УДК 004.9

Е.П.Охапкина
Л.И.Воронова
Москва, Россия

РАЗРАБОТКА ФРАГМЕНТА ОНТОЛОГИИ ДЛЯ МНОГОАГЕНТНОЙ СИСТЕМЫ МОДЕРАЦИИ СООБЩЕНИЙ ПОЛЬЗОВАТЕЛЕЙ

Аннотация. Социальные сети прочно вошли в жизнь современного человека. Удобство их использования и многофункциональность с каждым днем завоевывает все большее количество людей. Однако наряду с явными преимуществами социальные сети содержат в себе немало проблем. Одной из таких проблем является публикация в открытом доступе постов провокационного и антисоциального содержания с целью воздействия на общественное сознание. Зачастую такие посты содержат призывы к межнациональной розни, обсуждение антиобщественных мероприятий, пропаганду насилия, пропаганду радикальных движений.

Как правило, модерация постов в социальных сетях сводится к поиску таких нарушений как распространение спама, флуда, размещение постов и комментариев не по теме или содержащих информацию рекламного характера.

В связи с экспоненциальным ростом размещаемой пользователями информации стала актуальной проблема автоматизации процесса модерации. Для решения этой задачи ведутся активные исследования как зарубежными, так и отечественными разработчиками.

Авторы статьи предлагают использовать технологии многоагентных систем (МАС) и программных агентов для построения системы мониторинга сообщений пользователей в сети интернет с целью выявления угроз безопасности. При этом самостоятельной задачей является разработка тематических онтологий, содержащих описание угроз в выбранной предметной области.

В статье описана концептуальная модель многоагентной системы модерации пользовательских сообщений и структура онтологии МАС, на основе которой разработан фрагмент онтологии, описывающий типы и направленность угроз безопасности.

Онтология представляет собой описание частично упорядоченного множества понятий, которые должны использоваться агентами, выявляющими угрозы безопасности. Специализация каждого агента отражается подмножеством понятий, некоторые из которых могут быть общими для нескольких агентов. Приведенный фрагмент онтологии позволяет агентам в процессе коммуникации оперировать и однозначно воспринимать рабочую терминологию.

В качестве средства для разработки онтологии был выбран онторедактор Protégé, в качестве способа моделирования — создание онтологий на языке OWL.

Показано, что применение агентного подхода значительно упрощает разработку программного обеспечения, поскольку дает возможность работать с данными

E.P.Okhapkina
L.I.Voronova
Moscow, Russia

DEVELOPING THE ONTOLOGY FRAGMENT FOR THE MULTI-AGENT MODERATION SYSTEM OF USERS' POSTS

Abstract. Social networks have become a part of our modern life. They are easy to use and versatile, which helps to attract more and more people every day. However, along with the obvious benefits social networks pose a number of problems. One of these problems is the open access publication of provocative and anti-social posts used to influence the public. Often such posts contain calls for ethnic discord, anti-social activities, promotion of violence, and radical religious views.

As a rule, the moderation of posts in social networks is reduced to the search for spam, flooding, off-topic posts and comments, usually containing advertising.

Due to the exponential growth of the data posted by users the automated moderation process has become an urgent problem. Both foreign and domestic developers are involved in active research to solve this problem.

The authors suggest using the technology of multi-agent systems (MAS) and software agents to build a system for monitoring users' posts on the Internet in order to identify security threats. At the same time, an independent task is to develop thematic ontologies containing a description of threats in the selected domain.

This article describes a conceptual model of multi-agent system for users' posts moderation and a structure of MAS ontology used as a basis for a fragment of ontology describing the type and direction of security threats.

The ontology is a description of a partially ordered set of concepts to be used by agents detecting security threats. Each agent specialty is reflected through a subset of concepts, some of which may be shared by several agents. The described fragment of ontology allows agents to use one and same terms in the communication process and clearly perceive working terminology.

The authors have chosen Protégé ontology editor as a tool for developing the ontology and OWL language as a modeling technique.

It is shown that the use of agent-based approach greatly simplifies software design, because it enables working with data as knowledge, taking into account the context through the use of ontologies.

как со знаниями, учитывая при этом контекст посредством использования онтологий.
Ключевые слова: онтологии; агенты; многоагентные системы.

Key words: ontology; agents; multi-agent systems.

