

С.А. Малышко

## Информационная технология анализа медиaproстранства на основе системы *NEWSCAPE*

Описана информационная технология анализа медиaproстранства на основе системы *NEWSCAPE*, формализованы основные этапы сравнительного анализа объектов в информационном пространстве; предложены их математически вычисляемые характеристики, положенные в основу многофакторного анализа объектов посредством информационно-аналитической системы.

**Ключевые слова:** *NEWSCAPE*, анализ медиaproстранства, система поддержки принятия решений, информационное пространство, большие данные.

Описано інформаційну технологію аналізу медиaproсторору на основі системи *NEWSCAPE*, формалізовано основні етапи порівняльного аналізу об'єктів в інформаційному просторі; запропоновано їх математично обчислювальні характеристики, покладені в основу багатofакторного аналізу об'єктів за участі інформаційно-аналітичної системи.

**Ключові слова:** *NEWSCAPE*, аналіз медиaproсторору, система підтримки прийняття рішень, інформаційний простір, великі дані.

**Введение.** Современные системы поддержки принятия решений (СППР) в качестве основных компонент содержат информационные хранилища данных, а также подсистемы, реализующие соответствующие методы и средства извлечения, обработки и загрузки данных. Источником данных, помимо специализированных баз данных (БД), может выступать открытое информационное пространство (медиапространство). В общем случае термин *информационное пространство*, как первичное понятие, не имеет строго научного определения и считается общеупотребимым понятием [1]. В данной статье под информационным пространством будем понимать медиапространство, образованное совокупностью сообщений от выделенных источников информации. В качестве таких источников рассматривается множество медийных ресурсов физических и юридических лиц, электронных версий традиционных СМИ, онлайн-ресурсов (сайты, порталы) и «новых медиа» (социальные сети, блоги).

Всеобщая компьютеризация и масштабное развитие глобальных сетей электронных коммуникаций привели к возникновению огромного медиaproстранства, вследствие чего, задача поиска необходимой информации переросла в проблему отсеивания нерелевантных данных и анализа результатов поиска [2]. На первый план при анализе информационного пространства выходят задачи поиска закономерностей и взаимосвязей, отслеживания тенденций и трендов, прогнозирование динамики

числовых показателей исследуемых объектов в медиaproстранстве. Разрабатываемые информационно-аналитические системы (ИАС), которые служат основой многих развитых СППР, позволяют увидеть «простое и понятное» в огромном, сложном и хаотическом информационном пространстве.

Единой универсальной стандартизированной технологии мониторинга и анализа информационного пространства пока не существует, и есть сомнения в том, что она появится в обозримом будущем. Медиaproстранство продолжает количественно увеличиваться и качественно расширяться, что приводит к росту актуальности и практической востребованности исследуемой в статье проблематики. На базе разрабатываемой ИАС *NEWSCAPE* [3, 4] создаются современная эффективная информационная технология и основанный на ней онлайн-сервис, аккумулирующие различные математические методы и алгоритмы, ориентированные на мониторинг и многофакторный анализ медиaproстранства. Часть идей, методов и алгоритмов, используемых в ИАС, – это оригинальные разработки, начало которым было положено свыше 20 лет назад [5]. Сравнительный обзор существующих систем мониторинга медиaproстранства выходит за рамки данной публикации вследствие разнообразия и специфики программных продуктов такого рода, а также особенностей понятия *открытые данные* в различных странах. В качестве примера подобного рода разработок отметим технологию

интеграции новостей в отечественной системе *InfoStream* [6].

Современные научно-технические разработки сотрудников Института кибернетики им. В.М. Глушкова НАН Украины в сфере математического моделирования и информационных технологий, развитие и реализованные усилиями объединенного коллектива высококвалифицированных специалистов, и представляет собой современный программно-аппаратный комплекс ИАС *NEWSCAPE*, реализующий основные функции мониторинга медиапространства. Следует отметить, что технология мониторинга и анализа в *NEWSCAPE* продолжает активно развиваться и совершенствоваться в соответствии с потребностями практики и современными требованиями к эффективности программного обеспечения, ориентированного на широкий круг пользователей. Расширяется также и аналитический инструментарий системы, основанный на оригинальном математическом аппарате.

