

НАУКОВО-ПРАКТИЧНИЙ ІНФОРМАЦІЙНИЙ ЖУРНАЛ
НАУКОВО-ТЕХНІЧНА
ІНФОРМАЦІЯ
3(29)'2006

ПЕРЕДПЛАТНИЙ ІНДЕКС 22879

ЗАСНОВНИКИ:

Міністерство освіти і науки України
Український інститут науково-технічної
і економічної інформації (УкрІНТЕІ)

Засновано 11 березня 1999 р.
Свідоцтво про реєстрацію
Міністерства інформації України:
серія КВ, № 3720

Постановою президії ВАК України
від 9 лютого 2000 р. № 2-02/2
журнал «НТИ» включено до переліку №4
наукових фахових видань України
в галузі технічних наук

Постанова президії ВАК України
від 13.12.2000 р. № 2-01/10
ухвалила зараховувати статті,
опубліковані в журналі «НТИ», як фахові
в галузі економічних наук

РЕДАКЦІЙНА КОЛЕГІЯ:

Головний редактор

В. Д. Пархоменко, докт. техн. наук
Заст. головного редактора

Г. І. Калитич, докт. екон. наук
Члени редакційної колегії:

І. М. Астрелін, докт. техн. наук
О. І. Волков, канд. техн. наук
В. І. Воронков, канд. техн. наук
А. П. Гончаренко
В. В. Камишин, канд. техн. наук
Ю. М. Канигін, докт. екон. наук
В. Г. Лукомський, канд. техн. наук
Б. А. Малицький, докт. екон. наук
С. І. Мороз, докт. техн. наук
Л. О. Мусіна, канд. екон. наук
В. В. Петров, докт. техн. наук
В. Я. Рубан, докт. техн. наук
О. О. Саверченко
О. Я. Савченко, докт. фіз.-мат. наук
П. М. Цибульов, докт. техн. наук
Д. М. Черваньов, докт. екон. наук
А. Е. Юзефович, докт. екон. наук

Рекомендовано до друку
Вченою радою Українського інституту
науково-технічної і економічної інформації
та редакційною колегією журналу

ВІД РЕДАКЦІЇ:

*Журнал НТИ друкує матеріали, що відповідають
профілю видання.*

*Редакція рецензує наукові статті, листується з
читачами на сторінках журналу та на сайті.*

*Відповідальність за достовірність інформації, що
міститься в друкованих матеріалах, несуть автори.*

*Усі права застережені. Передруки та переклади
дозволені лише за згодою авторів і редакції.*

*Статті журналу "НТИ" рецензувалися, обговорювалися
на редколегії журналу та рекомендовані
для публікації Вченою радою Українського інституту
науково-технічної і економічної інформації.*

Передплатити

журнал НТИ можна за каталогом Укрпошти
(індекс 22879) у відділі маркетингових досліджень по
забезпеченню реалізації інформаційної продукції
УкрІНТЕІ, тел. 528-23-45.

Адреса редакції:
03680, Київ, вул. Горького, 180, УкрІНТЕІ,
тел. 528-25-57
<http://www.uintei.kiev.ua>

Відповідальний за випуск

В.В. Камишин

Редактор Н.М. Кучеренко

Технічний редактор Л.М. Басова

Комп'ютерна верстка Б.О. Грабовський

Комп'ютерний набір Н.В. Дудник

Підписано до друку 10.08.2006 р.
Формат 60 × 84 1/8. Друк. арк. 7,8. Обл.-вид. арк. 9,9
Тир. 300 прим. Зам. 248в.
03680, Київ, вул. Горького, 180.
Видавничо-поліграфічне відділення УкрІНТЕІ

©УкрІНТЕІ, 2006

ЗМІСТ

НАУКОВО-ІНФОРМАЦІЙНА ДІЯЛЬНІСТЬ

<i>Бугай С.М., Лисак Н.В.</i> Практичний досвід інформаційно-аналітичної підтримки процедур вироблення і прийняття управлінських рішень	3
<i>Степаненко В.Л., Кривулькін І.М.</i> Інформація — стратегічний важіль ефективного управління страховим фондом документації України	7
<i>Пархоменко О.В.</i> Інформаційна аналітика як напрям наукового дослідження	11
<i>Чорнобай Р.В.</i> Стан науково-технічного потенціалу України	14
<i>Желюк Т.Л.</i> Механізм аналітичного обґрунтування ефективності виконання управлінських рішень в органах державного управління	19
<i>Круп'як Т.П.</i> Інноваційний процес: аналіз у рамках теорії інформаційної економіки	25
<i>Гава Ю.В.</i> Структура інтелектуального капіталу	29
ЕКОНОМІЧНІ ПРОБЛЕМИ НАУКОВО-ТЕХНОЛОГІЧНОГО ТА ІННОВАЦІЙНОГО РОЗВИТКУ	
<i>Царук О.В.</i> Державний борг у системі статистики державних фінансів: проблеми гармонізації	33
<i>Гура Н.О.</i> Проблеми співвідношення витрат на житлово-комунальні послуги і тарифів на них	36
<i>Майдаченко І.В.</i> Інвестиційна діяльність страхових компаній — шлях до економічного зростання	41
<i>Гуляєва Л.П.</i> Інвестиційна привабливість банківського бізнесу та його ефективність ...	45
ІНФОРМАЦІЙНІ РЕСУРСИ І ТЕХНОЛОГІЇ	
<i>Гончаренко А.П., Каретнікова Л.Ф., Войтенко В.В., Кушнір, Г.Л., Мельник М.П.</i> Автоматизована інформаційна система реєстрації, пошуку й аналізу науково-технічних заходів	49
<i>Симоненко Т.В.</i> Розвиток системи інформаційно-бібліотечного забезпечення науки	52
<i>Герасимов Б.М., Камишин В.В., Любарський С.В.</i> Сучасні тренажерні системи навчання спеціалістів з інформаційних технологій	56
<i>Буров О.Ю.</i> Гіперкомплексна система як інформаційна модель функціональної системи діяльності людини	62
<i>Ланде Д.В.</i> Дослідження лінгвостатистичних характеристик інформаційних потоків новин у рамках магістерського курсу "Теорія інформаційного пошуку"	65
РЕФЕРАТИ	71

