

# ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

---

---

УДК 681.3.01:681.327.1

Е.В. Бритвина<sup>1</sup>, В.В. Крылов<sup>1</sup>, Ю.А. Мальков<sup>2</sup>

## АЛГОРИТМ МАКСИМИЗАЦИИ РЕЛЕВАНТНОСТИ, ИСПОЛЬЗУЮЩИЙ ГРАФОВЫЕ МОДЕЛИ ДАННЫХ

Нижегородский государственный технический университет им. Р.Е. Алексеева<sup>1</sup>,  
Институт прикладной физики РАН<sup>2</sup>

Предлагается алгоритм поиска максимально релевантного элемента из конечного множества ответов по заданному конечному множеству запросов и заданной функции релевантности, имеющий существенно низкую вычислительную сложность, чем полный перебор. Для этого используется предварительное формирование на множестве ответов координатной системы, индуцированной функцией релевантности, и после на этой основе графа метризованного тесного мира. Проведенное численное моделирование подтвердило предложенный эвристический алгоритм. Целевым приложением метода является таргетирование рекламных сообщений в телекоммуникационных системах с установлением соединения.

*Ключевые слова:* релевантность, поиск информации, дискретная оптимизация графовая модель данных, графы тесного мира.

### Введение

Термин “релевантность” имеет много определений и в общем смысле наиболее близок по значению к термину “адекватность” [1]. Для информационных систем релевантность обычно связывается с семантическим соответствием поисковых запросов и найденных документов. В задачах управления релевантность характеризует меру соответствия управляющих сигналов целям управления. [2]. Проблема максимизации релевантности возникает в значительном числе задач проектирования компьютерных систем обработки информации.

В настоящей работе мы рассматриваем задачу максимизации релевантности с позиций дискретной оптимизации [3] и полагаем, что все семантические аспекты отражены в способе вычисления функционала, значение которого и отождествляется с оценкой релевантности. При такой постановке задачи ее решение становится пригодным для использования при проектировании широкого круга систем, где требуется найти наилучшее реагирование на каждый из входных наборов данных без учета их темпоральной компоненты.

Основной целевой задачей, которая привела к разработке изложенного далее алгоритма, явилась задача гладкого встраивания в процесс установления соединения в мультимедийных телекоммуникационных системах дополнительных информационных сообщений [4], специфицированных для каждого из вызывающих абонентов.

Процедура выбора сообщений в соответствии с профилем каждого из абонентов рассматривается как задача выбора релевантного мультимедиа-сообщения из доступной базы данных таких сообщений. Наличие значительного числа обслуживаемых абонентов и малый интервал времени, доступный для встраивания сообщения, требуют чрезвычайно производительных алгоритмов максимизации релевантности с неограниченными возможностями масштабирования. Эти требования существенно ограничивают возможность применения известных алгоритмов и вызвали необходимость новой разработки.

### Постановка задачи

Пусть задано два конечных множества:

- $X = \{x_i\}_{i=1}^N$ , имеющее семантический смысл запросов некоторой предметной области;
- $Y = \{y_j\}_{j=1}^M$ , имеющее семантический смысл ответов на запросы в этой же предметной области.

Определено отношение  $X \circledast Y$  называемое отношением релевантности, такое, что каждому элементу из множества запросов соответствует подмножество так называемых релевантных ответов из множества

$$\forall x_i \in X \Rightarrow Y_i \subseteq Y.$$

Отношение может быть задано некоторой матрицей из  $N$  строк, соответствующих всем возможным запросам, и  $M$  столбцов, соответствующих всем возможным ответам.

$$\begin{matrix} r_{11} & r_{12} & \dots & r_{1M} \\ r_{21} & r_{22} & \dots & r_{2M} \\ r_{31} & r_{32} & \dots & r_{3M}. \end{matrix}$$

Если значения элементов матрицы релевантности равны либо 0 либо 1, то речь идет о том, что задано булева релевантность, то есть для любой пары запрос-ответ о их релевантности можно утверждать только «да» или «нет».

