

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ НАН УКРАИНЫ

НАЦИОНАЛЬНАЯ АКАДЕМИЯ ПРАВОВЫХ НАУК УКРАИНЫ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ ИНФОРМАТИКИ И
ПРАВА

НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ УКРАИНЫ
«КИЕВСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ»

УЧЕБНО-НАУЧНЫЙ ЦЕНТР ИНФОРМАЦИОННОГО ПРАВА И
ПРАВОВЫХ ВОПРОСОВ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
ФАКУЛЬТЕТА СОЦИОЛОГИИ И ПРАВА

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И БЕЗОПАСНОСТЬ

МАТЕРИАЛЫ XV МЕЖДУНАРОДНОЙ
НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

ВЫПУСК 15

Киев – 2015

*Рекомендовано к печати ученым советом
Института проблем регистрации информации НАН Украины
(протокол № 12 от 15 сентября 2015 г.)*

**Информационные технологии и безопасность. Материалы XV
Международной научно-практической конференции ИТБ-2015.** – К.:
ИПРИ НАН Украины, 2015. – 250 с. ISBN: 978-966-2344-45-5

В сборник вошли материалы докладов, представленных на XV
Международной научно-практической конференции «Информационные
технологии и безопасность» (ИТБ-2015, 21 октября 2015 года, г. Киев,
Украина).

В сборнике представлены статьи, посвященные вопросам внедрения
информационных технологий, актуальным проблемам обеспечения
информационной и кибербезопасности, противодействия информационной
агрессии и кибертерроризму, проведения информационно-аналитических
исследований на основе контента сети Интернет, правового обеспечения
информационной безопасности.

Для специалистов в области информационных технологий,
информационной безопасности, информационного права, а также для
аспирантов и студентов старших курсов высшей школы соответствующих
специальностей.

Редакционная коллегия:

*А.Г. Додонов, д.т.н., профессор; В.Г. Пилишчук, д.ю.н., профессор, член-корр.
НАПрН Украины; А.М. Богданов, д.т.н., профессор; Д.В. Ландэ, д.т.н.,
с.н.с.; В.В. Мохор, д.т.н., профессор; Н.А. Ожеван, д.ф.н., профессор; В.Н.
Фурашев, к.т.н., с.н.с.; Е.С. Горбачик, к.т.н., с.н.с.; М.Г. Кузнецова, к.т.н.,
с.н.с.*

ISBN 978-966-2344-45-5

- © Институт проблем регистрации информации НАН Украины, 2015
- © Научно-исследовательский институт информатики и права НАПрН Украины, 2015
- © Учебно-научный центр информационного права и правовых вопросов информационных технологий ФСП НТУУ «КПИ», 2015
- © Коллектив авторов

ЖИВУЧІСТЬ ІНФОРМАЦІЙНИХ ПОТОКІВ У СУЧАСНИХ ІНФОРМАЦІЙНИХ СЕРЕДОВИЩАХ

Березін Б.¹, Ланде Д.¹,

Павленко О.²

¹ІПРІ НАН України,

²Відкритий міжнародний університет розвитку людини "Україна"

Постановка проблеми, її актуальність та аналіз публікацій Сучасний інформаційний простір є динамічною системою, яка складається з інформаційних елементів (документів) пов'язаних між собою за тематикою. Більшість інформаційних ресурсів, які на сьогодні публікують та зберігають інформацію в цифровому вигляді, розміщують інформацію он-лайн чи на цифрових носіях та мають потокову структуру [1]. Прикладами таких ресурсів є соціальні мережі, стрічки новин, блоги. Публікацію наукових статей також можна розглядати як інформаційний потік (ІП) – тобто послідовність документів, які відповідають певній тематиці та можуть розміщуватися на веб-серверах інституційних репозитарієв. Нарешті, документи, що розміщуються для довготермінового або постійного зберігання в електронних, комп'ютерних бібліотеках, архівах, страхових фондах тощо, теж мають звичай форму ІП. Такому тематичному ІП ставлять у відповідність часовий ряд, який складається з кількості повідомлень або документів опублікованих за певний проміжок часу. Поняття живучості інформаційних об'єктів в Інтернет-середовищі чи на носіях має на увазі їх здатність своєчасно виконувати свої функції (інформування) в умовах дії дестабілізуючих факторів. Такими факторами можуть бути їх усунення з інформаційного простору, втрата ними властивостей актуальності, доступності [2,3].