АВТОРИ ЦЬОГО НОМЕРА

- Бугай Сергій Миколайович** — к.е.н., заст. нач. Управління регіональної політики, зав. відділу розвитку регіонів Кабінету Міністрів України
- Буров Олександр Юрійович** — к.т.н., провідний наук. співробітник НДІ проблем військової медицини
- Войтенко Валентина Вікторівна** — провідний інженер УкрІНТЕІ
- Гава Юрій Васильович** — аспірант Київського національного університету ім. Тараса Шевченка
- Герасимов Борис Михайлович** — д.т.н., професор Військового інституту Київського національного університету ім. Тараса Шевченка
- Гончаренко Алевтина Петрівна** — перший заступник директора УкрІНТЕІ
- Гуляєва Людмила Петрівна** — асистент кафедри фінансів та кредиту Академії праці і соціальних відносин
- Гура Надія Олександрівна** — к.е.н., доцент Київського національного університету ім. Тараса Шевченка
- Желюк Тетяна Леонтіївна** — к.е.н., доцент Тернопільського державного економічного університету
- Камишин Володимир Вікторович** — к.т.н., заступник директора УкрІНТЕІ
- Каретнікова Лариса Хомівна** — зав. відділенням УкрІНТЕІ
- Кривулькін Ігор Михайлович** — к.ф.-м.н., заст. нач. відділу Державного департаменту страхового фонду документації, м. Харків
- Круп'як Тарас Пилипович** — аспірант Київського національного університету ім. Тараса Шевченка
- Кушнір Ганна Леонтіївна** — к.т.н., с.н.с., провідний наук. співробітник УкрІНТЕІ
- Ланде Дмитро Володимирович** — к.т.н. заст. директора Інформаційного центру "Електронні вісті"
- Лисак Наталія Володимирівна** — к.ф.-м.н. наук, консультант Інформаційно-аналітичної фірми "ЕКСОР"
- Любарський Сергій Володимирович** — ст. викладач Військового інституту телекомунікацій та інформатизації Національного технічного університету України "КПІ"
- Майдаченко Ірина Володимирівна** — нач. відділу врегулювання збитків ЗАТ "Європейський страховий альянс"
- Мельник Микола Петрович** — гол. спеціаліст ГІОЦ м. Києва
- Пархоменко Олексій Володимирович** — аспірант УкрІНТЕІ
- Симоненко Тетяна Василівна** — аспірантка Національної бібліотеки України імені В.І. Вернадського
- Степаненко Володимир Леонідович** — к.е.н., голова Державного департаменту страхового фонду документації, м. Харків
- Царук Олександр Васильович** — аспірант Київського національного університету ім. Тараса Шевченка
- Чорнобай Руслан Васильович** — с.н.с. УкрІНТЕІ

ДОСЛІДЖЕННЯ ЛІНГВОСТАТИСТИЧНИХ ХАРАКТЕРИСТИК ІНФОРМАЦІЙНИХ ПОТОКІВ НОВИН У РАМКАХ МАГІСТЕРСЬКОГО КУРСУ “ТЕОРІЯ ІНФОРМАЦІЙНОГО ПОШУКУ”



Д. В. Ланде,
канд. техн. наук

Проблематика

Сучасні навчальні курси інформатики, зокрема, теорії інформаційного пошуку не повною мірою підкріплені методикою проведення практичних занять. Студенти, навіть маючи доступ до Інтернету, не завжди можуть переконатися в дієвості таких основних закономірностей інформаційних потоків [1], як наприклад, закони Зіпфа, Бредфорда або Хіпса, не говорячи вже про дослідження взаємозв'язків понять та явищ, їх кластеризації, що впливає з їхнього подання, наприклад, в Інтернеті. Максимум, що можна запропонувати для практичних занять у разі використання загальнодоступних пошукових систем — це пошук з використанням різних інформаційно-пошукових систем, вивчення розмаїтості мов запитів і форм подання результатів.

Як основна система полігону для проведених автором практичних занять використовувалася спеціалізована система контент-моніторингу InfoStream, розроблена в Інформаційному центрі “Електронні вісті” (ElVisti). Ця система, що застосовується для вирішення завдань автоматизованого збирання інформації з відкритих веб-сайтів і забезпечення доступу до неї в пошукових режимах, у цей час охоплює понад 2 тис. джерел інформації — більш як 40 тис. унікальних повідомлень новин на добу. Отже, як основне інформаційно-технологічне середовище для організації практичних занять розглядалася система InfoStream, навчально-аналітичний модуль якої надає можливість:

- досліджувати динаміку тематичних інформаційних потоків новин в Інтернеті;
- вибирати набори джерел інформації, необхідні для задоволення інформаційних потреб;
- досліджувати тематичну стабільність джерел інформації;
- визначати взаємозв'язок джерел інформації;
- визначати кореляції між окремими поняттями, досліджуючи інформаційні потоки за певний час;

- визначати взаємозв'язок тематичних рубрик і взаємозв'язок термінів, що відповідають тематиці інформаційного потоку;
- формувати рейтинги популярності певних понять в Інтернеті.

Інформаційний портрет

Портрет можна розглядати як модель об'єкта (або суб'єкта) реального світу, виражену його найбільш пізнаваними рисами. Засоби системи InfoStream включають так званий інформаційний портрет [2], [3], реалізований у вигляді статистично значимої сукупності характеристик інформаційного потоку. Побудова інформаційного портрета виконується на основі емпіричних і статистичних методів, основу яких складають частотно-лінгвістичні алгоритми.

Інформаційний портрет — це адаптивний механізм уточнення запитів, що будується після того, як користувач введе запит або активізує режим пошуку за певною тематичною рубрикою. За допомогою інформаційного портрета в системі InfoStream можуть деталізуватися та уточнюватися критерії пошуку. Інформаційний портрет реалізовано як таблиця на екрані з результатами пошуку (рис. 1).

Для уточнення запити параметром з інформаційного портрета досить активізувати гіперпосилання, що відповідає цьому параметрові. Для уточнення запити відразу декількома словами з інформаційного портрета використовується механізм віконця для вибору (checkbox), що перебувають поруч зі словами з інформаційного портрета.

Крім того, інформаційний портрет являє собою досить універсальний інструмент для вивчення характеристик різних документальних добірок.

Информационный портрет	
Уточнить запрос	
Рубрики (35)	
Языки (3)	
Размер (3)	
Цифровая насыщенность (3)	
Страны источников (15)	
Источники (50)	
Тональность (2)	
География (50)	
Персоны (50)	
Слова (80)	
AND	NOT
<input type="checkbox"/> АКТИВ	<input type="checkbox"/>
<input type="checkbox"/> АКЦ	<input type="checkbox"/>
<input type="checkbox"/> БАНК **	<input type="checkbox"/>
<input type="checkbox"/> БАНКНОТ	<input type="checkbox"/>
<input type="checkbox"/> БАНКОВСК	<input type="checkbox"/>
<input type="checkbox"/> БИЗНЕС	<input type="checkbox"/>
<input type="checkbox"/> ВАЛЮТ	<input type="checkbox"/>
<input type="checkbox"/> ВАЛЮТН	<input type="checkbox"/>
<input type="checkbox"/> ВКЛАД	<input type="checkbox"/>
<input type="checkbox"/> ВЫПУСК	<input type="checkbox"/>
<input type="checkbox"/> ГОДОВ	<input type="checkbox"/>
<input type="checkbox"/> ГОСУДАРСТВЕН	<input type="checkbox"/>
<input type="checkbox"/> ГРИВЕН	<input type="checkbox"/>
<input type="checkbox"/> ГРИВН	<input type="checkbox"/>
<input type="checkbox"/> ГРУПП	<input type="checkbox"/>
<input type="checkbox"/> ДАН	<input type="checkbox"/>
<input type="checkbox"/> ДЕНЕЖН	<input type="checkbox"/>

Рис. 1. Фрагмент інформаційного портрета, побудованого за тематикою “Банківська діяльність”

Підтвердження класичних закономірностей

Існують дві основні закономірності, що відносяться до розподілу слів у текстових документах (закон Зіпфа) і джерел, що належать до певної теми (закон Бредфорда), які свідчать про гіперболічну форму рангових розподілів. За допомогою інтерфейсу системи InfoStream досліджувалися рангові розподіли джерел (веб-сайтів), що належать до комп'ютерної тематики. При цьому для досягнення однорідності аналізованого масиву документи видавалися без урахування змістовного дублювання. Крім того, розмір документів обмежувався 10 тис. символів.

