип СИО, может принимать значения из множества {*ter*, *hea*, *por*, *dru*, *cru*, *none*}, *ter* — СИО, содержащий призывы к терроризму и экстремизму; *hea* — СИО, содержащий информацию, вредную для здоровья людей (особенно детей), морального и духовного развития; *por* — СИО с пропагандой порнографии; *dru* — СИО, содержащий информацию о путях распространения наркотиков и призывы к суициду; *cru* — СИО, содержащий призывы к насилию (войне); *none* — СИО не является нежелательным (*hltype* равен *n*);

state — состояние компрометации СИО, может принимать значения *compr*, если СИО скомпрометирован вредоносной информацией, и *nonc* — если не скомпрометирован;

ioaud — аудитория СИО, представляющая собой массив ссылок, которые связаны с отправителем посредством сообщений и которые являются получателями объектов (может быть нулевым);

saud — вещественное число (при наличии счетчика посетителей СИО) или экспертная оценка субъектов, являющихся получателями СИО (может быть 0).

Модель контрмеры *rm* из множества контрмер *RM* задается в следующем виде:

$$rm = <rm\_class, rm\_type, rm\_cost, rm\_role, rm\_ef, rm\_cd>, \quad (9)$$

где *rm_class* — класс контрмеры (барьер, маскировка, информирование или принуждение); *rm_type* — размер СИО (малый, средний или большой); *rm_cost* — стоимость контрмеры; *rm_role* — роль СИО; *rm_ef* — эффективность контрмеры; *rm_cd* — побочный ущерб от реализации контрмер.

Модель контрмеры используется для определения модели противодействия. Противодействие влияет на состояние информационной системы: {*IO*, *IC*} становится $\{IO^l, IC^l\}$, где *l* — номер контрмеры. Для *j* информационных объектов из IO^l ($j = 0,\ldots,N$, где *N* – номер элемента в *ioaud* нежелательного СИО, на которого воздействует контрмера), СИО удаляется, или модифицируются следующие их параметры: *role* принимает значение *r* или *u*, *hltype* становится равным *n*, *type* становится равным *none*, *state* становится равным *nonc*. Для *d* связей из IC^l ($d = 0,\ldots,D$, где *D* — номер связи между нежелательным СИО и связанным объектом, на который воздействует контрмера) информационная связь удаляется.

Модели (8) и (9) используются для формализации алгоритма противодействия нежелательной информации. Входными данными этого алгоритма являются: размер СИО, роль СИО, абстрактный тип СИО (параметр *hltype*) и детальный тип СИО (параметр *type*). Алгоритм включает две фазы. На первой фазе производится анализ аудитории нежелательной информации. На второй производится анализ и выбор меры противодействия. Для учета аудитории нежелательной информации на первой фазе производится поиск связанных объектов и изменение их состояния на скомпрометированное. Затем, с учетом этих скомпрометированных объектов и их трафика (с помощью счетчиков), вычисляются размер и возраст аудитории.

Анализ контрмеры на второй фазе заключается в вычислении ее эффективности (параметр *rm_ef*) и стоимости (*rm_cost*). Эффективность вычисляется как отношение СИО-получателей, которые не будут скомпрометированы в случае реализации противодействия, к общему количеству получателей. Следует отметить, что при оценке эффективности не учитываются возможные случаи самокомпрометирования, когда получатель попадает под действия средств защиты, например, ловушек;

учет таких случаев относится к дальнейшим исследованиям. Стоимость задается экспертами вручную. Кроме того, учитываются класс средств противодействия (который выбирается в зависимости от типа вредоносной информации) и размер информационного объекта.

Для выбора контрмеры на второй фазе используются предварительно сформированные правила. В качестве примера приведем одно из правил выбора контрмеры, используемое в разработанном алгоритме и основанное на параметрах моделей (8) и (9): «если *role = s* и *size = sm* и *type = ter* и размер аудитории меньше 3000 и возраст аудитории больше 18, то тогда выбрать контрмеры, у которых *rm_class* равно *disguise* или *informing* либо *rm_type* равно *small*».

5. Реализация и экспериментальная оценка системы. Для проведения экспериментальной оценки системы с помощью компонента распределенного сетевого сканирования был сформирован набор данных, содержащий категорированный веб-контент. Интеллектуальные сканеры были размещены на четырех хостах. Характеристика вычислительной базы сканеров приведена в таблице 2.

Таблица 2. Характеристика вычислительной базы сканеров

№	Процессор	Тактовая частота (ГГц)	ОЗУ (Гб)	ОС
1	Intel Core i5-8250U	1,8	8	Windows 10
2	Intel Core i7-7700HQ	2,8	12	Windows 10
3	Inter Core i7-8665U	1,9	16	Windows 10
4	AMD Ryzen 5 3500U	2,1	8	Windows 10

Исходный набор данных, представляющий множество сетевых адресов, получен из общедоступных категорированных списков веб-страниц, включающих в себя URLBlacklist, MESD blacklists [36], Shallalist [37] и DMOZ [38]. В нем содержатся адреса веб-страниц, маркированных 23 категориями контента, в том числе относящимися к нежелательной информации (таблица 3).

Для создания сбалансированного набора множества сетевых адресов была проверена доступность веб-ресурсов различных категорий, так как открытые категорированные списки могут содержать много устаревших данных. Для проведения эксперимента было введено ограничение на 2000 доступных веб-страниц для каждой категории, за исключением «алкоголь» и «политика», для которых в исходных списках содержится меньшее количество маркированных данных. Итоговое экспериментальное множество адресов веб-страниц включает в себя 44 866 URL, информация о которых записана в общую базу данных. Для каждой веб-страницы определены следующие признаки:

– *id* – идентификатор веб-страницы;

– *url* – сетевой адрес веб-страницы;

– *category* – категория веб-страницы;

– *domain* – домен веб-страницы;

– *status* – доступность веб-страницы;

– *content_type* – тип контента;

– *content_length* – размер контента;

– *language* – язык текстового содержимого веб-страницы;

– *text_length* – размер текстового содержимого веб-страницы;

– *local_hyperlinks_count* – количество гиперссылок на ресурсы того же домена, содержащиеся на веб-странице.

– *external_hyperlinks_count* – количество гиперссылок на внешние ресурсы, содержащиеся на веб-странице.

Таблица 3. Категории веб-контента

No.	Название категории	Нежелательная информация	Число веб-страниц
1	Для взрослых (adult)	✓	2000
2	Агрессия (aggression)	✓	2000
3	Алкоголь (alcohol)	✓	1386
4	Искусство (arts)	×	2000
5	Бизнес (business)	×	2000
6	Компьютеры (computers)	×	2000
7	Сервисы знакомств (dating)	✓	2000
8	Наркотики (drugs)	✓	2000
9	Игры (games)	×	2000
10	Азартные игры (gamling)	✓	2000
11	Хакерство (hacking)	✓	2000
12	Медицина (health)	×	2000
13	Дом (home)	×	2000
14	Для детей (kids)	×	2000
15	Новости (news)	×	2000
16	Политика (politics)	×	1480
17	Досуг (recreation)	×	2000
18	Ссылки (reference)	×	2000
19	Религия (religion)	✓	2000
20	Наука (science)	×	2000
21	Шоппинг (shopping)	×	2000
22	Общество (society)	×	2000
23	Спорт (sports)	×	2000

Пример отображения перечисленных признаков в базе данных представлен на рисунке 4.

