

Системна інформатизація інформаційного супроводження процесів: статистичне дослідження уваги Інтернет-ЗМІ до визначеної проблематики

Вступ

Як було визначено в роботі [1], інформаційне суспільство – це безперешкодний і гарантований доступ різних верств населення до інформації, знань. Динаміка та обсяги представлення новинної інформації в Інтернет створюють потужний інформаційний потік [2]. Причому потік досить неоднорідний, який може характеризуватися великою кількістю параметрів, серед яких виділяються такі, як джерела інформації (наприклад, Web-сайти) і тематики. Інформація, яка сьогодні циркулює у цьому потоці, є невичерпним за обсягом ресурсом, а також відображенням громадської думки у віртуальному просторі. За своєю природою інформація, що породжується і розповсюджується у мережі Інтернет, має бути віднесена до категорії електронних засобів інформації. Більшість інформаційних агентств, друкованих ЗМІ, радіо та телеканалів не тільки мають свої представництва в Мережі, але й створюють там повноцінні інформаційні ресурси.

Саме тому питання системної інформатизації інформаційного супроводження соціально-економічних, соціально-політичних процесів, процесів аналізу та прийняття рішень набувають надзвичайної актуальності.

Свій погляд та пропонуємо рішення автори демонструють на прикладі статистичних досліджень уваги Інтернет-ЗМІ до проблематики євроатлантичної інтеграції України. Цілком зрозуміло, що пропонує мий підхід є досить уніфікованим та може застосовуватися при вирішенні питань інформаційного супроводження будь-яких соціально-політичних та соціально-економічних процесів, процесів аналізу та прийняття рішень.

Євроатлантична інтеграція – питання далеко не безперечне, це процес, який має враховувати думку всіх верств суспільства. Як один з інструментів дослідження уваги суспільства до цієї проблематики пропонується засоби аналізу статистичних показників щодо обсягу публікацій за названою тематикою в мережі Інтернет.

Для проведення необхідних досліджень щодо уваги Інтернет-ЗМІ до проблематики інтеграції України з НАТО авторами було застосовано бази даних сервера інформаційної підтримки прийняття рішень, встановленого у Національному центрі з питань євроатлантичної інтеграції України [3]. Однією з основних складових цього сервера є комплекс контент-моніторингу інформації з мережі Інтернет, побудований на основі застосування системи InfoStream (<http://www.infostream.ua>), яка розроблена в Інформаційному центрі «Електронні вісті». При цьому на сервері реалізовано повномасштабне інформаційне сховище, що враховує особливості євроатлантичної проблематики, накопичує та надійно зберігає інформацію для використання в аналітичній роботі. Комплекс контент-моніторингу виконує основну “чорнову” роботу зі збору інформації з мережі Інтернет та забезпечує створення та постійне поповнення документального сховища оперативними повідомленнями, які застосовувалися, у тому числі і для проведення статистичних досліджень.

Типова задача комплексу контент-моніторингу - це побудова діаграм динаміки появи понять у часі. На Рис. 1, наприклад, проілюстровано, як за допомогою цієї діаграми відслідковується поява україномовних повідомлень щодо євроатлантичної інтеграції України у лютому 2006 року.

Теоретичні засади дослідження - фрактали та часові ряди

У той час, як для традиційних засобів наукової комунікації підходи до статистичних досліджень інформаційних масивів з погляду теорії фракталів були вперше досліджені Ван Рааном [4], що аналізував масиви статей і зв'язки, утворені цитуванням, інформаційні потоки повідомлень із Інтернет до останнього часу не асоціювалися із фракталами. Це пов'язано із проблемами ідентифікації інформаційних потоків як фрактальних множин, а також із труднощами знаходження основ для побудови кластерів - повідомлень у політематичних потоках, що породжують багаторазове цитування.