Результати ранжирування джерел за кількістю вхідних у них документів дали змогу, зокрема, підтвердити для інформаційних потоків з Інтернету справедливість закону Бредфорда [3], [4], що в первісному вигляді відносився до тра-

диційних "паперових" періодичних видань. Джерела розподілилися за трьома множинами, рівними за кількістю релевантних документів: R_1 , R_2 , R_3 . При цьому R_1 — це найбільш рейтингові джерела, які безпосередньо відносяться до певної тематики; R_2 — множина джерел, що кореспондуються з комп'ютерною тематикою; R_3 — джерела, які лише частково торкаються даної теми.

Якщо прийняти позначення, що $|A|$ — це кількість елементів множини A , то пропорція Бредфорда записується в такий спосіб:

$$|R_1| : |R_2| : |R_3| = n : n^2 : n^4.$$

У цьому випадку значення n становить приблизно п'ять джерел.

На рис. 2. наведено рангову діаграму розподілу джерел за кількістю вхідних до них повідомлень за комп'ютерною тематикою за певний період (сім днів).

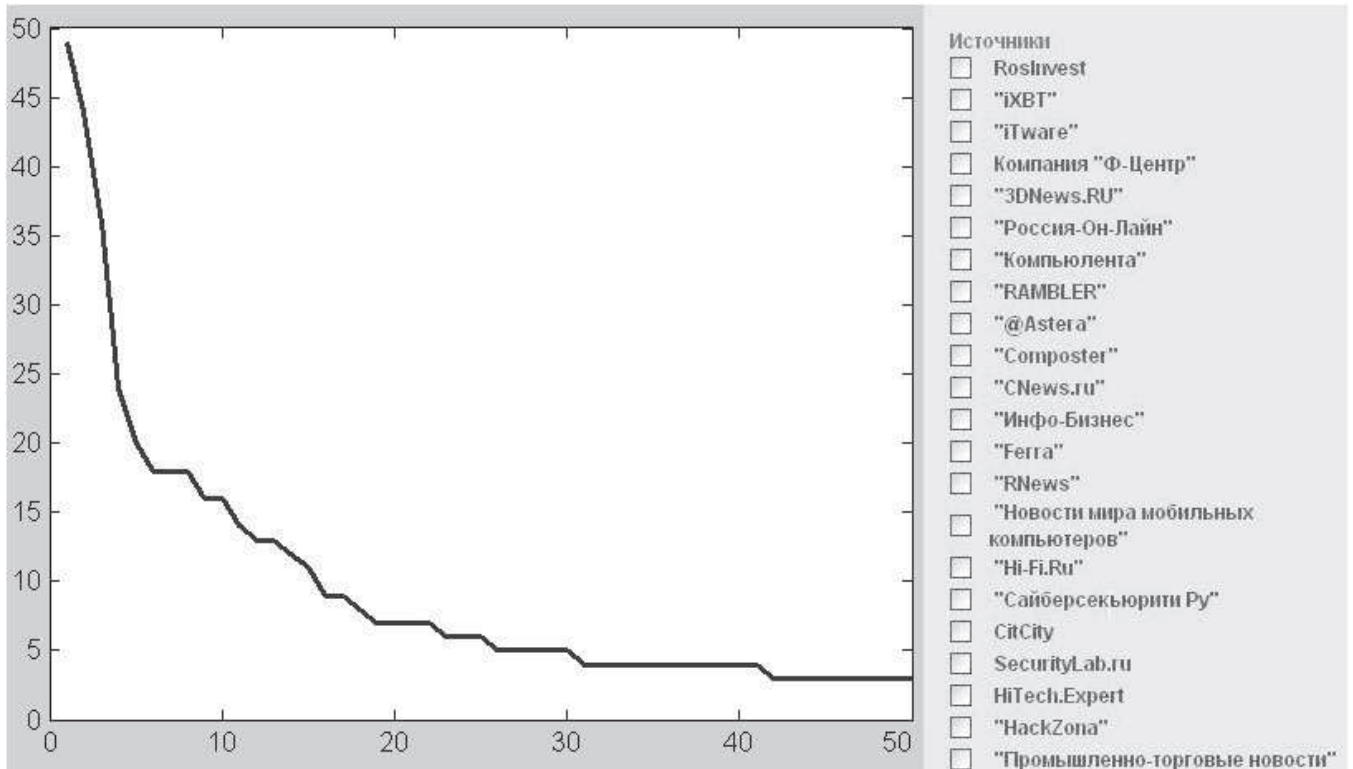


Рис.2. Кількість документів, що відповідають ранговим джерелам

Динаміка надходження повідомлень від джерел

Дослідження, аналогічне проведеному в попередньому пункті, але з уточненням за датами, дає змогу одержати тривимірне зображення представлення динаміки надходження повідомлень від джерел за заданою тематикою. На рис. 3. наведено діаграму надходжень від 30 джерел, що містять релевантні комп'ютерній тематиці документи за один місяць. Варто звернути увагу на значне зменшення публікацій у вихідні дні.

Тематична спрямованість запитів

У рамках практичних занять для деякої фіксованої тематичної рубрики визначався набір найбільш рейтингових джерел — веб-сайтів, що найбільш часто публікують інформацію з даної тематики протягом певного періоду часу (наприклад протягом тижня) [5].

Побудований у такий спосіб набір даних використовувався далі як еталонний. При цьому отримані "сирі" дані ранжирувалися за 10-бальною шкалою (тобто, кожному джерелу, залежно від кількості релевантних тематичній рубриці документів із цього джерела, привласнювався нормалізований ранг в інтервалі від 1 до 10).

