

УДК 623.647:623.644:623.618

В. Е. Лісцин, С. В. Бєлай

ПІДХОДИ ДО РОЗВ'ЯЗУВАННЯ ЗАДАЧ ПРОГНОЗУВАННЯ РОЗВИТКУ КРИЗОВИХ СИТУАЦІЙ В РЕГІОНАХ УКРАЇНИ ЗА ДОПОМОГОЮ ГЕОІНФОРМАЦІЙНОЇ СИСТЕМИ ВНУТРІШНІХ ВІЙСЬК “ІНСТРУМЕНТ”

Пропонується підхід до прогнозування соціальних процесів та розвитку кризових ситуацій соціального характеру за допомогою інструментів геопросторового аналізу спеціалізованої геоінформаційної системи внутрішніх військ.

Ключові слова: кризові ситуації соціального характеру, кодування ознак, геопросторовий аналіз, геоінформаційні системи.

Постановка проблеми. Недавні події на Близькому Сході, які отримали назву “арабська весна”, масові заворушення, протести та виступи в країнах Європейського Союзу в черговий раз демонструють необхідність мати в державній системі ефективні інструменти моніторингу і запобігання розвитку подібних явищ. Найважливішою умовою запобігання соціальним конфліктам є своєчасний аналіз і розпізнавання ситуації, передуючої соціальним виступам. З цією метою в розвинених країнах широко застосовуються різноманітні методи збирання, статистичного аналізування і комп’ютерного оброблення інформації про суспільно-протестну активність різних шарів суспільства. Створюються бази даних, кодовані та індексовані з метою прискорення доступу і аналізу, що містять ознаки соціальних явищ.

У статті надаються пропозиції щодо створення такої системи для зберігання, відображення на карті, статистичного аналізування та подальшого прогнозування розвитку кризових явищ соціального характеру в регіонах держави.

Аналіз останніх публікацій. Розглянемо декілька характерних прикладів реально існуючих сьогодні комп’ютерних систем, що містять подібну інформацію. Це дозволить одержати відповідь на поставлене питання.

Деякі з таких баз даних відкриті і загальнодоступні для використання. Наприклад, інформація про всі (незалежно від тематики та чисельності) протестні дії на території України [1], що зібрана за допомогою міжнародного фонду “Відродження”. Мінімальною структурою кодування інформації у цій системі на логічному рівні є, так звана, “протестна подія”. У супровідній документації наведена методика збирання даних, що віднесені до цієї тематики. Під час соціологічного опитування визначають критерії подій [2], які слід віднести до класу протестних, та як кодувати зібрану інформацію перед її внесенням у комп’ютер. Така методика мало чим відрізняється від класичних схем проведення та оброблення результатів соціологічного опитування.

Більш цікавим є те, що у документі [1] пропонується логічна структура бази даних, яка містить зібрану інформацію. Наприклад, структура, яка відведена для події, що кодується, містить поля “Oblast” (назва обласного центру) та “Location” (назва населеного пункту). Це дає можливість провести геокодування події, тобто прив’язати її до карти місцевості. Проте для більш детального аналізування розподілу подій на великомасштабній карті (карта міста або міських кварталів) воно не є ефективним. Водночас застосування будь-якої геоінформаційної системи (ГІС) дозволяє під час роботи вибирати різні масштаби електронних карт. Більш того, деякі результати оброблення геопросторової інформації суттєво залежать від обраного масштабу карти.

Слід також зазначити, що на час написання цієї статті вказана база не містила будь-яких інструментів статистичного аналізування отриманих результатів опитування. Отже, надаючи за допомогою Інтернету вільний доступ до інформації бази, розробники перекладають весь тягар оброблення цієї інформації на користувачів. З одного боку це оптимально з точки зору вибору алгоритмів оброблення, максимально пристосованих для досягнення конкретної мети, з іншого – обов’язково потребує розроблення необхідного комплексу програмних інструментів та методик інтерпретації результатів.