Разработанная информационная технология анализа медиапространства позволяет объединять в единое целое на основе качественных и количественных оценок и показателей информационных объектов программные системы и подсистемы, а также аккумулировать разрозненные базы данных, что позволяет повысить эффективность и стандартизировать процесс анализа информационного пространства.

### **О системе, основанной на технологии анализа медиапространства**

Напомним об особенностях ИАС *NEWSCAPE*, поскольку подробное ее описание приведено в [3]. Предлагаемая технология создания развитых информационных систем базируется на использовании современных веб-приложений, что обеспечивает оперативность и доступность ИАС в любых точках, имеющих подключение к интернету с возможностью использования различных терминальных устройств – компьютеры, планшеты, смартфоны. Системы подобного класса востребованы в крупных промышленных и финансовых структурах, исследующих и анализирующих свою деятельность и в медиапространстве. Как эф-

фективный инструмент поддержки принятия решений ИАС *NEWSCAPE* может быть использована центральными и местными органами власти для анализа экономической и социально-экономической информации.

Система обеспечивает процесс выработки эффективных решений различного уровня: от задач оперативного функционирования ситуационного центра до проблем стратегического планирования. *NEWSCAPE* также может быть эффективно использована и в современной информационной войне – как для контроля и планирования действий «своих», так и для выявления и анализа действий «чужих». Система ориентирована на использование как в автономном режиме, так и в качестве интеллектуальной компоненты СППР в трудноформализуемых и конфликтных ситуациях.

Реализованные на основе технологии анализа медиапространства программно-алгоритмические средства ИАС позволяют:

- обеспечить непрерывный автоматизированный мониторинг разнотипного медиапространства;
  - аккумулировать разнообразные разрозненные базы данных в единую информационную систему;
  - накапливать в структурированном виде результаты мониторинга для последующего анализа;
  - использовать современные методы количественного и качественного анализа информации;
  - обеспечить эффективную коллективную работу операторов и аналитиков.
- Архитектура и функциональное наполнение ИАС создавалось для достижения таких целей, как обеспечение возможности:
- оперативно получать необходимую и выявлять достоверную информацию;
  - отслеживать основные тенденции в медиапространстве;
  - проводить многофакторный анализ разнотипных объектов;
  - совершенствовать управление масштабными системами и подсистемами;

- моделировать и исследовать сложные социально-экономические процессы;

- формировать и анализировать предполагаемые прогнозные сценарии развития процессов.

Пользователям доступен развитый инструментарий для многофакторного исследования, обеспечивающий: анализ характеристик объектов в информационном пространстве, разнотипный контент-анализ (количественный, частотный, разностный), категорийный, морфологический анализ, анализ количественных характеристик текстов.

### **Технология анализа информационного пространства**

Анализа медиапространства с использованием, в том числе, актуальной версии системы *NEWSCAPE* основан на анализе категорий, которые формируются пользователями и характеризуют исследуемые структурные единицы – объекты, персоны, процессы, события и др. Опишем основные этапы анализа информационного пространства.

**Формирование подмножества медиапространства для анализа** – составление объединенного структурированного списка источников информации с возможностью выбора из них отдельных групп:

$$A \subset Q, \quad (1)$$

$$A = A_1 \cup \dots \cup A_n, \quad (2)$$

где  $Q$  – все медиапространство,  $A$  – подмножество для мониторинга информационного пространства,  $A_i$  – группы источников для мониторинга медиапространства,  $i = 1, \dots, n$ ,  $n$  – количество групп в данной сессии мониторинга.

Отметим, что группы источников информации могут формироваться по различным признакам, например по географической привязке, по собственности (группа официальных государственных сайтов), по популярности (группа наиболее рейтинговых сайтов), по содержанию (группа новостных сайтов) и пр. Пользователь системы может формировать персональные наборы групп и отдельных сайтов для решения специфических задач мониторинга.

**Извлечение информации из подмножества** – исправление ошибок при загрузке информаци-

онных блоков, отсечение информационных шумов, структурирование загруженной информации:  $A_i = B_{i1} \cup \dots \cup B_{mi}$ ,  $B_{ji} = C_{1j} \cup \dots \cup C_{pj}$ , где  $B_{ki}$  – объект из множества источников информации, входящих в группу  $A_i$  для мониторинга,  $k = 1, \dots, m_i$ ,  $m_i$  – количество источников информации в данной  $i$ -й группе,  $C_{kj}$  – информационные блоки, полученные от  $B_j$  источника информации,  $k = 1, \dots, p_j$ ,  $p_j$  – количество информационных блоков от  $j$ -го источника.