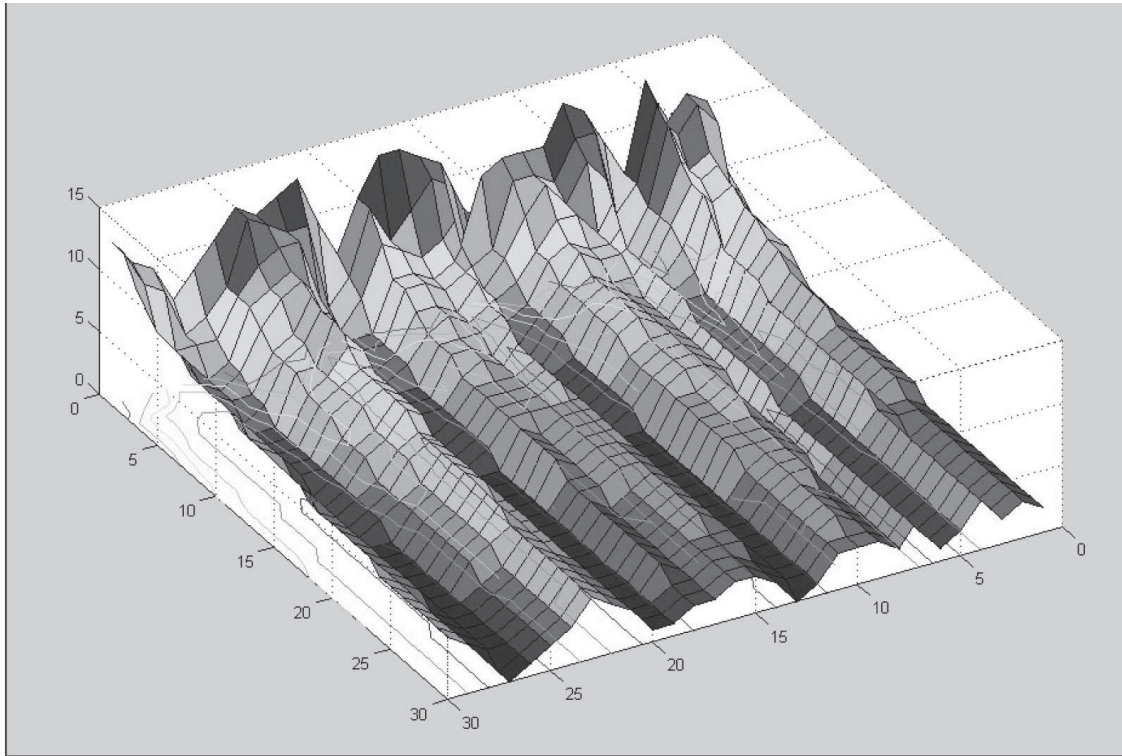


Рис. 3. Рангова діаграма розподілу надходжень від джерел за датами

Відхилення запиту від тематики обчислювалося за формулою:

$$\delta S = \frac{1}{N} \sum_{i=1}^N \Delta R_i,$$

де δS — лінійне відхилення (розкид) джерел; N — кількість джерел (у досліджуваному випадку 50),

ΔR_i — різниця рангів джерела за основною темою (рубрикою) та запитом користувача.

Так, при виборі основної теми, обумовленої як рубрикою "Комп'ютери", лінійне відхилення запиту "процесори" становить $\delta S = 5.24$ (рис. 4), а запиту "мобільні телефони" — $\delta S = 8.14$. З наведених результатів з очевидністю випливає, що "про-

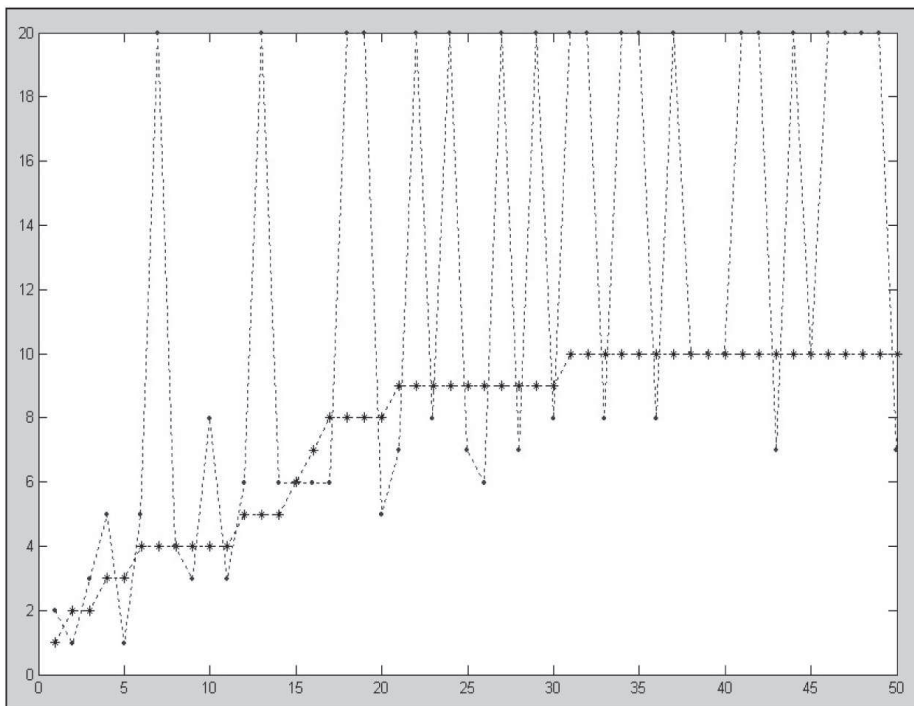


Рис. 4. Основна тема — рубрика "Комп'ютери", з нею порівнюються результати запиту "процесори"; $\delta S = 5.24$

цесори" значно ближче за списком джерел до основної теми, ніж "мобільні телефони".

Таким чином, отримана можливість оцінки тематичної близькості документальних масивів за характеристиками, не заданими первісно при побудові набору рубрик. Крім цього, подібний аналіз може проводитися незалежно від методів рубрикації, що використовувалися раніше.

Тематична стабільність джерел інформації

Необхідно враховувати, що на практиці велике значення має тематична стабільність документів. Рівень тематичної стабільності джерела визначається кореляцією наборів тематичних рубрик, яким відповідають документи з цього джерела в різні періоди часу.

Визначення тематичної стабільності документів виконувалося за таким алгоритмом:

проводився пошук документів у базі даних за певний період;

формувався таблиця, що включала в себе код джерела інформації, коди відповідних йому тематичних рубрик та їхню кількість у розрізі дат;

для кожного джерела розраховувався нормований показник тематичної стабільності R за формулою [6]:

$$R = \frac{1}{N} \sum_{i=1}^N \frac{1}{M * \max(rub_i)} \sum_{j=1}^M \left| rub_{ij} - \frac{\sum_{k=1}^M rub_{ik}}{M} \right|,$$

де N — кількість рубрик; M — кількість днів; $\max(rub_i)$ — нормована максимальна кількість входження рубрики i в документ за весь час; rub_{ij} — нормована кількість входження рубрики i за день j .

