И.В. Смирнов, И.В. Соченков, Р.Е. Суворов, И.А. Тихомиров

Фильтрация контента в интернете: современный уровень и перспективы[1]

Аннотация. В статье описан современный уровень и перспективы развития методов и средств контентной фильтрации в Интернете. Проанализированы принципы построения систем контентной фильтрации, выделены тенденции их развития. Рассмотрены подходы к преодолению недостатков существующих решений в данной области.

Ключевые слова: фильтрация контента, информационная безопасность, родительский контроль, анализ текстов, автоматическая классификация.

Введение

Существенную часть сайтов в Интернете составляют информационные ресурсы, содержание которых может быть классифицировано как нежелательное для определенных категорий пользователей. К таким ресурсам относятся порнографические сайты, онлайн-казино, сайты об оружии и наркотических веществах, сатанизме, насилии и т.д. Также в число нежелательных входят развлекательные сайты и порталы, содержащие ненормативную лексику и контрафактный контент. Согласно исследованиям [1, 2], количество порно-сайтов в Интернете в 2010 г. составило около 12 % и быстро увеличивается.

В современном обществе существует проблема неконтролируемого доступа пользователей к ресурсам Интернета [3-5]. Кроме этого в бизнес-среде необходимо пресекать нарушение корпоративного регламента использования Интернетом (просмотр информации, не имеющей отношения к непосредственным обязанностям сотрудников). По данным компании GFI [5], свыше 40% обращений к Интернету с рабочего места никак не связано с исполнением служебных обязанностей, что приводит к значительному снижению производительности труда. Кроме того, по данным исследовний [5], более 70% всего web-трафика порно-сайтов генерируется сотрудниками компаний в рабочее время с рабочих компьютеров.

Обеспечение социальной безопасности детей и подростков в Интернете (запрет доступа к нежелательной информации) также является важной задачей. Использование Интернета сотрудниками или учащимися, не связанное с учебной или служебной деятельностью, получило название «киберслэкинг» (от англ. cyberslacking – кибербездельничание). Посещение нежелательных сайтов ведет к возникновению уязвимостей в информационных системах и способствует заражению компьютеров пользователей вредоносным программным обеспечением (ПО) [2].

Задача выявления нежелательного контента в Интернете и блокировки доступа пользователей к нему решается с помощью систем контентной фильтрации (СКФ). В статье исследуется эволюция принципов функционирования СКФ Интернет-ресурсов и рассматривается развитие методов фильтрации web-трафика. Представлены также перспективы и направления развития методов контентной фильтрации в Интернете.

1. Современный уровень развития систем контентной фильтрации

Появление первых СКФ более 10 лет назад связано с ростом количества материалов неже-

[1] Работа выполнена при финансовой поддержке РФФИ (грант № 12-07-33012).

лательного содержания в Интернете [6]. Одной из первых программ, в которой реализована возможность фильтрации контента, является браузер Internet Explorer корпорации Microsoft. В этой программе существует возможность блокировать web-сайты при их принадлежности к темам сексуального характера, насилия и жестокости. Эта принадлежность определяется на основе самоидентификации сайтов по процедурам, разработанным в Internet Content Rating Association (ICRA) [7] (в настоящее время The Family Online Safety Institute - FOSI [8]). Кроме того, предусмотрена возможность явного задания пользователем списка нежелательных сайтов. Однако этот подход оказался малоэффективным на практике [9]. Экспоненциальный рост количества сайтов в Интернете, развитие поисковых машин, расширение аудитории пользователей Интернета обусловило потребность в специализированных программных средствах контентной фильтрации. Системные администраторы и домашние пользователи уже не могли вручную поддерживать списки запрещенных сайтов в актуальном состоянии. Это привело к созданию специализированных программ – систем контентной фильтрации, функционирующих как на компьютерах конечных пользователей, так и на промежуточном сетевом оборудовании, в частности, прокси-серверах или шлюзах доступа в Интернет (gateways).