**Конкретизация информации** – привязка информационных блоков по времени, территории, упомянутым физическим персонам и юридическим объектам. Проводится, при необходимости, дополнительный поиск по объектам. После обработки конкретизированный информационный блок представляет собой структурированное множество вида  $C_k = \{T, D_1, \dots, D_r\}$ , где  $T$  – текст информационного блока (совокупность символов, слов, предложений),  $r$  – количество идентификаторов, связанных с данным информационным блоком.

Примерами идентификаторов могут быть  $D_1$  – уникальный номер (идентификатор) информационного блока,  $D_2$  – дата/время создания данного информационного блока,  $D_3$  – точный адрес (ссылка) источника информации для блока и пр.

В зависимости от специфики проводимого информационного анализа набор идентификаторов для информационных блоков может изменяться.

**Вычисление базовых показателей** информационного присутствия, таких как количество информационных блоков, индексы цитирования и медиаприсутствия, показатели медиаактивности, рейтинги динамики и другие числовые оценки:

$S$  – общее количество информационных блоков в  $A$  – (1) – за выделенный период времени;

$N$  – количество найденных информационных блоков  $C_k$  в (2) по запросу (слову);

$K$  – индекс цитирования, равный количеству найденных слов по запросу (слову);

$M$  – показатель медиаактивности, равный  $N / S \times 100\%$ ;

$L$  – *індекс медіаприсутства* об'єкта, рівний  $K / N$ .

Пример вычисления базовых показателей показан на рис. 1.

**Вычисление относительных показателей** информационного присутствия, сравнительных индексов информационной активности и рейтингов информационной динамики, сравнительный временной анализ объектов (изотонные и антиизотонные функции) основано на следующих индексах и рейтингах:

- *индекс информационной активности* определяет количественную сравнительную оценку присутствия объекта в информационном пространстве;

- *рейтинг информационной динамики* определяет количественную относительную скорость изменения присутствия объекта в информационном пространстве;

- *индекс регулярности событий*, определяемый на основе анализа показателей повторяемости событий, связанных с объектом за исследуемый период;

- *рейтинг информационной уникальности* объекта определяется на основе анализа динамики изменения базовых количественных показателей объекта в информационном пространстве с учетом ретроспективы (количество и размеры «вершин и пиков» в информационном пространстве);

- *индекс информационной масштабности* объекта определяется на основе агрегирования и взвешивания базовых числовых показателей объекта за исследуемый период;

- *рейтинг информационного влияния* объекта определяется на основе анализа показателей взаимосвязанных дополнительных объектов с учетом рейтингов источников информации.

