

---

**INFORMATION TECHNOLOGY**

---

УДК 004(9+056.5)

ДМИТРО ЛАНДЕ,  
ЯРОСЛАВ КОНДРАТЕНКО**ОСОБЛИВОСТІ ПОБУДОВИ СИСТЕМ РОЗПОДІЛЕНОГО КОНТЕНТ-МОНІТОРИНГУ ГЛОБАЛЬНИХ ІНФОРМАЦІЙНИХ МЕРЕЖ**

Досліджено властивості інформаційних потоків з урахуванням їх належності до сегментів глобальних мереж. У статті розглядається процес здійснення контент-моніторингу з урахуванням постійного підвищення складності сегментів мережі Інтернет та пропонуються для застосування практичні методи побудови систем розподіленого контент-моніторингу. Особливої ваги це питання набуває при здійсненні діяльності, спрямованої на запобігання, своєчасне виявлення, припинення чи нейтралізацію реальних і потенційних загроз інформаційній безпеці держави. Організація ефективної протидії викликам та загрозам національній безпеці та її конкретним проявам у інформаційній сфері – одне з нагальних завдань, вирішення якого є важливим завданням в умовах розбудови громадянського суспільства та правової держави сьогодення. Охоплення та узагальнення великих динамічних контентних потоків, які безперервно генеруються в Інтернет-просторі, вимагає якісно нових методів та підходів до виконання заходів із забезпечення інформаційної безпеки. Вивчення інформаційних потреб та формування вимог для реалізації аналогічних систем свідчить, що в процесі здійснення контент-моніторингу мережі Інтернет важливим є застосування засобів розподіленого (сегментованого) контент-моніторингу глобальних мереж/медіа та використання повного набору інтерфейсів взаємодії сегментів глобальних мереж та інструментарію їх моніторингу (сканування). Складність сегментів мережі Інтернет, як програмно-технічних об'єктів є суттєвою, а не другорядною властивістю. Проектування програмно-технічних засобів розподіленого контент-моніторингу глобальних мереж/медіа, що абстрагуються від їх складності, часто абстрагуються і від їх сутності. Запропоновано вирішення таких задач на базі системи контент-моніторингу InfoStream. При побудові експертних систем розподіленого контент-моніторингу вбачається необхідність врахування рівня популярності (доступності) сегменту мережі Інтернет та його оперативності у формуванні контенту, різних підходів до накопичення та надання інформації, інтерфейсів взаємодії, тому для кожного “виокремленого” сегменту глобальної мережі Інтернет доцільно формувати власну базу знань та прийомів.

**Ключові слова:** контент-моніторинг, розподілені системи моніторингу, інформаційний потік, складні програмні об'єкти, експертні системи.

**Постановка проблеми.** В умовах сьогодення рівень складності завдань, що вирішуються за допомогою автоматизованих систем контент-моніторингу глобальних інформаційних мереж, постійно зростає: від автоматизованого збору, інтерпретації та реалізації інформації/контенту до управління, проектування, моделювання та прогнозування різноманітних процесів/подій. Водночас, доцільно зазначити, що важливою “виокремленою” складовою глобальної мережі Інтернет – є соціальні мережі/медіа. Їх кількість зростає, а обсяги інформації, що накопичується, стають гігантськими. Отже, при проектуванні автоматизованих систем контент-моніторингу доцільно враховувати особливості генерації інформаційних потоків сегментів глобальної мережі, виникає необхідність пошуку нетрадиційних підходів використання інформаційних технологій та математичних методів для оброблення та аналізування інформації отриманої від систем розподіленого контент-моніторингу [1], [2], [4], [7]. Зокрема, соціальних мереж/медіа, форумів, коментарів, окремих сервісів моніторингу, спеціалізованих банків даних, архівів.