В основе СКФ, разработанных за последние 10 лет, лежат достаточно простые принципы фильтрации: списки URL - ресурсов Интернета, классифицированных по различным категориям (списки доступа). Фильтрация реализуется на основе белых и черных списков [10, 11]. Белые списки обеспечивают достаточно надежную фильтрацию, но в этом случае теряется полнота доступа к информации в Интернете. Черные списки предусматривают систематизацию web-сайтов по рубрикам нежелательных тем и накладывают различные ограничения на доступ к web-сайтам в зависимости от категории пользователей Интернета (дети, взрослые, служащие) и категории web-сайта. На практике списки доступа нежелательных ресурсов могут содержать шаблоны (например, в виде регулярных выражений), позволяющие заблокировать группу Интернет-ресурсов, URL-адрес которых соответствует шаблону, а также запретить доступ к содержимому заданного формата (например, файлам mp3-аудио, avi-видео и т.д.).

Подход на основе черных списков дает пользователям бо́льшую свободу и, соответственно, доступность всей необходимой информации из Интернета, но содержит ряд важных недостатков. Первый из них состоит в потенциальной неполноте списков доступа. Чтобы поддерживать их в актуальном состоянии, необходимо постоянно анализировать миллиарды сайтов Интернета на предмет наличия нежелательной информации. Изначально эта задача решалась исключительно привлечением большого числа экспертов-оценщиков (асессоров) для ручной классификации сайтов. В последнее время имеет место тенденция автоматизировать процесс оценки сайтов путем предварительной классификации сайтов с применением методов машинного анализа контента.

Ежедневно в Интернете появляются тысячи новых сайтов, а на уже известных сайтах появляется новый контент. Темпы появления новой информации настолько велики, что она уже не может быть своевременно обработана асессорами и помещена в списки доступа. Ресурсы помещаются в списки доступа с запозданием, что делает их неполными и неактуальными.

Эпоха Web-2.0 и ставшие стандартами дефакто технологии и инструменты построения Интернет-сайтов, такие как Ajax, Adobe Flash и другие, требуют внесения изменений и в технологии фильтрации контента. Например, в настоящее время ни одна из СКФ не позволяет производить анализ информации, передаваемой с использованием технологий AJAX [10]. Таким образом, существующие решения на основе статических списков доступа не обеспечивают должного качества (полноты и точности) предотвращения доступа к нежелательной информации.

Исследование СКФ, разработанных известными российскими и зарубежными компаниями, проведенное авторами в 2011 г., показало, что ни одна из систем не обеспечивает 100% полноты фильтрации по теме «порнография». Так, некоторые СКФ позволяют обращаться к нежелательным сайтам напрямую по IP-адресу, хотя доменные имена этих сайтов присутствуют в списках доступа. Некоторые СКФ принудительно включают режим безопасного «социального» поиска у крупнейших поисковых машин Интернета, в частности, Yandex и Google. Этот режим предназначен для блокировки нежелательной информации в поисковой

выдаче, например, при обычном поиске или при поиске по картинкам и мультимедиа. Однако указанный режим реализован не у всех поисковых машин и не всегда дает надежные результаты [12]. В частности, результаты поиска картинок по ряду вполне допустимых запросов непосредственно содержат нежелательную информацию.

Еще одним фактором, который следует принимать во внимание при решении задачи контентной фильтрации, является растущая популярность использования анонимайзеров и прокси-серверов для доступа к web-страницам. Эти технологии позволяют работать в режиме web-прокси [13, 14], доступ к нежелательной информации при этом осуществляется опосредованно – через web-страницы анонимайзеров. Некоторые системы СКФ борются с этим явлением путем блокировки известных прокси-серверов и анонимайзеров, однако их списки не полны. В ходе исследования существующих СКФ авторам удавалось находить работоспособный неблокируемый анонимайзер с помощью поисковых систем за время, не превышающее 10–15 минут. Через найденный анонимайзер можно получить практически неограниченный доступ к контенту любой категории.

Итак, в существующих СКФ выделяются следующие основные факторы, препятствующие качественному решению задачи контентной фильтрации:

– неполнота списков доступа (время жизни сайта может составлять не более 30 дней);

– невозможность блокировки нежелательного содержания динамических страниц (например, выдачи поисковой машины);

– невозможность блокировки нежелательного содержания на отдельных страницах вполне допустимых web-сайтов;

– невозможность блокировки доступа к нежелательному содержанию через анонимайзеры.

На практике вышеуказанное означает, что при целенаправленном поиске нежелательной информации защиту СКФ удается обойти за 10-60 минут работы с поисковой системой.