Взаємозв'язок понять

У рамках практичних занять аналізувалися взаємозв'язки трьох видів понять: тематичних рубрик, джерел і ключових слів. Усі ці види взаємозв'язків являють собою таблиці, які відбивають близькість (спільне входження в документи) окремих понять. У загальному випадку таблиця взаємозв'язків понять — це симетрична матриця $A = \{a_{ij}\}$, елементи якої a_{ij} — коефіцієнти близькості пари понять i та j . Коефіцієнт a_{ij} відповідає кількості документів вхідного інформаційного потоку, що включають в себе поняття (рубрики, джерела або терміни, або словосполучення, що відповідають поняттю) i , а коефіцієнт a_{ij} — кількості документів у вхідному потоці, які одночасно відповідають поняттям i та j . При цьому для фільтрації можливого шуму значення елементів матриці з найменшими коефіцієнтами можуть обнулятися.

Для виявлення блоків — множин найбільш взаємозалежних понять — застосовуються алгоритми кластерного аналізу. Наприклад, для виявлення двох таких блоків необхідно виділити два поняття-полюси (що відповідають, наприклад, індексам k та l), що найбільше корелюють з іншими поняттями, але найменше зв'язані між собою.

Формально ці умови можна записати в такий спосіб:

$$\sum a_{ik} \rightarrow \max, \sum a_{il} \rightarrow \max,$$

$$\text{при } a_{kl} \rightarrow \min, k \neq l.$$

Інші поняття (наприклад, поняття i) віднесені до блоку k , якщо $a_{ik} > a_{il}$. У протилежному випадку поняття i буде віднесено до блоку l .

Взаємозв'язок рубрик

Таблиця взаємозв'язків тематичних рубрик (рис. 5) відображає близькість окремих тематик (відношення окремих документів до декількох тематик). На практичних заняттях таблиця взаємозв'язку рубрик будувалася як "вручну" на підставі аналізу даних, що містилися в інформаційному портреті, так і за допомогою спеціальної серверної програми, що входить до аналітичного блоку системи контент-моніторингу InfoStream.

Взаємозв'язок джерел

У разі проведення аналізу всього масиву інформаційних джерел велике значення має облік їхнього тематичного зв'язку. У рамках практичних занять таблиця взаємозв'язку джерел будувалася як "вручну" за інформаційним портретом, так і за допомогою спеціальної програми на сервері за таким алгоритмом:

на підставі результатів пошуку за запитом користувача з релевантних документів формується таблиця "джерело — тематична рубрика" (у разі проведення аналізу за допомогою інформаційного портрета; у випадку використання спеціальної серверної програми будується таблиця "джерело — терміни");

розраховується коефіцієнт взаємозв'язку джерел за кількістю співпадаючих рубрик (термінів);

формується вихідна таблиця взаємозв'язку джерел. Джерела з мінімальними зв'язками з іншими (наприклад, меншими ніж 2) не входять до розгляду;

методом k -means (k -середнє) виконується автоматичне перегруповування стовпців і рядків таблиці (в найпростішому вигляді приймається $k=2$);

виконується сортування в межах певних класів за діагональними елементами;

кластери поєднуються, таблиця взаємозв'язку джерел у цифровому вигляді виводиться в інтерфейсі користувача;

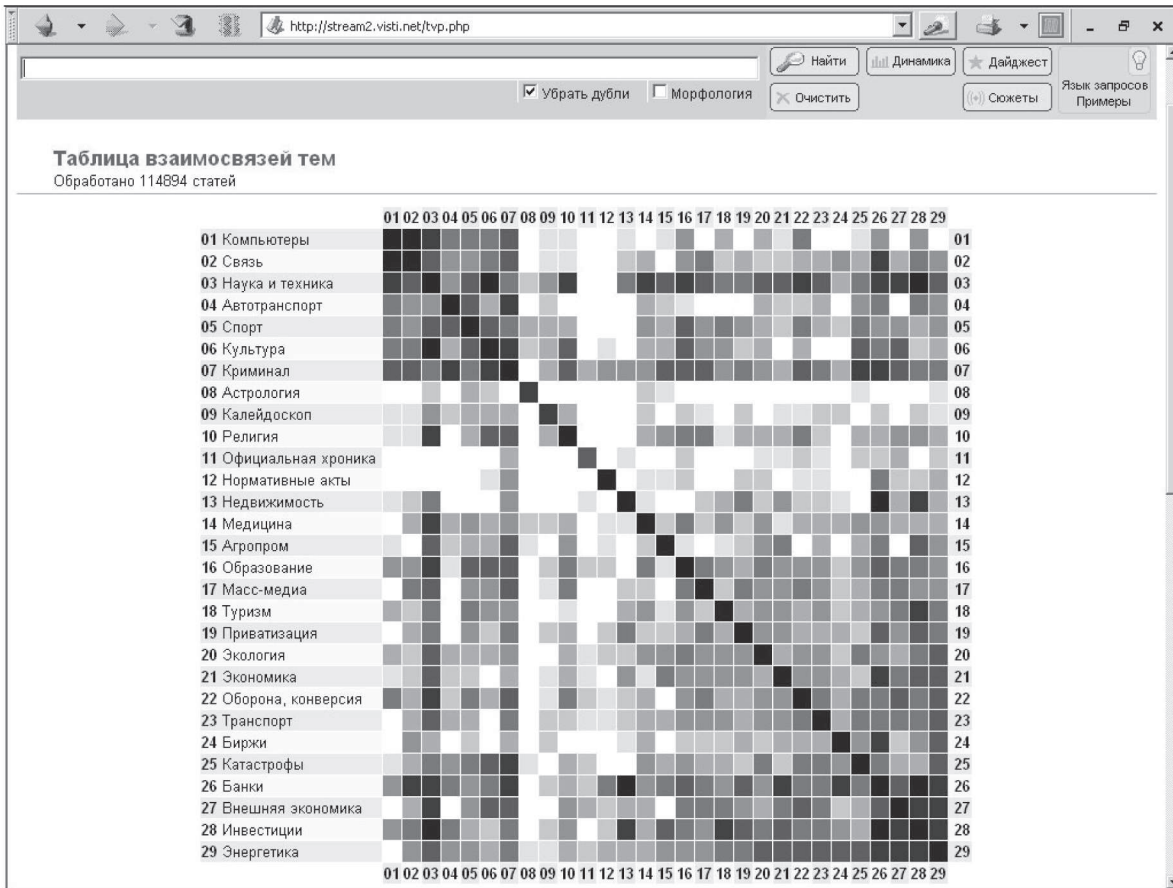


Рис. 5. Таблица взаємозв'язків рубрик

для візуалізації таблиці взаємозв'язку джерел підключаються спеціальні пакети програм (Matlab, Excel та інші).

На рис. 6 наведено тривимірне зображення таблиці взаємозв'язку джерел.

Взаємозв'язок джерел інформації завжди необхідно враховувати під час визначення необхідного списку джерел для задоволення цільових інформаційних потреб, мінімізації кількості цих джерел у разі максимального покриття всіх аспектів, що цікавлять користувачів.

