

**Морозов Владимир Петрович**

Morozov Vladimir Petrovitch

Воронежский государственный архитектурно-строительный университет

Voronezh state architectural and construction university

Профессор кафедры управления строительством

Professor of chair of management of construction

Кандидат технических наук

E-Mail: vp\_morozov@mail.ru

05.13.10 (Управление в социальных и экономических системах);

05.13.01 (Системный анализ, управление и обработка информации);

05.13.18 (Математическое моделирование, численные методы и комплексы программ)

## **Математическая модель подсистемы весовой обработки информации на заданном терминологическом портрете**

Mathematical model of a subsystem of weight information processing on the set terminological portrait

**Аннотация:** В статье представлена математическая модель подсистемы весовой обработки информации на заданном терминологическом портрете. Предложен один из возможных алгоритмов реализации механизма обработки терминологического портрета.

**The Abstract:** The mathematical model of a subsystem of weight information processing is presented in article on the set terminological portrait. One of possible algorithms of realization of the mechanism is offered processings of a terminological portrait.

**Ключевые слова:** информационная подсистема, математическая модель, механизм, терминологический портрет.

**Keywords:** information subsystem, mathematical model, mechanism, terminological portrait.

\*\*\*

В интересах реализации мониторинга внешней среды – одной из важных составляющих интегрированного менеджмента (ИМ) [1], должна быть разработана (спроектирована) соответствующая информационная система вообще, и ее составная часть – подсистема обработки информации, в частности. Этапу проектирования данной подсистемы должна предшествовать стадия теоретических исследований. В рамках данной статьи рассматривается один из важных аспектов теоретических исследований – разработка математической модели подсистемы обработки информации. Следует отметить, что автором уже разработана математическая модель весовой распределенной информационной системы на тезаурусе [2]. Отличительной особенностью данной статьи является дальнейшее развитие полученной математической модели применительно к новому предмету исследования – терминологическому портрету. Терминологический портрет отличается от тезауруса, весовой взаимовязкой терминов в нем и более высоким уровнем детализации рассматриваемой предметной области. Рассмотрим основные определения.

*Определение 1.* Терминологическим портретом будем называть конечное непустое множество  $\{Z\}$  терминов  $z$ , отвечающих следующим условиям:

1) имеется непустое подмножество  $\{Z_0\} \subset \{Z\}$ , называемое множеством терминов;

2) имеется симметричное, транзитивное рефлексивное отношение  $R \subset Z \times Z$ , такое, что:

$$a) z_1 \neq z_2 \wedge z_1 R z_2 \Rightarrow (z_1 \in Z \setminus Z_0) \vee (z_2 \in Z \setminus Z_0);$$

$$b) z_1 \in Z \setminus Z_0 \Rightarrow (\exists z \in Z_0)(z R z_1);$$

при этом отношение  $R$  называется синонимическим отношением, а термины  $z_1$  и  $z_2$ , отвечающие этому отношению, называются синонимическими терминами;

3) имеется транзитивное и несимметричное отношение  $K \subset Z_0 \times Z_0$ , называемое обобщающим отношением;

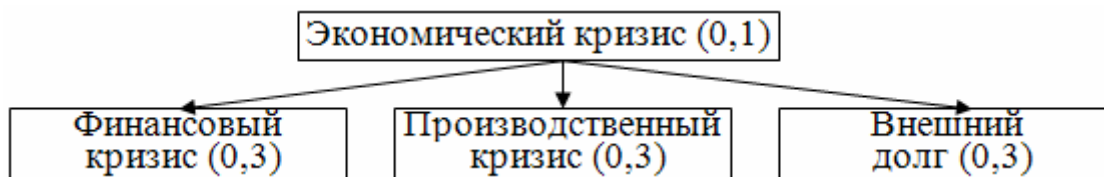
4) термины в терминологическом портрете имеют веса в зависимости от их степени значимости. Вес  $g_i$  определяет степень значимости термина  $z_i$ , т. е.:

$$Z(p) = \{\langle z_1, g_1 \rangle, \langle z_2, g_2 \rangle, \dots, \langle z_k, g_k \rangle\}.$$

При этом сумма всех весов ( $i: 1, \dots, N$ ) терминологического портрета равна 1, т.е.:

$$\sum_{i=1}^N g_i = 1.$$

На рисунке приведен пример графической интерпретации фрагмента терминологического портрета «Экономический кризис».