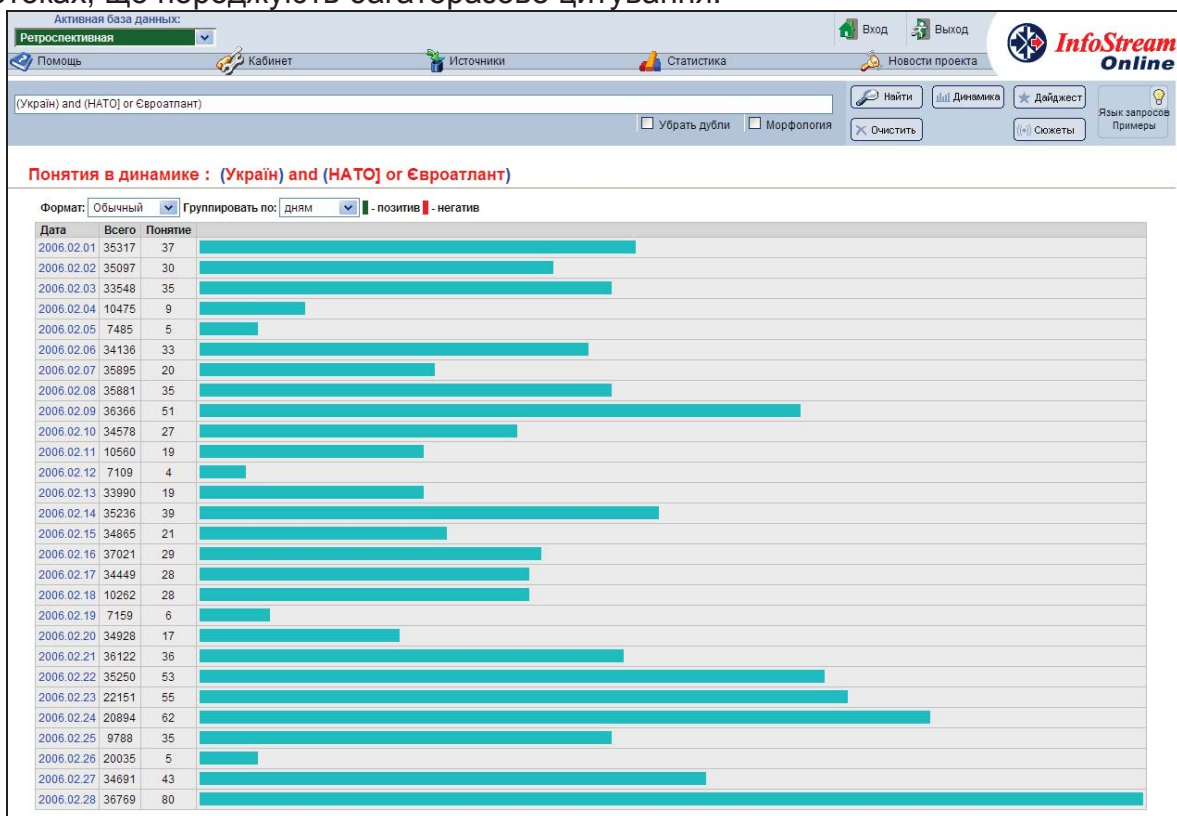


Рис. 1. Динаміка появи у часі поняття, визначеного запитом: «(Україн) and (НАТО) or Євроатлант»

З цієї ж причини у рамках даного дослідження розглядаються кількісні характеристики саме тематичного інформаційного потоку з питань євроатлантичної інтеграції, який характеризується ітеративністю при формуванні й цілком доступний як для кількісного, так і для якісного аналізу (маються на увазі можливості системи контент-моніторингу InfoStream).

Обсяги повідомлень у тематичних інформаційних потоках утворюють часові ряди. Для дослідження часових рядів сьогодні усе ширше використовується теорія фракталів [5-6], традиційна область застосування якої - фрактальна геометрія, обробка зображень та інше. Разом з тим часові ряди, породжувані тематичними інформаційними потоками, також володіють фрактальними властивостями і можуть розглядатися як стохастичні фрактали [7-8]. Цей підхід розширює область

застосування теорії фракталів на інформаційні потоки, динаміка яких описується засобами теорії випадкових процесів.