Розглянемо можливості іншої системи, яка вирішує проблему аналізування соціального стану на основі даних, отриманих із засобів масової інформації. Американська комп’ютерна система “Наутилус” спроможна на підставі даних, що містяться у повідомленнях агентств новин та газетних

статтях, прогнозувати розвиток подій у певному регіоні [3]. Коректність прогнозу цієї системи перевірена за допомогою ретроспективної моделі, у якій комп'ютер аналізував текстові повідомлення про події, що вже мали місце, та результат яких був відомий. Виконувався пошук достатньо простих ключових слів та фраз, що на думку розробників системи характеризують наявність або відсутність ознак стану, що аналізують. Ознаки, які додавали у базу даних, також підлягали геокодуванню. Було проаналізовано інформаційні потоки у країнах "арабської весни". У більшості випадків результати роботи ретроспективної моделі співпали з відомим розвитком подій, також було виявлене значне соціальне напруження та помітне погіршення суспільної атмосфери у країнах напередодні розвитку кризових явищ.

На відміну від першої системи, розглянутої у статті, інформація про структуру даних та розрахункові моделі комп'ютерної системи "Наутилус" вкрай обмежена. Повідомляється тільки, що було проаналізовано більше 100 000 ретроспективних джерел інформації, а апаратна структура процесора містила 1 024 ядра.

Мета статті – розглянути вимоги до системи, яка може бути використана для аналізування соціального стану та прогнозування розвитку кризових явищ в межах певної території, а також запропонувати інструменти для роботи такої системи сумісно з ГІС "Інструмент" внутрішніх військ МВС України.

Виклад основного матеріалу. Слід звернути увагу на декілька вкрай важливих моментів. По-перше, під час службово-бойової діяльності внутрішні війська безпосередньо виконують завдання із урегулювання та попередження розвитку кризових явищ, обумовлених протестними соціальними виступами. Саме тому, що будь-яку проблему краще вирішувати на випередження, вкрай необхідно мати простий, ефективний інструмент, який дозволив би аналізувати поточний стан соціальної напруженості в суспільстві, виходячи зі статистики інформаційних потоків, які циркулюють у засобах масової інформації та соціальних мережах. Простота такого інструмента обумовлена програмною та технічною базами, за допомогою яких має бути вирішена ця проблема. Цілком зрозуміло, що, виходячи з реальних економічних обставин сьогодення, не слід закладати в систему надто високі чи дорогі технічні показники, наприклад, процесор з 1 024 ядрами. Система повинна функціонувати на базі звичайного двоядерного процесора настільного комп'ютера або не дуже потужного сервера.

По-друге, враховуючи територіальні чинники неоднорідності накопичення і розвитку соціальних проблем, такі дослідження необхідно проводити для конкретних регіонів або області України, а під час роботи з соціальними мережами – в межах міста чи, навіть, окремих міських кварталів. Таким чином, ми підходимо до питань геопросторового статистичного аналізу соціальних процесів українського суспільства. Якщо реалізувати структуру геопросторових даних, що описує соціальні кризові явища, на сервері баз даних MS-SQL, то у нас з'явиться можливість формувати не тільки точкові події, а і регіони поширення явища. За допомогою методів, передбачених для спеціальних просторових типів даних GEOGRAPHY та GEOMETRY, у системі управління базами даних MS SQL Server 2008 можна створювати код на мові T-SQL, що реалізує операції з реляційними та просторовими даними, наприклад, виконує пошук точок перетину двох об'єктів карти або визначає відстані між населеними пунктами. Також можливо виконувати повноцінні і складні просторові запити, наприклад, визначити кількість подій з ознакою, яка цікавить нас, що відбулися у межах міського кварталу за певний період часу.

По-третє, цінність висновків, генерованих такою системою, напряму пов'язана з обсягом статистичної інформації, яка була зібрана на етапі аналізування інформаційних потоків. Не варто у будь-якому випадку обмежуватися двома-трьома десятками джерел повідомлень. Нагадаємо, що у системі ретроспективного аналізу "Наутилус" кількість таких повідомлень для кожного прогнозу перевищувала 100 000. У такому разі постає питання ефективності та швидкості оброблення запитів до електронній карти та бази даних, у якій містяться ознаки ситуацій, що аналізують. Якщо, наприклад, під час роботи на карті масштабу міського кварталу обсяги даних відносно невеликі, то після переходу на карти масштабу регіонів та областей без застосування спеціальних методик роботи з просторовими даними, суттєво збільшується час, потрібний системі для аналізування.