Более сложный случай определяется возможностью оценивать уровень релевантности. В этом случае можно ввести функцию релевантности  $rel$ :

$$X \times Y \rightarrow [0,1]: 0 \leq rel(x_i, y_j) \leq 1.$$

Эта функция аналогична используемой обычно функции “похожести” – similarity, используемой в тех случаях, когда запрос и ответ рассматриваются как элементы одного и того же множества. В таких случаях можно считать это множество наделенным структурой метрического пространства или близкого к такому. Для решения задач поиска в метрических пространствах создан целый ряд мощных инструментов, основанных на различных математических свойствах таких пространств. [5]. Целью данной статьи является решение задачи поиска для случая, когда множества запросов и ответов затруднительно представить элементами одного и того же метрического пространства.

Однако введение матрицы релевантности делает постановку задачи весьма близкой, поскольку позволяет абстрагироваться от природы строк и столбцов этой матрицы и рассматривать ее как матрицу некоторой метрической природы.

Заметим сразу, что для этой матрицы не выполняются многие свойства, характерные для метрических матриц (или матриц похожести), например матрица неквадратная, диагональные элементы необязательно равны нулю (единице, как для матриц похожести).

Тем не менее, решение задачи поиска может быть сведено к отысканию для каждой строки максимальных элементов, и тем самым идентификации номеров тех столбцов, соответствующие которым ответы наилучшим образом соответствуют запросу, относящемуся к данной строке.

Здесь может быть предложен очевидный алгоритм сортировки элементов для каждой строки с сохранением идентификатора столбца. Такой алгоритм может быть применен, если мощность множеств  $N$  и  $M$  не очень велика. Однако на практике, число возможных запросов и ответов оказывается весьма значительным.

Известно, что вычислительная сложность устойчивой сортировки массива из  $M$  чисел оценивается как  $O(M^2)$  для большинства алгоритмов или в лучшем случае как  $O(M \log M)$  [6].

Для каждого запроса потребуется исполнить алгоритм с сложностью, равной произведению сложности алгоритма сортировки на сложность алгоритма вычисления значения функции релевантности для всех вариантов ответа. Если обозначить сложность вычисления

значения функции релевантности  $\mu$ , то оценка общей сложности может быть дана следующим образом:

$$C_i = \mu \cdot M \cdot O(M \log M) = \mu O(M^2 \log M).$$

Дополнительной сложностью для реализации практически интересного алгоритма является необходимость поддержки динамического пополнения множеств ответов и запросов. В этом случае результаты предыдущей сортировки не смогут быть использованы для последующих идентичных запросов, что не позволит кэшировать результаты.

В этом случае требуется построение другого эффективного алгоритма. Для снижения вычислительной сложности можно также ограничиться нахождением только одного элемента с максимальным значением функции релевантности. В этом случае нет необходимости упорядочения всех возможных ответов, а искать только один. В этом случае задача переходит в категорию задач дискретной оптимизации. Сведем задачу максимизации релевантности к задаче дискретного программирования. Иногда вместо задачи на максимум выгоднее рассматривать задачу минимизации. С этой целью будем рассматривать вместо функции релевантности связанную с ней функцию расстояния релевантности:

$$d: X \rightarrow R; d(x_i, y_j) = -\ln(\text{rel}(x_i, y_j))$$

$$d(x, y): X \times Y \rightarrow [0, \infty)$$

Задача: найти алгоритм отыскания максимально релевантной точки

$$y_* \in Y; d(x_*, y_*) = \min_{y \in Y} d(x_*, y)$$

В изложенной постановке задача относится к классу трудноразрешимых задач дискретной оптимизации [3]. Точные методы для высокоразмерных задач в этом случае приводят к недопустимо большому числу вычислений, поэтому далее обсуждаются эвристические приближенные алгоритмы.