Для преодоления указанных недостатков в ряде СКФ реализованы примитивные методы определения тематики web-сайтов и web-страниц «на лету» - в момент обращения к ним приложения пользователя. Тематическая принадлежность определяется на основе содержания запрошенной страницы. Существующие системы реализуют примитивную фильтрацию контента по стоп-словам, присущим тем или иным категориям нежелательной информации. Зачастую это сводится к поиску в тексте web-страницы подстрок, совпадающих со словами из списка стоп-слов, что приводит к значительному числу ложных срабатываний на вполне допустимых web-страницах. Кроме того, такой подход не работает для страниц на русском языке в силу флективности последнего.

Следует упомянуть и о подходе к решению задачи контентной фильтрации на основе самоидентификации. Подход заключается в том, что web-сайт сам сообщает СКФ о своей тематической направленности, а СКФ в зависимости от настроек принимает решение о допустимости содержания для конечного пользователя. Подход был реализован ассоциацией ICRA (в настоящее время FOSI), а также в проекте Platform for Internet Content Selection (PICS) [15], который был позднее преобразован в инициативу Protocol for Web Description Resources (POWDER), поддерживаемую одноименной рабочей группой при консорциуме W3C [16]. В настоящее время рабочая группа POWDER прекратила свою деятельность. Эти решения не являются всеохватывающими. Подавляющее большинство сайтов не предоставляет информацию о своей тематике. При этом сайты с сомнительным содержанием могут предоставлять недостоверную информацию о теме, что делает применение в СКФ такого решения ненадежным [9].

Из вышесказанного следует, что, несмотря на наличие ряда технологий и методов контентной фильтрации, реализованных в нескольких десятках СКФ различных фирм, задачу ограничения доступа пользователей к нежелательной информации нельзя считать решенной. Перспективным направлением развития СКФ, по мнению авторов, является использование методов динамического определения категории Интернет-ресурсов по их содержимому в момент обращения к ним приложений пользователя, т.е. динамическая контентная фильтрация (ДКФ). Это стало особенно актуально с развитием порталов, которые содержат информацию различных категорий, изменяющуюся во времени и подстраивающуюся под настройки клиента [10, 11].

ДКФ основывается на всестороннем анализе содержания web-страниц и web-сайтов и выделении признаков, характеризующих их тематику. В настоящее время разрабатываются методы динамического анализа информации в различных форматах: гипертекстовом, графическом, видео и др. Однако существующие методы работают с невысокой полнотой и малой точностью [17, 18].

Далее в статье рассматриваются принципы построения СКФ и способы реализации перспективных методов ДКФ в существующих СКФ.

2. Принципы построения систем контентной фильтрации

Для качественного и полного решения задачи контентной фильтрации необходимо учитывать все особенности реализации СКФ, принципы их функционирования в сетевой среде, а также принимать во внимание методы работы пользователей в Интеренте. Поэтому оценка тенденций развития методов контентной фильтрации и повышение качества их работы невозможно без анализа архитектурных решений в области СКФ.

СКФ не функционируют обособленно от других компонентов сетевой инфраструктуры. Они могут функционировать в виде [19]:

– служб операционной системы, функционирующих на компьютерах пользователей;

– служб операционной системы, функционирующих на промежуточных сетевых узлах;

– модулей расширений или самостоятельных программно-аппаратных устройств, относящихся к промежуточному сетевому оборудованию.

СКФ могут выступать в роли клиента или сервера, а так же в обеих ролях сразу [10].

При функционировании СКФ в качестве промежуточного оборудования необходимо осуществлять интеграцию СКФ с другими сетевыми сервисами. Для этих целей разработаны стандартные протоколы Internet Content Adaptation Protocol (ICAP) и Open Pluggable Edge Services (OPES). Кроме того, некоторые производители создавали собственные протоколы для обеспечения взаимодействия конкретных продуктов друг с другом или со сторонним программным обеспечением. Сюда можно отнести протоколы Cisco Web Cache Coordination Protocol (WCCP), Check Point Content Vectoring Protocol (CVP) и другие [10]. Некоторые протоколы – ICAP и OPES – разработаны так, что могут использоваться для реализации как сервисов контентной фильтрации, так и других сервисов – переводчики, размещение рекламы, доставка данных, зависящая от политики их распространения, и т.п. Наиболее широкое применение протокол ICAP нашел в продуктах для защиты от вредоносного кода, поскольку он позволяет использовать эти проверки в различных продуктах и не зависит от платформы, на которой выполняется клиент ICAP. Протокол OPES в настоящее время становится более популярным, так как он устраняет ряд недостатков, присущих ICAP [10].