В роботі розглядається живучість ІП у різних видах сучасних середовищ: живучість ІП новин з веб-сторінок соціальної мережі Twitter; живучість ІП наукових публікацій (НП) з сторінок відповідних веб-репозитарієв, електронних бібліотек в Інтернеті; живучість ІП ресурсів на оптичних носіях, що довготерміново зберігаються в комп'ютерних та електронних архівах, страхових фондах. Існування ІП у різних середовищах передбачає подолання різних видів загроз: пошкодження носіїв інформації/обладнання, старіння носіїв/обладнання, старіння програмного забезпечення/форматів, помилки операторів, атаки, природні катастрофи, економічні помилки і т.і., тобто передбачає забезпечення живучості цих ІП. На сьогодні більшість досліджень в області інформаційних потоків присвячена новинним ІП. (В роботі [4] запропонована логістична модель для розгляду динаміки новинних ІП, в роботах [2-3] розглядається живучість повідомлень, роботи [1,5] стосуються аналізу новинних ІП). В той же час, живучість ІП НП, ІП на носіях детально не досліджені.

Метою роботи є розробка методологічних засад забезпечення живучості інформаційних потоків у різних сучасних середовищах на основі аналізу характеристик ІП та розробки відповідних моделей для оцінки живучості ІП.

Характеристики ІП та моделі для оцінки їх живучості

Аналіз показує, що серед основних особливостей вищезазначених ІП, які впливають на їх живучість, можна виділити доступність, републікацію, індексованість та поширеність форматів даних.

Доступність. Оцінка живучості ІП в Інтернет-середовищі залежить від доступності веб-серверів, на яких розміщені інформаційні елементи (ІЕ). Для визначення характеру цієї залежності було налагоджено моніторинг стану сайтів, на яких може бути розміщено контент. На основі априорного матеріалу було розраховано дані й побудовано розподіли випадкових значень показника доступності й недоступності сайтів, а також періодів їх доступності та недоступності [6] (відповідні ранжирувані розподіли можуть бути апроксимовані експоненціальною та степеневою функціями).

Оцінка живучості ІП на оптичних носіях залежить від їх доступності, тобто можливості читування файлів. Внаслідок особливостей технології запису інформації, оптичні диски вміщують деяку кількість помилок, число яких зростає в процесі довготермінового зберігання. Можливості системи корекції стандартного оптичного приводу обмежені, тому наявність помилок характеризує зниження функціональності носіїв інформації. Для визначення живучості інформації на оптичних дисках з точки зору доступності може використовуватися показник помилок. (Для DVD це PI Sum8, його максимально допустиме значення складає 280). Для колекції CD та DVD дисків було проведено тестування помилок. Аналіз розподілів виборок оптичних дисків, ранжируваних за кількістю помилок, показав можливість їх апроксимації степеневою функцією [7]. Доступність ІЕ на сапфірових дисках для архівного зберігання [8,9] може оцінюватися за допомогою показника помилок PI Sum8 (який характеризує відповідність записаної інформації основним вимогам стандарту DVD) з врахуванням часової стабільності металевого відбиваючого покриття.