Взаємозв'язок термінів

У багатьох випадках має сенс оцінювати спільне входження в документи різних термінів. Частота спільного входження термінів в окремому повідомленні, а також отримані за допомогою методів кластерного аналізу групи найбільш тісно взаємозалежних термінів можуть свідчити про деякі змістовні аспекти інформаційного потоку.

На відміну від таблиць взаємозв'язків рубрик або джерел, які можна

формувані на основі інформаційного портрета "вручну", таблиця взаємозв'язку термінів через значно більш можливу кількість елементів формується винятково за допомогою спеціальних програм, що функціонують на серверній платформі системи контент-моніторингу.

1 "@Aster" (srd00432)
 2 "Информационная безопасность" (srd00915)
 3 "RNews" (srd01505)
 4 CitCity (srd01926)
 5 Ежедневник "ComputerWorld/Украина" (srd00309)
 6 Журнал "BYTE" (srd01132)
 7 "Ferra" (srd01430)
 8 "Сайберсьюрпінг Ру" (srd02006)
 9 ПРАЙМ-TACC (srd00034)
 10 "RAMBLER" (srd00267)
 11 "СNews.ru" (srd00452)
 12 ПА "Алгоритм" (srd01938)
 13 "3DNews.RU" (srd00371)
 14 "Россия-Он-Лайн" (srd00259)
 15 "Компьюлента" (srd00133)
 16 "Инфо-Бизнес" (srd00136)
 17 "ITware" (srd00372)
 18 Компания "Ф-Центр" (srd01372)
 19 "СХВТ" (srd00204)
 20 RosInvest (srd01522)
 51 12 2 7 8 4 7 6 3 6 9 5 8 10 11 11 14 9 25 25 ;
 12 32 3 6 3 1 5 5 3 6 6 4 4 5 6 7 6 4 8 8 ;
 2 3 7 3 2 1 2 1 2 2 3 2 3 4 3 3 2 4 5 7 ;
 7 6 3 12 4 1 4 4 5 4 4 5 4 6 6 4 6 4 5 8 ;
 8 3 2 4 14 2 3 4 2 2 5 2 5 4 6 5 4 4 7 10 ;
 4 1 1 1 2 14 2 2 1 2 4 3 3 1 4 4 2 4 7 6 ;
 7 5 2 4 3 2 15 1 2 7 2 5 6 5 7 4 5 6 9 ;
 6 5 1 4 4 2 1 16 5 2 5 5 4 3 7 6 8 2 5 7 ;
 3 3 2 5 2 1 2 5 1 7 3 4 4 3 3 4 2 4 3 3 6 ;
 6 6 2 4 2 2 7 2 3 20 3 4 5 5 6 7 5 5 7 9 ;
 9 6 3 4 5 4 2 5 4 3 20 3 7 6 5 6 3 10 16 ;
 5 4 2 5 2 3 5 5 4 4 3 20 4 4 4 3 7 4 5 8 ;
 6 4 3 4 5 3 5 4 3 5 7 4 2 1 7 7 8 7 9 9 13 ;
 10 5 4 6 4 1 6 3 5 6 4 7 22 5 6 8 5 10 16 ;
 11 6 3 6 6 4 5 7 4 6 5 4 7 5 23 2 1 7 12 14 ;
 11 7 3 4 5 4 7 6 2 7 5 3 8 6 2 1 28 6 7 12 14 ;
 14 6 2 6 4 2 4 8 4 5 6 7 7 8 7 6 30 4 13 16 ;
 9 4 4 4 4 4 5 2 3 5 3 4 9 5 7 7 4 3 7 16 19 ;
 25 8 5 5 7 7 6 5 3 7 10 5 9 10 12 12 13 16 7 1 5 4 ;
 25 8 7 8 10 6 9 7 6 9 16 8 13 16 14 14 16 19 5 4 10 1 ;

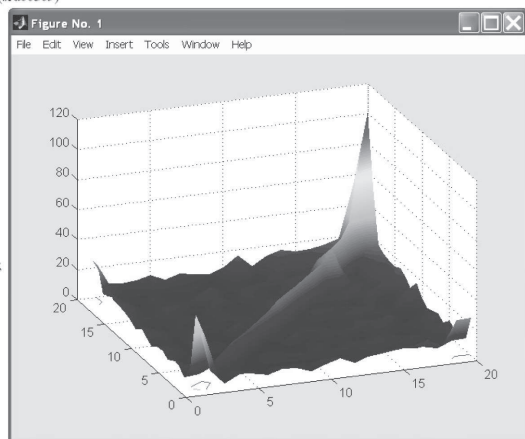


Рис. 6. Взаємозв'язок джерел, документи яких релевантні запиту

Так само, як і в разі автоматичної побудови таблиць взаємозв'язків рубрик і джерел, у цьому випадку з інформаційного портрета в інтерфейсі аналітика вибираються необхідні для дослідження елементи (у даному випадку — терміни), після чого користувач шляхом натискання на кнопку "зв'язки" формує таблицю взаємозв'язку термінів (рис. 7).

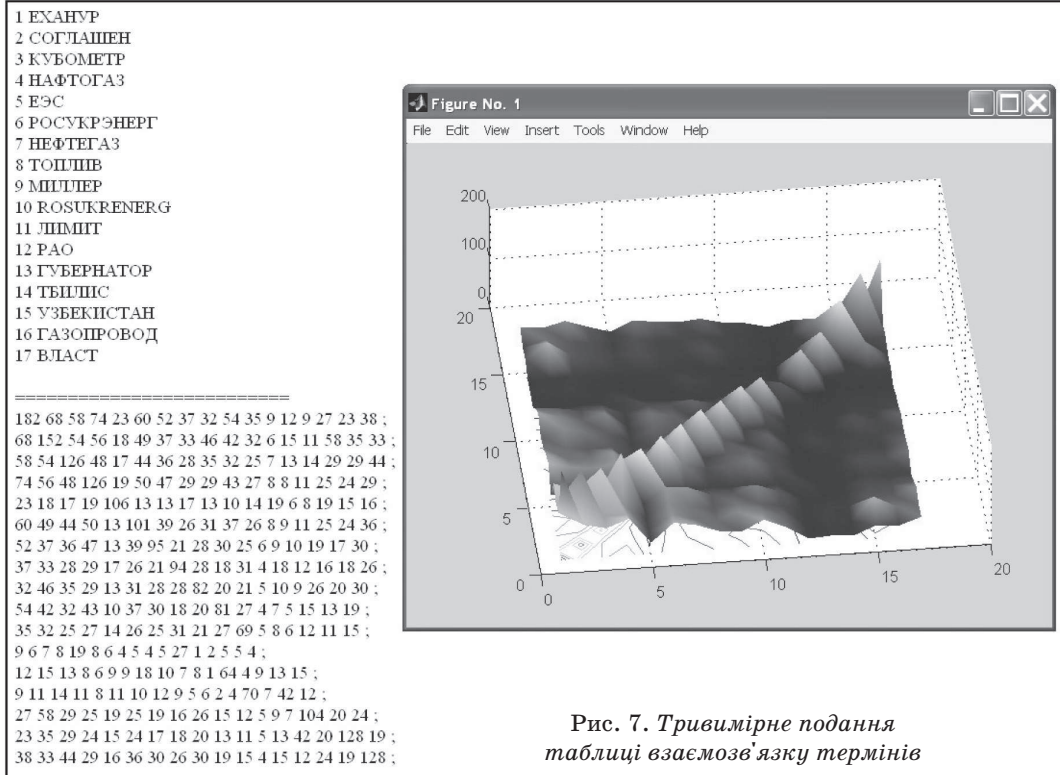


Рис. 7. Тривимірне подання таблиці взаємозв'язку термінів

На рис. 8 зображено результати розрахунку взаємозв'язків персон за другим алгоритмом, проведеного в режимі он-лайн за допомогою спеціальної програми, що функціонує на сервері.