У таких випадках допомагають просторові індекси. Їх застосування – це аналог загального процесу індексації у реляційних базах даних, коли створюється додатковий атрибут, який містить інформацію деякого стовпчика реляційної таблиці, впорядковану за обраним алгоритмом. Наприклад,

слова електронного англо-українського словника внесені у базу даних та впорядковані за алфавітом. Уявіть собі, наскільки б ускладнився пошук у такому словнику, якби слова були розташовані у вигляді хаотичного, невпорядкованого набору. Саме так створюють і просторові індекси. Регіон карти поділяють на окремі квадрати, кожному з яких присвоюють певний первинний індекс. Потім кожен зі створених квадратів ділять на квадратні регіони другого рівня. Для кожного з них створюють наступну частину індексу другого рівня. Процес повторюють кілька разів (рис.1). Об'єкти, які розташовані на карті, відображаються сірим кольором. Після закінчення процесу місцезнаходження кожного об'єкта буде визначено з точністю, що відповідає площині квадрата, створеного на останньому рівні індексації. Таким чином, пошук на карті і у базі деякого об'єкта буде обмежений, наприклад, квадратами 5 (рис. 1, а), 11 (рис. 1, б), 8 (рис. 1, в) та 15 (рис. 1, г). Це значно скорочує час, потрібний для виконання запитів.

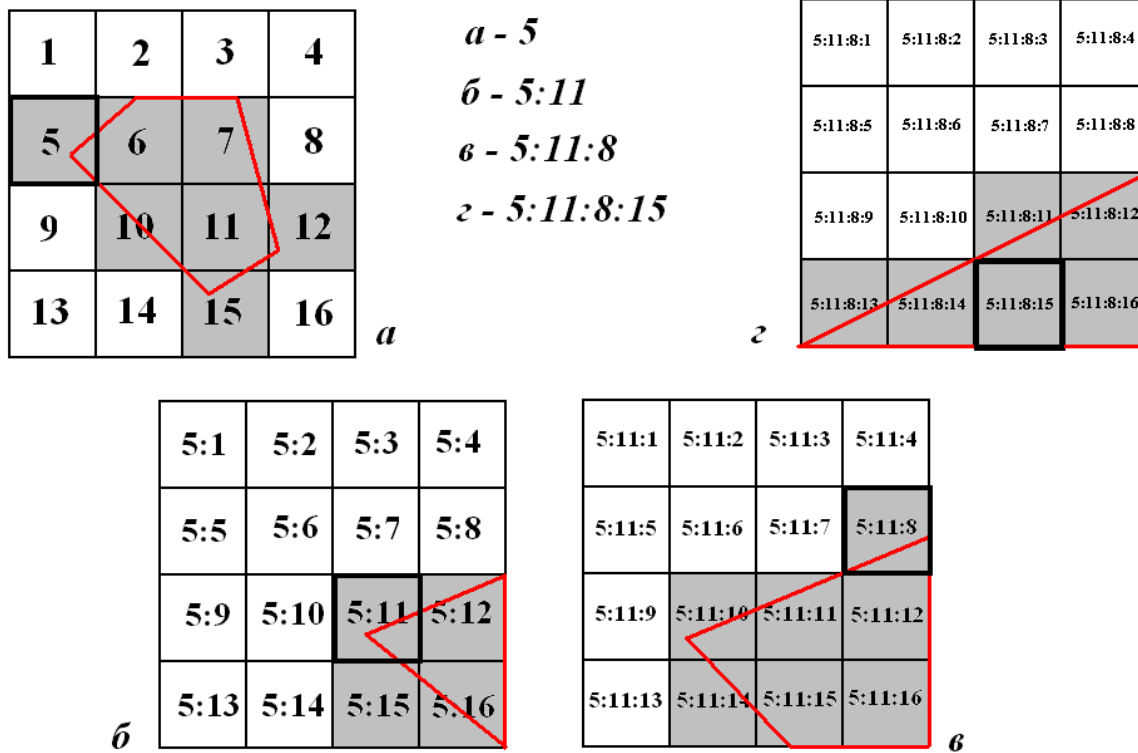


Рис. 1. Процес створення чотирирівневого просторового індексу для визначеного регіону карти

Зрозуміло, що чим більше рівнів містить просторовий індекс, тим швидше виконуватиметься пошук у геобазі даних. Проте слід пам'ятати, що збільшення кількості рівнів просторового індексу призводить до розростання самої бази просторових об'єктів. На сервері баз даних MS-SQL-2008 кількість рівнів просторового індексу обмежена чотирма.