Републікація. При користуванні інформаційними ресурсами в мережі Інтернет може відбуватися копіювання, розмноження їх версій, тобто републікація. Оцінка живучості ІП в Інтернет-середовищі в значній мірі залежить від кількості версій ІЕ. В ІП новин з соціальної мережі Twitter в якості показника републікації для твітів може бути використано значення ретвітів. Для визначення характеристик процесу републікації контенту ІП НП в Інтернет використовувалися запити до пошукової системи Google Scholar для отримання списку публікацій з заданої тематики на проязі періоду в один рік [6]. В отриманому списку публікацій аналізувалось поле "всі версії статті" та були проранжирувані кількості версій НП по вибраній тематичній в Інтернет за рік. Отримані таким чином розподіли для виборок розміром біля 100 НП в усі роки інтервалу 1998 – 2013 р.р. були апроксимовані за допомогою експоненціальної функції. Для ІП на оптичних

носіях в якості характеристики републікації може використовуватися кількість копій носіїв (які доцільно зберігати в різних сховищах).

Індексованість. Оцінка живучості ІП при довготерміновому зберіганні в Інтернет-середовищі також залежить від індексованості веб-серверів, на яких розміщені ІЕ потоків. Індексованість новинних ІП мережі Twitter забезпечується пошуковими системами Google, Yandex, Meta тощо. Для визначення характеру цієї залежності для ІП ІП було використано запити до пошукових систем Google та Scholar Google. Результати показують, що доля файлів, проіндексованих в Scholar Google по відношенню до файлів, проіндексованих Google, може змінюватися від одиниць до десятків відсотків, в залежності від конкретного електронного архіву. Відповідно, може змінюватися властивість ІП виконувати свої функції, тобто живучість. Для більшої живучості ІП на носіях, інформація про них повинна бути доступна в Інтернет або у спеціальних системах. Наприклад, для Страхового фонду документації (СФД) [9] ведеться державний реєстр документів СФД – автоматизована система з обмеженим доступом (на початок 2013 р. до неї було внесено більше 33 тис. документів).

Поширеність та стабільність форматів. Оцінка живучості ІП в Інтернет-середовищі або на носіях також може залежати від поширеності форматів, використаних для їх представлення. Тобто, на їх живучість може впливати старіння форматів, в яких представлені ці ІП (поява нових форматів, ІЗ і обладнання, несумісного зі старими форматами представлення даних). В роботі [6] наведено результати аналізу щодо поширеності форматів в Інтернет (html – 90%, pdf – 2%-3% тощо), а також поширеності версій формату pdf (1.0, 1.1, 1.2, 1.7 – одиниці відсотків, 1.3-1.6 десятки відсотків ресурсів).

Для забезпечення доступу до інформації через значний інтервал часу необхідно використання стабільних форматів представлення даних, що дозволяє істотно подовжити час між послідовними конверсіями електронних документів та зробити цей процес більш передбачуваним. Використання відкритих форматів (наявність вичерпного опису, відсутність правових обмежень і т.і.) гарантує, що за певних зусиль документи можна буде прочитати навіть через тривалий час. До міжнародних стандартів відкритих форматів документів відносять формати серії pdf.

Моделі для оцінки живучості ІП. Із-за протяжності у часі ІП його елементи мають мережеву структуру, тобто більш нові елементи потоку можуть посилатися на ресурси попередніх елементів. (У текстах новинних повідомлень – посилання на попередні новини, на веб-сторінки, відео тощо, у текстах наукових публікацій – посилання на публікації, презентації, додатки і т.і., в текстах геологічних звітів на оптичних носіях - посилання на публікації, на попередні звіти тощо.)

На основі наведених вище особливостей представлення ІП в середовищі Інтернет та на носіях, будемо вважати, що живучість ІЕ залежить від наступних основних факторів:

- живучості тексту ІЕ, на яку впливають кількість копій публікації в Інтернет-середовищі та на носіях (в загальному випадку будемо розглядати кількість версій публікації, тобто не лише повнотекстові копії, а й анотації статей); доступності серверів, на яких зберігаються копії та версії ІЕ; індексованості тексту публікації в універсальних, наукових та спеціальних пошукових системах; поширеності формату, в якому представлено текст тощо;
- частки доступних посилань, що використовуються в ІЕ;
- живучості ІЕ, на які є посилання в тексті розглядаемого ІЕ (теж залежить від кількості копій, версій, доступності серверів, індексованості, поширеності форматів тощо).