Сведения об авторах: Охапкина Елена Павловна, старший преподаватель кафедры информационных систем и моделирования; Воронова Лилия Ивановна, заведующая кафедрой информационных систем и моделирования.

About the authors: Elena Pavlovna Okhapkina, Senior Lecturer, Department of Information Systems and Modeling; Lilia Ivanovna Voronova, Head of the Department of Information Systems and Modeling.

Место работы: Российский государственный гуманитарный университет.

Place of employment: Russian State University for the Humanities.

Контактная информация: 121353, г. Москва, ул. Вяземская, д. 13, кв. 37; тел.: 9671548100.
E-mail: lenaokhapkina@mail.ru

Задача обеспечения информационной безопасности в сети Интернет многогранна. Речь идет об обеспечении безопасности не только отдельно взятых пользователей, но и безопасности государства в целом. Практически неконтролируемый рост количества ресурсов, посвященных самым разным темам, и объемы информации, генерируемые пользователями сети Интернет каждый день, затрудняют возможность мониторинга и выявления ресурсов и информации, представляющих угрозу. Одним из источников появления и распространения небезопасной информации являются социальные сети. Данный способ распространения подобного рода информации выбран неслучайно. Во-первых, общение, предоставляемое в социальных сетях, является социальной потребностью человека и именно в процессе диалога на пользователя можно оказать необходимое влияние. Во-вторых, личные страницы пользователей зачастую несут достаточно информации о своих владельцах для выбора подходящей тактики манипулирования. В-третьих, любая открытая социальная сеть — это огромный каталог пользователей с практически неограниченными возможностями для поиска единомышленников либо потенциальных жертв не только в отдельно взятом городе, но и по всему миру. Это далеко не все предпосылки для использования социальных сетей для реализации угроз обществу.

В этой связи актуальна задача разработки многоагентной системы мониторинга сети с целью выявления угроз безопасности и конкретных онтологий, содержащих описание угроз.

Зачастую модерация постов в социальных сетях сводится к поиску таких нарушений, как распространение спама, флуда, размещение постов и комментариев не по теме, как правило, содержащих информацию рекламного характера. Однако подобные сообщения могут содержать и более серьезные нарушения, например, такие как призывы к разжиганию межнациональной розни, обсуждение и организация антиобщественных мероприятий, пропаганда насилия и т.д. В связи с неуклонным ростом размещаемой пользователями информации выполнять модерацию в основном человеческими силами становится все труднее.

Для решения этой проблемы существует ряд разработок, как зарубежных, так и отечественных. Среди отечественных стоит отметить, например, CleanTalk — сервис защиты web-сайтов от спама [2]. Данный сервис позволяет выполнять автоматическую модерацию постов после установки клиентского модуля либо подключения по API. Затем все пользовательские сообщения или запросы на регистрацию проходят ряд проверок, среди которых анализ текстов сообщений на релевантность обсуждаемой статье или оставленным комментариям, сравнение с автозаполняемыми черными списками e-mail и IP адресов, проверка на наличие запрещенных слов, так называемых стоп-слов, список которых может быть расширен по желанию пользователя.

Среди зарубежных можно выделить Comment E-Mail Verification — плагин для модерации постов [9]. При публикации пользователем комментария на введенный при публикации адрес электронной почты высылается письмо со ссылкой-подтверждением. При переходе по ссылке пользователь подтверждает, что он реальный человек и указанный e-mail является действующим, а не сгенерированным, после чего подтвержденный комментарий публикуется в блоге.

Безусловно, подобный подход способен облегчить модерацию и решить часть проблем, но вышеперечисленные средства имеют и ряд недостатков. Например, при усложнении клиентского приложения возрастает его размер, а следовательно, и требования к клиентскому компьютеру, проверка по адресу электронной почты, как в случае Comment E-Mail Verification, не способна защитить от реальных людей, распространяющих спам, флуд и т.д. В данной работе предлагается использовать в этих целях технологию многоагентных систем и программных агентов.

Программные агенты могут применяться в самых различных областях. Области, в которых могут быть применены программные агенты в сети Интернет, показаны на рис. 1 [8].

Рис. 1. Области применения программных агентов в сети Интернет

Применение агентного подхода значительно упрощает разработку программного обеспечения (ПО), поскольку новые агенты могут использовать в своей работе других агентов путем наследования их функций и свойств или же просто посылая им свои запросы, а также упрощается процесс размещения ПО в условиях сети. Происходит это за счет автоматизации процессов перемещения программного кода, его установки и конфигурирования.