Современные СКФ, разработанные коммерческими фирмами, ориентированы на использование в сетях передачи данных, и, как правило, реализуют поддержку протоколов OPES или ICAP. Нередко СКФ реализуются в виде модулей-расширений для популярных продуктов, обеспечивающих работу сетевой инфраструктуры, например Microsoft Internet Security and Acceleration Server (в настоящее время переименован в Microsoft Forefront [20]). Такие СКФ работают в комплексе с другими средствами обеспечения сетевой безопасности и контроля: фильтрами электронной почты, детекторами спама, антивирусами и сканерами потенциальных утечек информации.

Существуют серверные решения с открытым исходным кодом, использующие прокси-сервер Squid [21] для фильтрации web-трафика, например, squidGuard [22], «Режик» [23], DansGuardian [24]. Вышеперечисленные средства, как правило, настраиваются таким образом, чтобы обеспечивать прозрачный прокси (transparent proxy) в локальной вычислительной сети (ЛВС), не заметный на стороне конечных пользователей [19]. Компонентная архитектура СКФ, использующей прокси-сервер squid, рассматривается в работах [12, 25].

Существуют аппаратные решения задачи контентной фильтрации – Cisco IOS® Content Filtering [26] и Cisco Service Control Engine (SCE) [27, 28]. В этом аппаратном продукте интегрированы функции межсетевого экрана, шлюзового интерфейса и СКФ. Система работает на основе статических списков доступа, но также поддерживает решения Websense [29] и Smartfilter [30] для актуализации этих списков.

Варианты реализации СКФ

Следующим подходом является использование фильтрации на основе служб доменных имен – DNS (Domain Name Service). Примерами реализации подхода служат система OpenDNS [31] и Comodo Secure DNS [32]. В основе этих решений лежит услуга (Softaware as a Service – SaaS) фильтрации доступа к ресурсам в виде сервиса DNS. При этом на стороне клиентов, в роли которых могут выступать как целые локальные сети, так и компьютеры отдельных пользователей, настраивается доступ к службе DNS - компании, предоставляющей услугу. При таком подходе процесс фильтрации прозрачен для пользователя. В упомянутых решениях реализована дополнительная фильтрация адресов тех web-ресурсов, которые несут вред вирусного заражения компьютеров пользователя.

Другой подход к решению задачи заключается в развертывании СКФ непосредственно на компьютерах пользователей. В этом случае СКФ функционирует подобно антивирусному программному обеспечению или системе фильтрации почтового спама. СКФ отслеживает все обращения приложений пользователя к ресурсам Интернета и блокирует доступ к нежелательной информации. Вариант СКФ на компьютерах конечных пользователей дает бо́льшую гибкость персонализации, но обладает меньшей защищенностью против взлома, нежели СКФ на промежуточном сетевом оборудовании. Поэтому необходимо защищать СКФ, функционирующие на компьютерах пользователей, против несанкционированного отключения и повреждения их программных компонентов с целью получения доступа к нежелательной информации. СКФ обычно реализуются в виде приложений-сервисов, отключение или удаление которых возможно только при наличии у пользователя прав администратора. Компонентная архитектура СКФ, функционирующих на компьютерах конечных пользователей, рассматривается в работе [19]. Примерами реализации СКФ на компьютерах пользователей являются системы NetPolice [33], Интернет-фильтр «Один дома» [34] и др. На рисунке показана обобщенная диаграмма вариантов реализации СКФ. Сопоставление характеристик СКФ в зависимости от варианта реализации приведено в таблице.

3. Перспективы развития систем контентной фильтрации

Проведенное исследование показало, что в настоящее время не существует технологии или продукта, который бы в достаточной мере решал задачу фильтрации нежелательного контента. Повышение качества контентной фильтрации следует ожидать, прежде всего, при реализации различных методов ДКФ для информации различных типов (гипертекст, графические изображения, видео и др.).

Наиболее перспективными, по мнению авторов, являются следующие направления развития СКФ.

1. Применение методов динамического выявления нежелательного содержания совместно со статическими списками доступа, пополняемыми за счет автоматически классифицируемых страниц.