id	url	category	domain	status	content_type	content_length	language	text_length	local_hyperlinks_count	external_hyperlinks_count
2771887	http://feeds.fee…	Arts	feedb…	OK	text/xml; char…	21732	en	21692	1	0
2771893	http://www.npr.…	Arts	npr.org	OK	text/xml;chars…	4919	en	19181	1	0
2771895	http://www.artis…	Arts	artistdi…	OK	text/html; char…	73785	en	73735	89	11
2771896	http://www.curli…	Arts	curlio.…	OK	text/html	49610	en	49610	176	1
2771897	http://www.cbc.…	Arts	cbc.ca	OK	text/html; char…	13570	en	44968	41	15
2771898	http://www.musi…	Arts	musicn…	OK	text/html; char…	88598	en	88598	1	63
2771899	http://www.xfm.…	Arts	xfm.co…	OK	text/html; char…	166234	en	166216	257	29
2771900	http://www.mi2…	Arts	mi2n.com	OK	text/html; char…	159297	en	159136	34	22
2771901	http://www.ukm…	Arts	ukmusi…	OK	text/html; char…	31368	en	31367	1	42
2771903	http://www.xs4…	Arts	xs4all.nl	OK	text/html	9836	en	24186	8	1
2771904	http://www.anti…	Arts	antimu…	OK	text/html	71920	en	71920	54	207
2771906	http://nextbigthi…	Arts	nextbi…	OK	text/html; char…	1389359	en	1389358	3981	104

Рис. 4. Признаки сетевых информационных объектов

Общий размер извлеченных HTML-кодов веб-страниц составляет 3,3 Гбайта. На рисунке 5 представлен размер текстового контента для каждой категории: количество строк таблицы в базе данных (*Rows*), средняя длина строки (*Avg Row Length*) и размер данных (*Data Length*). Объем всей базы данных составляет 8,26 Гбайта.

Name	Engine	Version	Row Format	Rows	Avg Row Length	Data Length
text_adult	InnoDB	10	Dynamic	546465	196	102.6 MiB
text_aggression	InnoDB	10	Dynamic	632059	291	175.6 MiB
text_alcohol	InnoDB	10	Dynamic	508918	320	155.6 MiB
text_arts	InnoDB	10	Dynamic	381114	356	129.6 MiB
text_business	InnoDB	10	Dynamic	385600	289	106.6 MiB
text_computers	InnoDB	10	Dynamic	507244	478	231.7 MiB
text_dating	InnoDB	10	Dynamic	606009	271	156.6 MiB
text_drugs	InnoDB	10	Dynamic	531781	257	130.6 MiB
text_gambling	InnoDB	10	Dynamic	228279	576	125.6 MiB
text_games	InnoDB	10	Dynamic	252228	405	97.6 MiB
text_hacking	InnoDB	10	Dynamic	225028	347	74.6 MiB
text_health	InnoDB	10	Dynamic	573686	289	158.6 MiB
text_home	InnoDB	10	Dynamic	591596	311	175.6 MiB
text_kids	InnoDB	10	Dynamic	380076	1999	724.6 MiB
text_news	InnoDB	10	Dynamic	548568	381	199.7 MiB
text_politics	InnoDB	10	Dynamic	369384	481	169.6 MiB
text_recreation	InnoDB	10	Dynamic	387911	371	137.6 MiB
text_reference	InnoDB	10	Dynamic	576658	332	182.6 MiB
text_religion	InnoDB	10	Dynamic	513433	303	148.6 MiB
text_science	InnoDB	10	Dynamic	361502	1234	425.6 MiB
text_shopping	InnoDB	10	Dynamic	462285	244	107.6 MiB
text_society	InnoDB	10	Dynamic	494495	330	155.6 MiB
text_sports	InnoDB	10	Dynamic	440567	306	128.6 MiB

Рис. 5. Размер текстового содержимого веб-страниц

Следует отметить, что анализ графических ресурсов веб-страниц выходил за рамки проведенных исследований. Основное внимание уделялось признакам СИО и их текстовому содержимому.

Компонент многоаспектной классификации СИО оценивался по четырем показателям: достоверности (*accuracy*), точности (*precision*), полноте (*recall*) и F-мере (*F-measure*). Использовалось семь классификаторов: простой мешок слов; взвешенный мешок слов; непрерывный

мешок слов, классификатор skip-gram, сверточная нейронная сеть, классификатор fastText [39] и случайный классификатор.

В таблице 4 представлены указанные показатели, вычисленные для каждого классификатора.

Таблица 4. Показатели достоверности, точности, полноты и F-меры, вычисленные для трехблочной кросс-валидации

Классификатор / Показатель	Простой мешок слов	Взвешенный мешок слов	Непрерывный мешок слов	Классификатор skip-gram	Сверточная нейронная сеть	Классификатор fastText	Случайный классификатор
Достоверность	63,16%	64,69%	16,02%	22,53%	25,81%	84,15%	5,28%
Точность	63,06%	66,2%	26,66%	36,8%	29,56%	80,48%	5,28%
Полнота	58,63%	61,04%	12,21%	18,53%	20,42%	78,83%	5,28%
F-мера	60,76%	63,52%	16,75%	24,65%	24,15%	79,65%	5,28%

Показатели точности, полноты и F-меры вычислялись для каждого класса в отдельности. Поэтому значения соответствующих показателей были усреднены по всем классам. Случайный классификатор с равной вероятностью генерировал числовую метку одного из 19 классов нежелательной информации. Эксперимент для этого классификатора проводился 100 раз, что позволило достаточно точно приблизить экспериментальные (5,28%) и теоретические (1/19 * 100 % ≈ 5,26 %) оценки показателей. Наилучшие результаты показал классификатор fastText (его достоверность равна 84,15%). На рисунке 6 представлены детальные показатели точности, полноты и F-меры, вычисленные для этого классификатора с использованием трехблочной кросс-валидации.

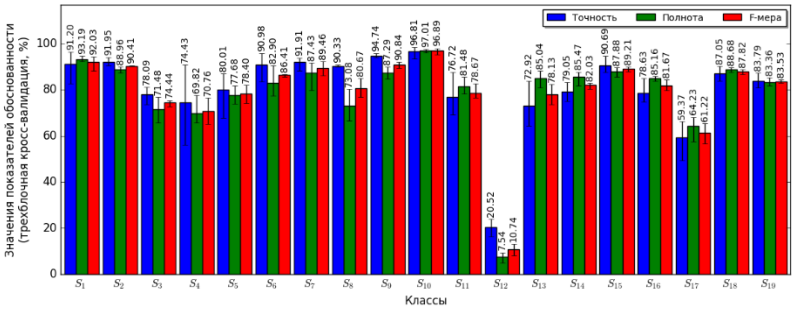

Рис. 6. Значения показателей точности, полноты и F-меры, вычисленные для классификатора fastText и трехблочной кросс-валидации

На рисунке 7 показана зависимость количества ошибок, показателя достоверности и среднеквадратичной ошибки сверточной нейронной сети от номера эпохи обучения. В экспериментах использовалась нейронная сеть с двумя слоями свертки (с функцией активации ReLU) и следующим за каждым из них слоем субдискретизации (с функцией max). Достоверность на обучающей выборке для нейронной сети не превосходит 32,55%. Этим в полной мере объясняется низкое значение соответствующего показателя (25,81%) на тестовой выборке.

Таким образом, эксперименты показали, что максимальная эффективность обнаружения нежелательной информации, определяемая показателями достоверности, точности, полноты и F-меры, достигается в ИСАОЦСК при использовании классификатора fastText. Однако исследования в области повышения эффективности будут продолжаться и дальше.