На етапі дослідження пропонується реалізувати:

- структуру бази даних, що містить атрибути соціальної активності об'єктів суспільства, індексовані за якісними ознаками;
- геопросторову атрибутивну складову, що дозволяє визначати регіони поширення соціальних явищ і ознак загрози соціального характеру, а також виконувати індексацію і вибирання за геопросторовими ознаками (координати, регіони поширення);
- сполучення бази даних соціальних явищ з ГІС “Інструмент” внутрішніх військ;
- механізм та інструменти наповнення геопросторової бази даних інформацією, одержаною з текстових джерел (Інтернет, засоби масової інформації);
- механізм формування висновку на основі статистичної інформації, що міститься в геопросторовій базі даних, прогнозів розвитку соціальної ситуації в різних регіонах України і відображення на карті результатів аналізу.

Джерелами поповнення такої геопросторової бази даних соціальних явищ в даному випадку можуть бути електронні і паперові засоби масової інформації, сайти і соціальні мережі Інтернету. Повноцінне автоматичне аналізування таких даних пов'язане з питаннями комп'ютерного оброблення текстів,

написаних звичайними мовами. Така реалізація надзвичайно ускладнила б поставлену задачу. Тому як альтернативне рішення пропонується використовувати регулярні вирази [4], які пропонують засоби розроблення пакету MS Visual Studio, і готові текстові шаблони пошуку. Структура системи, що пропонується, наведена на рис. 2.

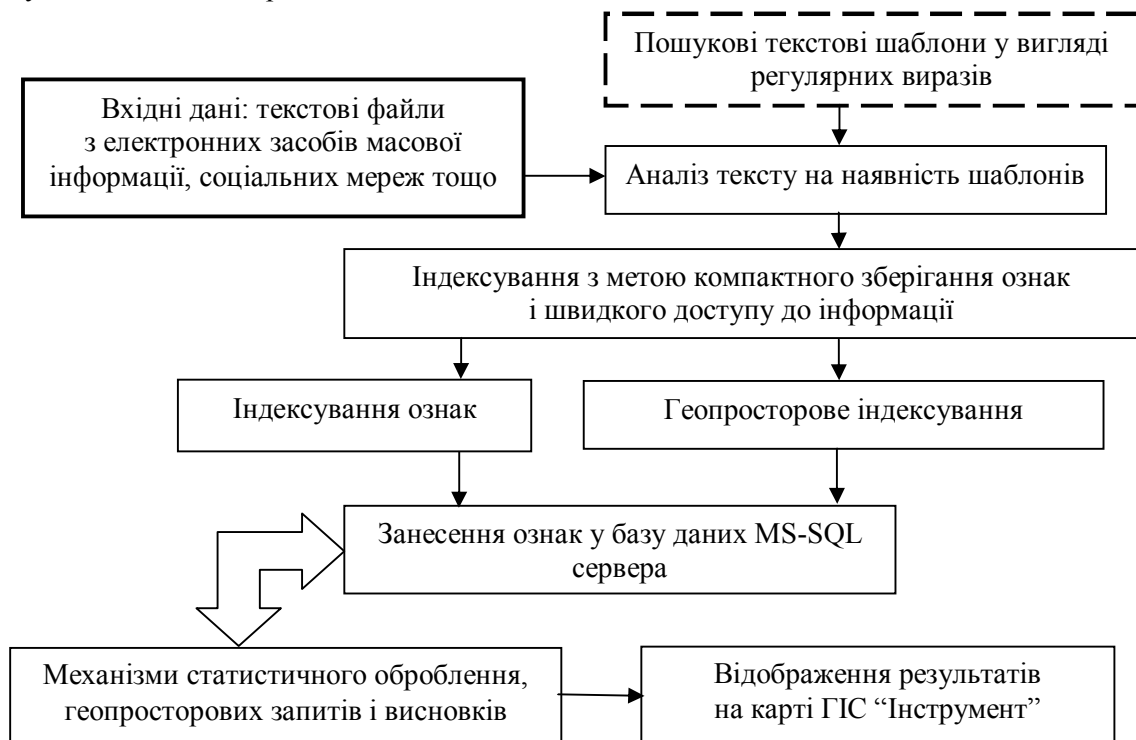


Рис. 2. Система моніторингу текстової інформації та аналізування соціального стану і розвитку кризових явищ

