

17. Амзин А. Мой сервер сильнее твоего // Лента.ру. – 21.01.2009. URL: <http://lenta.ru/articles/2009/01/21/cloud/> (дата обращения 8.12.2010).
18. Материал из Википедии – свободной энциклопедии. (Последнее изменение страницы: 22:01, 16 декабря 2010). URL: [http://ru.wikipedia.org/wiki/Облачное\\_хранилище\\_данных](http://ru.wikipedia.org/wiki/Облачное_хранилище_данных) (дата обращения 17.12.2010).
19. Apple iPad и другие планшеты могут привести к буму облачных сервисов // Computerworld.com. – 13.04.2010 URL: <http://www.astera.ru/news/?id=76669> (дата обращения 17.12.2010).
20. «Облачную» роль Google в России сыграют операторы. URL: <http://www.softpower-linux.org/blog/news/182.html> (дата обращения 17.12.2010).
21. Кудрявцев Ю. Будущее BI в облаках? – 30.07.2008 URL: <http://www.citcity.ru/19090/> (дата обращения 17.12.2010).
22. Коваленко О.С. Обзор проблем и состояний облачных вычислений [Электронный ресурс] // Информатика, вычислительная техника и инженерное образование. – 2011. – № 1 (3). – <http://digital-mag.ti.sfedu.ru/index.htm>.
23. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. – М.: Физматлит, 2010.
24. Гладков Л.А., Курейчик В.В., Курейчик В.М. Сороколетов П.В. Биоинспирированные методы в оптимизации. – М.: Физматлит, 2009.
25. Курейчик В.М. Биоинспирированный поиск с использованием сценарного подхода // Известия ЮФУ. Технические науки. – 2010. – № 7 (108). – С. 7-12.

Статью рекомендовал к опубликованию д.т.н., профессор Я.Е. Ромм.

**Коваленко Олеся Сергеевна** – Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге; e-mail: [kovalenko.olesja@gmail.com](mailto:kovalenko.olesja@gmail.com); 347928, г. Таганрог, пер. Некрасовский, 44, ГСП 17А; тел.: 88634360793; кафедра систем автоматизированного проектирования; аспирант.

**Курейчик Виктор Михайлович** – Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге; e-mail: [kur@tsure.ru](mailto:kur@tsure.ru); 347928, г. Таганрог, пер. Некрасовский, 44, ГСП 17А; тел.: 88634393260; зам. руководителя по научной и инновационной деятельности; профессор.

**Kovalenko Olesya Sergeevna** – Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”; e-mail: [kovalenko.olesja@gmail.com](mailto:kovalenko.olesja@gmail.com); GSP 17A, 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634360793; the department of computer aided design; postgraduate student.

**Kureichik Victor Michylovich** – Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”; e-mail: [kur@tsure.ru](mailto:kur@tsure.ru); GSP 17A, 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634393260; the deputy the head on scientific work and Innovations; professor.

УДК 004.912

**Р.Ю. Вишняков, Ю.М. Вишняков**

### **ОБ ОДНОЙ МОДЕЛИ СЕМАНТИЧЕСКОЙ КЛАССИФИКАЦИИ МЕТОДОВ ИНФОРМАЦИОННОГО ПОИСКА**

*Предлагается классификация методов информационного поиска на основе их возможностей интерпретировать семантику текстов. Для этого вводится понятия функционала смысловыразительности, морфологического подобия и рассматривается их представления в различных видах поиска. Показывается, что наиболее распространенные методы информационного поиска, моделирующие семантику на основе частотных характеристик слов, обладают худшими точностными характеристиками в сравнении с дескрипторными методами поиска. В тоже время дескрипторные методы поиска имеют ограни-*

ченную сферу применения и бедные языковые средства. Поэтому для повышения точностных характеристик информационного поиска предлагается использовать специализированные лингвистические процессоры для моделирования и интерпретирования семантики предложений естественного языка.