### Основні результати за період

Всього інформаційних блоків в системі за період: 198192

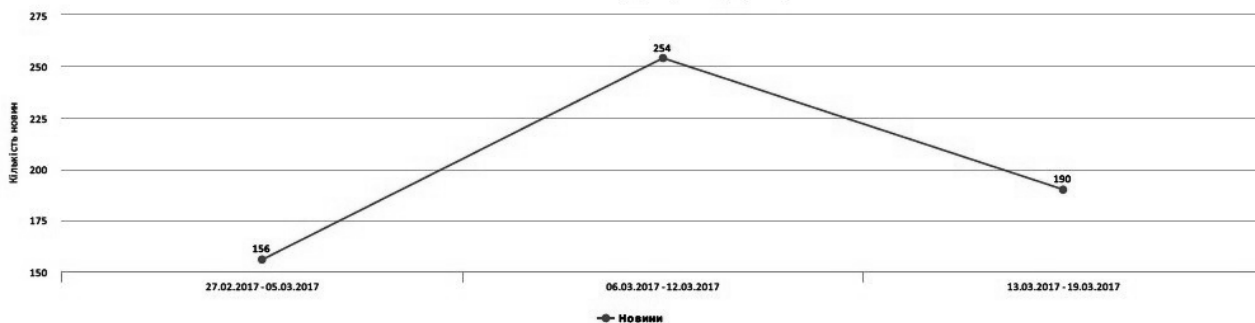
Всього новин: 600

Показник медіаактивності: 0.3

Індекс цитування: 1182

Індекс медіаприсутності: 1.97

Кількість новин в інформаційному просторі



### Додаткові результати за період (по днях)

|                          | 27.02.2017 - 05.03.2017 | 06.03.2017 - 12.03.2017 | 13.03.2017 - 19.03.2017 |
|--------------------------|-------------------------|-------------------------|-------------------------|
| Всього новин             | 156                     | 254                     | 190                     |
| Індекс цитування         | 286                     | 516                     | 380                     |
| Індекс медіаприсутності  | 1.83                    | 2.03                    | 2                       |
| Всього новин в системі   | 59548                   | 65815                   | 72829                   |
| Показник медіаактивності | 0.26                    | 0.39                    | 0.26                    |

**Оценивание информации** базируется на вычислении показателей агрессивности и рейтингов «протестных настроений», оценивании рейтингов персон и других объектов, упомянутых в информационных блоках, подсчете количества перепубликаций информационных блоков, определении тональности и выявлении манипуляционных признаков в информационных блоках. При этом в определении тональности по объекту анализа все соответствующее множество блоков разбивается на три группы: *позитивные, нейтральные* и *отрицательные* оценки по объекту.

**Построение информационных связей** между исследуемыми объектами – группировка взаимосвязанных объектов в подмножества, вычисление дополнительных характеристик взаимосвязей, исследование результатов операций объединения и пересечения сформированных подмножеств, определение независимых событий.

**Углубленный анализ информации** – на этом этапе исследования объектов в медиaprостранстве за период используются дополнительные математические алгоритмы и методы: контент-анализ, категорийный анализ с использованием статичных и динамичных категорий, разностный анализ, многофакторный сравнительный анализ. Подробное описание дополнительных уникальных методов анализа, реализованных в расширенной версии *NEWSCAPE*, выходит за рамки данной публикации.

**Оценивание и уточнение множества источников информации** – определение рейтингов актуальности и популярности источников, уточнение и корректировка исследуемого подмножества источников. На этом этапе анализа возможно проведение нескольких итераций для получения более достоверных результатов с учетом требований заказчика.

**Экспертные оценки предварительных результатов анализа информационного пространства** – на этом этапе анализа дополнительно привлекаются эксперты в соответствующей предметной области и происходит построение результирующих аналитических отчетов с использованием развитого специализированного функционала *NEWSCAPE*.

Пример дополнительных показателей приведен на рис. 2.

### **Применение системы, основанной на информационной технологии анализа медиaprостранства**

Разработанная версия ИАС *NEWSCAPE* может быть использована для решения различных сложных и трудноформализуемых задач, среди которых в первую очередь можно выделить:

- **Информационный мониторинг публичных персон, организаций, процессов и событий** – комплексный мониторинг и анализ заданного множества объектов в медиaprостранстве с построением сложных взаимосвязей и характеристик.

- Многокритериальный анализ информационного пространства, где источниками информации служат интернет-пространство, печатные СМИ, телевидение и радио. При этом система обеспечивает поддержку коллективной работы администраторов и операторов данных.

Существенная роль в анализе медиaprостранства для отдельных категорий пользователей принадлежит специализированным подсистемам:

- **паспортам регионов** – информационно-аналитическая подсистема для реализации сбора, обработки и хранения экономической, социальной и политической информации для анализа в разрезе областей и районов Украины.

- **Региональным проблемам** – подсистема оперативного сбора и управления статусами информации, непосредственно связанной с региональными разнотипными проблемами в разрезе областей и районов Украины.

- **Распоряжениям руководителя** – на основе подсистемы информационного мониторинга реализован автоматизированный модуль отслеживания на областном и районном уровнях степени отображения в информационном пространстве результатов действий, осуществляемых в рамках комплексных программ, утвержденных руководителем высокого уровня.