Тепер розглянемо, яким чином геопросторові відношення між об'єктами карти можливо використати для створення зв'язків та складання прогнозу розвитку кризових явищ. Кожна протестна подія, нанесена на карту, стає геокодованою, тобто може бути описана за допомогою координат карти. Взагалі, кожен запис у реляційній таблиці, що містить таку подію, необхідно розглядати як об'єкт карти. Значення атрибутів цього запису стають властивостями об'єкта. Для спрощення моделі розглядатимемо тільки дискретні властивості – такі, що мають фіксовану, наперед визначену кількість значень у реляційній таблиці. Для більшої наочності кожному такому дискретному значенню поставимо у відповідність унікальне значення кольору, за допомогою якого властивість може бути відображена на карті у точці географічного розташування об'єкта. Якщо за допомогою цих кольорів нанести на електронну карту місцевості усі можливі значення деякого атрибута реляційної таблиці, що містить геокодовані події, то у результаті отримаємо кольорову карту деякої дискретної змінної у вигляді додаткового тематичного шару (див. рис. 3).

Тепер ми можемо отримати кількісні характеристики для тих регіонів електронної карти, де географічно прив'язані події (або деякі їх властивості) мають тенденцію до створення кластерів. Для цього треба розрахувати показник просторової автокореляції. Він знаходиться у межах від -1 до $+1$, при цьому значення менше нуля вказують на ті регіони карти, де географічні об'єкти з однаковим значенням властивостей (або однакові події) розподілені по всьому регіону (див. рис. 4).

Суттєвою відмінністю застосування таких методів під час аналізування статистичних показників є те, що в явному вигляді враховуються просторові зв'язки об'єктів електронної карти, їх взаємне розташування, перекриття, спільні межі. Тобто оцінюється показник взаємозалежності між різними значеннями одного атрибута у точках, які просторово зміщені одна відносно одної. На відміну від звичайної одновимірної автокореляції, просторові показники розраховуються у двовимірному (тривимірному) просторі. Водночас звичайні (не просторові) статистичні методи обмежуються тільки розрахуванням показників групування значень вибраних атрибутів відносно точок карти.

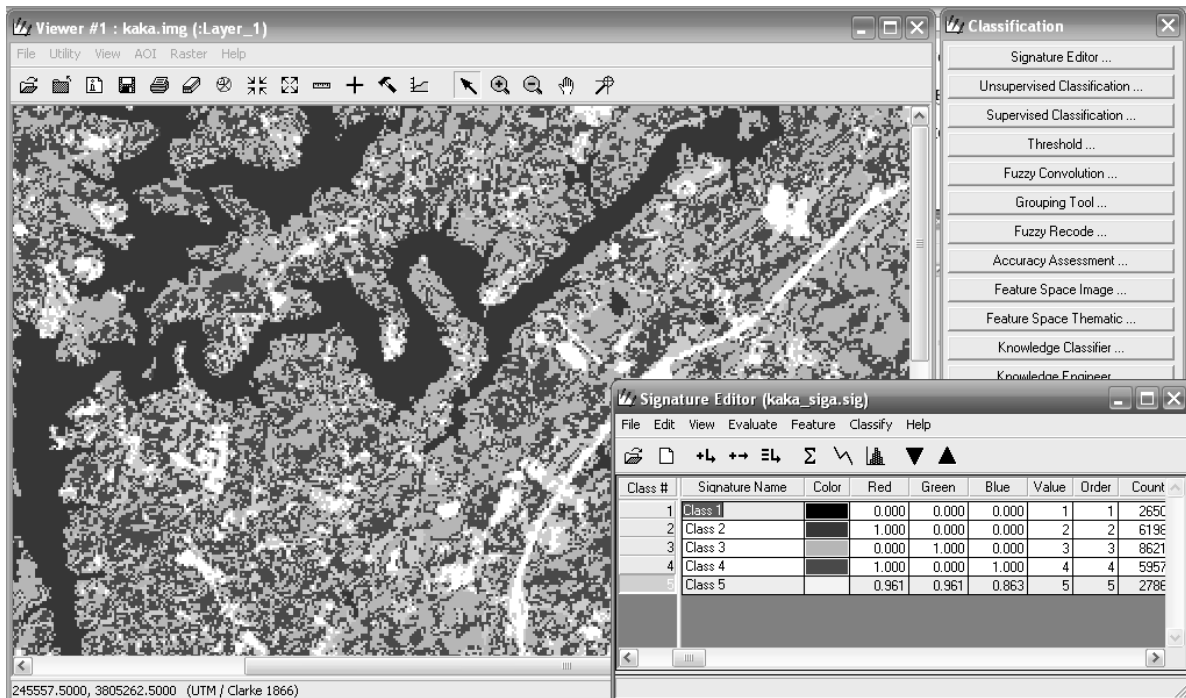


Рис. 3. П'ять геокодованих значень властивостей деяких об'єктів, нанесених на карту за допомогою ГІС ERDAS Imagine