З урахуванням вище наведеного, живучість ІЕ при довготерміновому зберіганні в Інтернет-середовищі або на носіях можна оцінювати на основі кількості версій ІЕ та її посилань в Інтернет/на носіях. Живучість ІЕ будемо представляти двома значеннями: живучістю тексту ІЕ та живучістю її посилань.

Живучість тексту ІЕ будемо оцінювати кількістю версій ІЕ з урахуванням доступності, індексації та поширеності формату:

$$ST = \sum_{i=1}^{VT} AT_i * IT_i * PFT_i \quad (1)$$

ST – живучість тексту ІЕ;

AT_i – доступність i -ї версії тексту ІЕ;

IT_i – індексованість в пошуковій системі i -ї версії тексту ІЕ;

PFT_i – поширеність формату i -ї версії тексту ІЕ в Інтернет.

Живучість посилань ІЕ будемо оцінювати усередненою кількістю версій посилань з урахуванням доступності та індексації, а також з урахуванням загальної частки доступних посилань.

$$SR = \frac{RRL}{RRC} * \frac{\sum_{j=1}^{RR} \sum_{i=1}^{VR_j} AR_{ij} * IR_{ij} * PFR_{ij}}{RRL} \quad (2)$$

SR – живучість ресурсів, на які посилляється ІЕ;

AR_{ij} – доступність i -ї версії j -го ресурсу, на який посилляється ІЕ;

IR_{ij} – індексованість в пошуковій системі i -ї версії j -го ресурсу, на який посилляється ІЕ;

PFR_{ij} – поширеність формату i -ї версії j -го ресурсу ІЕ в Інтернет.

VR_j – кількість версій j -го ресурсу, на який посилляється ІЕ;

RRL – кількість “живих” посилань, на який посилляється ІЕ;

RRC – загальна кількість Інтернет-посилань, які є в ІЕ.

Методологічні засади забезпечення живучості інформаційних потоків

В роботах [2,3,6] розглянуто кілька механізмів, що забезпечують живучість інформаційних потоків в Інтернеті. Зазвичай, ці механізми використовуються не в чистому вигляді, а комбінуються. Дзеркальне копіювання даних при розміщенні їх на цільовий ресурс. Тобто автор розміщує інформацію, яка копіюється хостинг-провайдером на деяку кількість дзеркальних серверів. (Приклад – відома служба WikiLeak, використовує кілька сотень серверів-дзеркал). Передрук інформації (републікації, «копіпаст») на інші сайти з метою їх інформаційного наповнення. Найбільш важлива і цікава інформація передруковується сотні разів, у той час як неактуальна, нецікава інформація практично не дублюється. Розміщена одного разу інформація назавжди потрапляє в архівні служби Інтернету типу Internet Archive. Інформація індексується глобальними пошуковими системами типу Google, Яндекс та залишається у їх кеш-пам'яті, звідки вона доступна користувачам.

І нарешті, інформація з веб-сайту може зберігатися на локальних комп'ютерах кінцевих користувачів, які отримали до неї доступ або безпосередньо, або через інтеграторів інформації.

Особливості оцінки живучості НП з використанням запропонованої вище моделі може бути показана на прикладі наукової періодики України, представленої на сайті Національної бібліотеки України імені В.І. Вернадського. Для вибірки НП окремого наукового видання з цього електронного архіву середнє значення кількості версій НП в Інтернет середовищі складає приблизно 2 (на основі даних пошукових систем Google Scholar та Google, аналізувалися 2008-2010 р.р.). Доступність серверів, на яких зберігаються версії НП видання наукової періодики України може оцінюватися за допомогою сервісу <http://uptimerobot.com>. Індексованість НП в електронному архіві наукових публікацій є незначною. Для порівняння оцінювалася також середня кількість версій НП в Інтернет-середовищі для одного з найбільших електронних архівів наукових публікацій (<http://arxiv.org>). Для вибірки за 2008-2010 р.р. вона складає біля 7. Враховуючи те, що кількість версій НП є однією з основних складових живучості, це показує порівняно невисокий рівень живучості НП періодичних видань України при довготерміновому зберіганні в Інтернет-середовищі. Виходячи із запропонованої структури живучості НП, серед основних шляхів її підвищення: збільшення кількості версій НП в Інтернет-середовищі; підвищення доступності серверів, де розміщуються версії НП; підвищення рівня індексованості НП в пошукових системах; використання більш нових форматів тощо.