**Аналіз останніх досліджень і публікацій.** У світовому інформаційно-технологічному середовищі активно триває процес удосконалення систем оперативного реагування і раннього попередження про виклики та загрози національній безпеці на базі моніторингу відкритої інформації мережі Інтернет. Вивчення інформаційних потреб для функціонування аналогічних систем свідчить, що в процесі здійснення контент-моніторингу мережі Інтернет важливим є застосування засобів розподіленого (сегментованого) контент-моніторингу глобальних мереж/медіа (побудова мережі інформаційних проксі-серверів) та використання повного набору інтерфейсів взаємодії сегментів глобальних мереж та інструментарію їх моніторингу (сканування) [1] - [5], [7].

Це обумовлено такими чинниками:

- швидкі темпи росту потреб у достовірній інформації для прийняття управлінських рішень;
- необхідність врахування рівня популярності (доступності) сегменту мережі Інтернет та його оперативності у формуванні контенту;
- різні підходи до накопичення та надання інформації сегментів глобальних мереж, постійне розширення програмних інтерфейсів взаємодії;
- зменшення навантаження на обчислювальні показники засобів інформатизації завдяки абстрагуванню модулів обробки та сегментів глобальних мереж;
- необхідність в автоматизованому управлінні, проектуванні, моделюванні та прогнозуванні подій/процесів з урахуванням розподіленості контенту відносно сегментів глобальних мереж.

**Метою статті** є дослідження особливостей побудови систем розподіленого контент-моніторингу глобальних інформаційних мереж для аналізування інформаційних потоків з урахуванням їх належності до сегментів глобальних мереж, прогнозування тенденцій подальшого розвитку технологій побудови означених систем.

**Виклад основного матеріалу дослідження.** Одним із важливих факторів застосування автоматизованих систем контент-моніторингу є нагальна необхідність виявлення спотвореної інформації. У сучасному інформаційному суспільстві, основною ознакою якого є глобалізація, завдяки максимальній швидкості генерації нової інформації та надання доступу до неї, Інтернет має велику перевагу перед іншими джерелами інформації. Враховуючи це, як ніколи, стає актуальним поняття інформаційної операції.

З урахуванням активного зростання користувачів соціальних засобів масової інформації мережі Інтернет (інтернет-засоби масової інформації, соціальні мережі/медіа, блоги, форуми, інтегратори новин), їх цільовий (тематичний) контент активно впливає на формування суспільної думки. Інтернет-ресурси, які публікують контент близький настрою суспільства перетворюються в основні джерела надання інформації.

Однак, реалії сьогодення такі, що Інтернет-ресурс потенційно може виявитись зовсім не тим, за кого він себе видає. Наприклад, він може використовуватись для здійснення шахрайських дій з метою заволодіння конфіденційною інформацією, грошима або ж власністю. А у розрізі системи забезпечення національної безпеки держави він може заподіяти шкоди життєво важливим інтересам особи, суспільства та держави, особливо у інформаційній сфері.

Організація ефективної протидії викликам та загрозам національній безпеці та її конкретним проявам у інформаційній сфері – одне з нагальних завдань, вирішення якого є важливою справою в умовах розбудови громадянського суспільства та правової держави сьогодення (інформаційних війн, протиборств). Національний інформаційний простір постійно піддається впливу чинників, які створюють небезпеку життєво важливим інтересам людини і громадянина, суспільства і держави в інформаційній сфері [1].

Особливої ваги набуває здійснення діяльності, спрямованої на запобігання, своєчасне виявлення, припинення чи нейтралізацію реальних і потенційних загроз інформаційній безпеці держави. Охоплення та узагальнення великих динамічних контентних потоків, які безперервно генеруються в Інтернет-просторі, вимагає якісно нових методів та підходів до виконання заходів із забезпечення інформаційної безпеки. Вирішення таких задач відбувається

завдяки застосуванню засобів автоматизації виявлення найбільш важливих складових загроз (інформаційних операцій) у контентних потоках. Для оперативного аналізу обстановки з метою виявлення інформаційних операцій застосовуються спеціалізовані системи моніторингу інформаційного простору (контент-моніторингу) [2].

