

9. Erten C., Harding P.J., Kobourov S.G., Wampler K., Yee G. Exploring the computing literature using temporal graph. <http://tgrip.cs.arizona.edu>.
10. Dittmann F., Bobda C. Temporal graph placement on mesh-based coarse grain reconfigurable systems using the spectral method // *From Specification to Embedded Systems Application*, vol.184, Springer, 2005. – P. 301-310.

**Берштейн Леонид Самойлович**

Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге  
E-mail: lsb@tti.sfedu.ru.  
347928, Таганрог, пер. Некрасовский, 44.  
Тел.: 88634371695.

**Боженюк Александр Витальевич**

E-mail: avb002@yandex.ru.

**Bershtein Leonid Samoilovich**

Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education «Southern Federal University».  
E-mail: lsb@tti.sfedu.ru.  
44, Nekrasovskiy, Taganrog, 347928, Russia.  
Phone: +78634371695.

**Bozhenyuk Alexander Vitalievich**

E-mail: avb002@yandex.ru.

УДК 681.3.06

**С.Л. Беляков, И.Н. Розенберг**

**КОМБИНИРОВАННАЯ АНАЛОГИЯ ПРИ ПОИСКЕ  
КАРТОГРАФИЧЕСКИХ МАТЕРИАЛОВ**

*Рассматривается задача поиска картографических материалов в интеллектуальных геоинформационных системах, использующих механизм аналогии. Предлагается использовать комбинированную аналогию как способ представления знаний о поиске. Приводится алгоритм нахождения контекстов комбинированной аналогии.*

*Геоинформационные системы; цифровые карты; компьютерные сети.*

**S.L. Belyakov, I.N. Rozenber**

**COMBINED ANALOGY WHEN LOOKING FOR CARTOGRAPHIC  
MATERIALS**

*In work the problem of search of maps and charts in the intellectual geoinformation systems using the mechanism of analogy is considered. It is offered to use the combined analogy as a way of representation of knowledge of search. The algorithm of a finding of contexts of the combined analogy is resulted.*

*Geographic information system; digital maps; computer network.*

Решение прикладных задач с помощью интеллектуальной геоинформационной системы предполагает использование процедуры картографического анализа. Одним из начальных его этапов является поиск и отбор необходимых картографических ресурсов. В настоящее время картографические ресурсы доступны из источников, размещённых в сети Интернет. Использование средств поиска по ключевым словам либо каталогам поисковых серверов имеет определённые недостат-

ки. Аналитик вынужден проходить все этапы поискового процесса – отбор ключевых слов, экспериментирование с различными поисковыми системами, просмотр ранее использовавшихся ссылок, каталогов. Если же накоплен опыт поиска в виде уже апробированных источников картографической информации, их повторное использование имеет несомненные преимущества. Особый интерес представляет накопление коллективного опыта использования картографических ресурсов в рамках сетевых сообществ и социальных сетей. Поиск картографических материалов в перечнях ранее найденных и использованных ресурсов членами сетевого сообщества делает актуальным развитие средств поиска по аналогии.

Решение заданной прикладной задачи предполагает поиск по аналогии в множестве  $E = \{e_1, e_2, \dots, e_N\}$  накопленных сообществом картографических ресурсов, хранящихся в персональной поисковой системе сообщества и доступных через поисковый сервис. Укажем на достоинства использования поискового сервиса на основе аналогии:

- ◆ если рассматривать поиск как один из инструментов генерации и анализа гипотез при решении прикладных задач, механизм рассматриваемого режима поиска приобретает особое значение. Аналогию относят к одному из способов правдоподобных рассуждений, применимых к порождению гипотез [1,2]. Вероятность получить полезную информацию, в некотором смысле аналогичную отсутствующей, выше при поиске в ограниченном смысловом пространстве данных поискового сервиса. Чем шире тематическая направленность информационных ресурсов, тем шире спектр аналогий, однако их разумность и полезность снижаются;
- ◆ анализ собственного опыта ( в данном случае коллективного опыта сетевого сообщества ) удачных прецедентов поиска как для человека, так и для интеллектуальной системы является логичным начальным шагом решения проблемы [3].