Доцільним напрямком збільшення кількості версій НП є використання архівних сервісів мережі Інтернет (webarchive.org, website.org та деяких

інших) поряд з розміщенням додаткових версій НП у відкритих репозитаріях, що створюються в університетах та інших наукових установах. Але, при використанні архівних сервісів для збільшення кількості версій НП важливо забезпечити їх індексацію, тобто представлення їх адрес у результатах обробки запитів пошуковими системами. Рішенням цієї проблеми може бути розміщення адрес копій в полях метаданих НП (наприклад, в міжнародному стандарті метаданих для архівних матеріалів ISAD (G) передбачені дані про наявність та місцезнаходження копій).

Інструментальні засоби моніторингу інформаційних потоків

Для отримання статистичних даних про ІП з метою побудови моделей живучості, створюються відповідні програмні засоби. Програма моніторингу для соціальної мережі Twitter розроблена на R-мові програмування та забезпечує збір даних про ретвіти та вибране з новинних повідомлень мережі. Для збору даних про ІП наукових публікацій на R-мові створюються програмні засоби доступу до сервісу Google Scholar Citations. При цьому було використано також набір дистрибутивів Denver для налагодження веб-додатків на локальному ПК. Для тестування показників помилок при зберіганні інформації на оптичних дисках було використано програму Opti Drive Control 1.51, пакет PlexTools Professional V2.12, та деякі інші засоби.

Висновки

В роботі визначено особливості представлення ІП в мережі Інтернет та на оптичних носіях, які впливають на живучість ІП: републікація; доступність серверів, на яких зберігаються ІП; індексація ІП в пошукових системах; поширеність форматів даних. На базі цих особливостей запропоновано моделі для оцінки живучості ІП в Інтернет-середовищі та на носіях, показано їх можливості на прикладі української наукової періодики.

Використання цих моделей для оцінки живучості ІП дозволяє сформулювати рекомендації по підвищенню їх живучості на основі збільшення кількості версій ІЕ, підвищення доступності та індексованості цих версій, впровадження сучасних форматів даних або конвертації форматів.

Література

1. Kleinberg J. Temporal dynamics of on-line information streams // Data Stream Management: Process-ing High-Speed Data Streams. — Springer, 2006. —18 p.
2. Додонов А.Г., Ландэ Д.В. Живучість інформаційних об'єктів в сети Інтернет // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2013): материалы III междунар. науч.-техн. конф. (Минск 21-23 февраля 2013 года) / - Минск: БГУИР, 2013. - С. 457-460.

3. Додонов А.Г., Ландэ Д.В. Живучість інформаційних повідомлень в Інтернет-середі //Комп'ютерні системи та мережні технології (CSNT-2012): тези доповідей V міжнародної науково-технічної конференції. м. Київ, 13-15 червня 2012 р. Національний авіаційний університет. - К.: НАУ, 2012. - С. 49-51.

4. Ландэ Д.В. Основы интеграции информационных потоков: Монография -К.: Инжиниринг, 2006. - 240 с.

5. Baranovskiy O. Recurrence plots as method for information flows analysis// The Advanced Science Journal, Volume 2015, Issue 2. – P.45-49

6. Ландэ Д.В., Березін Б.О. Підхід до оцінки живучості наукових публікацій при довготерміновому зберіганні в інтернет-середовищі // Ресстрація, зберігання і обробка даних, 2014. - Т. 16. - N 4. - С. 34-43.