Сравнение характеристик СКФ в зависимости от варианта реализации

Характеристика \ Вариант реалиизации СКФ	В выделенной сетевой подсистеме	На компьютерах пользователей
Способ реализации	На промежуточном оборудовании в ЛВС: 1. Модуль-расширение для прокси-сервера. 2. Модуль-расширение для службы DNS. 3. Интегрированное программно-аппаратное устройство (шлюз) для управления маршрутизацией. 4. Интегрированное программно-аппаратное устройство, работающее по принципу «прозрачного» прокси-сервера. В среде облачных вычислений как услуга: 1. Облачный сервис фильтрации DNS; 2. Облачный сервис фильтрации URL.	1.Служба операционной системы. 2. Модуль-расширение для web-браузера. 3. Административные настройки DNS. 4. Компонент сетевого экрана или антивирусного ПО.
Персонализация	Путем создания профилей пользователей и настроек прав доступа на сервере для групп пользователей.	Путем создания профилей и настроек прав доступа на компьютерах пользователей.
Риск получения доступа к нежелательной информации путем нарушения работоспособности СКФ	Низкий: только путем получения доступа к серверу СКФ или прокси-серверу.	Присутствует: возможно, несанкционированное получение прав администратора для управления настройками персонального компьютера.
Механизмы обновления ПО	Системными администраторами вручную. Возможно автоматическое обновление через Ин-	Как правило, автоматическое обновление через Интернет.
Варианты реализации динамической фильтрации контента	Требуется отдельная (как правило, распределенная) подсистема анализа контента, способная эффективно обрабатывать запросы пользователей обслуживаемого сегмента сети.	1. Динамический анализ контента на компьютере пользователя: незначительное снижение быстродействия и незначительное увеличение вычислительной нагрузки. 2. Динамический анализ контента на удаленных серверах: некоторое увеличение объемов передаваемого трафика, незначительное снижение быстродействия web-браузера.
Фильтрация защищенного HTTPS-трафика	Внедрение в браузеры на компьютерах пользователей специально созданных сертификатов для взаимодействия с промежуточным прокси-сервером. Прокси-сервер использует свой собственный сертификат для взаимодействия с запрошенным web-сервером. Передаваемый через прокси-сервер трафик расшифровывается и анализируется. Данные, передаваемые конечному пользователю, шифруются сертификатом прокси-сервера. Подробнее в [8].	1. Перехват расшифрованных данных непосредственно в приложениях. 2. Расшифровка трафика с использованием локально доступных сертификатов на компьютере пользователя.
Приоритетные области внедрения	В корпоративных и научно-образовательных сетях.	В среде домашних пользователей.

2. Разработка и совершенствование методов автоматической классификации гипертекстовых документов на основе статистических данных о встречаемости слов. В качестве признаков классификации могут применяться слова и устойчивые словосочетания ЕЯ, которые выделяются из текста путем машинного анализа. Подробнее о подходе см. в работах [12, 35].

3. Использование ссылочного ранжирования и учет ссылочной структуры web-страниц. Для сайтов, относящихся к категориям нежелательной информации, характерно наличие ссылок на другие сайты из той же категории – тематическая замкнутость. Кроме того, наличие ссылок на изображения и встраиваемый контент в совокупности с ключевыми терминами

ЕЯ может, например, повысить точность выявления порносайтов. Значительное количество ссылок на свободно доступные файловые хранилища может свидетельствовать о размещении на сайте контрафактного контента и другого нежелательного содержания.

4. Разработка точных и эффективных методов выявления нежелательного медиа-контента. Для выявления материалов порнографического характера перспективным представляется разработка методов, позволяющих, например, выявлять обнаженное человеческое тело на изображениях и видео. Некоторые успехи в области фильтрации видео достигнуты в проекте Licenzero [36, 37], что свидетельствует о возможности решения этой задачи с помощью компьютера. Фильтрация подобного контента необходима при работе с динамическими страницами поисковых машин (поиск по видео и изображениям), а также при работе с сайтами, предоставляющими услуги размещения фотографий и видеороликов. Заинтересованность в методах борьбы с нежелательным медиаконтентом проявляют администраторы вышеперечисленных сервисов, поскольку распространение определенных категорий нежелательной информации (в частности, детской порнографии) является правонарушением, преследуемым по закону [38].

5. Анализ поведения пользователей Интернета. Подобный подход успешно применяется при формировании выдачи поисковых машин Интернета, однако в области контентной фильтрации не нашел своего применения.