З іншого боку, теорія фракталів розглядається як підхід до статистичного дослідження, що дозволяє одержувати важливі характеристики інформаційних потоків, не вдаючись у детальний аналіз їхньої внутрішньої структури і зв'язків. Однією з основних властивостей фракталів є самоподібність (скейлинг). Як показано в роботах С. А. Іванова, для послідовності повідомлень тематичних інформаційних потоків у відповідності до скейлингового принципу, кількість повідомлень, резонансів на події реального миру пропорційна деякому ступеню кількості джерел інформації (кластерів) та ітераційно триває протягом певного часу. Так само, як і у традиційних наукових комунікаціях, множина повідомлень в Інтернет по одній тематиці в часі являє собою динамічну кластерну систему, що виникає в результаті ітераційних процесів. Цей процес породжується републікаціями, прямим або спільним цитуванням, різними публікаціями - відбиттями тих самих подій реального миру, прямими посиланнями тощо. Крім того, для більшості тематичних інформаційних потоків спостерігається збільшення їхніх обсягів, причому на коротких часових інтервалах - лінійний ріст, а на тривалих - експонентний.

Фрактальна розмірність у кластерній системі, що відповідає тематичним інформаційним потокам, показує ступінь заповнення інформаційного простору повідомлень протягом певного часу:

$$N_{публ} = \varepsilon^{\rho} N_k(t)^{\rho} \quad (1),$$

де $N_{публ}$ – розмір кластерної системи (загальне число електронних публікацій в інформаційному потоці); N_k – розмір - число кластерів (тематик або джерел), ρ - фрактальна розмірність інформаційного масиву; ε - коефіцієнт масштабування. У наведеному співвідношенні між кількістю повідомлень і кластерів проявляється властивість збереження внутрішньої структури безлічі при зміні масштабів його зовнішнього розгляду.

На думку С.А. Іванова, всі основні закони наукової комунікації, такі як закони Парето, Лотки, Бредфорда, Зіпфа, можуть бути узагальнені саме в рамках теорії стохастичних фракталів.

Показник Хьорста

Сьогодні у зв'язку з розвитком теорії стохастичних фракталів стає популярною така характеристика часових рядів, як показник Хьорста (H). У книзі Е. Федера [5] показано, що він пов'язаний із традиційною «клітинною» фрактальною розмірністю (ρ) простим співвідношенням:

$$\rho = 2 - H \quad (2).$$

Умова, за якою показник Хьорста пов'язаний із фрактальною «клітинною» розмірністю відповідно до формули (2), визначена Е. Федером у такий спосіб: «...розглядаються клітки, розміри яких малі в порівнянні як із тривалістю процесу, так і з діапазоном зміни функції; тому співвідношення справедливо, коли структура кривої, що описує фрактальну функцію, досліджується з високою розподільною здатністю, тобто в локальній межі». Ще однією важливою умовою є самоафінність функції. Не вдаючись у подробиці помітимо, що для інформаційних потоків ця властивість інтерпретується як самоподібність, яка виникає в результаті процесів їхнього формування. Можна помітити, що зазначені властивості притаманні не всім інформаційним потокам, а лише тим, які характеризуються достатньою потужністю та ітеративністю при формуванні. При цьому часові ряди, побудовані на підставі потужних тематичних

інформаційних потоків, цілком задовольняють цій умові. Тому при розрахунку показника Хьорста, фактично визначається і такий показник тематичного інформаційного потоку, як фрактальна розмірність.

Відомо, що показник Хьорста являє собою міру персистентності - схильності процесу до трендів (на відміну від звичайного броунівського руху). Значення $H > 1/2$ означає, що спрямована в певну сторону динаміка процесу в минулому, найімовірніше, спричинить продовження руху в тім же напрямку. Якщо $H < 1/2$, то прогнозується, що процес змінить спрямованість. $H = 1/2$ означає невизначеність - броунівський рух.

Для вивчення фрактальних характеристик тематичних інформаційних потоків вивчалися значення показника Хьорста за певний період для часових рядів, складених з кількості стосовних до них повідомлень. Показник Хьорста зв'язують із коефіцієнтом нормованого розмаху (R/S), де R – обчислений певним чином «розмах» відповідного часового ряду, а S – стандартне відхилення.