Моделирование рассуждений, использующих механизм аналогии, издавна применяется для реализации интеллектуальных систем [2-6]. Несмотря на свою логическую необоснованность [1], аналогия отражает «логику здравого смысла», используемую в практике решения прикладных задач. Формально аналогия строится на оценке сходства одних частей объектов и признании истинными заключений об объектах в целом или сходстве других частей. В данной работе предлагается следующая модель информационного пространства для поиска аналогий

$$O = \langle C, R, X, F, D \rangle,$$

где  $C$  – множество описаний классов картографических ресурсов,  $R$  – множество отношений между классами и между экземплярами картографических ресурсов,  $X$  – множество статических атрибутов (свойств) информационных ресурсов,  $F$  – множество функций, определяющих поведение картографических ресурсов,  $D$  – множество запросов на поиск картографических ресурсов. Если в контексте решения прикладной задачи установлена функция  $s(a, b)$  оценки сходства пары подмножеств  $a, b$  по атрибутам из  $\langle C, R, X, F, D \rangle$ , то сходство любой пары экземпляров  $o_1 \in O, o_2 \in O$  логически выводится как

$$s(a, b) \geq s^* \Rightarrow s(o_1, o_2) \geq s^*,$$

где  $s^*$  – требуемый уровень сходства.

Выбор базы аналогии – пары  $(a, b)$ , атрибутов из  $\langle C, R, X, F \rangle$  и функции  $s(a, b)$  – определяет содержательное многообразие рассуждений. Например,

- 1)  $a \in C, b \in C$  предполагает аналогию «природы» объектов: классы растровых и векторных изображений описывают аналогичные объекты в смысле визуального представления;
- 2)  $a \in X, b \in X$  предполагает аналогию свойств: частота употребления в атрибутивных данных ключевых слов позволяет утверждать об их аналогичности по содержанию;
- 3)  $a \in F, b \in F$  означает аналогию поведения: если видеоматериалы во время воспроизведения позволяют включать или отключать собственные субтитры, можно говорить об их аналогичности в смысле восприятия речи на иностранных языках.

Поиск реализуется отбором картографических ресурсов на основе гипотезы: если некоторый ресурс  $e_i \in E$  обладает информативностью  $I(e_i) > I^*$  и в контексте близости  $s(e_i, e^*) > s^*$  (здесь  $e^*$  – эталонный ресурс для сравнения), то по аналогии  $I(e_i) > I^*$  в этом контексте.

Для минимизации времени решения прикладной задачи должен быть построен механизм оценки аналогии (схожести) максимально содержательных информационных ресурсов.

Трудности решения данной проблемы состоят в следующем:

- ◆ понятие содержательности трудноформализуемо и понимается аналитиком субъективно. Фактически, следует экспериментальным образом сопоставить каждому пользователю сервиса информационную структуру, описывающую содержательность;
- ◆ отсутствием принципов определения контекста аналогии и эталонного ресурса  $e^*$  для оценки сходства. Как показал анализ известных методов [6], данный вопрос решается соответственно прикладной задаче;
- ◆ процедура рассуждения по аналогии описывается аналитиком неоднозначно, неопределённо и неполно. Это требует построения нечёткой модели вывода об аналогичности экземпляров ресурсов;
- ◆ параметры, по которым определяется сходство, также подвержены влиянию НЕ-факторов [7].

По перечисленным причинам непосредственное применение механизмов поиска по аналогии [4] не представляется возможным. Поэтому в данной работе предлагается новый метод комбинированной аналогии, являющийся развитием метода структурной аналогии [4,5].

В основе предлагаемого метода комбинированной аналогии лежит предположение о том, что аналогия между картографическими ресурсами имеет составную конструкцию, т.е. характеризуется вектором аналогий

$$A = (a_1, a_2, \dots, a_M).$$

В качестве  $a_i$  выступают независимые аналогии, целостно отражающие определённые аспекты вида и поведения объектов. Предлагаемое структурирование достаточно естественно для картографических ресурсов: рассматривая аналогию между парой  $e_i, e_j \in E$ , эксперт учитывает сходство по источнику ресурса, тематической

направленности, качеству представления данных, уровню достоверности, степени актуальности, привычности терминологии, ссылочной ценности, и т.д.

В терминах теории структурной аналогии [4], любая аналогия  $A$  имеет контекст  $U_A$ , и множество

$$r_{ab} = U_A \cap a \cap b$$

включает в себя элементы, которые используются для оценивания близости объектов  $a$  и  $b$  функцией  $s(a, b)$ . В качестве контекста выступает подмножество атрибутов

$$U_A = \langle C_A, R_A, X_A, F_A \rangle, C_A \subseteq C, R_A \subseteq R, X_A \subseteq X, F_A \subseteq F, D_A \subseteq D.$$