Использование удаленного доступа и управления возможно с различных мобильных устройств, таких как КПК и сотовые телефоны. Кроме того появляется возможность работать с данными, как со знаниями, учитывая при этом контекст посредством использования онтологий.

Возможна асинхронная обработка данных по следующим схемам:

1) запрос агенту на обработку → отключение от сети → подключение к сети → результат обработки;

2) запрос агенту → выполнение других работ → уведомление о завершении обработки и результат [3].

Значительно сокращается время на администрирование за счет способности агентов к коммуникации и перемещению их кода. Присутствует возможность персонифицировать обработку данных, ориентируя ее на предпочтения конкретного пользователя.

Таким образом, программный агент является повторно используемым программным компонентом, который взаимодействует с другими агентами посредством передачи сообщений. Возможность повторного использования позволяет использовать агента в различных сервисах.

Как правило, агент состоит из двух частей — декларативной, описательной части агента и процедурной, являющейся совокупностью продукций, объединенных в класс [5]. В общем виде схема работы агента представлена на рис. 2.

Рис. 2. Схема работы агента

Для эффективного решения проблем модерации пользовательских постов необходимо использование целого ряда агентов, а следовательно, необходимо создание многоагентной системы (МАС). Разработанная концептуальная модель многоагентной системы представлена на рис. 3.