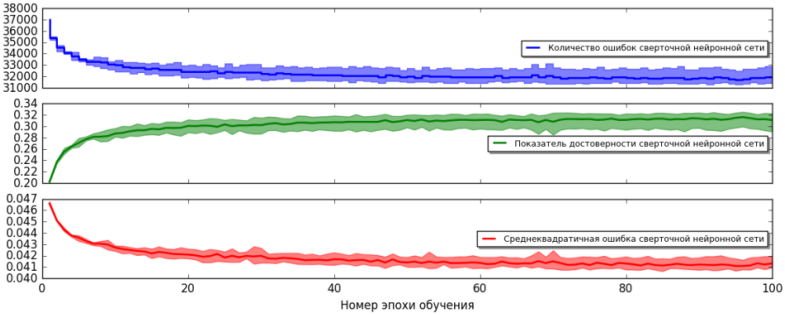

Рис. 7. Зависимость количества ошибок, показателя достоверности и среднеквадратичной ошибки сверточной нейронной сети от номера эпохи обучения

Здесь представляет интерес подход, основанный на комплексировании различных классификаторов, предложенный в [40]. При таком подходе итоговая эффективность классификации становится выше, чем эффективность отдельного классификатора, подлежащего комплексированию. Для того чтобы реализовать этот подход в ИСАОЦСК, необходимо проверить работу различных схем комплексирования (мажоритарной, взвешенной и т.д.). Авторы относят это к направлениям дальнейших исследований.

Рассмотрим теперь примеры реализации компонента устранения неполноты и противоречивости и связанного с ним компонента принятия решений. Для реализации первого этапа, ориентированного на использование нечетких множеств, используются методы обработки не-

четких знаний (вычисления дизъюнктивной суммы). Положим, что задан исходный состав множества нечетко заданных признаков и сформулированы нечетко заданные мнения экспертов — начальные функции принадлежности нечетких множеств, характеризующие предварительный, нечетко заданный состав множества признаков СИО:

$$\tilde{\tilde{J}} = [\Delta\tilde{\tilde{j}}_{\text{терр}} \big| \mu(\Delta\tilde{j}_{\text{терр}});\ \Delta\tilde{\tilde{j}}_{\text{дет}} \big| \mu(\Delta\tilde{j}_{\text{дет}});\ \Delta\tilde{\tilde{j}}_{\text{порн}} \big| \mu(\Delta\tilde{j}_{\text{порн}});\ \Delta\tilde{\tilde{j}}_{\text{нарк}} \big| \mu(\Delta\tilde{j}_{\text{нарк}});\ \Delta\tilde{\tilde{j}}_{\text{войн}} \big| \mu(\Delta\tilde{j}_{\text{войн}})]^{\text{т}}, \quad (10)$$

где $\Delta\tilde{j}_{\text{терр}}(k)$ — признак СИО, характеризующий аномальное отклонение в трафике среднего количества информации, содержащей публичные призывы к осуществлению террористической и экстремистской деятельности; $\Delta\tilde{j}_{\text{дет}}(k)$ — признак СИО, характеризующий аномальное отклонение в контенте среднего количества информации, причиняющей вред здоровью, нравственному и духовному развитию людей (особенно детей); $\Delta\tilde{j}_{\text{порн}}(k)$ — признак СИО, характеризующий аномальное отклонение среднего количества информации, нацеленной на пропаганду порнографии; $\Delta j_{\text{нарк}}(k)$ — признак СИО, характеризующий аномальное отклонение среднего количества информации, содержащей данные о способах разработки, изготовления и использования наркотических средств и совершения самоубийства, а также нецензурную брань, а $\Delta\tilde{j}_{\text{войн}}(k)$ — признак СИО, характеризующий аномальное отклонение в контенте среднего количества прямых призывов к насилию и жестокости (войне), этнической и религиозной ненависти либо вражде; символ μ — функция принадлежности нечеткого множества, принимающая значения от 0 до 1.

Дизъюнктивная сумма двух нечетких множеств $\tilde{X}$ и $\tilde{Y}$, характеризующих мнения первого и второго экспертов о степени принадлежности признаков СИО к множеству опасных признаков, имеет следующий вид:

$$\tilde{X} \oplus \tilde{Y} = (\tilde{X} \cap \bar{\tilde{Y}}) \cup (\bar{\tilde{X}} \cap \tilde{Y}), \quad (11)$$

где $\bar{\tilde{X}}$ и $\bar{\tilde{Y}}$ — дополнения этих нечетких множеств.

Тогда функция принадлежности для *i*-го признака имеет вид:

$$\forall j_i \in \overline{1,\ldots,5}: \quad \begin{array}{l} \mu_{\tilde{X}\oplus\tilde{\bar{Y}}}(j_i) = \max\{[\min\{\mu_{\tilde{X}}(j_i), 1-\mu_{\tilde{Y}}(j_i)\}]; \\ \quad [\min\{1-\mu_{\tilde{X}}(j_i), \mu_{\tilde{Y}}(j_i)\}]\}. \end{array}, \qquad (12)$$

Мнение первого (X) из двух экспертов об оценке и категоризации каждого признака как признаков нежелательной информации можно представить в виде следующего нечеткого множества:

$$\tilde{X} = \{\Delta\tilde{j}_{\text{терр}}|\,0{,}3;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}1;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}1;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}2\}.$$

Аналогичное мнение второго (Y) эксперта можно представить в виде следующего нечеткого множества:

$$\tilde{Y} = \{\Delta\tilde{j}_{\text{терр}}|\,0{,}7;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}9;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}4;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}4\}.$$

Для этих нечетких множеств их дополнения равны:

$$\bar{\tilde{X}} = \{\Delta\tilde{j}_{\text{терр}}|\,0{,}7;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}9;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}9;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}8\};$$

$$\bar{\tilde{Y}} = \{\Delta\tilde{j}_{\text{терр}}|\,0{,}3;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}1;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}6;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}6\},$$

а пересечения этих нечетких множеств имеют вид:

$$\tilde{X} \cap \bar{\tilde{Y}} = \{\Delta\tilde{j}_{\text{терр}}|\,0{,}3;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}1;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}1;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}2\};$$

$$\bar{\tilde{X}} \cap \tilde{Y} = \{\Delta\tilde{j}_{\text{терр}}|\,0{,}7;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}9;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}4;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}4\}.$$

В итоге объединение этих нечетких множеств дает следующие конечные результаты дизъюнктивного суммирования, характеризующие совокупное мнение двух экспертов об оценке и категоризации каждого признака как признака нежелательной информации:

$$\tilde{X} \oplus \tilde{Y} = (\tilde{X} \cap \bar{\tilde{Y}}) \cup (\bar{\tilde{X}} \cap \tilde{Y}) =$$
$$= \{\Delta\tilde{j}_{\text{терр}}|\,0{,}7;\ \Delta\tilde{j}_{\text{дет}}|\,0{,}9;\ \Delta\tilde{j}_{\text{порн}}|\,0{,}4;\ \Delta\tilde{j}_{\text{нарк}}|\,0{,}5;\ \Delta\tilde{j}_{\text{войн}}|\,0{,}4\}.$$

В случае, когда экспертов больше двух, формулируется мнение третьего эксперта, итоговое совокупное мнение двух первых экспертов выступает в роли отдельного мнения, и цикл повторяется заново до тех пор, пока не иссякнут эксперты. Тогда получим совокупное, единое мнение экспертов на основе обработки нечетких знаний.