Показник Хьорста обчислюється за наступним алгоритмом. Спочатку обчислюється середнє значення вимірюваної змінної (у нашому випадку кількість повідомлень в інформаційному потоці) за N днів:

$$\langle \xi \rangle_N = \frac{1}{N} \sum_{t=1}^N \xi(t) \quad (3)$$

Потім розраховується відхилення, що накопичуються, ряду *вимірів* $\xi(t)$ від середнього $\langle \xi \rangle_N$:

$$X(t, N) = \sum_{u=1}^t (\xi(u) - \langle \xi \rangle_N) \quad (4)$$

Після цього розраховується різниця максимального й мінімального відхилень, що накопичилися, яка називається «розмахом»:

$$R(N) = \max_{1 \leq t \leq N} X(t, N) - \min_{1 \leq t \leq N} X(t, N) \quad (5)$$

Стандартне відхилення розраховується за відомою формулою:

$$S = \left(\frac{1}{N} \sum_{t=1}^N (\xi(t) - \langle \xi \rangle_N)^2 \right)^{1/2} \quad (6)$$

У свій час Хьорст експериментально виявив, що для багатьох часових рядів справедливо:

$$R/S = (N/2)^H \quad (7)$$

Саме коефіцієнт H і є показником Хьорста.

Обчислювальний експеримент

Як експериментальна база для дослідження фрактальних властивостей тематичних інформаційних потоків за заданою проблематикою використався комплекс контент-моніторингу InfoStream. Цей комплекс, що застосовується для рішення завдань автоматизованого збору новинної інформації з відкритих Web-сайтів і забезпечення доступу до неї в пошукових режимах, у даний час охоплює понад 2000 джерел інформації - більше 40000 унікальних повідомлень на добу. У ретроспективних базах даних комплексу накопичено понад 25 млн. повідомлень.

Тематика досліджуваного інформаційного потоку, визначалася запитом до системи InfoStream:

«(Київ or Київ or Україн or Украин) and (НАТО) or Євроатлантик or Євроатлантик)».

Ретроспективний період дослідження становив весь 2005 рік, тобто 365 днів ($N = 365$). У результаті пошуку за наведеним запитом було знайдено 54933 релевантних документів. На підставі обробки цих даних була отримана повна картина експериментальних даних - часовий ряд за зазначений період (Рис.1).

Для цього часового ряду по формулі (6) було обчислено стандартне відхилення ($S=127.62$). Одночасно, за допомогою механізму формування основних сюжетів, що входить до складу системи InfoStream, були визначені основні події, які сприяли виникненню пікових значень часового ряду.

На Рис. 3 представлено динаміку накопичення відхилення, що була обчислена відповідно до формули (4) і дозволила відповідно до формули (5) визначити «розмах» цього параметра ($R = 7676.15$).

І нарешті, для значення $N = 365$ по формулі (7) був обчислений показник Хьорста, що виявився рівним 0,79, що свідчить про великий рівень персистентності всього часового ряду. Крім того, були виконані розрахунки показників Хьорста для всіх значень N , починаючи з 10, результати яких наведені на Рис. 4.