Результаты работы ИАС *NEWSCAPE* доступны группам аналитиков и лицам, принимающим решения как в виде интерактивного оперативного доступа к системе, так и в виде

Регіональний розподіл інформації по об'єкту

|                        | 27.02.2017 - 05.03.2017 | 06.03.2017 - 12.03.2017 | 13.03.2017 - 19.03.2017 | Сума | %     |
|------------------------|-------------------------|-------------------------|-------------------------|------|-------|
| Загальноукраїнські ЗМІ | 120                     | 215                     | 164                     | 499  | 83.17 |
|                        | 234                     | 455                     | 330                     | 1019 | 86.21 |
| м. Київ                | 13                      | 7                       | 5                       | 25   | 4.17  |
|                        | 18                      | 17                      | 13                      | 48   | 4.06  |
| Харківська             | 8                       | 7                       | 9                       | 24   | 4     |
|                        | 14                      | 14                      | 17                      | 45   | 3.81  |
| Миколаївська           | 6                       | 4                       | 3                       | 13   | 2.17  |
|                        | 7                       | 5                       | 5                       | 17   | 1.44  |
| Інші                   | 9                       | 21                      | 9                       | 39   | 6.5   |
|                        | 13                      | 25                      | 15                      | 53   | 4.48  |

Розподіл інформації по ЗМІ

|                        | 27.02.2017 - 05.03.2017 | 06.03.2017 - 12.03.2017 | 13.03.2017 - 19.03.2017 | Сума | %     |
|------------------------|-------------------------|-------------------------|-------------------------|------|-------|
| ИТАР-ТАСС              | 0                       | 33                      | 37                      | 70   | 11.67 |
|                        | 0                       | 65                      | 53                      | 118  | 9.98  |
| Экономические Известия | 12                      | 17                      | 21                      | 50   | 8.33  |
|                        | 20                      | 38                      | 55                      | 113  | 9.56  |
| В "Взгляд.ру"          | 11                      | 15                      | 16                      | 42   | 7     |
|                        | 25                      | 38                      | 22                      | 85   | 7.19  |
| Інші                   | 133                     | 189                     | 116                     | 438  | 73    |
|                        | 241                     | 375                     | 250                     | 866  | 73.27 |

Звіт сформовано:  
22.03.2017 16:37

Регіональний розподіл інформації по об'єкту  
Основні результати за період  
Сторінка 2 з 2

Рис. 2

регулярних формалізованих інформаційних звітів. Додаткові аналітичні групи аналогічно можуть оперативно генерувати різні події та предметні звіти за запитом та забезпечувати колективну роботу групи аналітиків високого рівня в межах функціонування центрів прийняття рішень різного рівня.

**Висновок.** В наведеному статті об'єктивно викладено нову технологію аналізу об'єктів в медіапросторі, на основі якої формалізовано 10 основних етапів порівняльного аналізу об'єктів в інформаційному просторі та запропоновано 12 їх математички визначених характеристик. Детальне описання математички методів та обчислювальних алгоритмів виходить за межі даної публікації. Ці результати лежать в основі аналітичних засобів системи *NEWSCAPE*, а її практичне застосування підтвердило їх ефективність. і запропоновано 12 математички

визначених характеристик. Ці результати лежать в основі при створенні аналітичних засобів системи «*NEWSCAPE*», а її практичне застосування підтвердило їх ефективність.

З початку 2010 рр. в системі накоплено більше 30 мільйонів інформаційних блоків, більше 200 тисяч статей друкованих ЗМІ та тисячі новинних відеосюжетів. В системі також доступна структурована по областях та районах України економічна, соціальна та політична інформація. Використовуються та відкриті агреговані дані по юридичним об'єктам, підприємствам та фізичним особам [7].

В перспективі планується розвиток системи в наступних напрямках: агрегування зростаючої кількості даних з різних джерел інформації (думки експертів, телебачення, радіо, соціальні мережі, реклама), стандартизація та підключення додаткових баз даних та знань, удосконалення існуючого та розробка нової математичко-

го аппарата, аналитического инструментария и информационных технологий анализа больших данных. Это позволит использовать сложные математические модели и методы для повышения степени адекватности формируемых оценок и прогнозов, а также для формирования оперативных и качественных аналитических отчетов для лиц, принимающих ответственные решения.

1. *Boisa Max H.* Information space: a framework for learning in organizations, institutions and culture, – Охон: Routledge, 2016. – 568 p.
2. *Майер-Шенбергер В., Кукьер К.* Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим, 2014. – М.: Манн, Иванов и Фербер, 2013. – 240 с.
3. *Мальшко С.А.* Мониторинг и анализ медиапространства – Информационно-аналитическая система «NEWSCAPE» // УСиМ. – 2015. – № 2. – С. 88–93.