Критерием оценки СИО в рассмотренном случае выступает «граничное», пороговое значение функции принадлежности, описывающей важность (предпочтительность) включения признаков СИО в состав множества признаков нежелательной информации, например, на уровне $\mu^{\text{тр}} \geq 0{,}65$.

Завершающим шагом экспериментальной оценки признаков СИО в условиях неопределенности для данного этапа является отбор конкретных признаков в состав множества признаков нежелательной информации. Графики значений функции принадлежности, описывающей критерий оценки и категоризации в условиях нечеткости, полученные для рассмотренного примера, представлены на рисунке 8.

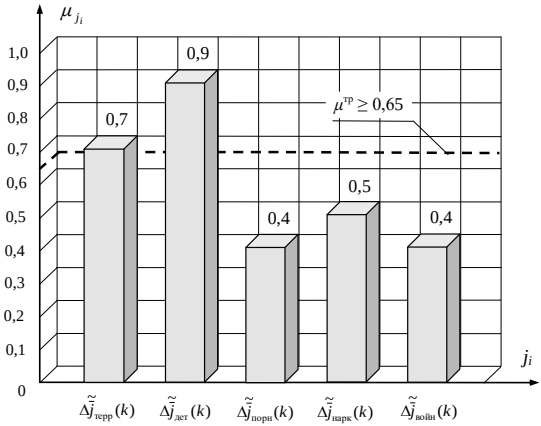

Рис. 8. Результаты вычислительного эксперимента по оценке и категоризации признаков нежелательной информации в условиях нечеткости

Из рисунка 8 видно, что признаки $\Delta\bar{j}_{\text{порн}}(k)$ $\Delta j_{\text{нарк}}(k)$ и $\Delta\bar{j}_{\text{войн}}(k)$ не превосходят опасный уровень и не являются нежелательными и потенциально вредоносными. Предпочтение по опасности отдано $\Delta\bar{j}_{\text{терр}}(k)$ и $\Delta\bar{j}_{\text{дет}}(k)$, то есть аномальному отклонению среднего количества информации, содержащей призывы к терроризму и причиняющей вред здоровью детей. Расчеты характеризуют вес (важность, предпочтительность) конкретного нечеткого признака. Предложенные значения функций принадлежности можно интерпретировать как прогноз гарантированной предпочтительности включения конкретного признака в состав множества признаков нежелательной информации.

Рассмотрим второй пример, демонстрирующий возможности компонента устранения неполноты и противоречивости с точки зрения использования ИНС. В рамках этого этапа осуществляется нейросетевая процедура устранения неполноты и противоречивости оценки и категоризации признаков СИО. Основу этого этапа составляет двухслойная ИНС. Сущность данного этапа заключается в том, что определяется хотя бы один признак, гарантированно включаемый в состав множества признаков нежелательной информации, причем этот признак однозначно включается в состав этого множества.

Искусственная нейронная сеть, применяемая в нашем случае, имеет традиционную двухслойную структуру и является классической нейронной сетью прямого распространения. Данный тип сети из-за своего широкого распространения в рамках тривиальных вычислений, выбран как простой и эффективный инструмент устранения неполноты и противоречивости оценки небольшого количества признаков нежелательной информации [41].

При этом число нейронов в слоях двухслойной искусственной нейронной сети соответствует количеству выбранных (анализируемых) признаков нежелательной информации, числу элементов вектора входных признаков, и может составлять от 1 до 50 нейронов для простой сети прямого распространения.

С помощью двухслойной искусственной нейронной сети формируется вектор входных признаков $\{\vec{C}_{\text{вх}}^{\,l}\}$, который учитывает неполные и противоречивые взаимосвязи всех признаков (по мнению L экспертов). По итогам решения задачи нейросетевого преобразования на выходе двухслойной ИНС получаем выходной вектор признаков СИО с коэффициентами (элементами), характеризующими вес (уровень опасности) этих признаков. Результаты этих вычислений позволяют (с учетом неполноты и противоречивости исходных данных) оценить и категорировать эту информацию, как нежелательную.

Предлагаемая модель выбора важных (значимых) признаков нежелательной информации в условиях неполноты и противоречивости позволяет избавиться от субъективных оценок и приобретать знания эмпирически, опираясь на мнения экспертов.

Пусть эмпирические данные имеют вид протокола:

$$\{\vec{C}_{\text{вх}}^{\,l}, \quad l = 1, \ldots, L\}, \tag{13}$$

где $\vec{C}^{l}_{\text{вх}} = (C^{l}_{\text{вх}\,1}, C^{l}_{\text{вх}\,2}, \ldots, C^{l}_{\text{вх}\,J})$ — вектор входных признаков (в терминах ИНС — входной вектор $\vec{A}$), который учитывает неполные и противоречивые взаимосвязи всех $j = 1, \ldots, J$ признаков нежелательной информации, по мнению l-го из множества L экспертов.

Показательным примером может служить вектор, характеризующий важность для каждого из пяти рассмотренных ранее признаков $\Delta\bar{j}_{\text{терр}}$, $\Delta\bar{j}_{\text{дет}}$, $\Delta\bar{j}_{\text{порн}}$, $\Delta\bar{j}_{\text{нарк}}$ и $\Delta\bar{j}_{\text{войн}}$:

$$\vec{A} = \vec{C}^{1}_{\text{вх}} = (1, 0, 0, 1, -1). \qquad (14)$$

Вектор (14) является символьной записью следующего выражения: «В соответствии с мнением первого эксперта первый признак СИО $C_{\text{вх}\,1}$ (имеет физический смысл $\Delta\bar{j}_{\text{терр}}$) и четвертый признак $C_{\text{вх}\,4}$ ($\Delta\bar{j}_{\text{нарк}}$) являются важными, существенными, значимыми. Пятый признак $C_{\text{вх}\,5}$ (имеет физический смысл $\Delta\bar{j}_{\text{войн}}$) является «не важным», не существенным. По остальным признакам ($C_{\text{вх}\,2}$ и $C_{\text{вх}\,3}$) мнение первого эксперта отсутствует».

Предположим, что в данный момент гарантированно важным, существенным и значимым признаком является признак $C_{\text{вх}\,5}$, характеризующий $\Delta\bar{j}_{\text{войн}}$. Другие признаки — неопределенны. Тогда в целях получения обоснованных результатов оценки смыслового содержания контента для поиска и обнаружения нежелательной информации, необходимо реконструировать недостающие компоненты вектора важных, существенных и значимых признаков.