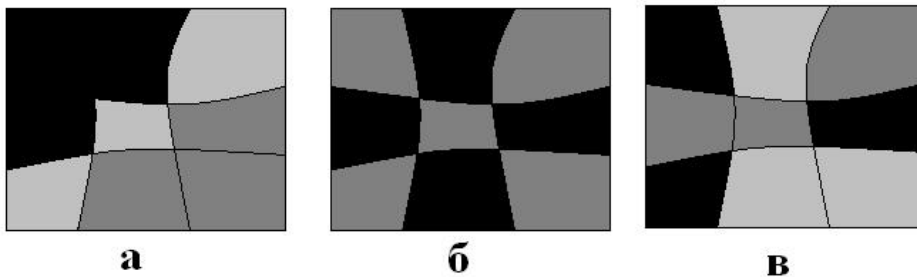


Рис. 4. Ілюстрація просторової автокореляції: *a* – позитивна, об'єкти групуються на карті у вигляді кластерів; *б* – негативна, об'єкти рівномірно розташовані в обраному регіоні карти; *в* – показник кореляції дорівнює нулю, об'єкти розташовані на карті хаотично

Показник просторової автокореляції [5] може бути визначений, наприклад, за допомогою І-коефіцієнта Морана (Patrick A. P. Moran):

$$I = \frac{n}{S} \left[\frac{\sum_{i=1}^n \sum_{j=1}^n \mathcal{M}_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right], \quad (1)$$

де n – загальна кількість осередків на кольоровій карті ознак, в яких оцінюють значення деякого атрибута; i, j – одновимірні індекси осередків; x – поточне значення ознаки чи атрибута; \bar{x} – середнє значення атрибута, що аналізують у межах регіону; \mathcal{M}_{ij} – вагові коефіцієнти для кожної пари осередків на карті ознак; $\mathcal{M}_{ij} = 1$, якщо i та j – сусідні осередки, у протилежному випадку $\mathcal{M}_{ij} = 0$; S визначається за формулою:

$$S = \sum_{i=1}^n \sum_{j=1}^n \mathcal{M}_{ij}, \quad i \neq j. \quad (2)$$

Висновки

1. Одним з основних завдань штабів військових частин та з'єднань внутрішніх військ є моніторинг кризових ситуацій соціального характеру в зонах їх відповідальності та проведення відповідних заходів протидії з метою недопущення виникнення загроз національній безпеці.
2. Командуванню військ необхідно мати простий, але ефективний інструмент, який дозволив би проводити аналіз поточного стану соціальної напруженості в суспільстві, використовуючи дані статистики інформаційних потоків, які циркулюють у засобах масової інформації та соціальних мережах.
3. Запропонований у статті підхід планується реалізувати у рамках наукових досліджень з використанням ГІС-технологій, методів оптимізації виконання операцій з об'єктами карти та побудови просторових індексів, кінцевою метою яких є побудова системи прогнозування кризових явищ соціального характеру.

Список використаних джерел

1. Протести, перемоги і репресії в Україні: результати моніторингу, жовтень 2009 – вересень 2010 [Текст] – К. : Центр дослідження суспільства, 2011. – 64 с.
2. Ukrainian protest and coercion data codebook. November 1, 2011 [Текст]. – К. : Центр дослідження суспільства, 2011. – 20 с.
3. Суперкомпьютер способен предсказывать будущее [Электронный ресурс]. – Режим доступа : <http://www.segodnya.ua/news/14287696.html>. – Загл. с экрана.
4. Фридл, Дж. Регулярные выражения / Дж. Фридл; пер. с англ. – С Пб. : Питер, 2003. – 464 с.
5. Prachi Misra Sahoo. Statistical techniques for spatial analysis. I.A.S.R.I., Library Avenue, New Delhi – 110 012. 2009, 11 pp.

Стаття надійшла до редакції 20.06.2012 р.