Такі системи забезпечують:

- оперативність, яку не можна отримати від традиційних пошукових систем (час індексації мережевого контенту може варіюватись від доби до декількох тижнів);
- повноту (як з точки зору джерел, так і з надання матеріалів джерел), яку не забезпечують агрегатори новин;
- задіяння аналітичних засобів автоматизованого проектування, моделювання і прогнозування процесів/подій.

Величезна кількість багатомовних інформаційних ресурсів обумовлює складність їх використання при здійсненні інформаційно-аналітичної роботи. При збиранні, аналізі таких даних виникають проблеми обробки надвеликих обсягів даних, пошуку і навігації в динамічних інформаційних потоках. На цей час для вирішення цих проблем застосовуються такі технологічні концепції як Big Data (великі дані), Complex Networks (складні мережі), Cloud Computing (хмарині обчислення), Data/Text Mining (глибинний аналіз даних і тексту) [2], [3].

Проблеми розмірності і динаміки багатомовних інформаційних ресурсів в глобальних мережах потребують здійснення фундаментальних досліджень в областях дискретної математики (теорії графів, мереж), розпізнавання образів (класифікація, кластерний аналіз), лінгвістики, цифрової обробки сигналів, вейвлет і фрактального аналізу.

Хоча сучасний розвиток технологій дозволяє у деяких випадках знаходити необхідну інформацію в мережах, але досі залишаються невирішеними проблеми подальшої аналітичної обробки цієї інформації, виокремлення необхідних фактографічних даних, виявлення тенденцій розвитку в окремих предметних областях, взаємозв'язків об'єктів, подій, розпізнавання змістовних аномалій, прогнозування тощо. Більшість із цих проблем – актуальні питання семантичної обробки надвеликих динамічних текстових масивів інформації [3].

Навіть окремі спроби практичного вирішення цих проблем сьогодні обумовлюють успіх проектів, таких як пошукові системи Google, Yandex, Baidu, системи моніторингу соціальних мереж типу Keyhole, Brandwatch, CyberAlert, аналітичні системи типу Palantir, Centrifuge.

Один із підходів до вирішення таких задач пропонується на базі системи контент-моніторингу InfoStream [4]. Модернізація та масштабування (формування багатомовних повнотекстових баз даних, моделювання інформаційних потоків у глобальних комп'ютерних мережах) цієї складної системи відбувається з урахуванням програмно-технічного виокремлення “сутності” сегментів глобальних мереж/медіа, тобто розгортається мережа інформаційних проксі-серверів розподіленого (сегментованого) контент-моніторингу (див. рис. 1).

Такий підхід дозволяє подолати складність розподіленого контент-моніторингу сегментів глобальних мереж як програмних об'єктів [5]. Складність програмних об'єктів більше залежить від їх розмірів, ніж, можливо, для будь-яких інших конструкцій, що створюються людиною, так як ніяка пара складових частин не схожа між собою (хоча б не нижче рівня операторів). Якщо вони схожі, то їх можна об'єднати в один відкритий чи закритий програмний модуль. Аналогічно, масштабування програмних об'єктів – це не просто збільшення розмірів функціональних елементів, це обов'язково збільшення числа різноманітних елементів, що безпосередньо залежить від постійного розширення програмних інтерфейсів взаємодії сегментів глобальних мереж/медіа. У більшості випадків ці елементи взаємодіють нелінійно, і складність комплексного (цілого) зростає значно швидше, ніж лінійно [5].

Складність програмно-технічних об'єктів є суттєвою, а не другорядною властивістю. Тому проектування програмно-технічних засобів розподіленого контент-моніторингу глобальних мереж/медіа абстрагується як від їх складності, так і від їх сутності. На цей час обсяги інформаційних ресурсів в глобальних мережах перевищують трильйон документів, що

розміщуються тільки на веб-сайтах. Динаміка наповнення соціальних мереж дозволяє оцінювати відповідні ресурси, які у десятки разів більші. При цьому інформаційні ресурси мережі Інтернет репрезентативно представляють ресурси із інших засобів масової інформації (ЗМІ), зокрема, газет, радіо, телебачення. Значна кількість із цих ресурсів містить різноманітні експертні оцінки, деяка частина призначена для здійснення впливу на свідомість людей для , маніпулювання ними. Виходячи з цього, урахування інформації з цих джерел має велике значення як для виявлення напрямків розвитку економіки, науки, технологій, тощо, так і для вирішення задач у сферах безпеки і оборони [1].

