

Д. Ланде, доктор технічних наук, завідувач відділу спеціалізованих засобів моделювання Інституту проблем реєстрації інформації НАН України

Аналіз інформаційних потоків у глобальних комп'ютерних мережах (за матеріалами наукової доповіді на засіданні Президії НАН України 25 січня 2017 р.) // Вісник НАН України. – 2017. – № 3. – С. 46–54.

Вступ

У глобальних мережах міститься величезна кількість інформації, за допомогою якої можна вирішувати найрізноманітніші завдання. Однак при цьому важливо вибрати саме ту інформацію, яка цікавить користувача.

Велика кількість інформаційних ресурсів у глобальних мережах містить різні експертні оцінки, певна частина яких пов'язана з реалізацією інформаційних впливів, здійсненням спрямованих інформаційних операцій, веденням інформаційних війн. Такі матеріали можуть бути проаналізовані, узагальнені, на їх основі можна створювати бази для подальшого прийняття рішень, які відрізняються від традиційних експертних оцінок як за обсягами, так і за рівнем об'єктивності.

Крім того, у мережах може розміщуватися інформація, пов'язана з організацією протиправної діяльності, тероризмом. Відомі також ефекти так званої *мережевої мобілізації*, впливу на людську свідомість, управління і маніпулювання громадською думкою. Аналізуючи інформаційні потоки, можна знайти і так звані *інформаційні резервації* – частини інформаційного простору, які характеризуються замкненістю, обмеженістю тематики даних.

Отже, врахування інформації з мережевих джерел відіграє важливу роль як для виявлення напрямів розвитку економіки, науки, технологій та інших сфер життя, так і для вирішення конкретних завдань у сферах безпеки людини, суспільства, держави.

Деякі параметри інформаційного простору

На сьогодні кількість інформаційних ресурсів у глобальних мережах перевищує сотні трильйонів документів. У 2014 р. система Google вже індексувала в мережі 60 трлн документів, причому ці документи розміщуються не лише на веб-сайтах. За даними відомого інтернет-сервісу Netcraft (<http://netcraft.com>), у 2014 р. кількість веб-сайтів у мережі вже перетнула позначку в мільярд, а нині становить понад 1,7 млрд (рис. 1).

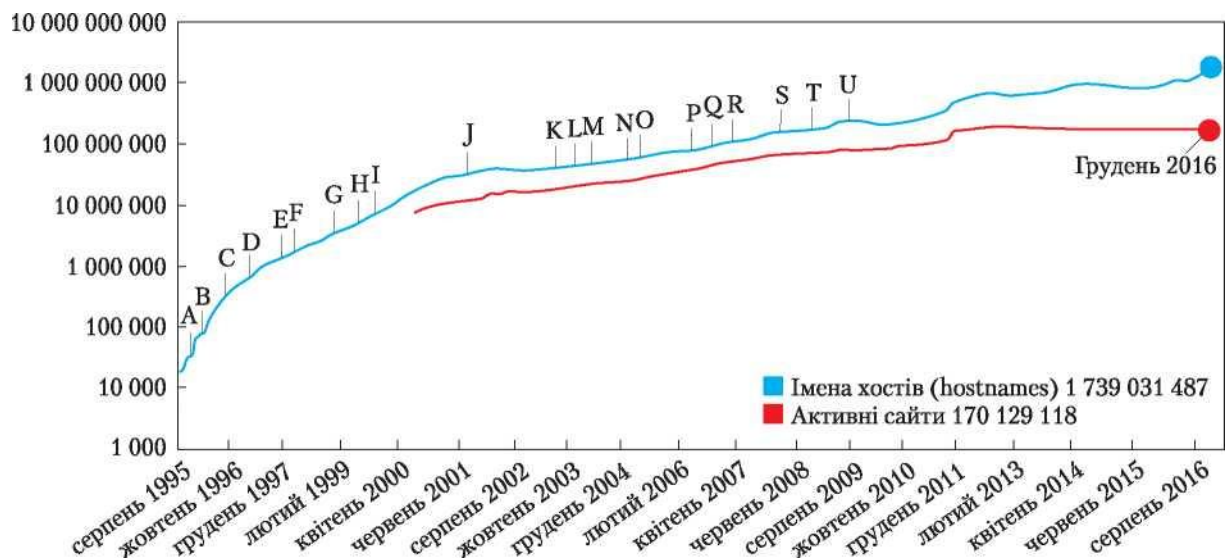


Рис. 1. Динаміка кількості сайтів в усіх доменах (за даними Netcraft, логарифмічна шкала)

У 2014 р. у виданні *Supercomputing Frontiers and Innovations* з'явилась публікація, в якій стверджувалося, що приблизний обсяг даних у мережі Інтернет сягає 1024 байтів, тобто один йотабайт. Лише в одній соціальній мережі Facebook активні користувачі генерують більш як 4 петабайти даних за добу. За даними компанії CISCO, обсяг інтернет-трафіку в 2016 р. досяг одного зеттабайта, тобто 1 099 511 627 776 гігабайтів.

Технологічні рішення

При збиранні й аналізі таких даних виникають проблеми, пов'язані з *обробкою надвеликих обсягів даних, пошуком і навігацією в динамічних інформаційних потоках*. Величезна кількість багатомовних інформаційних ресурсів зумовлює складність їх використання при здійсненні інформаційно-аналітичної роботи. Для вирішення цих проблем сьогодні застосовуються такі технологічні концепції, як Big Data (великі дані, рис. 2), Complex Networks (складні мережі), Cloud Computing (хмарні обчислення), Data/Text Mining (глибинний аналіз даних і тексту).