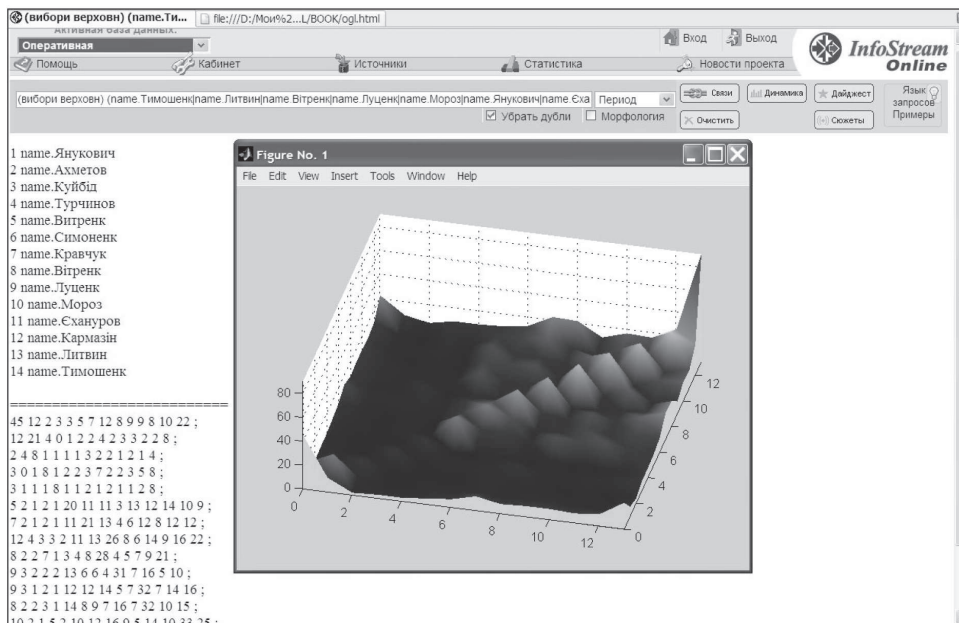


Рис. 8. Подання взаємозв'язку персон за другим алгоритмом

Взаємозв'язок персон

Взаємозв'язок персон оцінюється за допомогою двох алгоритмів шляхом:

- обчислення частоти спільного входження до одних документів згадувань про різних персон;
- розрахунку кореляцій наборів ключових слів, що входять у документи, в яких зустрілися згадування про персон.

Висновки

У комп'ютерному класі, підключеному до Інтернету, студенти одержують навички аналітичної роботи з потоками текстової інформації, освоюють не тільки зовнішні інтерфейси інформаційно-пошукових систем, але й можливості використання таких систем у сполученні з популярними математичними пакетами. Важливо, що основну увагу студентів у процесі виконання практичних завдань приділено не написанню програм обробки даних, доступних лише адміністраторам інформаційно-пошукових серверів, а змістовному аналізу інформації, одержуваної із зовнішнього інтерфейсу аналітичного блоку системи контент-моніторингу.

Автор сподівається, що навички, одержувані сту-

дентами в процесі роботи із системою InfoStream, надалі будуть використані ними під час пошуку та узагальнення інформації багатьох типів (не тільки текстів) за допомогою найрізноманітніших пошукових систем. Формалізація запитів, пошук значимих елементів, урахування лінгвостатистичних особливостей, поетапне уточнення критеріїв пошуку та багато інших підходів і прийомів повинні сприяти підвищенню ефективності інформаційно-аналітичної діяльності в будь-якій галузі.

ЛІТЕРАТУРА

1. *Брайчевский С.М., Ландэ Д.В.* Современные информационные потоки: актуальная проблематика // Научно-техническая информация. — 2005. — № 11. — С. 21—33.

2. *Григорьев А.Н., Ландэ Д.В.* Адаптивный интерфейс уточнения запросов к системе контент-мониторинга InfoStream // Тр. междунар. конф. "Диалог'2005". — С. 109—111.

3. *Ландэ Д.В.* Поиск знаний в Internet. Профессиональная работа — М.: "Вильямс", 2005. — 272 с.

4. *Чурсин Н.Н.* Популярная информатика. К.: Техника, 1982. — 158 с.

5. *Ландэ Д.В., Брайчевский С.М.* Определение тематической направленности запросов путем анализа наборов рейтинговых источников // Открытые информационные и компьютерные интегрированные технологии: Сб. науч. трудов. — Вып. 29. — Харьков: аэрокосмический ун-т "ХАИ", 2005. — С. 41—48.

6. *Ландэ Д.В., Григорьев А.Н., Брайчевский С.М.* Стабильность источников информации как один из параметров информационных потоков. // Компьютерная лингвистика и интеллектуальные технологии: Тр. междунар. конф. Диалог'2006 — Москва: Наука, 2006. — С. 332—334.



РЕФЕРАТИ

Практичний досвід інформаційно-аналітичної підтримки процедур вироблення і прийняття управлінських рішень / Бугай С.М., Лисак Н.В. // Научно-техническая информация. — 2006 — № 3. — С. 3—6.

Розкрито практичний досвід інформаційно-аналітичної підтримки процедур вироблення і прийняття управлінських рішень щодо економічного і соціального розвитку регіонів. Розроблено методику розрахунку та діагностики, яка надає можливість отримати узагальнену оцінку роботи райдержадміністрацій за тенденцією зміни основних показників статистичної звітності. Усі розрахунки та побудова картографічного матеріалу виконано на базі автоматизованої системи "Аналіз соціально-економічних показників діяльності місцевих органів виконавчої влади" на платформі пакету прикладних програм "Логос 3.0", розробником якої є ТОВ "Інформаційно-аналітична фірма ЕКСОР".

Інформація — стратегічний важіль ефективного управління страховим фондом документації України / Степаненко В.Л., Кривулькін І.М. // Научно-техническая информация. — 2006 — № 3. — С. 7—10

Розкрито роль інформації в державній системі страхового фонду документації України. Висвітлена мета діяльності Державного департаменту СФД та державної системи в цілому. На основі теоретичних робіт з державного управління здійснено аналіз роботи державної системи СФД. Наголошено на важливості створення СФД в Україні для забезпечення потреб мобілізаційної підготовки та сталого функціонування економіки держави, відбудови об'єктів систем життєзабезпечення населення, проведення аварійно-рятувальних та аварійно-відбудовних робіт під час ліквідації наслідків надзвичайних ситуацій.