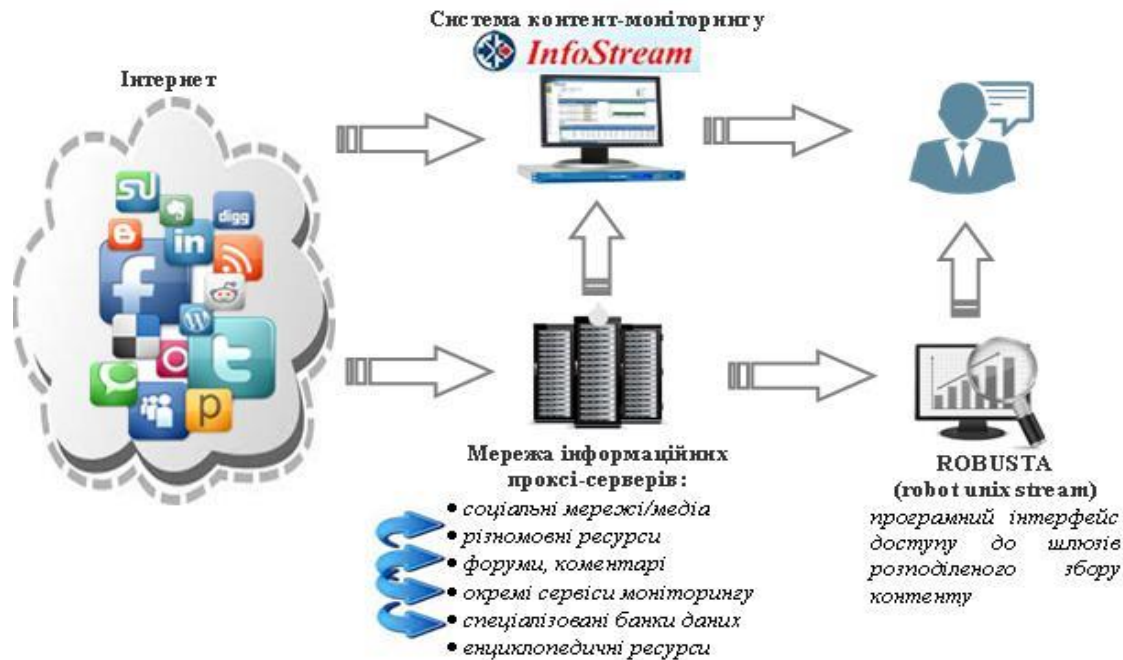


Рисунок 1 – Структура розподіленого (сегментованого) контент-моніторингу

Визначення розподіленої системи в технічному сенсі тісно пов'язане з передаванням інформації і організацією оброблення даних. Існує щонайменше п'ять критеріїв, якими можна охарактеризувати розподілену систему: архітектура апаратного забезпечення, оброблення та нагромадження даних, управління системою, її інваріантність. Загальна пам'ять, за допомогою якої різні процеси можуть обмінюватися даними, в розподіленій системі майже не використовується. Тому класичні методи синхронізації і комунікації виключаються.

Розподіленою системою моніторингу називають [6] програмно-апаратне рішення щодо збирання, оброблення та відображення інформації, яке складається із компонентів, що функціонують на фізично віддалених і незалежних одне від одного гетерогенних вузлах та надається користувачам єдиною об'єднаною системою. Це сукупність автономних процесів, що об'єднані в комунікаційні підмережі для нагромадження даних і діють спільно для розв'язання поставленої задачі. За допомогою мережі відбувається координація розподілених процесів і обмін інформацією.