### Решение задачи релевантности методом локальной дискретной оптимизации со случайным поиском на графах

Основной идеей этого раздела является попытка свести задачу на поиск релевантного ответа из  $Y$  на запрос из  $X$  при известной функции релевантности  $\text{rel}(\cdot, \cdot)$  к нескольким задачам локального поиска на множестве  $Y$ . В основу метода положим оснащение этого множества графом со специальными свойствами, известными как MSW (MetriZedSmallWorld) для поиска локального минимума [7].

Классический подход применения метода локальной оптимизации сводится к наделению множества  $Y$  любой произвольной метрикой, назначением начальной точки и радиуса окрестности локального поиска. Алгоритм поиска минимума в этом случае сводится к поиску локального минимума в выбранной окрестности и перехода к новой начальной точке, в качестве которой назначается найденная на предыдущем локальном шаге точка минимума.

Как показано в [8,9], если на дискретном множестве существует функция расстояния, то на нем может быть построен граф, обладающий свойствами тесного мира (MSW граф). То есть расстояние между любыми вершинами такого графа в среднем оценивается логарифмической величиной от размера графа, числа его вершин. При этом существует простой алгоритм жадного поиска, приводящий в среднем за логарифмическое число шагов к отысканию ближайших соседей для любого вновь выбранного элемента  $y_0$  из множества (поглощающего)

$$\bar{Y} \supset Y,$$

наделенного той же функцией расстояния. Выбрав произвольно начальную точку из  $Y$ , можно найти в  $Y$  ближайших соседей для  $y_0$ , совершив логарифмически связанное с размером  $Y$  число локальных поисков.

Сведем задачу поиска элемента из множества ответов  $Y$  релевантного заданному запросу из  $X$  к задаче поиска на MSW графе.

Пусть заданы элемент и MSW граф  $x_* \in X$  на  $Y$ . Задана также вычислимая функция расстояния релевантности между  $X$  и  $Y$ :

$$d(x, y): X \times Y \rightarrow [0, \infty)$$

*Задача:* найти алгоритм отыскания максимально релевантной точки

$$y_* \in Y; d(x_*, y_*) = \min_{y \in Y} d(x_*, y).$$

Основная идея алгоритма может быть представлена следующим образом. Сначала назначим случайным образом точку  $y_0$  в  $Y$  – одну из вершин графа MSW, построенного на  $Y$ .

Найдем расстояние релевантности между этой точкой и  $x_*$ .

$$d(x_*, y_0) = d_0$$

Рассмотрим все вершины графа, являющиеся соседями  $y_0$ , и найдем расстояния релевантности до них. Выберем ту вершину, расстояние релевантности до которой от  $x_*$  минимально. Сравним это расстояние с значением  $d_0$ : если найденное расстояние меньше, то выбираем новую соответствующую ему вершину за новую начальную точку поиска и повторяем поиск максимально релевантной точки среди ее соседей, если нет, то выбираем точку  $y_0$  в качестве решения задачи – возвращаем как максимально релевантную запросу  $x_*$  в данном цикле поиска. После этого выбирается другая случайная точка  $y_0$  в  $Y$ , и процесс поиска повторяется. Число случайных поисков может быть выбрано заранее или определяться адаптивно в зависимости от разброса найденных. Среди всех найденных точек локального минимума выбираются те, которые имели минимальное расстояние релевантности, и это подмножество объявляется решением. В большинстве случаев такое подмножество состоит из единственного элемента и в этом случае найденное решение строго соответствует поставленной задаче.

Запишем алгоритм формально в виде программного псевдокода:

Greedy\_Search(q: object, y<sub>enter\_point</sub>: object)

```

1  ycurr ← yenter_point;
2  dmin ← d(x, ycurr); vnext ← NIL;
3  foreach yfriend ∈ ycurr.getFriends() do
4  if dfr ← d(x, yfriend) < dmin then
5    dmin ← dfr;
6    ynext ← yfriend;
7  if ynext = NIL then return ycurr;
8  else return Greedy_Search(x, ynext);
```

В приведенном подходе остается неизвестным и принципиально важным вопрос о построении MSW-графа на множестве  $Y$ . Очевидно, что поскольку на этом множестве не задана функция метрики, то мы можем назначить ее произвольно. Но также очевидно, что от ее выбора будет сильно зависеть эффективность работы алгоритма. Мы предлагаем здесь выбрать в качестве функции метрики некоторую функцию внутреннего расстояния на  $Y$ , индуцированного расстоянием релевантности.