Інформаційна аналітика як напрям наукового дослідження / Пархоменко О.В. // Научно-техническая информация. — 2006. — № 3. — С. 11—13

У статті обґрунтовано можливість віднесення інформаційно-аналітичної діяльності до наукової сфери, також розглянуто особливості ведення інформаційно-аналітичних досліджень і викладено послідовність інформаційно-аналітичного методу дослідження.

Стан науково-технічного потенціалу України / Чорнобай Р.В. // Научно-техническая информация. — 2006. — № 3. — С. 14—19

У статті розглядаються питання розвитку складових науково-технічного потенціалу. Проведений аналіз дозволяє запропонувати ряд заходів щодо покращення побудови системи моніторингу стану науково-технічного потенціалу в Україні.

Механізми аналітичного обґрунтування ефективності виконання управлінських рішень в органах державного управління / Желюк Т.Л. // Научно-техническая информация. — 2006. — № 3. — С. 19—24

У статті проаналізовано організаційний, правовий, функціональний механізми аналітичного забезпечення оцінки ефективності управлінської діяльності з використанням інструментарію інституційного, економіко-стратегічного та змішаного функціонального огляду та адміністративного аудиту, що нададуть можливість підняти ефективність управлінської діяльності в окремому державному органі й у державному секторі загалом до рівня європейських стандартів.

Інноваційний процес: аналіз у рамках теорії інформаційної економіки / Круп'як Т.П. // Научно-техническая информация. — 2006. — № 3. — С. 25—29

У статті відзначаються основні недоліки традиційного підходу до розуміння інновацій і надаються рекомендації щодо аналізу інноваційного процесу з позицій інформаційного підходу. Наголошується на актуальних тенденціях у стратегії розвитку інновацій.

Структура інтелектуального капіталу / Гава Ю. В. // Науково-технічна інформація. — 2006. — № 3. — С. 29–32

Розглянуто підходи, щодо визначення структури інтелектуального капіталу, систематизовано елементи даної категорії, запропоновано нову структуру категорії.

Державний борг у системі статистики державних фінансів: проблеми гармонізації / Царук О.В. // Науково-технічна інформація. — 2006. — № 3. — С. 33–36.

Розглянуто основні стандарти статистики державних фінансів. Визначено загальні концептуальні відмінності в методології обчислення зовнішнього боргу та обліку державних коштів, що використовується в Україні та прийнятій МВФ. Визначені напрямки гармонізації статистики державних фінансів в Україні.

Проблеми співвідношення витрат на житлово-комунальні послуги і тарифів на них / Гура Н.О. // Науково-технічна інформація. — 2006. — № 3. — С. 36–41.

Визначено проблемні питання формування витрат і доходів підприємств житлово-комунального господарства, проаналізовано їхню структуру, розмір та співвідношення витрат і доходів. Обґрунтовано необхідність підвищення тарифів на житлово-комунальні послуги з урахуванням платоспроможності населення і визначено взаємопов'язані шляхи оптимізації тарифів і зниження витрат на послуги підприємств житлово-комунального господарства.

Інвестиційна діяльність страхових компаній — шлях до економічного зростання / Майдаченко І. В. // Науково-технічна інформація. — 2006. — № 3. — С. 41–44.

Розкрито одну з функцій страхових компаній — інвестиційну діяльність та її значення для зростання економіки України. Подано дані про частку інвестиційного доходу та прибутку в загальному обсязі доходів і прибутків за окремими страховими компаніями та за регіональним розподілом.

Інвестиційна привабливість банківського бізнесу та його ефективність / Гуляєва Л.П. // Науково-технічна інформація. — 2006. — № 3. — С. 45–48.

У статті за допомогою моделі Дюпона здійснено аналіз факторів, що формують ефективність та інвестиційну привабливість банківської діяльності. Досліджується роль кожного з них у діяльності українських банків з метою визначення можливих напрямків підвищення ефективності банківської справи в Україні.

Автоматизована інформаційна система реєстрації, пошуку й аналізу науково-технічних заходів / Гончаренко А.П., Каретнікова Л.Х., Войтенко В.В., Кушнір Г.Л., Мельник М.П. // Науково-технічна інформація. — 2006. — № 3. — С. 49–52.

Стаття присвячена створенню бази даних "Науково-технічні заходи України", яка у вигляді автоматизованої інформаційної системи введена в експлуатацію вперше.

Наведено матеріали щодо структурної організації бази даних, пошукової системи та правил доступу користувачів до бази.

Розвиток системи інформаційно-бібліотечного забезпечення науки / Симоненко Т. В. // Науково-технічна інформація. — 2006. — №. — С. 52–56.

Визначено напрямки розвитку системи інформаційно-бібліотечного забезпечення наукових досліджень в умовах інтенсивного використання комп'ютерно-телекомунікаційних технологій.

Сучасні тренажерні системи навчання спеціалістів з інформаційних технологій / Герасимов Б.М., Камішин В.В., Любарський С.В. // Науково-технічна інформація. — 2006. — № 3. — С. 56–62.

Розглянуто принципи побудови сучасних тренажерних систем навчання спеціалістів з інформаційних технологій, наведено основні етапи створення таких систем, показано схеми взаємодії складових при функціонуванні тренажерних систем, запропоновано застосування конкретних методів, що можуть бути застосовані під час розробки тренажерних систем та визначені основні етапи розробки тренажерів. Запропонована структура сучасного тренажера та його типова функціональна схема.

Гіперкомплексна система як інформаційна модель функціональної системи діяльності людини / Бузов О.Ю. // Науково-технічна інформація. — 2006. — № 3. — С. 62–64.

Запропоновано опис працездатності оператора як багатоякісної системи, що складається з підсистем, які описують різні за якістю складові працездатності. З цією метою використано теоретичні та методичні положення теорії гіперкомплексних динамічних систем та відповідну ГДС-модель.

Така модель дає змогу зробити адекватний формалізований опис трирівневої моделі формування працездатності оператора, розробленої автором, в якій різноякісними елементами виступають множини параметрів професійної придатності оператора, його професійного старіння та поточної працездатності.

Дослідження лінгвостатистичних характеристик інформаційних потоків новин у рамках магістерського курсу "Теорія інформаційного пошуку" / Ланде Д.В. // Науково-технічна інформація. — 2006. — № 3. — С. 55–71.

Наведено опис тематики й змісту практичних занять у рамках магістерського курсу "Теорія інформаційного пошуку".

Для проведення занять був організований доступ студентів до системи контент-моніторингу інформаційних потоків новин з Інтернету InfoStream, забезпечена взаємодія з популярними програмами обробки статистичної інформації.