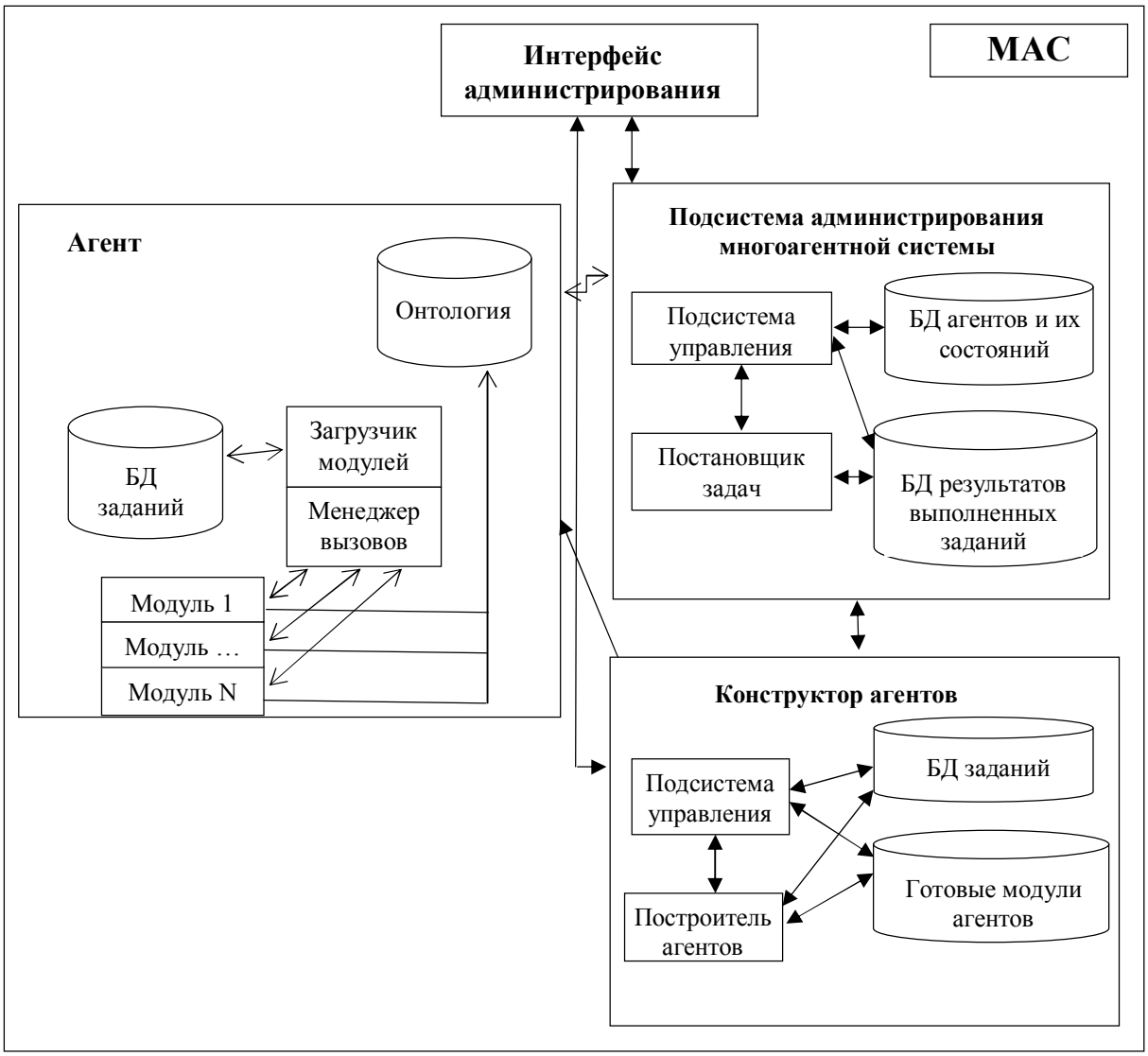

Рис. 3. Концептуальная модель многоагентной системы для модерации пользовательских постов

Модель состоит из следующих частей:

1) агент, который включает готовые модули, генерируемые конструктором агентов, базу данных заданий, делегированных агенту, онтологию, которой агент руководствуется при модерации, менеджера вызовов соответствующих модулей агентов в зависимости от задания, загрузчика модулей;

2) подсистема администрирования многоагентной системы, включающая подсистему управления, непосредственно связанную с постановщиком задач, БД существующих агентов и их состояний и БД результатов выполненных заданий;

3) конструктор агентов, используемый для создания модулей новых агентов, которые затем могут быть использованы модулем Агент, подсистему управления, позволяющую управлять процессом создания новых модулей, и БД заданий;

4) интерфейс администрирования многоагентной системы, позволяющий управлять ее частями.

Одной из наиболее важных частей большинства многоагентных систем является онтология. Онтологии применяются для структурирования информации и являются своего рода посредником между человеко- и машинно-ориентированными уровнями ее представления. В этом случае онтология интерпретируется как система соглашений о некоторой области интересов для достижения поставленных целей [7]. При разработке многоагентных систем и приложений очень важно, чтобы агенты в процессе коммуникации оперировали одними и теми же терминами и однозначно их воспринимали. Онтологии являются довольно универсальным средством, позволяющим описывать предметные области. Без использования онтологий работа любой многоагентной системы или приложения сильно усложнится необходимостью вводить сложные механизмы распознавания и классификации поступающей информации.

В области компьютерной безопасности онтологии находят применение в семантическом управлении доступом к интеллектуальным ресурсам, построении репозиториев и логического вывода в системах управления информацией и событиями безопасности (SIEM-системах), обеспечении безопасности персональных данных, обеспечении безопасности СУБД и т.д.

В рамках задачи, исследуемой в данной статье, онтология представляет собой описание частично упорядоченного множества понятий, которые должны использоваться агентами, выявляющими угрозы безопасности. Онтология должна определять подмножество понятий, которые используют агенты МАС для кооперативного решения поставленных задач, и являться основой для взаимодействия агентов [1]. Каждый агент использует определенный фрагмент общей онтологии предметной области. Специализация каждого агента отражается подмножеством понятий, некоторые из которых могут быть общими для нескольких агентов [10].

Онтология МАС состоит из следующих понятий:

- «предметная область агентов, обеспечивающих безопасность»;
- «типы и направленность угроз»;
- «функционирование агентов».

«Предметная область агентов защиты, обеспечивающих безопасность» задает функциональности и области ответственности каждого агента. «Типы и направленность угроз» характеризуют возможные виды и направленность угроз безопасности. «Функционирование агентов» определяет, каким образом агенты должны реализовывать обнаружение угроз безопасности, защиту от них и противодействие им. Функционирование агентов включает в себя понятие взаимодействия агентов, которое является инструментом кооперации и осуществляется средствами языка общения. Взаимодействие агентов системы защиты строится с помощью языка общения в соответствии с описанными онтологиями. На базе онтологии создаются сценарии поведения агентов, определяется содержимое базы знаний

агентов, которая определяет действия агентов по поиску и устранению угроз безопасности. Наглядно структура онтологии МАС представлена на рис. 4.

Рис. 4. Структура онтологии МАС для модерации сообщений пользователей

Исходя из вышеописанной структуры онтологии МАС, был разработан фрагмент онтологии, описывающий типы и направленность угроз. В качестве средства для разработки онтологии был выбран онторедактор Protégé. Protégé представляет два пути моделирования онтологий [4]:

- онтологическое представление знаний с помощью фреймов (Protégé 3.x);
- создание онтологий на языке OWL.