Розглянемо використання методів розподіленого контент-моніторингу на прикладі системи InfoStream, у якій реалізовано розподіл моніторингу за мовами (українська, англійська, німецька, італійська, іспанська, французька, російська), за сегментами мережі Інтернет (youtube.com, twitter.com, livejournal.ru, ok.ru). Враховуючи відмінну особливість системи, що виражається у обслуговуванні вузького кола споживачів зі специфічною сферою завдань, які потребують оперативного вирішення, окремі інформаційні проксі-сервери обробляють внутрішні (корпоративні) масиви користувачів. Такий підхід забезпечує зменшення навантаження на обчислювальні ресурси засобів інформатизації за рахунок абстрагування модулів обробки та сегментів глобальних мереж. Крім того [7], у системі накопичено більш як 20-річний досвід моніторингу Інтернет-преси, який у подальшому також можливо опрацювати з урахуванням сегментованого підходу.

Враховуючи той факт [1], що в умовах сьогодення існує нагальна необхідність в удосконаленні систем оперативного реагування і раннього попередження про виклики та загрози національній безпеці на базі моніторингу відкритої інформації мережі Інтернет, вітчизняними науковцями теоретично обґрунтовано і створено засоби аналізу динаміки тематичних інформаційних потоків, зокрема, вперше застосовано вейвлет-аналіз до задач виявлення інформаційних операцій.

При цьому доцільно зазначити, що прийоми аналізування інформаційних потоків та виявлення інформаційних операцій досить різняться, особливо з урахуванням підходів до накопичення та надання інформації сегментів глобальних мереж, постійного розширення їх програмних інтерфейсів взаємодії. А при створенні експертних систем автоматизованого управління, проектування, моделювання та прогнозування подій/процесів використовують взагалі інші методи, що не схожі ні з першими, ні з другими, зокрема, штучний інтелект.

Експертна система – це програмний засіб, що містить програмний генератор висновків, базу правил та призначений для прийому вхідних даних і допущень, дослідження логічних наслідків своїх рекомендацій, також повторного переосмислення своїх суджень [5].

Доцільно зазначити, що якість таких систем безпосередньо залежить від правильного наповнення та функціонування баз знань і правил, які точніше відображають реальність. В даній ситуації при побудові експертних систем розподіленого контент-моніторингу, вбачається необхідність врахування рівня популярності (доступності) сегменту мережі Інтернет та його оперативності у формуванні контенту, різних підходів до накопичення та надання інформації, інтерфейсів взаємодії, тому для кожного виокремленого сегменту глобальної мережі Інтернет доцільно формувати власну базу знань та прийомів.

**Висновки.** Унікальність запропонованої технології полягає у вирішенні складності розподіленого контент-моніторингу сегментів глобальних мереж як програмних об'єктів. З одного боку складність програмних об'єктів більше залежить від їх розмірів, ніж, можливо, для будь-яких інших конструкцій, що створюються людиною, з іншого – від їх унікальності, з точки зору призначення інформаційних потоків.

Запропонований підхід до побудови систем контент-моніторингу дозволить здійснити поглиблений аналіз динаміки тематичних інформаційних потоків. Це надасть можливість підвищити рівень оперативності у прийнятті управлінських рішень, завдяки своєчасному виявленню інформаційних операцій (впливів).