Двухслойная ИНС реконструирует недостающие компоненты вектора $\vec{A}$. Рассмотрим этот процесс на примере. Предположим, нас интересуют составляющие вектора, характеризующего важность всех признаков при условии, что обязателен для включения в список опасных признаков именно пятый признак из всей совокупности признаков. Иными словами, значение $C_{\text{вх}\,5}$, характеризующее важность этого признака, равно «1». Нормируем приращения всех признаков относительно шкалы активационной функции. Пусть активационная функция имеет следующий ступенчатый вид:

$$f(C_{\text{вх}}) = \begin{cases} 1, & C_{\text{вх}} \geq 1; \\ 0, & 0 \leq C_{\text{вх}} < 1; \\ -1, & C_{\text{вх}} < 0. \end{cases} \quad (15)$$

Простая ступенчатая функция активации нейронов выбрана из множества возможных (ступенчатая, линейная, сигмоидальная, гиперболический тангенс, функция ReLu и др.) с учетом того, что в нашем, в сущности «бинарном», случае принятия решения о важности конкретного признака нежелательной информации для определения границы активации достаточно определить, превышает ли значение этого признака некоторое пороговое значение [27, 29]. В этом случае $C_{\text{вх }5}$, характеризующее важность признака $\Delta\bar{j}_{\text{войн}}$, будет соответствовать значению выхода 5-го нейрона, равному 1, а входной вектор примет вид $\vec{A}$ = (0, 0, 0, 0, 1). Тогда выходной вектор $\vec{B}$ = (b_1, b_2, b_3, b_4, b_5) двухслойной ИНС последовательно принимает следующие значения:

$\vec{B}$ (0) = f([0; 0; 0; 0; 1]) = [0, 0, 0, 0, 1];

$\vec{B}$ (1) = f([0,667; – 0,333; 1; 1; 0]) = [0, – 1, 1, 1, 1];

$\vec{B}$ (2) = f([3; – 0,667; 4; 4; 7]) = [1, – 1, 1, 1, 1];

$\vec{B}$ (3) = f([3; – 1,667; 4,667; 4,333; 7,667])= [1, – 1, 1, 1, 1];

$\vec{B}$ (4) = f([3; – 1,667; 4,667; 4,333; 7,667])= [1, – 1, 1, 1, 1];

$\vec{B}$ (5) = f([3; – 1,667; 4,667; 4,333; 7,667])= [1, – 1, 1, 1, 1].

Полученные результаты характеризуют суммарную предпочтительность включения данных признаков в состав множества опасных и могут быть представлены графически (рис. 9).

Из рисунка 9 видно, что двухслойная ИНС в интересах оценки смыслового содержания контента для поиска и обнаружения нежелательной информации стабилизировалась уже после третьего шага. Таким образом, с помощью такой ИНС, состоящей из двух слоев нейронов, можно осуществить оценку и краткосрочное нормативное прогнозирование веса (важности, значимости, опасности) признаков в условиях неполноты и противоречивости исходных данных. Результаты решения задачи в рамках второго примера позволяют с высокой степенью объективности, опираясь на накопленные в нейронной сети данные, сформировать вектор существенных признаков СИО и

корректно выбрать объем и номенклатуру признаков нежелательной информации. При этом следует заметить, что в состав множества опасных признаков гарантированно войдут такие признаки, как $C_{\text{вх }1}$ (имеет физический смысл $\Delta\bar{j}_{\text{терр}}$), $C_{\text{вх }3}$ ($\Delta\bar{j}_{\text{порн}}$), $C_{\text{вх }4}$ ($\Delta\bar{j}_{\text{нарк}}$) и $C_{\text{вх }5}$ (имеет физический смысл $\Delta\bar{j}_{\text{войн}}$), и не войдет признак $C_{\text{вх }2}$ ($\Delta\bar{j}_{\text{дет}}$).

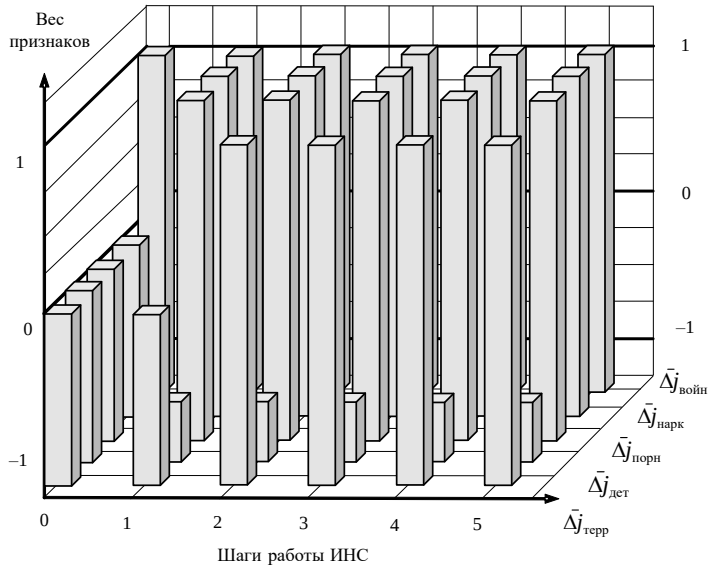

Рис. 9. Диаграмма зависимости веса признаков нежелательной информации от шага (цикла) вычисления новых состояний нейронов выходного слоя ИНС

Результаты реализации рассмотренных этапов работы компонента устранения неполноты и противоречивости результатов классификации СИО, а также результаты его экспериментальной оценки показывают, что использование обоих этапов в совокупности позволяет устранить неопределенность любого вида при формировании множества опасных, явных признаков для принятия решений в интересах выявления и противодействия нежелательной информации. Возможным очередным шагом исследований, нацеленным на совместное решение задач устранения как нечеткости, так и неполноты и противоречивости признаков нежелательной информации, может быть создание нейро-

нечетких математических моделей и алгоритмов обработки и интерпретации данных [33]. Данные механизмы, несмотря на сложность реализации, потенциально применимы, учитывая, что количество входов (признаков нежелательной информации) небольшое, а нечеткие алгоритмы и ИНС работают достаточно эффективно.

6. Заключение. В настоящей статье предложен новый тип интеллектуальных систем, ориентированный на аналитическую обработку цифрового сетевого контента в интересах защиты от нежелательной информации. Проведенный анализ состояния исследований в этой области показал, что автоматизированное обнаружение и противодействие нежелательной информации в цифровом сетевом контенте остается открытой проблемой. Предложенная архитектура ИСАОЦСК содержит три уровня, на которых располагаются компоненты распределенного сканирования сети, многоаспектной классификации сетевых ИО, устранения неполноты и противоречивости, принятия решений и визуализации. Рассмотрены модели и алгоритмы функционирования наиболее характерных компонентов системы, таких как компонент распределенного сканирования, компонент многоаспектной классификации, компонент устранения неполноты и противоречивости и компонент принятия решений. Для распределенных сетевых сканеров разработаны решения по реализации таких функций, как обнаружение и загрузка веб-контента, структурная категоризация, вычисление частотных характеристик и построение карт переходов СИО. В состав компонента многоаспектной классификации сетевых ИО предложено включить модули фильтрации, извлечения признаков, предобработки признаков и классификации СИО. При этом в функционировании этого компонента выделяются режим обучения и режим анализа. При обучении выполняется настройка классификаторов с помощью последовательностей обучающих векторов. При анализе определяется класс ИО, включая характер и степень вредоносности сетевого контента. Для устранения неполноты и противоречивости результатов классификации используются методы обработки нечетких знаний и обработка исходных данных с помощью ИНС. Принятие решений по противодействию нежелательной информации основывается на предложенных моделях информационной системы, СИО и контрмеры.

Экспериментальная оценка предложенных решений по построению и функционированию ИСАОЦСК показала, что предлагаемая система вполне отвечает предъявляемым к ней требованиям. Так, достоверность классификации нежелательных СИО в наборе данных, сформированном с помощью распределенных сетевых сканеров, достигала

84 процентов. Этот результат был получен в реальном масштабе времени при объеме набора данных, превышающем 8 Гбайт. Предложенные этапы работы компонента устранения неполноты и противоречивости результатов классификации СИО позволяют устранить в исходных данных для классификации СИО неопределенности любого вида в интересах принятия решений по выявления и противодействию нежелательной информации.