### Построение функции внутреннего расстояния, индуцированного расстоянием релевантности

Основная идея подхода состоит в том, что в множестве  $X$  некоторым образом выбирается подмножество из  $M$  “представительных” элементов, далее называемый базисом в  $X$ . Обозначим это подмножество

$$X_B = \{x_{B1}, x_{B2}, \dots, x_{BM}\}$$

определим для каждой точки из  $Y$  расстояния релевантности

$$y_k \in Y, d(y_k, x_{Bi}) = d_{ki}, \forall i = 1, M$$

Назовем полученный для каждого  $k$  набор чисел координатами элемента  $y_i$  в системе координат  $X_B$ . Таким образом, на множестве  $Y$  индуцируется координатная сетка размерности  $M$ . Теперь определим функцию внутреннего расстояния на  $Y$  с помощью этой координатной системы, задавая ее значения следующим образом

$$d_y(y_i, y_j) = \sqrt{\sum_{k=1}^M (d_{ik} - d_{jk})^2}$$

Такая функция удовлетворяет всем основным аксиомам расстояния, поскольку дает нулевое значение для совпадающих точек в  $Y$  и положительна для несовпадающих. Мы не будем давать здесь анализ выполнения неравенства треугольника для индуцированной функции расстояния, поскольку это не является существенным для дальнейшего, однако заметим, что при правильном выборе множества  $X_B$  введенная функция расстояния удовлетворит всем аксиомам метрики.

Именно введенная таким образом индуцированная исходным отношением релевантности функция расстояния будет использоваться для построения необходимого для эффективного поиска решений графа MSW. На рис. 1 приведена иллюстрация построения координатного пространства и графа на нем.

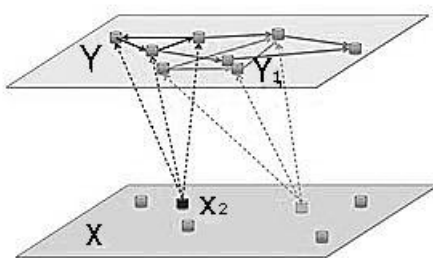


Рис. 1. Построение координатного пространства на  $Y$

Обсудим теперь выбор множества  $X_B$ . Очевидно, что с вычислительной точки зрения следует выбирать базис в  $X$  минимально возможной мощности. Однако, если число элементов базиса мало, то индуцированная им координатная сетка на  $Y$  будет плохо различать элементы, в том смысле, что координаты всех их окажутся весьма близкими и порождаемая топология на  $Y$  окажется весьма слабой для построения качественных решений с помощью графа MSW. Число соседей в графе окажется значительным, что приведет к катастрофическому росту вычислительных затрат из-за трудностей нахождения глобального экстремума. В силу весьма слабых предположений о структуре множеств  $X$  и  $Y$ , а также матрицы релевантности, нахождение регулярных методов выбора базиса представляется маловероятным.

Предлагаемый подход основывается на теоретико-системном предположении, что оптимальный базис должен обеспечивать достаточный уровень “разнообразия” координат элементов  $Y$ , который может быть измерен с помощью значения энтропии системы точек, соответствующих этим элементам в индуцированном координатном пространстве. Выбор базиса тогда может осуществляться итеративным образом, начиная с случайного выбора  $X_B$  при достаточно малом размере  $M$ , вычисления координат для всех элементов  $Y$  и оценки энтропии полученной системы точек. Далее базис  $X_B$  расширяется, увеличивается значение  $M$ , вычисляется энтропия системы точек в новом базисе и сравнивается с предыдущим значением. Правило остановки может быть сформулировано как достижение точки с малым приращением энтропии, подобно тому, как это делают в задачах на адаптацию по минимуму оценки

ошибки. Вычисление энтропии для множества  $Y$  с индуцированным внутренним расстоянием производится по формуле

$$H_B(Y) = - \sum p_i \log p_i$$

где вероятности оцениваются относительным числом точек из  $Y$ , попадающих в общую окрестность для каждой из них. Для этого вычисления используется любая из известных программ расчета энтропии системы  $M$ -мерных векторов, например [10].