Выбранная среда разработки является наиболее привлекательной по следующим причинам [4]:

- основана на Java, что означает кросплатформенность;
- полная совместимость с OWL 2 (с версии 4.1);
- возможность визуализации проектируемой онтологии;
- поддержка правил, основанных на языке SWRL;
- расширяемость плагинами;
- возможность подключить системы логического вывода.

В данном случае при разработке был выбран второй путь создания онтологий и версия Protégé 4.3. Язык web-онтологий OWL разработан для использования приложениями, обрабатывающими содержимое информации, а не только представляющими эту информацию.

За основу при разработке фрагмента онтологии, описывающего типы и направленность угроз, была взята статья А.Г.Никитина «Виды и классификации экстремистской деятельности: некоторые правовые аспекты» [6]. В ходе анализа статьи была получена следующая иерархия классов, описывающая типы и направленность угроз (рис. 5).

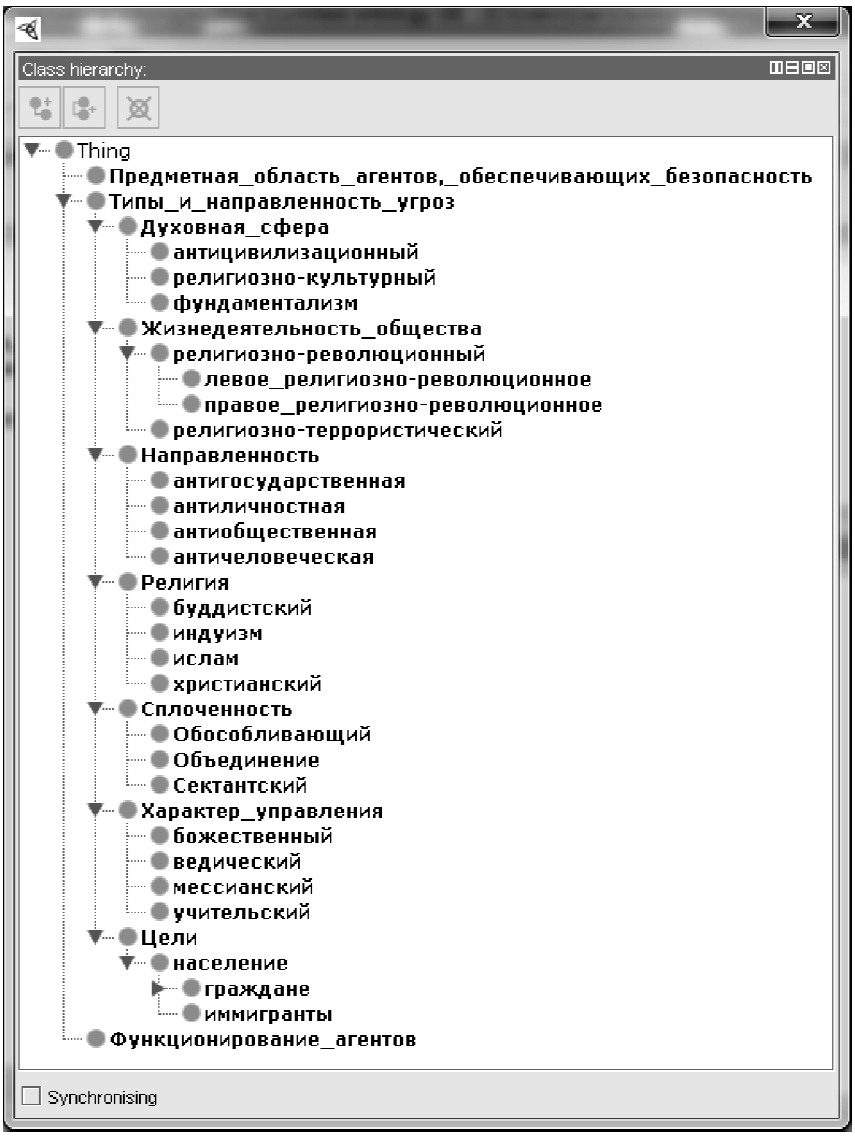

Рис. 5. Иерархия классов фрагмента онтологии, описывающая типы и направленность угроз

Таким образом, использование разработанной концептуальной модели позволит создать гибкую систему для модерации пользовательских постов в социальных сетях и удобную систему ее администрирования, позволяющую оперативно реагировать на введение новых критериев для модерации и изменение уже существующих, а также создавать качественно новых агентов и делегировать им вновь поставленные задачи. Разработанная иерархия классов онтологии в дальнейшем будет доработана, а также будут разработаны фрагменты онтологии, описывающие предметные области агентов, обеспечивающих безопасность и функционирование агентов.