*Методы классификации; семантическая интерпретация; смысловыразительность; морфологическое подобие; лингвистический процессор; обработка текстов на естественном языке; информационный поиск; автоматическая обработка текстов.*

**R.Yu. Vishnyakov, Yu.M. Vishnyakov**

### **MODEL OF THE SEMANTIC CLASSIFICATION OF INFORMATION RETRIEVAL METHODS**

*This article proposes a classification methods of information retrieval on the basis of their ability to interpret the semantics of texts. We introduce the concept of expression-meaning functionality, morphological similarity, and their representation in different methods of searches. We show that in comparison with the descriptor search methods most common approaches of semantic simulation of information retrieval based on the frequency characteristics of words occurrence, have low accuracy. However, the descriptor search methods are limited of scope and have poor language means. Therefore, to increase the accuracy of information retrieval we propose to use the specialized linguistic-processors to simulate and interpret the semantics of natural language sentences.*

*Classification methods; semantic interpretation; expression-meaning; morphological similarity; linguistic-processor; natural language processing; information retrieval; text mining.*

**Проблема.** Повышение точности информационного поиска является актуальной проблемой, а ее решение в настоящее время связывается с семантической интерпретацией текстов документов и поискового запроса. В используемых сегодня подходах по организации информационного поиска релевантность определяется на основе частоты встречаемости слов в документах, которую в грубом приближении можно считать выражением семантики текстов. Поэтому можно утверждать, что в этих видах поиска семантическая составляющая слов, не говоря уже о более сложных грамматических конструкциях, фактически не учитывается. Рассматривая естественный язык (ЕЯ) как таковой, следует отметить, что ЕЯ является коммуникационной средой, которая моделирует человеческое сознание и представляет собой сложную, но достаточно компактную знаковую систему. Грамматические конструкции в ЕЯ и их семантические значения находятся в неоднозначных отношениях, которые часто называют полисемичными. Однако именно эта полисемия делает язык открытым и способным к развитию [1]. Если бы ЕЯ был устроен так, что каждому новому смыслу давал новое название, он разрастался бы невероятно. Так, лексический состав ЕЯ представляет собой конечное и достаточно ограниченное множество слов, каждое из которых имеет множество семантических значений. При семантической интерпретации предложения ЕЯ происходит сужение семантических значений слов и формирование на их основе синтетического семантического значения всего предложения в целом. Исходя из этого обстоятельства, в настоящей работе предлагается понятие функционала смысловыразительности и на его основе классифицируются различные виды информационного поиска по предельным возможностям с точки зрения семантической интерпретации поисковых запросов и текстов документов.

**Метод решения.** Пусть  $L$  – некоторый язык и в нем существует последовательность  $\alpha$  слов  $x_1, x_2, \dots, x_n$ , представляющая собой некоторое осмысленное выражение (словосочетание, предложение, абзац, статью, книгу и т.д.), т.е.  $\alpha = x_1, x_2, \dots, x_n$ . Пусть  $S(x_i)$  – множество смыслов (семантических значений) слова  $x_i$ , тогда смысл (семантическое значение) выражения  $\alpha$  определим в виде функционала следующего вида:

$$S(\alpha) = \Phi(S(x_1), S(x_2), \dots, S(x_n)), \quad (1)$$

который будем называть функционалом смысловыразительности цепочки  $\alpha$ . Данный функционал вычленяет из множества смысловых значений слов определенные семантические значения и на их множестве строит смысловое значение всего выражения  $\alpha$ . Характерная особенность функционала  $S(\alpha)$  состоит в том, что он формирует новый смысл, как правило, не разлагаемый на смыслы отдельных слов и не сводящийся ни к одному из этих смыслов. Скорее всего, это определенным образом организованная совокупность узких смыслов слов, которая позволяет выявить (проявить) в человеческом сознании новый смысл  $S(\alpha)$ .

Введем понятие морфологического подобия слов  $\alpha$  и  $\beta$ , под которым будем понимать функцию  $p(\alpha, \beta)$ , значения которой определены на интервале  $[0, 1]$ . Если  $p(\alpha, \beta) = 1$ , то слова равны с точностью до буквы. В случае, когда  $0 < p(\alpha, \beta) < 1$ , слова подобны по каким-либо признакам. Например, имеют одну и ту же основу, или являются однокоренными, или и т.д. имеют какие-либо общие морфологические признаки. В том случае, когда  $p(\alpha, \beta) = 0$ , общие морфологические признаки отсутствуют.

Теперь рассмотрим формирование функционала смысловыразительности в различных видах информационного поиска, для чего определим следующие понятия. Пусть  $\beta$  поисковый запрос, а  $D = \{d_1, d_2, d_3, \dots, d_n\}$  – документное пространство, на котором реализуется поиск. Рассмотрим следующие ситуации.