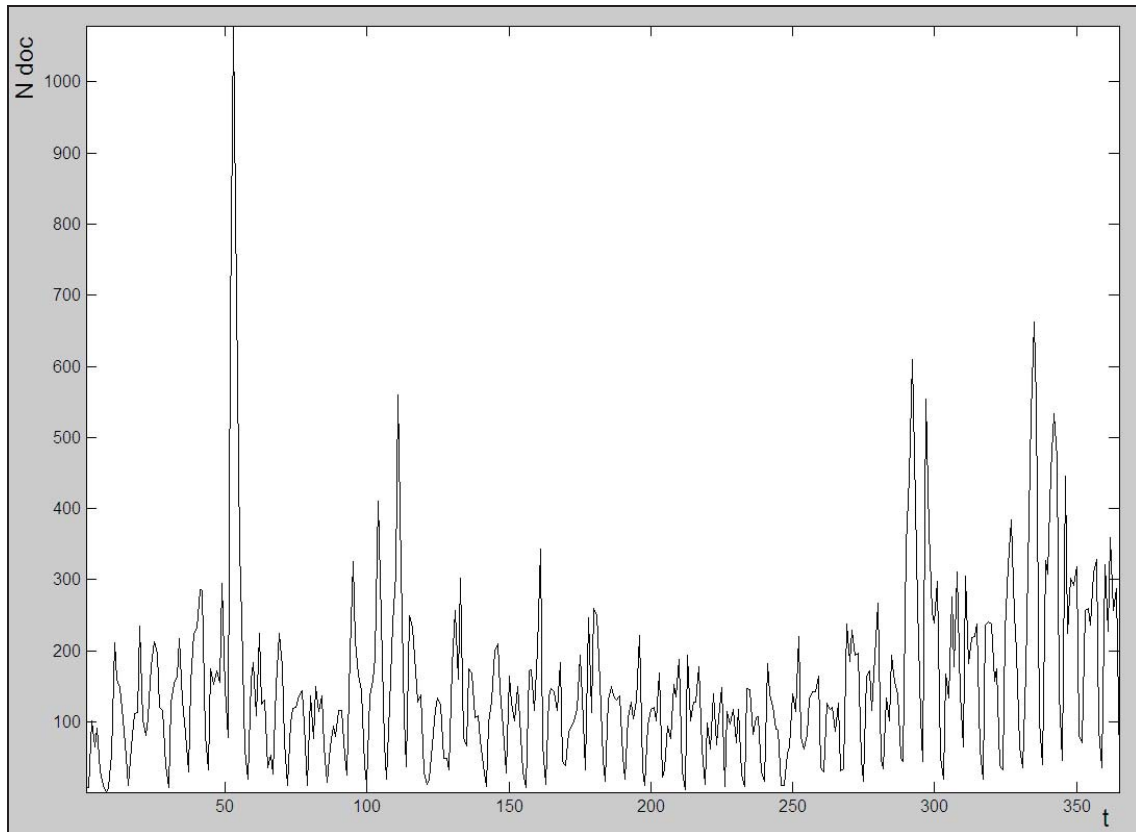


Рис. 2. Часовий ряд динаміки появи поняття за весь період. Пікові значення: Прес-конференція Президента України з генеральним секретарем Північноатлантичного альянсу у Брюсселі (лютий 2005 р.); засідання комісії Україна-НАТО на рівні міністрів закордонних справ (квітень 2005 р.); виступ міністра закордонних справ України на Форумі з безпеки РСАП (травень 2005 р.); зустріч Президента України з Генеральним секретарем НАТО (жовтень 2005 р.); конференція «Національна безпека України в контексті євроатлантичної інтеграції» (листопад 2005 р.)