Направления дальнейших исследований связываются с усовершенствованием моделей и алгоритмов функционирования предложенной системы, расширяя область ее применения на обработку графического и мультимедийного веб-контента, а также обнаружение и противодействие недостоверной (фейковой) новостной информации.

Литература

1. *Scott J.* Social Network Analysis: Developments, Advances, and Prospects // Social Network Analysis and Mining. 2011. vol. 1. no. 1. pp. 21-26.
2. *Jebari C.* A pure URL-based genre classification of web pages // Proceedings of the 25th International Workshop on Database and Expert Systems Applications. 2014. pp. 233–237.
3. *Kotenko I., Chechulin A., Komashinsky D.* Categorisation of Web Pages for Protection against Inappropriate Content in the Internet // International Journal of Internet Protocol Technology (IJIPT). 2017. vol. 10. no. 1. pp. 61-71.
4. *Vaismoradi M., Turunen H., Bondas T.* Content Analysis and Thematic Analysis: Implications for Conducting a Qualitative Descriptive Study // Nursing & Health Sciences. 2013. vol. 15. no. 3. pp. 398-405.
5. *Defranco J.F., Laplante Ph.A.* A Content Analysis Process for Qualitative Software Engineering Research // Innov. Syst. Softw. Eng. 2017. vol. 13. no. 2-3. pp. 129-141.
6. *Boettger R.K., Palmer L.A.* Quantitative Content Analysis: Its Use in Technical Communication // IEEE Transactions on Professional Communication. 2010. vol. 53. no. 4. pp. 346-357.
7. *Linhares R.N., Costa A.P.* The use of qualitative data analysis software in brazilian educational papers // Proceedings of the International Conference in Engineering Applications (ICEA). 2019. pp. 1–7.
8. *Pashakhanlou H.* Fully Integrated Content Analysis in International Relations // International Relations. 2017. vol. 31. no. 4. pp. 447-465.
9. *Timmermans S., Iddo T.* Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis // Sociological Theory. 2012. vol. 30. no. 3. pp. 167-186.
10. *Gunawan T.S., Abdullah N.A.J., Kartiwi M., Ihsanto E.* Social network analysis using python data mining // Proceedings of the 8th International Conference on Cyber and IT Service Management (CITSM). 2020. pp. 1–6.
11. UCINET documentation. URL: sites.google.com/site/ ucinetsoftware/document (дата доступа: 29.07.2021).
12. *Du W.* Toward semantic social network analysis for business big data // Proceedings of the 14th International Conference on Semantics, Knowledge and Grids (SKG). 2018. pp. 1–8.

13. *Li H., Zhang Z., Xu Y.* Web page classification method based on semantics and structure // Proceedings of the 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). 2019. pp. 238–243.
14. *Patil A., Pawar B.* Automated classification of web sites using Naive Bayessian algorithm // Proceedings of the International Multi-Conference of Engineers and Computer Scientists. 2012. vol. 1. pp. 466–467.
15. *Kotenko I., Chechulin A., Shorov A., Komashinsky D.* Analysis and evaluation of web pages classification techniques for inappropriate content blocking // Proceedings of the 14th Industrial Conference on Data Mining (ICDM 2014). Lecture Notes in Artificial Intelligence. 2014. vol. 8557. pp. 39–54.
16. *Shibu S., Vishwakarma A., Bhargava N.* A Combination Approach for Web Page Classification using Page Rank and Feature Selection Technique // International Journal of Computer Theory and Engineering. 2010. vol. 2. no. 6. pp. 897-900.
17. *Xu Z., Yan F., Qin J., Zhu H.* A web page classification algorithm based on link information // Proceedings of the 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. 2011. pp. 82–86.
18. *Hashemi M.* Web Page Classification: A Survey of Perspectives, Gaps, and Future Directions // Multimed. Tools Appl. 2020. vol. 79. pp. 11921-11945.
19. *Patel A.D., Pandya V.N.* Web page classification based on context to the content extraction of articles // Proceedings of the 2nd International Conference for Convergence in Technology (I2CT). 2017. pp. 539–541.
20. *Arya C., Dwivedi S.K.* News web page classification using URL content and structure attributes // Proceedings of the 2nd International Conference on Next Generation Computing Technologies (NGCT). 2016. pp. 317–322.
21. *Safae L., Habib B. E., Abderrahim T.* A Review of machine learning algorithms for web page classification // Proceedings of the 5th International Congress on Information Science and Technology (CiSt). 2018. pp. 220–226.
22. *Aydın K.E., Baday S.* Machine learning for web content classification // Proceedings of the Innovations in Intelligent Systems and Applications Conference (ASYU). 2020. pp. 1–7.
23. *Petprasit W., Jaiyen S.* E-commerce web page classification based on automatic content extraction // Proceedings of the 12th International Joint Conference on Computer Science and Software Engineering (JCSSE). 2015. pp. 74–77.
24. *Belmouhcine A., Idrissi A., Benkhalifa M.* Web Classification Approach Using Reduced Vector Representation Model Based on HTML Tags // Journal of Theoretical and Applied Information Technology. 2013. vol. 55. no. 1. pp. 137-148.
25. *Kotenko I., Chechulin A., Komashinsky D.* Evaluation of text classification techniques for inappropriate web content blocking // Proceedings of the IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2015). 2015. pp. 412–417.
26. *Novozhilov D., Kotenko I., Chechulin A.* Improving the categorization of web sites by analysis of html-tags statistics to block inappropriate content // Proceedings of the 9th International Symposium on Intelligent Distributed Computing (IDC'2015). 2016. pp. 257–263.
27. *Mishra M., Srivastava M.* A view of artificial neural network // Proceedings of the International Conference on Advances in Engineering & Technology Research (ICAETR - 2014). 2014, pp. 1–3.
28. *Mehlig B.* Artificial Neural Networks. University of Gothenburg, Sweden. 2019.
29. *Burghardt F., Garbe R.* Introduction of artificial neural networks in EMC // Proceedings of the IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity (EMC, SI & PI). 2018. pp. 165–169.

30. *Parashchuk I.B.* System formation algorithm of communication network quality factors using artificial neural networks // Proceedings of the 1st IEEE International Conference on Circuits and System for Communications (ICCSC'02). 2002. pp. 263–266.
31. *Pandey K., Bhanacharjee S., Lau S., Tushir M.* A Comparative study of fuzzy systems and neural networks for system modeling and identification // Proceedings of the 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES). 2018. pp. 876–880.
32. *Агеев С.А., Саенко И.Б.* Управление безопасностью защищенных мультисервисных сетей специального назначения // Труды СПИИРАН. 2010. № 2(13). С. 182–198.
33. *Kotenko I., Parashchuk I., Omar T.* Neuro-fuzzy models in tasks of intelligent data processing for detection and counteraction of inappropriate, dubious and harmful information // Proceedings of the 2nd International Scientific-Practical Conference Fuzzy Technologies in the Industry. 2018. pp. 116–125.
34. *Нугуманова А.Б., Бессмертный И.А., Пецина П., Байбурин Е.М.* Обогащение модели Bag of words семантическими связями для повышения качества классификации текстов предметной области // Программные продукты и системы. 2016. №. 2 (114). С. 89–99.
35. *Mikolov T., Chen K., Corrado G., Dean J.* Efficient estimation of word representations in vector space // arXiv preprint arXiv:1301.3781. 2013. pp. 1–12.
36. SquidGuard – Blacklists. URL: www.squidguard.org/blacklists.html (дата доступа: 29.07.2021).
37. Shalla Secure Services. Shalla's Blacklists. URL: www.shallalist.de/ (дата доступа: 29.07.2021).
38. DMOZ. Archive. URL: dmoz-odp.org/ (дата доступа: 29.07.2021).
39. *Joulin A., Grave E., Bojanowski P., Mikolov T.* Bag of tricks for efficient text classification // arXiv preprint arXiv:1607.01759. 2016. pp. 1–5.
40. *Браницкий А.А., Котенко И.В.* Обнаружение сетевых атак на основе комплексирования нейронных, иммунных и нейронечетких классификаторов // Информационно-управляющие системы. 2015. № 4 (77). С. 69–77.
41. *Паращук И.Б., Башкирцев А.С., Михайличенко Н.В.* Анализ уровней и видов неопределенности, влияющей на принятие решений по управлению информационными системами // Информация и космос. 2017. № 1. С. 112–120.