Случай 1. Пусть поиск осуществляется только на основе морфологического подобия слов поискового запроса и документа без обращения к семантической составляющей слов. Поисковые запросы составляются в виде отдельных слов или их логических комбинаций, а сама поисковая процедура состоит в отыскании в текстах документов вхождений слов на условиях морфологического подобия из поискового запроса. Очевидно, что, задавая характер функции морфологического подобия, можно реализовать различные виды поиска. Также очевидно, что в класс поисков на основе морфологического подобия включаются модели булева поиска, дублинского ядра, векторной модели, вероятностные модели и т.д., кроме дескрипторной модели поиска.

При морфологическом подобии семантическая составляющая запроса не учитывается, она предполагается равной множеству смысловых значений входящих в поисковый запрос слов. Если поисковый запрос образует слово  $x$ , то в выдачу включаются все документы, содержащие морфологически подобные слова по всему множеству смысловых значений слова  $x$ , т.е.  $\Phi(x) = S(x)$ . Если в запрос входят слова  $x$ ,  $y$  и  $z$ , то по всем документам осуществляется поиск объединенного множества смысловых значений всех слов. Исходя из этого, функционал смысловыразительности для поисков на основе морфологического подобия определяется следующим выражением:

$$\Phi(x, y, z) = S(x) \cup S(y) \cup S(z). \quad (2)$$

Попытка понизить шумовую составляющую в результатах поиска приводит к введению внешних ограничений на выдачу, которые как бы моделируют семантику запроса. Так, выдвигается следующая гипотеза о релевантности документа запросу. Из двух документов более релевантен тот, у которого слово запроса встречается большее число раз (чаще). К сожалению, уязвимость такой гипотезы более чем очевидна, что наблюдается при поиске в Интернет, когда в выдачу включаются совершенно разномысловые документы, но содержащие примерно одинаковое количество поисковых слов.

Случай 2. Пусть информационный поиск осуществляется на основе принципа координации [2]. Координация предполагает, что предварительно все понятия проклассифицированы и организованы в упорядоченные иерархии по каким-либо признакам, а также каждой классификации приписано свое уникальное имя (дескриптор).

К сожалению, при дескрипторном поиске грамматическая составляющая запроса играет второстепенную роль и представляется практически только перечислением слов из лексического словаря, включающего имена (идентификаторы) классов классификации. Лексика языка поисковых запросов представляется множеством имен (ключей) понятий, а предложение поискового запроса составляется из одного или нескольких таких имен. При этом отметим, что поисковый запрос представляет собой пересечение значений всех дескрипторов, то есть в дескрипторах выделяется общая часть. Таким образом, если в поисковый запрос входят слова  $x$ ,  $y$  и  $z$ , то функционал смысловыразительности такого запроса определяется выражением вида:

$$\Phi(x, y, z) = S(x) \cap S(y) \cap S(z). \quad (3)$$

Например, данное обстоятельство можно проиллюстрировать гипотетической информационно-справочной системой железнодорожного сообщения, в которой на запрос «Москва, Владивосток, Скорый» выводятся расписания всех скорых поездов, отправляющихся из Москвы до Владивостока.

Таким образом, формирование функционала смысловыразительности в поисках на основе морфологического подобия идет путем простого объединения смысловых значений слов (термов) в запросах, что и определяет их невысокую точность. В дескрипторных видах поисков функционал смысловыразительности формируется пересечением смыслов отдельных дескрипторов и поэтому его точность высокая. Но дескрипторные виды поиска узко специализированы и требуют хорошо организованного документного пространства. Также следует заметить, что редактирование такого документного пространства представляет определенные трудности.

Рассмотренные случаи семантической классификации информационных поисков показывают, что моделирование семантики частотными свойствами слов в текстах не удовлетворяет по точности запросы потребителей. Координатный принцип поиска обеспечивает высокую точность, но его языковые средства очень бедны и далеки от ЕЯ. Поэтому оптимальные решения в информационном поиске возможны путем введения специализированных лингвистических процессоров, которые могли бы моделировать и интерпретировать семантику предложений ЕЯ [3–5].