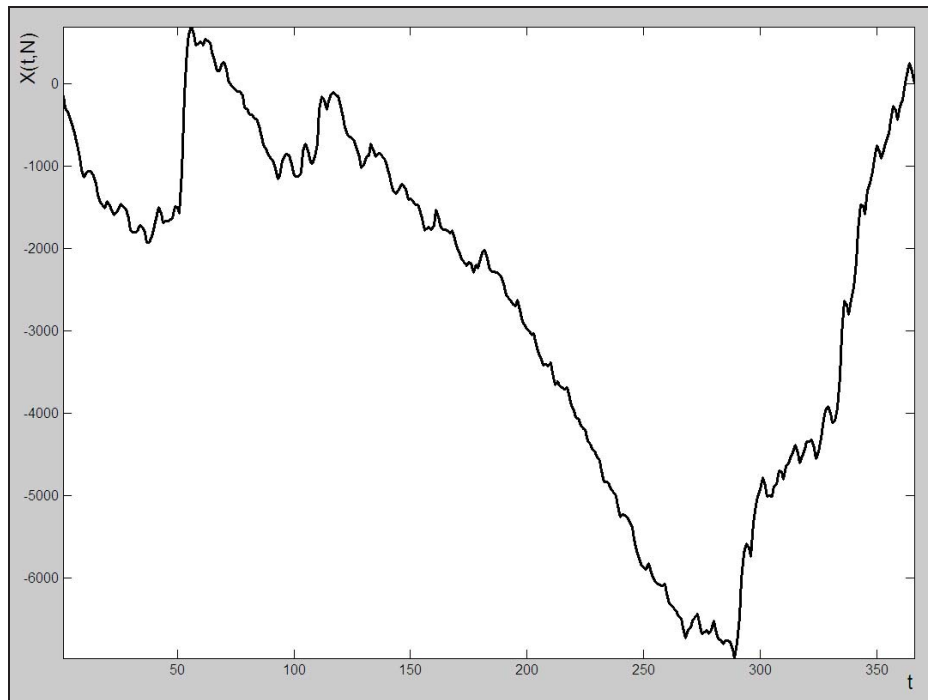


Рис. 3. Динаміка накопичення відхилення

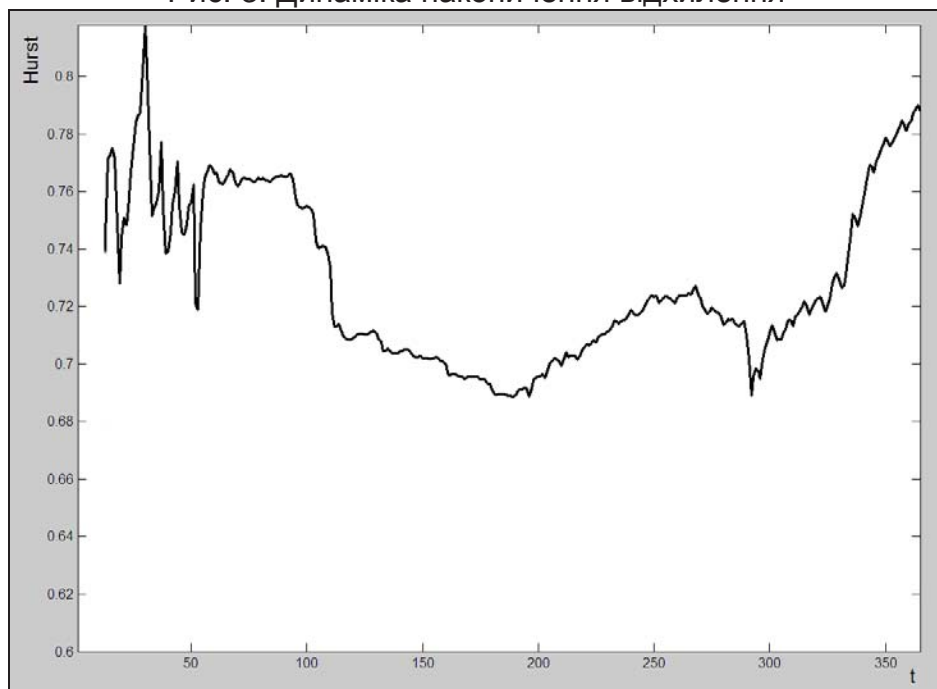


Рис. 4. Значення показника Хьорста для різних часових інтервалів

Інтерпретація результатів

Вивчення такої характеристики, як показник Хьорста дозволяє прогнозувати динаміку інформаційних потоків, повідомлення яких відбивають процеси, що відбуваються в реальному світі.

Наведені дані підтвердили припущення про ітеративність процесів в інформаційному просторі. Републікації, цитування, прямі посилання тощо породжують самоподібність, що проявляється в стійких статистичних розподілах і відомих емпіричних законах. Треба зауважити, що скейлінговий принцип ґрунтується також на подібності ментальності авторів, що публікують повідомлення в Інтернет. Разом з тим різні адміністративні, рекламні, PR-кампанії

ведуть до стрибкоподібних змін у стабільних статистичних закономірностях, різким скачкам і перекручуванням у порівнянні зі стандартними статистичними розподілами.