Приведенная процедура выбора базиса  $X_B$  является весьма громоздкой, хотя и позволяет найти достаточно хорошую систему координат в  $Y$ . На практике оказывается возможным получать неплохие результаты при просто случайном выборе  $M$  элементов из  $X$ , где  $M$  следует выбирать большим или равным внутренней размерности  $X$ . Далее проиллюстрируем методику выбора базиса и результаты построения MSW-графа на  $Y$  на примере.

### Пример построения системы релевантного выбора на основе предложенного алгоритма

В качестве задачи релевантного выбора рассмотрим подбор релевантного аудиофайла по параметрам исходящего вызова для использования в системе подобной предложенной авторами в патенте [4]. В упрощенной постановке рассмотрим множество  $X$  как набор двоичных векторов длиной 11, каждый бит которого определяет наличие в исходящем звонке одного из 11 признаков – так называемый фингерпринт вызова.

Таким образом, множество запросов содержит потенциально  $2^{11}$  различающихся элементов. Для обозначения каждого из них можно использовать их нумерацию, т.е. идентифицировать целым числом от 1 до 2048. Пусть для воспроизведения имеется 1024 аудиофайлов образующих множество  $Y$ , каждый из которых мы будем обозначать просто целым числом от 1 до 1024 или представлять 10-мерным двоичным вектором. Для нахождения значения функции расстояния релевантности между элементами множеств  $X$  и  $Y$  будем использовать тестовую функцию

$$d(x_k, y_n) = |k^2 - 2kn + n^2 - 146k - 146n + 5328|$$

Отметим, что для такой функции существует набор целых значений  $k$  и  $n$ , для которых расстояние обращается в нуль, и этот факт будет использован при тестировании алгоритма.

$$\begin{aligned} k &= 73t^2 + 145t + 72 \\ n &= 73t^2 - t; t = 0, 1, 2, \dots \end{aligned}$$

На рис. 2 показан типичный фрагмент графика такой функции расстояния в непрерывных координатах.

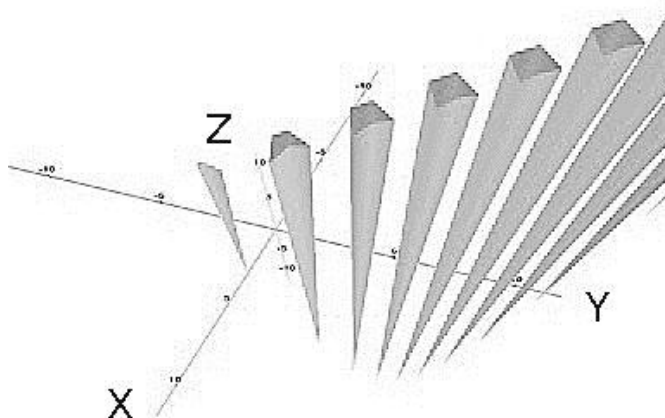


Рис. 2. График функции расстояния релевантности в непрерывных координатах  
Выберем из множества  $X$  базис

$$x_{Bi} = 2^i; i = 0, 1, 2$$

Найдем координаты точек из  $Y$  по этому трехмерному базису

$$y_n^i = |2^{2i} - 2^{i+1}n + n^2 - 146 \cdot 2^i - 146n + 5328| \quad i = 1, \dots, 12; n = 1, \dots, 1024$$

Например, для первой получаем

$$y_1 = (5036, 4891, 4607).$$

Располагая координатами точек в  $Y$  можно построить MSW-граф. Фрагмент этого графа изображен на рис. 3.