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Д. В. Ланде, "Аналіз інформаційних потоків у глобальних комп'ютерних мережах". [Електронний ресурс]. Доступно: <http://www.nas.gov.ua/text/pdfDocumentsMeetingPresidiumNASU/170125.PDF>. Дата звернення: Січ. 25, 2017.
- [2] А. Г. Додонов, Д. В. Ландэ, В. В. Прищепа, и В. Г. Путятин, *Конкурентная разведка в компьютерных сетях*. Киев, Украина: ИПРИ НАН Украины, 2013.
- [3] H. Liu, A. Gegov, and M. Cocea, *Rule Based Systems for Big Data. A Machine Learning Approach*. Heidelberg, Germany: Springer, 2016.  
doi: 10.1007/978-3-319-23696-4.
- [4] А. Н. Григорьев, Д. В. Ландэ, С. А. Бороденков, Р. В. Мазуркевич, и В. Н. Пацьора, *InfoStream. Мониторинг новостей из Интернет: технология, система, сервис*. Киев, Украина: Старт-98, 2007.
- [5] Ф. П. Брукс, *Проектирование процесса проектирования: записки компьютерного эксперта*. Москва, Россия: Вильямс, 2012.
- [6] В. К. Галіцин, О. П. Суслов, та Н. К. Самченко, *Системи моніторингу*. Київ, Україна: КНЕУ, 2015.
- [7] InfoStream Client. [Online]. Available: <http://infostream.ua/client/retro/>. Accessed on: March 19, 2017.

Стаття надійшла до редакції 30 березня 2017 року.

## REFERENCE

- [1] D. V. Lande, "Analysis of information flows in global computer networks". [Online]. Available: [http://www.nas.gov.ua/text/pdfDocumentsMeeting\\_PresidiumNASU/170125.PDF](http://www.nas.gov.ua/text/pdfDocumentsMeeting_PresidiumNASU/170125.PDF). Accessed on: Jan. 25, 2017.
- [2] A. H. Dodonov, D. V. Lande, V. V. Pryshchepa, and V. H. Putiatyn, *Competitive intelligence in computer networks*. Kyiv, Ukraine: Institute for Information Recording of NAS of Ukraine, 2013.
- [3] H. Liu, A. Gegov, and M. Cocea, *Rule Based Systems for Big Data. A Machine Learning Approach*. Heidelberg, Germany: Springer, 2016.  
doi: 10.1007/978-3-319-23696-4.
- [4] A. N. Grigorev, D. V. Lande, S. A. Borodenkov, R. V. Mazurkevich, and V. N. Patsera, *InfoStream. Monitoring of news from the Internet: technology, system, service*. Kyiv, Ukraine: Start-98, 2007.
- [5] F. P. Bruks, *The Design of Design: Essays from a Computer Scientist*. Moscow, Russia: Viliams, 2012.
- [6] V. K. Halitsyn, O. P. Suslov, and N. K. Samchenko, *Monitoring systems*. Kyiv, Ukraine: KNEU, 2015.
- [7] InfoStream Client [Online]. Available: <http://infostream.ua/client/retro/>. Accessed on: March 19, 2017.

ДМИТРИЙ ЛАНДЭ,  
ЯРОСЛАВ КОНДРАТЕНКО

## ОСОБЕННОСТИ ПОСТРОЕНИЯ СИСТЕМ РАСПРЕДЕЛЕННОГО КОНТЕНТ-МОНИТОРИНГА ГЛОБАЛЬНЫХ ИНФОРМАЦИОННЫХ СЕТЕЙ

Исследовано свойства информационных потоков с учетом их принадлежности к сегментам глобальных сетей. В статье рассматривается процесс осуществления контент-мониторинга с учетом постоянного повышения сложности сегментов сети Интернет и предлагаются для применения практические методы построения систем распределенного контент-мониторинга. Особое значение этот вопрос приобретает при осуществлении деятельности, направленной на предотвращение, своевременное выявление, пресечение или нейтрализацию реальных и потенциальных угроз информационной безопасности государства. Организация эффективного противодействия вызовам и угрозам национальной безопасности и ее конкретным проявлениям в информационной сфере – одна из насущных задач, решение которой является важным делом в условиях развития гражданского общества и правового государства современности. Охват и обобщение больших динамических контентных потоков, которые непрерывно генерируются в Интернет-пространстве, требует качественно новых методов и подходов к выполнению мероприятий по обеспечению информационной безопасности. Изучение информационных потребностей и формирование требований к построению аналогичных систем показывает, что в процессе осуществления контент-мониторинга сети Интернет важным является применение средств распределенного (сегментированного) контент-мониторинга глобальных сетей/медиа, а также использования полного набора интерфейсов взаимодействия сегментов глобальных сетей и инструментария их мониторинга (сканирования). Сложность сегментов сети Интернет, как программно-технических объектов является существенным, а не второстепенным свойством. Проектирование программно-технических средств распределенного контент-мониторинга глобальных сетей/медиа, которые абстрагируются от их сложности, часто абстрагируются и от их сущности. Предложено решение таких задач на базе системы контент-мониторинга InfoStream. При построении экспертных систем распределенного контент-мониторинга, усматривается необходимость учета уровня популярности (доступности) сегмента сети Интернет и его оперативности в формировании контента, различных подходов к накоплению и предоставлению информации, интерфейсов взаимодействия, поэтому для каждого "выделенного" сегмента глобальной сети Интернет целесообразно формировать собственную базу знаний и приемов.