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

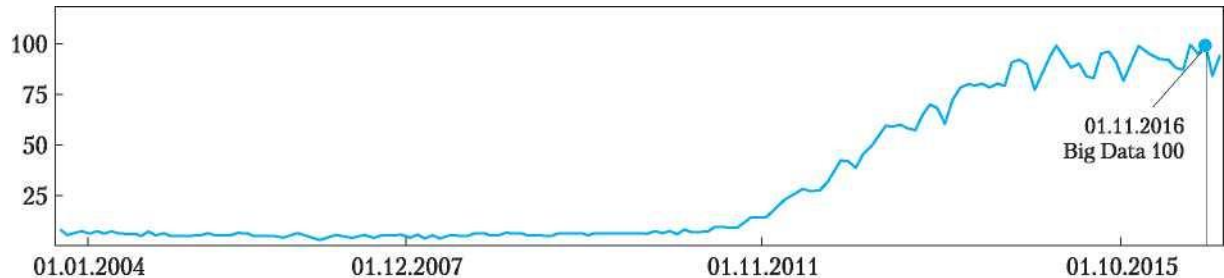


Рис. 2. Аналіз терміна Big Data за допомогою Google Trends

У світі й досі залишаються невирішеними завдання ефективної аналітичної обробки інформації з глобальних мереж, оперативного вилучення необхідних фактографічних даних, виявлення трендів в окремих предметних областях, розпізнавання змістових аномалій, прогнозування тощо. Більшість із зазначених завдань – це актуальні проблеми семантичної обробки надвеликих динамічних масивів інформації. Навіть спроби часткового практичного вирішення цих проблем зумовили успішність таких проектів, як пошукові системи Google, Yandex, Baidu, системи моніторингу соціальних мереж (SMM) типу Keyhole, Brandwatch, CyberAlert, аналітичні системи типу Palantir, Centrifuge, i2 та ін. В Україні до таких систем можна віднести, зокрема, системи, створені на базі наукових результатів, отриманих в Інституті проблем реєстрації інформації (ІПРІ) НАН України, – система контент-моніторингу InfoStream, аналітична система X-SCIF, система сканування ресурсів соціальних медіа Robusta тощо.

Розробки Інституту проблем реєстрації інформації НАН України

В ІПРІ НАН України було теоретично обґрунтовано і створено засоби:

- моделювання інформаційних потоків у глобальних комп'ютерних мережах, зокрема мультиагентну модель розповсюдження інформації;
- розподіленого контент-моніторингу глобальних мереж;
- формування багатомовних повнотекстових баз даних;
- аналізу динаміки тематичних інформаційних потоків, зокрема, вперше застосовано вейвлет-аналіз до задач виявлення інформаційних операцій;
- прогнозування розвитку подій на основі фрактального аналізу;
- формування мереж взаємозв'язку понять, що екстрагуються із тестових масивів, і аналізу цих мереж;
- автоматичного формування моделей предметних областей.

Мультиагентна модель розповсюдження інформації. Для моделювання тематичних інформаційних потоків як полігон для подальших досліджень в ІПРІ НАН України створено мультиагентну модель поширення інформації в соціальних мережах ¹. Для цього формується близький до реальності віртуальний інформаційний простір, населений віртуальними агентами, з якими асоціюються окремі повідомлення в соціальній мережі і які інкапсулюють у собі гіперпосилання на інформаційні ресурси мережі Інтернет. Передбачається, що окремі агенти можуть самозароджуватися; породжувати нових агентів шляхом репостингу (repost); «вмирати» – зникати з простору агентів; отримувати лайки (like) від інших агентів. Кожен агент має «потенціал», залежний від часу його життя, авторитетності (гіперпосилань, проставлених на нього) і плодючості (кількості породжених безпосередньо ним агентів).

Варіювання відповідними параметрами моделі дає можливість змоделювати різноманітні профілі поведінки інформаційних сюжетів. На рис. 3 наведено приклад можливої динаміки мультиагентної системи.

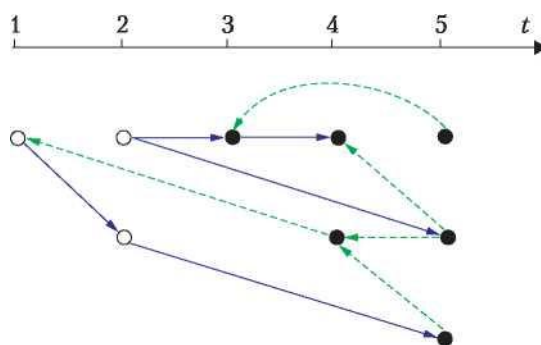


Рис. 3. Приклад динаміки мультиагентної системи

У результаті проведених досліджень було реалізовано програму еволюції простору агентів, досліджено еволюцію мультиагентної системи, знайдено аналогії з реальними тематичними інформаційними потоками. Виявлено статистичні закономірності, що стосуються життєвого циклу окремих повідомлень, розподіл яких відповідає розподілу Вейбулла. Дані

¹ Lande D.V., Hraivoronska A.M., Berezin B.O. Model of information spread in social networks. *European Journal of Natural History*. 2016. (5): 41; Lande D., Dodonov V., Kovalenko T. Corporate system of monitoring of network information resources on the basis of a multiagent approach. *Information Technology and Security*. 2016. 4(1): 4. [Ланде Д., Додонов В., Коваленко Т. Корпоративная система мониторинга сетевых информационных ресурсов на основе мультиагентного подхода. *Information Technology and Security*. 2016. Т. 4, № 1. С. 4–12.]

модельовання було підтверджено шляхом порівняння з реальною мережею мікроблогів Twitter.

Повний текст ([PDF](#))