4. *Hulianytskyi L., Malyshko S.* Big Data in Information Analytical System «NEWSCAPE» / In Data Stream Mining & Processing // Proc. of the 2016 IEEE First Int. Conf. on Data Stream Mining & Processing (23–27 Aug. 2016, Lviv, Ukraine). – Lviv, 2016. – P. 382–386.
5. *Гуляницкий Л.Ф., Сергиенко И.В., Мальшко С.А.* О программных средствах поддержки принятия решений в задачах группового выбора // УСиМ. – 1993. – № 5. – С. 90–97.
6. *Ландэ Д.В., Снарский А.А., Безсуднов И.В.* Интернетика: Навигация в сложных сетях: модели и алгоритмы. – М.: Либроком, 2009. – 264 с.
7. *Електронний парламент України: досвід створення / Л.Б. Баран, В.В. Вишневський, К.Д. Гуляев та ін.* – К.: Юстон, 2015. – 452 с.

Поступила 02.11.2017  
Тел. для справок: +38 044 526-5585 (Київ)  
E-mail: smsmsm1365@gmail.com  
© С.А. Мальшко, 2017

UDC 004.4

S.A. Malyshko

PhD of Physical and Mathematical Sciences, Senior Science Researcher, Glushkov Institute of Cybernetics of NAS of Ukraine, Kyiv, Glushkov ave., 40, Kyiv, 03187, smsmsm1365@gmail.com

## Information Technology of Analysis of Media Space Based on the System «NEWSCAPE»

**Keywords:** NEWSCAPE, media space analysis, decision support system, information space, large data.

**Introduction.** The information space is understood as a media space, formed by a set of messages from the selected sources of data. Many media resources of individuals and legal entities, electronic versions of traditional media, online resources (websites, portals) and “new media” (social networks, blogs) are considered as the sources of the information messages.

**Purpose.** The paper describes the main features and the peculiarities of information technology of the media space analysis on the basis of the “NEWSCAPE” system, which form the basis for multifactor analysis of the objects using the information-analytical system.

**Methods.** The basic stages of the information space analysis are: the formation of a subset of the media space for analysis, the extraction of information from a subset, the information specification, the benchmarks calculation, the calculation of the relative indicators, the information evaluation, the information links construction, in-depth analysis of information, assessment and refinement of the information sources variety, analysis of the information space.

**Results.** The main stages of the comparative analysis of objects in the media space are formalized and the mathematically calculated characteristics are presented. These results form the basis for the creation of analytical tools for the NEWSCAPE information and analysis system, and its practical application has confirmed their effectiveness.

**Conclusion.** It is planned to develop the system in the following areas: aggregating more and more data from diverse sources of information (expert opinions, television, radio, social networks, advertising), standardizing and connecting additional databases and knowledge, improving the existing and developing a new mathematical apparatus, analytical tools and information technologies for analysis of large data. This will make it possible to use complex mathematical models and methods to increase the adequacy of the assessments and forecasts that are being generated, and also to create operational and qualitative analytical reports for decision-makers.

1. *Elenin G.G.* Nanotekhnologii, nanomaterialy, nanoustrojstva, <http://spkurdyumov.ru/mathmethods/nanotexnologii-nanomaterialy-nanoustrojstva/> (In Russian).
2. *Yakovlev Yu.S.* Novye paradigmy postroeniya komp'yuternyh sistem. Informacijni tekhnologii ta komp'yuterna inzheneriya, 2012, N 1, P. 51–61. (In Russian).
3. *Osnovnye koncepty komp'yuteroi budushchego*, [http://zoom.cnews.ru/rnd/article/item/kompyutery\\_budushchego\\_osnovnye\\_kontsepty](http://zoom.cnews.ru/rnd/article/item/kompyutery_budushchego_osnovnye_kontsepty) (In Russian).
4. *Zajcev V., Shishlova A.* Moletronika, <https://www.nkj.ru/archive/articles/5199/> (In Russian).
5. *DNK-komp'yutery*. Part 1., <http://www.geometria.by/blogs/technology/50548> (In Russian).
6. *DNK-komp'yutery*, [http://www.odessapassage.com/passage/magazine\\_details.aspx?lang=eng&id=33256](http://www.odessapassage.com/passage/magazine_details.aspx?lang=eng&id=33256)
7. *Rybak Ehlya.* Geneticheskie i kletochnye biokomp'yutery, <http://www.computer-museum.ru/technlg/genecomp.htm> (In Russian).