Совместно с вышеприведенными методами подобный анализ применим для автоматизации процесса наполнения и актуализации списков доступа. В качестве исходных данных выступает статистика посещения пользователями web-страниц, анонимно собираемая поисковыми машинами Интернета, СКФ и прокси-серверами. Анализ пользовательских сессий, содержащих обращения как к известным ресурсам нежелательной тематики, так и к еще не классифицированным, поможет выделять потенциально нежелательные ресурсы. Для эффективной работы метода необходимы данные о пользовательском поведении, которые доступны в крупных сетях или в среде пользователей поисковых машин.

Очевидно, что рассмотренные выше подходы к динамическому выявлению нежелательной информации применимы в СКФ различных типов. Совместное применение методов ДКФ должно повысить точность и полноту фильтрации и решить проблемы, связанные с доступом к нежелательной информации через прокси-серверы и анонимайзеры. При этом задача контентной фильтрации будет решаться одновременно на нескольких уровнях: на компьютерах конечных пользователей, на уровне поставщиков услуг доступа в Интернет, а также непосредственно на информационных сервисах (поисковых машинах, мультимедиа-порталах, файл-хостингах и др.). Стандартом «де-факто» станет автоматическое обновление компонентов СКФ через Интернет, поскольку методы ДКФ будут совершенствоваться, а необходимые для их работы данные – постоянно актуализироваться разработчиками.

Заключение

В статье исследован современный уровень развития методов и средств контентной фильтрации в Интернете. Проанализированы принципы построения и функционирования СКФ. Современные СКФ не решают задачу предотвращения доступа к нежелательной информации эффективно и качественно, поэтому необходимо разрабатывать новые методы и подходы к контентной фильтрации.

Подводя итоги исследования тенденций в области контентной фильтрации, подчеркнем, что в перспективе произойдет интеграция СКФ с другими технологиями Интернет: поисковыми машинами, файлообменными, информационными и медиа-сервисами. При этом традиционный подход на основе списков доступа будет существенно расширен за счет внедрения методов динамической фильтрации информации. Это позволит повысить качество контентной фильтрации, учесть современные тенденции развития Интеренета, такие как Web 2.0, использование анонимайзеров, прокси-серверов и т.д.

Ряд упомянутых в статье методов ДКФ реализован и апробирован в прототипе системы динамической контентной фильтрации, разработанной в Институте системного анализа РАН. В частности, реализован метод автоматического определения тематики гипертекстовых документов на основе машинного анализа текста ЕЯ [12, 19, 35]. В 2009 и 2010 гг. метод был успешно проверен в дорожках автоматической

классификации web-страниц по независимым оценкам в рамках семинара РОМИП [39, 40]. Прототип системы успешно справляется с задачей контентной фильтрации в небольшой локальной сети на одном прокси-сервере. Современное аппаратное обеспечение позволяет применять разработанный метод для решения задачи контентной фильтрации без существенного снижения быстродействия.