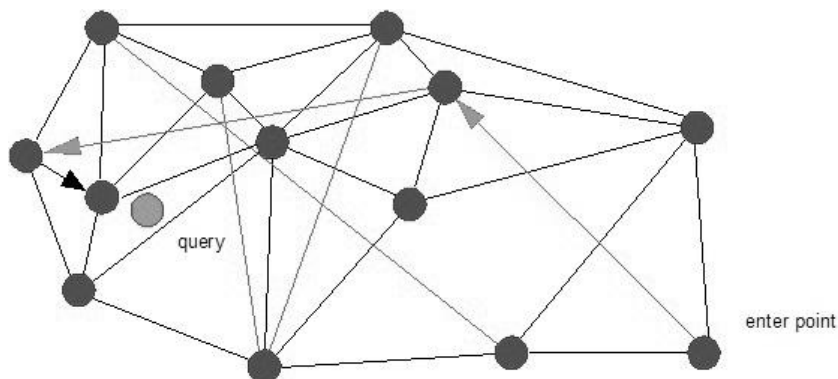


Рис. 3. Фрагмент MSW графа для примера

С помощью программы Metrized Small World Prototype (<http://www.meralabs.com/en/downloads/>) было произведено нахождение максимально релевантных точек в  $Y$  для нескольких заданных запросов из  $X$ . Напомним, что для такого поиска потребовалось количество вычислений функции расстояния существенно меньше размера  $Y$ . При этом часть запросов была взята с заведомо единичной релевантностью. В табл. 1 приведены результаты экспериментов.

Таблица 1

Нахождение максимально релевантных точек в  $Y$   
для нескольких заданных запросов из  $X$

#X	10	122	290	402	654	1164	1820
#Y1	137	6	654	818	290	654	1022
#Y2	138	7	653	817	291	655	1022
#Y3	136	5	655	819	289	653	1021
#Y4	139	8	652	822	292	656	1020
$d(x, y1)$	5	96	0	264	0	0	227200
$d(x, y2)$	104	281	581	421	873	1165	227200
$d(x, y3)$	112	475	583	951	875	1167	228943
$d(x, y4)$	215	656	1160	3024	1744	2328	230688

Как следует из анализа табл. 1, алгоритм находит четыре кандидата на роль максимально релевантного ответа, среди которых всегда находится точное значение, обеспечивающее нулевое значение функции расстояния релевантности, если такой ответ существует. Как видно, при достижении последнего элемента из множества ответов значения расстояния перестают меняться, и для всех остальных запросов при заданной функции расстояния максимально релевантным считается последний по счету ответ.

Для сравнения полностью аналогичный эксперимент был проведен для иного выбора базиса. Если взять в качестве базиса единственную точку с номером 1, т.е. формировать одномерное пространство  $Y$ , то результаты будут выглядеть следующим образом (табл. 2).

Таблица 2

Нахождение максимально релевантных точек в  $Y$  для запросов из  $X$ .

#X	10	122	290	402	654	1164
#Y1	149	6	654	818	290	654
#Y2	150	7	653	817	291	655
#Y3	151	5	655	819	289	653
#Y4	152	8	652	822	292	656
d(x,y1)	1435	96	0	264	0	0
d(x,y2)	1568	281	581	421	873	1165
d(x,y3)	1703	475	583	951	875	1167
d(x,y4)	1840	656	1160	3024	1744	2326

Некоторые ответы при таком примитивном выборе базиса находятся хуже, однако все опорные решения, для которых значение функции расстояния равно нулю оказываются определенными верно. На этом примере было проведено несколько подобных вычислительных экспериментов. В экспериментах изучалось влияние выбора базиса и его размерности на полноту найденных решений. Обнаружилось, что случайный выбор элементов из  $X$  всегда дает лучшие результаты, чем регулярный. Для данного примера существенное увеличение размерности слабо влияет на результат, что свидетельствует о низкой внутренней размерности задачи (близкой к единице).

## Выводы

В работе представлен весьма эффективный метод поиска максимально релевантных элементов для произвольных метрических отношений между парой конечных множеств, что представляет актуальную задачу для многих приложений, связанных, например, с рекомендациями и рекламным таргетированием. Основной идеей метода является выделение в множестве запросов некоторого базисного подмножества и наделение множества ответов координатной системой, которая строится путем ассоциации расстояний релевантности до каждого из элементов базиса с проекциями точек из множества ответов.