У результаті експерименту також підтверджена наявність високого рівня статистичної кореляції в інформаційних потоках на тривалих часових інтервалах. Зокрема, на розглянутому прикладі показана висока персистентність процесу, що говорить про загальне середнє збільшення публікації щодо визначеної тематики, періодичну появу піків, пов'язаних, як правило, з загальними переговорними заходами: форумами, конференціями та зустрічами керівництва держави з функціонерами НАТО.

Природно, описані результати досліджень можуть використатися не тільки для наведеного тематичного інформаційного каналу. Свого дослідження чекають кластери, породжувані у відповідності й з іншими принципами, наприклад, близькими по напрямкам джерелами інформації (Web-сайтами, мережними ЗМІ, «живими журналами» тощо).

Програмно-технологічні рішення, що базуються на сучасних статистичних методах, надають структурований огляд інформації щодо суспільних процесів, забезпечують високий рівень об'єктивності і неупередженості при підготовці і прийнятті рішень внаслідок отримання, врахування і обробки великих масивів інформації за визначеною темою.

Разом з тим для прийняття будь-яких серйозних рішень, у тому числі, пов'язаних з євроатлантичною інтеграцією, необхідне використання комплексних систем, що дозволяють компонувати та узагальнювати отриману з різних джерел інформацію щодо об'єкта досліджень.

Складні та відповідальні завдання щодо підготовки і прийняття рішень на сучасному етапі розвитку України, прагнення України „вписатися” у сучасний світовий простір, в тому числі і в євроатлантичному напрямі, вимагають серйозного опрацювання та інформаційної підтримки, для забезпечення якої необхідна обробка широкого спектра інформаційних джерел, значна частина з яких у даний час представлені в мережі Інтернет.

Список літератури

1. Фурашев В.М., Ланде Д.В., Григор'єв О.М., Фурашев О.В. Електронне інформаційне суспільство України: погляд у сьогодення і майбутнє: Монографія. – К.: Преса України, 2005. – 166 с.
2. Современные информационные потоки: актуальная проблематика / С.М.Брайчевський, Д.В.Ландэ // Научно-техническая информация. Сер. 1, 2005. - № 11. - С. 21-33
3. Ланде Д.В., Фурашев В.М., Григор'єв О.М. Програмно-апаратний комплекс інформаційної підтримки прийняття рішень: Науково-методичний посібник. - Київ: Інжиніринг, 2006. – 48 с.
4. Fractal geometry of information space as represented by cocitation clustering / Van Raan A. F. J. // Scientometrics. -1991. – Vol. 20, № 3. – P. 439-449.
5. Федер Е. Фракталы / -М.: Мир, 1991, -254 с.
6. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа. - М.: "Вильямс", 2005. - 272 с.
7. Стохастические фракталы в Информатике / Иванов С.А. // Научно-техническая информация. Сер. 2, 2002. - № 8. - С. 7-18
8. Статистический анализ документальных информационных потоков. / Иванов С.А., Круковская Н.В. // Научно-техническая информация. Информ. Процессы и системы. Сер. 2. – 2004. - № 2. – С. 11-14.

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
НАЦИОНАЛЬНЫЙ АЭРОКОСМИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Е. ЖУКОВСКОГО
"ХАРЬКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ"**

**ОТКРЫТЫЕ ИНФОРМАЦИОННЫЕ
И КОМПЬЮТЕРНЫЕ ИНТЕГРИРОВАННЫЕ
ТЕХНОЛОГИИ**

Сборник научных трудов

Выпуск 31

**Харьков
«ХАИ»
2006**

ББК 32.973.3
О 83
УДК 629.7.01; 681.327.11

Утверждено к печати ученым советом Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ», протокол № 9 от 24.05.2006 г.

Редакционная коллегия: А. И. Бабушкин, П. В. Балабуев, В. К. Борисевич, В. Е. Гайдачук, А. Г. Гребеников, В. В. Кириченко, Б. М. Конорев, В. С. Кривцов, А. К. Мялица, Н. В. Ефремова, В. И. Рябков, И. Б. Сироджа, В. Г. Сухоребрий, О. Е. Федорович, В. Н. Фурашев

Открытые информационные и компьютерные интегрированные технологии: Сб. науч. трудов. Вып. 31. - Харьков: Нац. аэрокосмический ун-т "ХАИ", 2006. – 288 с.