Литература

1. Количество порносайтов в сети стремительно растет // Центр исследования компьютерной преступности / [Электронный ресурс] Режим доступа: http://www.crime-research.ru/news/16.06.2006/2585/, 16.06.2006, Проверено 2013-02-01.
2. Wondracek G., Holz T, Platzer C., Kirda E., Kruegel C. Is the Internet for Porn? An Insight Into the Online Adult Industry // Proceedings of the Ninth Workshop on the Economics of Information Security (WEIS 2010) / [Электронный ресурс] Режим доступа: http://www.sba-research.org/wp-content/uploads/publications/weis2010_wondracek.pdf, 2010. Проверено 2013-02-01.
3. Открытое письмо к администрации и акционерам социальной сети “ВКонтакте”, [Электронный ресурс] Режим доступа: http://www.vedomosti.ru/cgi-bin/get_document.cgi/vedomosti_23-03-2011.pdf?file=2011/03/23/257032_2313597632, 2011-03-23. Проверено 2013-02-01.
4. Из-за доступности порносайтов растет количество педофилов // Интернет-издание «Газета.СПб» / [Электронный ресурс] Режим доступа: http://www.gazeta.spb.ru/31072-0/, 2008. Проверено 2013-02-02.
5. The importance of an acceptable use policy // GFI White Paper / [Электронный ресурс] Режим доступа: http://www.gfi.com/whitepapers/acceptable_use_policy.pdf, Проверено 2012-12-29.
6. Фильтруем порнографию // Волгоград / [Электронный ресурс] Режим доступа: http://www.volgograd.ru/theme/hitech/komp/sovet/23956.pub, 2005. Проверено 2012-12-29.
7. Web-сайт Internet Content Rating Association / [Электронный ресурс] Режим доступа: http://www.icra.org/. Проверено 2012-10-15.
8. Web-сайт The Family Online Safety Institute / [Электронный ресурс] Режим доступа: http://www.fosi.org/. Проверено 2012-10-15.
9. Archer P. ICRAfail: A lesson For the Future. / [Электронный ресурс] Режим доступа: http://philarcher.org/icra/ICRAfail.pdf. Проверено 2012-10-15.
10. Отт А. Современные тенденции в области контентной фильтрации / [Электронный ресурс] Режим доступа: http://alexott.net/ru/writings/cf/index.html. Проверено 2012-12-29.
11. Отт А. О контентной фильтрации // А. Отт, Jet Info №10(261), «Инфосистемы Джет», 2006. [Электронный ресурс] Режим доступа: http://www.jetinfo.ru/Sites/info/Uploads/2006.10.9E9241B3F9C846AB9AFEF38A01175122.pdf. Проверено 2012-12-29.
12. Соченков И.В. Система динамической контентной фильтрации на основе автоматического классификатора гипертекстовых документов. // XI конференция по искусственному интеллекту с международным участием КИИ-2008. Труды конференции. Т. 3, М.: ЛЕНАНД, 2008. С. 337-344.
13. Web-сайт Прокся.ру – анонимайзер / [Электронный ресурс] Режим доступа: http://www.proxya.ru. Проверено 2013-01-23.
14. Web-сайт http://proxer.ru/ / [Электронный ресурс] Режим доступа: http://proxer.ru/. Проверено 2013-01-23.
15. Web-сайт Platform for Internet Content Selection / [Электронный ресурс] Режим доступа: http://www.w3.org/PICS/. Проверено 2012-10-14.
16. Web-сайт Protocol for Web Description Resources (POWDER): POWDER Working Group / [Электронный ресурс] Режим доступа: http://www.w3.org/2007/powder/. Проверено 2012-10-14.
17. McCullagh D. Report criticizes Google's porn filters // CNET News, 2010 / [Электронный ресурс] Режим доступа: http://news.cnet.com/Report-criticizes-Googles-porn-filters/2100-1032_3-996417.html?tag=mncol;2n. Проверено 2013-01-25.
18. Web-сайт The Internet Content Filter Review / [Электронный ресурс] Режим доступа: http://internet-filter-review.toptenreviews.com/. Проверено 2013-02-04.
19. Соченков И.В. Контентный фильтр для социально-безопасного доступа в Интернет. // Теория и практика системного анализа: Труды I Всероссийской научной конференции молодых ученых. Т. 1. – Рыбинск: РГАТА им. П.А. Соловьева. 2010. С. 73–78.
20. Web-сайт Microsoft Forefront / [Электронный ресурс] Режим доступа: http://technet.microsoft.com/en-us/forefront/ee175814.aspx. Проверено 2012-10-15.
21. Web-сайт Squid: Optimising Web Delivery / [Электронный ресурс] Режим доступа: http://www.squid-cache.org/. Проверено 2012-10-14.
22. Web-сайт Welcome to squidGuard / [Электронный ресурс] Режим доступа: http://www.squidguard.org/. Проверено 2012-10-14.
23. Web-сайт «РЕЖИК» / [Электронный ресурс] Режим доступа: http://www.rejik.ru/. Проверено 2012-10-14.
24. Web-сайт DansGuardian / [Электронный ресурс] Режим доступа: http://dansguardian.org/. Проверено 2012-10-14.
25. Белов С.Д., Жижимов О.Л., Федотов А.М., Осипов Г.С., Тихомиров И.А., Соченков И.В. Комплексная защита крупных корпоративных сетей передачи данных // Третья Международная конференция «Системный анализ и информационные технологии» САИТ-2009 (14-18 сентября 2009 г., Звенигород, Россия): Труды конференции. М.: 2009.
26. Cisco IOS Content Filtering / [Электронный ресурс] Режим доступа: http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6586/ps6643/prod_qas0900aecd804abb06.pdf. Проверено 2013-02-04.
27. Cisco SCE 2000 Series Service Control Engine / [Электронный ресурс] Режим доступа: http://www.cisco.com/en/US/products/ps6151/index.html. Проверено 2013-02-04.