Далее полученное координатное пространство рассматривается как евклидово и на нем строится граф метризованного тесного мира. С помощью этого графа удастся сократить вычислительную сложность поиска релевантного элемента в  $Y$  в пределе до логарифмической (при большом размере  $Y$ ).

Проведенные численные эксперименты подтвердили эффективность предложенного в статье эвристического метода. С точки зрения теории дискретной оптимизации в работе удалось построить эффективный метод оснащения множества поиска топологией (системой окрестностей для каждой точки) для решения задачи нахождения элемента, минимизирующего заданную двухместную функцию. В качестве направления будущей работы представляет интерес получить более строгое обоснование для выбора базисного подмножества в  $X$ , используя алгебраическую структуру матрицы расстояний релевантности, поскольку ее столбцы являются координатами каждого из элементов  $Y$ .

## Библиографический список

1. **Передерий, В.И.** Математические модели и алгоритмы принятия релевантных решений / В.И. Передерий, А.П. Еременко. 2008. №2(22).
2. **Ковалев, М.М.** Дискретная оптимизация. Целочисленное программирование / М.М. Ковалев. – М.: Едиториал УРСС, 2003. – 192 с.
3. Пат. на изобретение RU 2461879 - Способ доставки целевой рекламы и/или информации абоненту посредством инфокоммуникативных сетей и система для его осуществления / Пономарев, Д.М., Крылов В.В., Бритвина Е.В.



4. **Pavel Zezula**, Similarity Search: The Metric Space Approach, 2005, New York, NY 10013,USA: Springer. – 220 pp.
5. **Savage, J.E.**, Models of Computation. Exploring the Power of Computing Addison-Pub., 1998. – 672 pp.
6. **Krylov, V.V.**, Logvinov A.A., Ponomarenko A.A., Ponomarev D.M. „Single-attribute Distributed Metrized Small World Data Structure, IEEE, 2009 Eighth IEEE/ACIS International Conference on Computer and Information Science / 1-3 june 2009/ Shanghai, China.
7. **Krylov, V.V.**, Logvinov A.A., Ponomarenko A.A. Malkov Y.A., Metrized Small World Approach for Nearest Neighbor Search SYRCoSE, 12-13, Ekaterinburg, May, 2011.
8. **Krylov, V.V.**, Logvinov A.A., Ponomarenko A.A. Malkov, Scalable Distributed Algorithm for Approximate Nearest Neighbor Search Problem in High Dimensional General Metric Spaces, Similarity Search and Applications, 5th International Conference, SISAP-2012, Toronto, ON, Canada, August, 2012, Proceedings.
9. **Y. Hu**, Y. Wang, D. Li, S. Havlin, and Z. Di, Physical Review Letters, Possible Origin of Efficient Navigation in Small Worlds, 106, 108701 (2011).

*Дата поступления  
в редакцию 06.052013*

**E.V. Britvina<sup>1</sup>, V.V. Krylov<sup>1</sup>, U.A. Malkov<sup>2</sup>**

## **RELEVANCE MAXIMIZING ALGORITHM BASED ON GRAPH DATA MODEL**

Nizhny Novgorod state technical university n.a. R.E. Alexeev<sup>1</sup>,  
Institute of Applied Physics of the Russian Academy of Sciences<sup>2</sup>

The article describes an algorithm for finding the most relevant element in a finite set of answers. The search is performed on a finite set of queries and with a given relevancy function which has significantly lower computational complexity than exhaustive search. To achieve the performance goal we created a relevance function induced coordinate system on the set of answers and, subsequently, created a metrized small world graph on its basis. The computer simulations we performed confirmed the validity of the proposed heuristic algorithm. The planned application of this method is the creation of a targeted advertisement system for telecommunication systems with connection path establishment.

*Key words:* graph data model based, relevance maximizing algorithm.