Для комбинированной аналогии

$$U_{a_i} = \langle C_{a_i}, R_{a_i}, X_{a_i}, F_{a_i}, D_{a_i} \rangle,$$

$$C_{a_i} \subseteq C, R_{a_i} \subseteq R, X_{a_i} \subseteq X, F_{a_i} \subseteq F, D_{a_i} \subseteq D,$$

$$C_{a_i} \cap C_{a_j} \neq \emptyset, R_{a_i} \cap R_{a_j} \neq \emptyset, X_{a_i} \cap X_{a_j} \neq \emptyset, F_{a_i} \cap F_{a_j} \neq \emptyset, D_{a_i} \cap D_{a_j} \neq \emptyset, i, j = \overline{1, |A|},$$

$$s(x, y) = (s_{a_1}(x, y), s_{a_2}(x, y), \dots, s_{a_M}(x, y)).$$

Приведённая модель позволяет формально описать содержательность как набор контекстов  $\hat{A} \subseteq A, \hat{A} \neq \emptyset$ . Набор  $\hat{A}$  специфичен для каждого пользователя системы, все пользователи имеют разные контексты  $\hat{A}$ . Величину  $|\hat{A}|$  можно использовать как меру содержательности: чем больше частных аналогий «просматривается» в найденном материале, тем больше вероятность восполнить недостаток информации. Пустой контекст соответствует бессодержательной аналогии, применение которой практически бесполезно.

Можно заключить, таким образом, что минимизация времени решения прикладной задачи сводится к построению максимальных по мощности контекстов пользователей системы. Существование решения в таком случае определяется тем, существует для конкретного пользователя ли набор контекстов  $\hat{A}$ .

Как показывает анализ, при нахождении набора контекстов  $\hat{A}$  возникают следующие задачи:

- ◆ задача отбора  $\hat{A}$  из множества возможных контекстов. Формально,  $\hat{A} \subseteq C \times R \times X \times F \times D$ , однако принцип отбора полезных элементов неясен;
- ◆ определения функции близости  $s(a, b)$ . Сходство на основе набора аналогий  $\hat{A}$  означает, что  $s(a, b)$  зависит от многих параметров и имеет

различный вид как для отдельных пользователей, так и в разных областях пространства аналогий. Получить адекватное аналитическое выражение для  $s(a, b)$  не представляется возможным из-за недостатка данных.

Поэтому должен использоваться некий практически полезный подход.

Отбор полезных контекстов в набор  $\hat{A} \subseteq C \times R \times X \times F \times D$  предлагается строить на основе анализа онтологии. Как известно [8], описание онтологий ис-

пользует понятие классов. Язык OWL, предназначенный для создания Web-онтологий ресурсов, наделён синтаксическими конструкциями для определения классов и их отношений. Таким образом, можно без ограничения общности считать, что между классами существуют отношения, рассматриваемые в объектно-ориентированном анализе:

- ◆ наследование;
- ◆ агрегирование (композиция);
- ◆ ассоциирование.

С каждым отношением  $r \in R$  между классами может быть связана смысловая аналогия. Например, всё, что относится к автомобилям вообще, в определённой степени может относиться к грузовым либо легковым автомобилям; всё, что связано с легковыми автомобилями, в той или иной степени имеет отношение к конкретной марке авто, а также к ассоциирующимся с этими марками объектами: запчастями, услугами по ремонту и т.д. Таким образом, может быть построен нечёткий ориентированный граф [9] аналогий  $\tilde{G} = (X, \tilde{U})$ , в котором множество вершин  $X$  соответствует классам онтологии, а множество нечётких рёбер

$$\langle x_i, x_j \rangle \in X^2, \tilde{U} = \{\mu(\langle x_i, x_j \rangle) / \langle x_i, x_j \rangle\}, i, j = \overline{1, |X^2|}$$

соответствует степени аналогии между парой классов  $\langle x_i, x_j \rangle$ . Здесь

$\mu(\langle x_i, x_j \rangle)$  является степенью принадлежности ребра нечёткому множеству рёбер  $\tilde{U}$ . Граф  $\tilde{G} = (X, \tilde{U})$  строится на основе опроса экспертов-аналитиков сетевого сообщества и является формальным представлением глубинных знаний об осмысленных аналогиях прикладной области. В общем случае для построения графа  $\tilde{G} = (X, \tilde{U})$  по описанию на OWL используется набор экспертных правил  $r_{OWL} = \langle C_{OWL}, A_{OWL}, D_{OWL} \rangle$ , где  $C_{OWL}$  – условие правила,  $A_{OWL}$  – множество фактов, добавляемых правилом,  $D_{OWL}$  – множество фактов, удаляемых правилом.