**Ключевые слова:** контент-мониторинг, распределенные системы мониторинга, информационный поток, сложные программные объекты, экспертные системы.

DMYTRO LANDE,  
YAROSLAV KONDRATENKO

## **FEATURES OF CONSTRUCTION SYSTEMS OF DISTRIBUTED CONTENT-MONITORING OF GLOBAL INFORMATION NETWORKS**

The properties of information flows are examined taking into account their belonging to segments of global networks. The article describes how to implement content-monitoring with taking to account continuous increasing complexity of the Internet segments and available for use practical methods of construction of distributed content monitoring systems. Particular importance attaches to this issue in carrying out activities aimed at preventing, timely detection, termination or neutralization of real and potential threats to state information security. Organization of effective counteraction to the challenges and threats to national security and its specific manifestations in the information sphere is one of the urgent tasks, the solution of which is an important task in terms of building civil society and the rule of law. Capture and synthesis of large dynamic content streams that are continuously generated in the Internet space requires qualitatively new methods and approaches to implement information security measures. Study of information needs and requirements formation for the implementation of similar systems suggests that the usage of distributed (segmented) content monitoring of global networks/media and full set of interaction interfaces of segments of global networks and their monitoring tools is important in the process of content monitoring (scanning) of the Internet. The complexity of segments of the Internet as a software and technical objects is essential, not a secondary property. Designing software and hardware for distributed content monitoring of global networks / media, abstracting from both their complexity and their essence. The solution of such problems is based on the InfoStream content monitoring system. When building expert systems of distributed content monitoring, perceived the need to consider the level of popularity (availability) segment of the Internet and its efficiency in shaping the content, of different approaches to collecting and providing information interfaces interaction, so each “singled-out” segment of the global Internet is expedient of forming own knowledge base and receptions.

**Keywords:** content-monitoring, distributed monitoring systems, information flow, complex software facilities, expert systems.

**Дмитро Володимирович Ланде**, доктор технічних наук, старший науковий співробітник, завідувач відділом спеціалізованих засобів моделювання, Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна.

E-mail: dwlande@gmail.com.

**Ярослав Анатолійович Кондратенко**, аспірант, Інститут спеціального зв'язку та захисту інформації Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна.

E-mail: yaruhka@gmail.com.

**Дмитрий Владимирович Ландэ**, доктор технических наук, старший научный сотрудник, заведующий отделом специализированных средств моделирования, Институт проблем регистрации информации Национальной академии наук Украины, Киев, Украина.

**Ярослав Анатолиевич Кондратенко**, аспирант, Институт специальной связи и защиты информации Национального технического университета Украины “Киевский политехнический институт имени Игоря Сикорского”, Киев, Украина.

**Dmytro Lande**, doctor of technical science, senior researcher, head of the specialized modeling tools department, Institute for information recording of National academy of science of Ukraine, Kyiv, Ukraine.

**Yaroslav Kondratenko**, postgraduate student, Institute of special communication and information protection of National technical university of Ukraine “Igor Sikorsky Kyiv polytechnic institute”, Kyiv, Ukraine.