28. Совместное решение с Cisco для операторов связи. Сайт Центра Анализа Интернет-Ресурсов / [Электронный ресурс] Режим доступа: http://www.cair.ru/pages.php?page=19. Проверено 2013-02-04.
29. Security on the social web. Websense makes it safe to be social. / [Электронный ресурс] Режим доступа: http://www.websense.com/content/social-web-security-solutions.aspx. Проверено 2013-02-04.
30. McAfee SmartFilter. Защита и контроль для Веб 2.0 / [Электронный ресурс] Режим доступа: http://www.mcafee.com/ru/products/smartfilter.aspx. Проверено 2013-02-04.
31. Premium DNS. Cloud-based Internet Security / [Электронный ресурс] Режим доступа: http://www.opendns.com/. Проверено 2013-02-04.
32. Comodo Secure DNS / [Электронный ресурс] Режим доступа: http://www.comodo.com/secure-dns/. Проверено 2013-02-04.
33. Сообщество пользователей безопасного Интернета «NetPolice» / [Электронный ресурс] Режим доступа: http://www.netpolice.ru/. Проверено 2013-02-04.
34. Интернет-фильтр «Один дома» / [Электронный ресурс] Режим доступа: http://www.centertelecom.ru/branches/voronezh/services_and_tariffs/people/Pages/filtr.aspx. Проверено 2013-02-04.
35. Тихомиров И.А., Соченков И.В. Метод динамической контентной фильтрации сетевого трафика на основе анализа текстов на естественном языке. Вестник НГУ, Информационные технологии, т. 6, Вып. 2, Новосибирск, 2008. С. 94-100.
36. Licenzero: простые движения. // Блог «Хабрахабр» – «Data Mining», 2011 / [Электронный ресурс] Режим доступа: http://habrahabr.ru/blogs/data_mining/116625/. Проверено 2013-01-25.
37. Licenzero: порно детектед. // Блог «Хабрахабр» – «Data Mining», 2011 / [Электронный ресурс] Режим доступа: http://habrahabr.ru/blogs/data_mining/116173/. Проверено 2013-01-25.
38. Бахарев И., Ковалевский А. iFolder обменялся с милицией детским порно // Интернет-издание «Газета.Ru», 2010 / [Электронный ресурс] Режим доступа: http://www.gazeta.ru/business/2010/03/18/3340048.shtml. Проверено 2013-01-25.
39. Смирнов И.В., Соченков И.В., Тихомиров И.А. Система интеллектуального поиска и анализа информации «Exactus» на РОМИП-2009 // Труды РОМИП-2009. Санкт-Петербург: НУ ЦСИ, 2009. 198 с.
40. Киселев А.А., Смирнов И.В., Тихомиров И.А., Соченков И.В.. Система интеллектуального поиска и анализа информации Exactus на РОМИП-2010 // Труды российского семинара по оценке методов информационного поиска РОМИП'2010. - Казань: Казан. ун-т, 2010. С. 49-69.

Смирнов Иван Валентинович. Старший научный сотрудник Института системного анализа РАН. Окончил Российский университет дружбы народов в 2003 году. Автор 34 научных работ. Область научных интересов: искусственный интеллект, компьютерная лингвистика, машинное обучение.
E-mail: ivs@isa.ru

Соченков Илья Владимирович. Инженер-исследователь Института системного анализа РАН. Окончил Российский университет дружбы народов в 2009 году. Автор 20 научных работ. Область научных интересов: интеллектуальные методы поиска и анализа информации, обработка больших массивов данных, защита сетей, контентная фильтрация, компьютерная лингвистика. E-mail: sochenkov@isa.ru

Суворов Роман Евгеньевич. Инженер-исследователь Института системного анализа РАН. Окончил Рыбинский государственный авиационный технический университет им П.А. Соловьёва в 2012 году. Автор 2 научных работ. Область научных интересов: тематическая классификация текстовых документов, контентная фильтрация, интеллектуальные динамические системы. E-mail: rsuvorov@isa.ru

Тихомиров Илья Александрович. Старший научный сотрудник Института системного анализа РАН. Окончил Рыбинскую государственную авиационную технологическую академию в 2002 году. Кандидат технических наук. Автор 47 научных работ. Область научных интересов: искусственный интеллект, компьютерная лингвистика, поисковые системы, информационная безопасность, интернет-системы. E-mail: tih@isa.ru