Котенко Игорь Витальевич — д-р техн. наук, профессор, главный научный сотрудник, руководитель лаборатории, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: безопасность компьютерных сетей, в том числе управление политиками безопасности, разграничение доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей, анализ и верификация протоколов безопасности и систем защиты информации, защита программного обеспечения от взлома и управление цифровыми правами, технологии моделирования и визуализации для противодействия кибер-терроризму. Число научных публикаций — свыше 500. ivkote@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т.: +7(812)328-3337, факс: +7(812)328-4450.

Саенко Игорь Борисович — д-р техн. наук, профессор, ведущий научный сотрудник, лаборатория проблем компьютерной безопасности, СПб ФИЦ РАН; профессор кафедры, Военная академия связи. Область научных интересов: автоматизированные информационные системы, информационная безопасность, обработка и передача данных по каналам связи, теория моделирования и математическая статистика, теория информации. Число

научных публикаций — свыше 400. ibsaen@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т.: +7(812)328-3337, факс: +7(812)328-4450.

Браницкий Александр Александрович — к.т.н., старший научный сотрудник, лаборатория проблем компьютерной безопасности, СПб ФИЦ РАН. Область научных интересов: безопасность компьютерных сетей, искусственный интеллект, функциональное программирование. Число научных публикаций — 50. branitskiy@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т.: +7(812)328-3337, факс: +7(812)328-4450.

Паращук Игорь Борисович — д-р техн. наук, профессор, ведущий научный сотрудник лаборатории проблем компьютерной безопасности, СПб ФИЦ РАН; профессор кафедры, Военная академия связи. Область научных интересов: анализ качества и эффективности автоматизированных информационных систем, центры обработки данных, безопасность информации, теория оценивания, теория моделирования и математическая статистика, теория информации, теория передачи данных. Число научных публикаций — свыше 250. parashchuk@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т.: +7(812)328-3337, факс: +7(812)328-4450.

Гайфулина Диана Альбертовна — младший научный сотрудник, лаборатория проблем компьютерной безопасности, СПб ФИЦ РАН. Область научных интересов: безопасность компьютерных сетей, искусственный интеллект. Число научных публикаций — 20. gaifulina@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т.: +7(812)328-3337, факс: +7(812)328-4450.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ, проект № 18-29-22034 мк.

DOI 10.15622/ia.20.4.1

I. KOTENKO, I. SAENKO, A. BRANITSKIY,
I. PARASHCHUK, D. GAIFULINA

INTELLIGENT SYSTEM OF ANALYTICAL PROCESSING OF DIGITAL NETWORK CONTENT FOR HIS PROTECTION AGAINST INAPPROPRIATE INFORMATION

Kotenko I., Saenko I., Branitskiy A., Parashchuk I., Gaifulina D. **Intelligent System of Analytical Processing of Digital Network Content for His Protection Against Inappropriate Information.**

Abstract. Currently, the Internet and social networks as a medium for the distribution of digital network content are becoming one of the most important threats to personal, public and state information security. There is a need to protect the individual, society and the state from inappropriate information. In scientific and methodological terms, the problem of protection from inappropriate information has an extremely small number of solutions. This determines the relevance of the results presented in the article, aimed at developing an intelligent system of analytical processing of digital network content to protect against inappropriate information. The article discusses the conceptual foundations of building such a system, revealing the content of the concept of inappropriate information and representing the overall architecture of the system. Models and algorithms for the functioning of the most characteristic components of the system are given, such as a distributed network scanning component, a multidimensional classification component of network information objects, a component for eliminating incompleteness and inconsistency, and a decision-making component. The article presents the results of the implementation and experimental evaluation of system components, which demonstrated the ability of the system to meet the requirements for the completeness and accuracy of detection and counteraction of unwanted information in conditions of its incompleteness and inconsistency.

Keywords: Intelligent System, Digital Network Content, Inappropriate Information, Classification, Fuzzy Knowledge, Decision Making.

Kotenko Igor — Ph.D., Dr.Sci., Professor, Head of Laboratory, Laboratory of Computer Security Problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: computer network security, including security policy management, access control, authentication, network security analysis, intrusion detection, firewalls, deception systems, malware protection, verification of security systems, digital right management, modeling, simulation and visualization technologies for counteraction to cyber terrorism; The number of publications — over 500. ivkote@comsec.spb.ru, www.comsec.spb.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3337, fax: +7(812)328-4450.

Saenko Igor — Ph.D., Dr.Sci., Professor; Leading research scientist, Laboratory of Computer Security Problems, SPC RAS; Professor of the department, the Military academy of communications. Research interests: automated information systems, information security, processing and transfer of data on data links, theory of modeling and mathematical statistics, information theory. The number of publications — over 400. ibsaen@comsec.spb.ru, www.comsec.spb.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3337, fax: +7(812)328-4450.

Branitskiy Alexander — PhD, Senior researcher, Laboratory of Computer Security Problems, SPC RAS. Research interests: security of computer networks, artificial intelligence, functional

programming. The number of publications — 50. branitskiy@comsec.spb.ru, http://www.comsec.spb.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3337, fax: +7(812)328-4450.

Parashchuk Igor — Ph.D., Dr.Sci., Professor; Leading research scientist, Laboratory of Computer Security Problems, SPC RAS; Professor of the department, the Military academy of communications. Research interests: analysis of the quality and efficiency of automated information systems, data processing centers, information security, estimation theory, modeling theory and mathematical statistics, information theory, data transmission theory. The number of publications — over 250. parashchuk@comsec.spb.ru, http://www.comsec.spb.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3337, fax: +7(812)328-4450.

Gaifulina Diana — Junior researcher, Laboratory of Computer Security Problems, SPC RAS. Research interests: security of computer networks, artificial intelligence. number of publications — 20. gaifulina@comsec.spb.ru, http://www.comsec.spb.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-3337, fax: +7(812)328-4450.

Acknowledgements. This research was funded by RFBR according to the research project No. 18-29-22034 mk.