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Налимов В.В.* Вероятностная модель языка. О соотношении естественных и искусственных языков. – М.: Наука, 1979. – 303 с.
2. *Черный А.И.* Введение в теорию информационного поиска. Монография. – М.: Наука, 1975. – 239 с.
3. *Вишняков Ю.М., Вишняков Р.Ю.* Проблемы семантического информационного поиска // Труды международных научно-технических конференций «Интеллектуальные системы» (AIS'06) и «Интеллектуальные САПР» (CAD-2006). Научное издание в 3-х томах. Т. 2. – М.: Физматлит, 2006. – С. 308-314.
4. *Вишняков Р.Ю.* Об одном подходе к интеллектуализации информационно-поисковых систем // Известия ТРТУ. – 2007. – № 1 (73) – С. 170-173.
5. *Вишняков Р.Ю.* Смысловыразительность и проблемы семантического информационного поиска // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'08) и «Интеллектуальные САПР» (CAD-2008). Научное издание в 3-х томах. – М.: Физматлит, 2008.

Статью рекомендовал к опубликованию д.т.н., профессор В.П. Карелин.

**Вишняков Ренат Юрьевич** – Технологический институт Южного федерального университета в г. Таганроге; e-mail: rvishn.sfu.edu@gmail.com; 347928, г. Таганрог, пер. Некрасовский, 44, ГСП 17А; тел.: +78634314485; кафедра системного анализа и телекоммуникаций; доцент.

**Вишняков Юрий Муссович** – e-mail: vishn@tsure.ru; факультет автоматики и вычислительной техники; декан.

**Vishnyakov Renat Yur'evich** – Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”; e-mail: rvishn.sfu.edu@gmail.com; 17A, 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634314485; the department of system analysis and telecommunication; associate professor.

**Vishnyakov Yuriy Mussovich** – e-mail: vishn@tsure.ru; the college of automation and computer engineering; dean.

УДК 681.3.06: 681.323 (519.6)

**Я.Е. Ромм, А.И. Тренкеншу**

**ИДЕНТИФИКАЦИЯ ФИГУР ГРАФИЧЕСКОГО АНАЛИЗА И ВЫДЕЛЕНИЕ  
ОСНОВНЫХ ТРЕНДОВ ФИНАНСОВЫХ РЫНКОВ С ПРИМЕНЕНИЕМ  
СХЕМ СОРТИРОВКИ**

*Изложен метод компьютерной идентификации фигур графического анализа финансовых рынков с помощью экстремальных признаков на основе алгоритмов сортировки. Представлен алгоритм выделения и идентификации фигур графического анализа на различных финансовых рынках, основанный на локализации экстремумов цен закрытия. Охарактеризована процедура, позволяющая выделять основной тренд на различных финансовых рынках. Приведены примеры работы процедуры по выделению тренда на финансовых рынках. Поставлен вопрос о сходимости итерационного процесса, позволяющего выделять основные тенденции на финансовых рынках.*

*Фигуры технического анализа; алгоритм идентификации; сортировка; финансовые рынки; распознавание; выделение тренда; локализация экстремумов.*

**Ya.E. Romm, A.I. Trenkenshu**

**IDENTIFICATION OF TECHNICAL PATTERNS AND SELECTION OF THE  
MAIN TRENDS OF FINANCIAL MARKETS USING SORTING SCHEMES**

*The method of software-based identification of technical patterns of financial markets with help of extremum characteristics based on sorting algorithms is recounted. The algorithm of selection and recognition of technical patterns on different financial markets based on daily returns extremum localization is presented. The procedure, permitting to select the main trends on different financial markets, is described. Examples of the procedure work of selection trends on financial markets is given. The question about iterative process convergence, permitting to select the main veins on financial markets, is posed.*

*Technical patterns; identification algorithm; sorting; financial markets; recognition; trend selection; extremums localization.*

**Постановка задачи.** Выявление различных графических моделей и трендов на графиках инструментов финансовых рынков (валютных пар, акций, фьючерсов и т.д.) является неотъемлемой частью графического анализа, позволяющего иногда достаточно точно предсказать дальнейшее движение цены. В [1] было показано, что применение метода локализации экстремумов на основе сортировки с сохранением обратной индексации [3, 4] к оцифрованным данным значений цен закрытия валютных пар рынка FOREX позволяет выделять на ценовых графиках основные модели графического анализа при вариации значений радиусов окрестностей для вычисления экстремумов.