ISBN 966-662-131-2

Освещены вопросы аэродинамики, прочности, конструкции и технологии производства авиационной техники, методы интегрированного проектирования, подготовки производства, инженерного анализа с помощью систем CAD/CAM/CAE. Представлены методы обеспечения безопасности конструкций при длительной эксплуатации, системы информационного обеспечения жизненного цикла самолетов, системы обеспечения безопасности жизнедеятельности для летательных аппаратов, вопросы экономического анализа рисков, методы поддержки, принятия управленческих решений.

Для специалистов промышленных и научных организаций, преподавателей и студентов.

Адрес редакционной коллегии: 61070, г. Харьков, ул. Чкалова, 17,
Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ»,
тел. (057) 707-43-63

Свидетельство о государственной регистрации КВ №7975 от 09.10.2003 г.
Журнал печатает статьи на украинском и русском языках

ISBN 966-662-131-2

© Национальный аэрокосмический университет
им. Н.Е. Жуковского «ХАИ», 2006

Содержание

В. С. Кривцов, П. О. Науменко. Метрологическое обеспечение первичного формообразования авиационных деталей на фрезерных станках с ЧПУ	5
А. Н. Нарожный. Компьютерная поддержка процесса определения летно-технических характеристик самолета. Часть 3	10
В. Ф. Шмырев. Оптимизация параметров осевого воздухозаборника турбореактивного двигателя, расположенного под крылом самолета	20
А. Г. Гребеников, В. Н. Клименко. Исследование влияния радиального натяга, осевой затяжки болтов и поверхностного упрочнения элементов срезного соединения из титанового сплава ВТ6 на их усталостную долговечность	41
В. Г. Сухоребрый, М. И. Никифорова. Оценка эффективности ускоренного статистического моделирования для определения проектной вероятности устойчивости ракеты-носителя	55
О. Г. Приймаков, Ю. О. Градиський. Оцінка довговічності авіаційної техніки	62
Д. В. Ланде, В. М. Фурашев. Системна інформатизація інформаційного супроводження процесів: статистичне дослідження уваги Інтернет-ЗМІ до визначеної проблематики	70
А. Н. Застела, И. П. Печенежский. Разработка многокритериальной модели штамповки коробчатых деталей из листа эластичной средой	77
С. С. Левин, А. В. Лоян, Ю. К. Чернышев. Трехмерное имитационное моделирование газодинамических процессов в СПД в предпусковом состоянии	82
С. А. Агаркова, А. А. Бояркин, И. П. Внуков, Б. И. Паначевный. Приводные машины компьютерно-интегрированных технологических процессов	95
Т. Є. Ударцева. Метод визначення працездатності авіаційних операторів на основі аналізу показників ергографії	99
В. В. Борисов, В. П. Зинченко. Анализ актуальных проблем информационной технологии передачи данных при моделировании нагрузок, действующих на конструкцию сложного технического объекта	104
П. Е. Пустовойтов, Эль Саед Абделаал Эль Саед Мохамед. Методика обоснования требований к техническим характеристикам узла компьютерной сети для обеспечения информационной гарантобезопасности	129
Ю. В. Стасев, В. В. Баранник, Е. А. Бридня. Информационная модель апертурного представления изображений	133
В. О. Повгородний. Математическая модель для оценки эффективности системы технического обслуживания и ремонта, включая температурные нагрузки	139
С. А. Калкаманов, Р. Н. Чигрин, В. П. Закурдаев. К вычислению вектора потока в методах граничных элементов	143
В. А. Дорошук, В. Ж. Яценюк, І. І. Олійник, Ф. Ф. Мисик. Радіолокатор з каналом розпізнавання класу повітряної цілі на основі адаптивного решітчастого фільтру	149
Е. П. Киричук, И. В. Варфоломеева. Разработка квантовой модели поддержки принятия решений технолога при выборе типа оборудования для резки листа на полосы как ядра интеллектуальной информационной технологии	156