References

1. Scott J. Social Network Analysis: Developments, Advances, and Prospects. *Social Network Analysis and Mining*. 2011. vol. 1. no. 1. pp. 21-26.
2. Jebari C. A pure URL-based genre classification of web pages. Proceedings of the 25th International Workshop on Database and Expert Systems Applications. 2014. pp. 233–237.
3. Kotenko I., Chechulin A., Komashinsky D. Categorisation of Web Pages for Protection Against Inappropriate Content in the Internet. *International Journal of Internet Protocol Technology (IJIPT)*. 2017. vol. 10. no. 1. pp. 61-71.
4. Vaismoradi M., Turunen H., Bondas T. Content Analysis and Thematic Analysis: Implications for Conducting a Qualitative Descriptive Study. *Nursing & Health Sciences*. 2013. vol. 15. no. 3. pp. 398-405.
5. Defranco J.F., Laplante Ph.A. A Content Analysis Process for Qualitative Software Engineering Research. *Innov. Syst. Softw. Eng*. 2017. vol. 13. no. 2–3. pp. 129-141.
6. Boettger R.K., Palmer L.A. Quantitative Content Analysis: Its Use in Technical Communication. *IEEE Transactions on Professional Communication*. 2010. vol. 53. no. 4. pp. 346-357.
7. Linhares R.N., Costa A.P. The use of Qualitative Data Analysis Software In Brazilian Educational Papers. Proceedings of the International Conference in Engineering Applications (ICEA). 2019. pp. 1–7.
8. Pashakhanlou H. Fully Integrated Content Analysis in International Relations. *International Relations*. 2017. vol. 31. no. 4. pp. 447–465.
9. Timmermans S., Iddo T. Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis. *Sociological Theory*. 2012. vol. 30. no. 3. pp. 167-186.
10. Gunawan T.S., Abdullah N.A.J., Kartiwi M., Ihsanto E. Social Network Analysis using Python Data Mining. Proceedings of the 8th International Conference on Cyber and IT Service Management (CITSM), 2020, pp. 1–6.
11. UCINET documentation. Available at: sites.google.com/site/ucinetsoftware/document (accessed: 29.07.2021).

12. Du W. Toward semantic social network analysis for business big data. Proceedings of the 14th International Conference on Semantics, Knowledge and Grids (SKG). 2018. pp. 1–8.
13. Li H., Zhang Z., Xu Y. Web page classification method based on semantics and structure. Proceedings of the 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). 2019. pp. 238–243.
14. Patil A., Pawar B. Automated classification of web sites using Naive Bayessian algorithm. Proceedings of the International Multi-Conference of Engineers and Computer Scientists. 2012. vol. 1. pp. 466–467.
15. Kotenko I., Chechulin A., Shorov A., Komashinsky D. Analysis and evaluation of web pages classification techniques for inappropriate content blocking. Proceedings of the 14th Industrial Conference on Data Mining (ICDM 2014). Lecture Notes in Artificial Intelligence. 2014. vol. 8557. pp. 39–54.
16. Shibu S., Vishwakarma A., Bhargava N. A Combination Approach for Web Page Classification Using Page Rank and Feature Selection Technique. *International Journal of Computer Theory and Engineering*. 2010. vol. 2. no. 6. pp. 897-900.
17. Xu Z., Yan F., Qin J., Zhu H. A web page classification algorithm based on link information. Proceedings of the 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. 2011. pp. 82–86.
18. Hashemi M. Web Page Classification: A Survey of Perspectives, Gaps, and Future Directions. *Multimed. Tools Appl.* 2020. vol. 79. pp. 11921-11945.
19. Patel A.D., Pandya V.N. Web page classification based on context to the content extraction of articles. Proceedings of the 2nd International Conference for Convergence in Technology (I2CT). 2017. pp. 539–541.
20. Arya C., Dwivedi S.K. News web page classification using URL content and structure attributes. Proceedings of the 2nd International Conference on Next Generation Computing Technologies (NGCT). 2016. pp. 317–322.
21. Safae L., Habib B. E., Abderrahim T. A review of machine learning algorithms for web page classification. Proceedings of the 5th International Congress on Information Science and Technology (CiSt). 2018. pp. 220–226.
22. Aydın K.E., Baday S. Machine learning for web content classification. Proceedings of the Innovations in Intelligent Systems and Applications Conference (ASYU). 2020. pp. 1–7.
23. Petprasit W., Jaiyen S. E-commerce web page classification based on automatic content extraction. Proceedings of the 12th International Joint Conference on Computer Science and Software Engineering (JCSSE). 2015. pp. 74–77.
24. Belmouhcine A., Idrissi A., Benkhalifa M. Web Classification Approach Using Reduced Vector Representation Model Based on HTML Tags. *Journal of Theoretical and Applied Information Technology*. 2013. vol. 55. no. 1. pp. 137-148.
25. Kotenko I., Chechulin A., Komashinsky D. Evaluation of text classification techniques for inappropriate web content blocking. Proceedings of the IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2015). 2015. pp. 412–417.
26. Novozhilov D., Kotenko I., Chechulin A. Improving the categorization of web sites by analysis of html-tags statistics to block inappropriate content. Proceedings of the 9th International Symposium on Intelligent Distributed Computing (IDC'2015). 2016. pp. 257–263.
27. Mishra M., Srivastava M. A view of artificial neural network. Proceedings of the International Conference on Advances in Engineering & Technology Research (ICAETR - 2014). 2014, pp. 1–3.
28. Mehlig B. Artificial Neural Networks. University of Gothenburg, Sweden. 2019.

29. Burghardt F., Garbe R. Introduction of artificial neural networks in EMC. Proceedings of the IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity (EMC, SI & PI). 2018. pp. 165–169.
30. Parashchuk I.B. System formation algorithm of communication network quality factors using artificial neural networks. Proceedings of the 1st IEEE International Conference on Circuits and System for Communications (ICCSC'02). 2002. pp. 263–266.
31. Pandey K., Bhanacharjee S., Lau S., Tushir M. A comparative study of fuzzy systems and neural networks for system modeling and identification. Proceedings of the 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES). 2018. pp. 876–880.
32. Ageev S.A., Saenko I.B. [Security management of protected multi-service networks for special purposes]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2010. № 2(13). pp. 182–198. (In Russ.).
33. Kotenko I., Parashchuk I., Omar T. Neuro-fuzzy models in tasks of intelligent data processing for detection and counteraction of inappropriate, dubious and harmful information. Proceedings of the 2nd International Scientific-Practical Conference Fuzzy Technologies in the Industry. 2018. pp. 116–125.
34. Nugumanova A.B., Bessmertny I.A., Petsina P., Bayburin E.M. [Enriching the Bag of words model with semantic links to improve the quality of classification of domain texts]. *Programmnye produkty i sistemy – Software products and systems*. 2016. № 2 (114). pp. 89–99. (In Russ.).
35. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space. *arXiv preprint* arXiv:1301.3781. 2013. pp. 1–12.
36. SquidGuard – Blacklists. Available at: www.squidguard.org/blacklists.html (accessed: 29.07.2021).
37. Shalla Secure Services. Shalla's Blacklists. Available at: www.shallalist.de/ (accessed: 29.07.2021).
38. DMOZ. Archive. Available at: dmoz-odp.org/ (accessed: 29.07.2021).
39. Joulin A., Grave E., Bojanowski P., Mikolov T. Bag of tricks for efficient text classification. *arXiv preprint* arXiv:1607.01759. 2016. pp. 1–5.
40. Branitskiy A.A., Kotenko I.V. [Network attack detection based on combination of neural, immune and neuro-fuzzy classifiers]. *Informacionno-upravljajushhie sistemy – Information management systems*. 2015. № 4 (77). pp. 69–77. (In Russ.).
41. Parashchuk I.B., Bashkircev A.S., Mihajlichenko N.V. [Analysis of the levels and types of uncertainty affecting decision-making in the management of information systems]. *Informacija i kosmos – Information and space*. 2017. № 1. pp. 112–120. (In Russ.).