Будем считать, что правила  $r_{OWL} = \langle C_{OWL}, A_{OWL}, D_{OWL} \rangle$  обеспечивают полноту описания аналогий в том смысле, что любому концепту  $c_i \in C, i = \overline{1, |C|}$  всегда может быть сопоставлен класс  $x_j \in X, j = \overline{1, |X|}$ . Тогда справедливы следующие утверждения.

*Утверждение 1.* Каждому дескриптору ресурса  $\tilde{D}_k(E')$  соответствует подграф  $g_{\tilde{D}_k} \subseteq \tilde{G}$ .

*Утверждение 2.* Путь  $L(x_i, x_j)$  из вершины  $x_i$  в вершину  $x_j$  в графе аналогий  $\tilde{G} = (X, \tilde{U})$  соответствует аналогии между любой парой концептов, сопоставляемых классам  $x_i$  и  $x_j$ .

Справедливость утверждения вытекает из способа построения графа аналогий.

**Определение 1.** Весом пути  $L(x_i, x_j)$  из вершины  $x_i$  в вершину  $x_j$  является величина  $V_{L(x_i, x_j)} = \min_{\langle x_n, x_m \rangle \in L(x_i, x_j)} \mu(\langle x_n, x_m \rangle)$ .

**Определение 2.** Степенью аналогии класса  $x_j$  классу  $x_i$  является вес пути  $L(x_i, x_j)$ .

**Утверждение 3.** Контекст  $a_i \in A$  аналогичен контексту  $a_j \in A$ , если справедливо высказывание

$$(\forall c_m \in a_i \exists x_p \in \tilde{X} : c_m \rightarrow x_p) \& (\forall c_n \in a_j \exists x_q \in \tilde{X} : c_n \rightarrow x_q) \& (L(x_p, x_q) \neq \emptyset).$$

Ограничив вес пути некоторым числом  $\alpha > 0$ ,

$$L(x_i, x_j) > \alpha,$$

получим набор аналогичных контекстов уровня  $\alpha$ .

Приведённые рассуждения позволяют генерировать набор контекстов  $\hat{A}$  на основе дескриптора запрашиваемого ресурса  $\tilde{D}_K(E')$ . Контексты зависят только от внутренней структуры онтологии и не персонализированы, т.е. остаётся не известным вид функции близости  $s(a, b)$ .

В основе предлагаемой методики лежит гипотеза о стабильности комбинированной аналогии и постоянстве значимости частных аналогий, которыми пользуется аналитик. Это означает, что персональная функция близости имеет вид

$$\hat{s}(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q), \hat{a}_i \in \hat{A},$$

и обладает свойством

$$\frac{d\hat{s}(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q)}{d\hat{a}_i} = \text{const}, i = \overline{1, Q}.$$

Считая, что функция  $\hat{s}(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q)$  растёт с увеличением степени близости, можно заключить, что практический интерес представляют частные аналогии, для которых выполняется соотношение

$$\frac{d\hat{s}(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q)}{d\hat{a}_i} = L_{\hat{a}_i} > 0. \quad (1)$$

Значение  $L_{\hat{a}_i}$  следует рассматривать как чувствительность функции степени близости по аналогии  $\hat{a}_i$ . Содержательно это означает, что аналитик, проводя аналогии между информационными ресурсами в некотором контексте, использует частную аналогию  $\hat{a}_i$  как существенно важную. Например, при анализе последних достижений в некоторой области деятельности важно учитывать фактор аналогий времени. Когда речь идёт о достижениях отдельных научных школ, существ-

венно более важную роль играют аналогии тематики исследований, количественного и качественного состава группы исследователей. Аналогии размещения самих информационных ресурсов роли не играют.

Для экспериментального выявления набора контекстов смешанной аналогии заданного аналитика предлагается следующий алгоритм:

- 1) формируется начальный набор контекстов  $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q$  на основе имеющейся онтологии и имеющихся для аналитика дескрипторов ресурсов;
- 2) из накопленного аналитиком опытного материала о поиске конструируются тестовые пары  $e^T = \{e_1^T, e_2^T, \dots, e_K^T\}, e_i^T = \{e_m, e_n\}, e_m, e_n \in E, m \neq n$ ;
- 3) проводится тестирование аналитика, результатом которого является набор значений  $s_{e_m}^{\ominus}, m = \overline{1, K}$ ;
- 4) для каждого контекста из набора  $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q$  подсчитывается число значений  $n_{\hat{a}_i}$ , чувствительных к контексту, по правилу: если пара  $e_m$  признана аналитиком аналогичной и выполнено (1), значение  $n_{\hat{a}_i}$  увеличивается на 1;
- 5) для каждого контекста из набора  $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_Q$  вычисляется величина  $\frac{n_{\hat{a}_i}}{K}$ , характеризующая значимость контекста;
- 6) отбрасываются контексты со значимостью  $\frac{n_{\hat{a}_i}}{K} < \delta$ , где  $\delta$  – порог значимости, выбираемый из ограничения средней трудоёмкости ручного поиска  $U_\delta$  информационного ресурса вне поисковой системы, которая рассчитывается как