ЛИТЕРАТУРА

1. Анализ онтологии взаимодействия МАС при решении задач обеспечения информационной безопасности. URL: http://progrm.ru/?p=260

2. Антиспам приложения для Web-сайтов от CleanTalk. URL: http://cleantalk.ru

3. Болдырев Е.В., Кирякова Г.С., Шилкин А.В. Агентный подход к проектированию сетевых приложений поиска информации // Вычислительные технологии. Т. 10. Специальный выпуск. 2005.

4. Болотова Л.С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях: Учебник. М., 2012.

5. Грибова В.В., Клещев А.С., Крылов Д.А., Москаленко Ф.М., Тимченко В.А., Шалфеева Е.А. Агентный подход к разработке интеллектуальных Интернет-сервисов // Труды Конгресса по интеллектуальным системам и информационным технологиям «IS&IT'12». М., 2012. Т. 1.

6. Никитин А.Г. Виды и классификации экстремистской деятельности: некоторые правовые аспекты // Татищевские чтения: актуальные проблемы науки и практики: Материалы VI Международной научно-практической конференции (16—19 апреля 2009 г.). Ч. 1. Тольятти, 2009.

7. Охапкина Е.П., Лукоянов И.А., Воронов В.И., Воронова Л.И. Разработка и внедрение поискового робота для анализа интересов клиентов // Студенческий научный форум: VI Международная студенческая электронная научная конференция (15 февраля — 31 марта 2014 г.). URL: http://www.scienceforum.ru/2014/495/4758

8. Таненбаум Э., Ван Стеен М. Распределенные системы. Принципы и парадигмы. СПб., 2003.

9. Comment E-Mail Verification. URL: http://wordpress.org/extend/plugins/comment-email-verify

10. Okhapkina E.P, Voronova L.I. The development of the ontology for a multi-agent subsystem analyzing user posts in social networks in order to identify security threats to society // International Journal of Applied and Fundamental Research. 2013. № 2. URL: http://www.science-sd.com/455-24393

REFERENCES

1. Analyzing MAC Interaction ontology in solving the problem of information security. URL: http://progrm.ru/?p=260

2. CleanTalk — Spam filter for blogs and forums. URL: http://cleantalk.ru

3. Boldyrev E.V., Kiriakova G.S., Shilkin A.V. Agent-based approach to designing network information retrieval applications // Computational Technologies. V. 10. Special Issue. 2005.

4. Bolotov L.S. Artificial intelligence systems: models and technologies based on knowledge: Reference book. Moscow, 2012.

5. Gribova V.V., Kleschev A.S., Krylov D.A., Moskalenko F.M., Timchenko V.A., Shalfeeva E.A. Agent-based approach to developing intelligent Internet services // Proceedings of the Congress on Intelligent Systems and Information Technology «IS & IT'12». Moscow, 2012. V. 1.

6. Nikitin A.G. Types and classifications of extremist activity: legal aspects // Tatischevsky Readings: Topical Problems of Science and Practice: Proceedings of VI International Scientific and Practical Conference (April 16—19, 2009). Part 1. Togliatti, 2009.

7. Okhapkina E.P., Lukoyanov I.A., Voronov V.I., Voronova L.I. Developing and applying a web crawler to analyze customers interests // Student Scientific Forum: VI International Student Scientific Conference (February 15 — March 31. 2014). URL: http://www.scienceforum.ru/2014/ 495/4758

8. Tanenbaum E., Van Steen M. Distributed systems. Principles and paradigms. St. Petersburg, 2003.

9. Comment E-Mail Verification. URL: http://wordpress.org/extend/plugins/comment-email-verify

10. Okhapkina E.P., Voronova L.I. The development of the ontology for a multi-agent subsystem analyzing user posts in social networks in order to identify security threats to society // International Journal of Applied and Fundamental Research. 2013. № 2. URL: http://www.science-sd.com/455-24393