7. Березін Б.О., Ландэ Д.В. Дослідження стану оптичних носіїв при довгостроковому зберіганні цифрової інформації // Студії з архівної справи та документознавства / Держ. архів. служба України, УНДІАСД. - К., 2012. - Т. 20. - С. 133-139.

8. Петров В.В., Семиноженко В.П. Новітня технологія довготривалого зберігання інформації на сапфірових оптичних дисках/Вісн. НАН України, 2014, № 4 – С. 24-32.

9. Березін Б.О., Ландэ Д.В., Шиховець О.В. Живучість інформаційних об'єктів при довготерміновому зберіганні із використанням архівних носіїв //Информационные технологии и безопасность: основы обеспечения информационной безопасности: Материалы международной научной конференции ИТБ-2014. - К.: ИПРИ НАН Украины, 2014. - С. 40-45.

10. Грищенко М.О., Полонський О.І., Піскачов О.І. Створення страхового фонду документації – одна із складових національної безпеки держави // Страховий фонд документації, 2014. - N 2. - С. 25-31.

СОДЕРЖАНИЕ

Додонов О.Г., Ландэ Д.В., Коваленко Т.В.

АРХІТЕКТУРА СИСТЕМИ ІНФОРМАЦІЙНОЇ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ МОНИТОРИНГУ ІНФОРМАЦІЙНОГО ПРОСТОРУ..... 3

Фурашев В.

ПРО ДЕЯКІ ПРИНЦИПОВІ МОМЕНТИ У СФЕРІ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ..... 7

Авраменко М.

ОСОБЛИВОСТІ ПРАВОВОГО ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ В СУЧАСНИХ УМОВАХ..... 9

Архипов О., Архипова Є.

РИЗИКОВИЙ ПІДХІД ДО ВИЗНАЧЕННЯ ОБ'ЄКТУ ОПТИМАЛЬНИХ ІНВЕСТИЦІЙ У БЕЗПЕКУ ІНФОРМАЦІЇ..... 12

Балагура І.В., Ландэ Д.В.

ВИЗНАЧЕННЯ ЕКСПЕРТНИХ ГРУП ФАХІВЦІВ НА ОСНОВІ АНАЛІЗУ РЕФЕРАТИВНОЇ БАЗИ ДАНИХ..... 18

Березін Б., Ландэ Д., Павленко О.

ЖИВУЧИСТЬ ІНФОРМАЦІЙНИХ ПОТОКІВ У СУЧАСНИХ ІНФОРМАЦІЙНИХ СЕРЕДОВИЩАХ..... 24

Бояринова Ю.Б., Городько Н.А.

О МОДЕЛІРОВАНИИ СИСТЕМЫ УПРАВЛЕНИЯ КАЧЕСТВОМ ИНФОРМАЦИОННЫХ УСЛУГ..... 31

Брицький С.О., Ландэ Д.В.

ВИКОРИСТАННЯ ГЕОГРАФІЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ З ВІДКРИТИМ КОДОМ ДЛЯ ВІДОБРАЖЕННЯ ТА ГЕОАНАЛІЗУ ІНФОРМАЦІЙНИХ ПОТОКІВ СИСТЕМ КОНТЕНТ-МОНИТОРИНГУ..... 40

Верголяс О.О.

ПЕРСПЕКТИВИ КРИМІНАЛІЗАЦІЇ СТАТТІ 173-1 КОДЕКСУ УКРАЇНИ ПРО АДМІНІСТРАТИВНІ ПРАВОПОРУШЕННЯ..... 42

Галушко М.М., Самаріна М.О., Ільницька Г.Т.

ДО ПИТАННЯ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ СИСТЕМИ ОРГАНІЗАЦІЇ УПРАВЛІННЯ ВИЩИМИ НАВЧАЛЬНИМИ ЗАКЛАДАМИ..... 47