$$U_\delta = U_{\max} \left( 1 - \sum_{\hat{a}_i: \frac{n_{\hat{a}_i}}{K} > \delta} \frac{n_{\hat{a}_i}}{K} \right).$$

Здесь  $U_{\max}$  – максимально возможное значение трудоёмкости.

Таким образом, в данной работе описан подход к поиску картографических материалов на основе аналогии. Применение описанного подхода позволит повысить эффективность решения прикладных задач, связанных с использованием картографических данных.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ивин А.А.* Логика. Учебное пособие. – М.: Знание, 1998. – 232 с.
2. *Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В.* Достоверный и правдоподобный вывод в интеллектуальных системах. – М.: Физматлит, 2008. – 564 с.
3. *Поспелов Д.А.* Моделирование рассуждений. Опыт анализа мыслительных актов. – М.: Наука, 1989.
4. *Варшавский П.Р., Еремеев А.П.* Поиск решений на основе структурной аналогии для интеллектуальных систем поддержки принятия решений // Известия РАН. Теория и системы управления. – 2005. – № 1. – С. 97-109.
5. *Carbonell J.G.* Learning By Analogy: Formulating and Generalizing Plans from Past Experience // Machine Learning: An Artificial Intelligence Approach. V 1. Tioga; Palo Alto, 1983. – P. 137-161.
6. *Люгер Д.Ф.* Искусственный интеллект: стратегии и методы решения сложных проблем. Пер. с англ. – М.: Вильямс, 2003.

7. Нариньяни А.С. НЕ-факторы : неточность и недоопределённость - различие и взаимосвязь // Изв. РАН «Теория и системы управления». – 2000. – № 5. – С. 51-55.
8. Гладун А. Рогозина Ю. Онтологии в корпоративных системах // Корпоративные системы. – 2006. – № 1. – С. 28-32.
9. Берштейн Л.С., Боженик А.В. Нечёткие графы и гиперграфы. – М.: Научный мир, 2005.

**Беляков Станислав Леонидович**

Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: beliacov@yandex.ru.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371743.

**Розенберг Игорь Наумович**

ОАО «Научно-исследовательский и проектно-конструкторский институт инженеров железнодорожного транспорта» (НИИАС).

E-mail: I.kudreyko@gismps.ru.

109029, Москва, ул. Нижегородская, д. 27, стр. 1.

Тел.: 84959677701.

**Beliacov Stanislav Leonidovich**

Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education «Southern Federal University».

E-mail: beliacov@yandex.ru.

44, Nekrasovskiy, Taganrog, 347928, Russia.

Phone: +78634371743.

**Rozenberg Igor Naymovich**

Public corporation “Research and development institute of railway engineers”.

E-mail: I.kudreyko@gismps.ru.

27/1, Nizhegorodskaya, Moscow, 109029, Russia.

Phone: +74959677701.

УДК 681.03.06

**Д.А. Диденко**

**РАЗРАБОТКА МОДЕЛИ ОЦЕНКИ КАЧЕСТВА ИНФОРМАЦИИ В ГИС**

*Рассмотрена проблема оценки качества информационной базы ГИС в условиях нечеткости информации о требуемом качестве. Проведен анализ известных методик и моделей оценки качества, выявлены их недостатки. Предложена гибридная модель оценки качества информационной базы ГИС, основанная на нечеткой модели пользователя.*

*Качество информации; ГИС; оценка качества; гибридная модель; нечеткая модель; пользователь; модель ответа, интеллектуальная система.*

**D. A. Didenko**

**WORKING OUT OF MODEL OF THE ESTIMATION OF QUALITY OF THE INFORMATION IN GIS**

*The problem of an estimation of quality of information base GIS in the conditions of an illegibility of the information on demanded quality is considered. The analysis of known techniques and models of an estimation of quality is carried out, their lacks are revealed. The hybrid model of an estimation of quality of the information base GIS, based on indistinct model of the user is offered.*

*Quality of the information; GIS, a quality estimation; hybrid model; indistinct model; the user; answer model; intellectual system.*