*Рис. Фрагмент терминологического портрета «Экономический кризис»*

Данный пример показывает, что в рамках рассматриваемого терминологического портрета, термины «Финансовый кризис», «Производственный кризис» и «Внешний долг» по степени значимости – равнозначны и имеют веса равные 0,3. Однако, каждый из них, по степени значимости, выше обобщающего термина «Экономический кризис», который имеет вес равный 0,1. Формальная запись данного терминологического портрета представляется следующим образом: Экономический кризис (0,1): финансовый кризис (0,3)  $\wedge$  производственный кризис (0,3)  $\wedge$  внешний долг (0,3).

*Определение 2.* Подсистемой весовой обработки информации на заданном терминологическом портрете будем называть четверку -  $(Z, D, M, \delta)$ , где  $Z$  – терминологический портрет  $Z_0 \subset T$ ;  $D$  – коллекция документов;  $M$  – множество запросов;  $\delta$  – отображение,  $\delta: M \times [0, 1] \rightarrow 2^{D \times [0, 1]}$  сопоставляет каждой паре (запрос, точность подобия) множество пар (документ, мера корреляции).

Остальные определения (2-5) и рассуждения, приведенные в [2] приемлемы для данного случая. Отличие заключается в механизме распределения запросов по подсистемам поиска. Применительно к тезаурусу [2], нагрузка на информационные подсистемы поиска равномерная. Применительно к терминологическому портрету, нагрузка на информационные подсистемы поиска избирательная, зависящая от весов терминов. В данном случае, ответ подсистеме

мы обработки информации  $Q$  на запрос  $\bar{m} = z_i \cdot g_i$  с точностью  $\rho = c$ , будет определяться в соответствии с выражением (1):

$$Q = \sum_{i=1}^n \delta_i(z_i \cdot g_i, c) = \sum_{i=1}^n \{(d, \alpha) : d \in D \wedge z_i \cdot g_i \nabla_c t(d) \wedge \alpha = \mu(z_i \cdot g_i, t(d))\} \subseteq D \times [0, 1], \quad (1)$$

где  $\nabla_c$  - отношение правдоподобия. При определении ответа  $N = \delta(\bar{m}, c)$ , отношение подобия  $\nabla_c$  осуществляет выбор документов, точность подобия которых не менее  $c$ . Мера корреляции  $\mu(\bar{m}, t(d))$  показывает, какая часть информации в документе соответствует ответу на вопрос. Если документ  $d \in D$  с мерой корреляции  $\alpha$  включен в ответ, т.е.  $(d, \alpha) \in \delta(\bar{m}, c)$ , то верно неравенство  $\alpha \geq c$ .

Один из возможных алгоритмов реализации механизма обработки терминологического портрета заключается в следующем. На начальном этапе информационные подсистемы отыскивают и обобщают информацию до определенного порогового значения  $\alpha \geq c$ , для термина с максимальным весом. Затем для более меньшего веса, и т. д, вплоть до термина с наименьшим весом.

Разработанная математическая модель подсистемы весовой обработки информации на заданном терминологическом портрете (1), позволяет составлять ответы на запросы более гибко с учетом неопределенности описания, как документов, так и запросов по сравнению с простой распределенной информационной подсистемой, а также перераспределять нагрузку на локальные информационные подсистемы в зависимости от значимости обслуживаемых терминов.

## ЛИТЕРАТУРА

1. Морозов, В.П. Методология и технология интегрированного менеджмента строительной организации. - Воронеж: Воронежский ГАСУ. - 2012. – 150 с.
2. Морозов, В.П. Математическая модель весовой распределенной информационной системы на тезаурусе // Системы управления и информационные технологии. 2012. №3 (49). – С. 54 – 56.

**Рецензент:** Кустов Андрей Игоревич, заведующий кафедры прикладной информатики и математики, кандидат физико-математических наук, доцент, автономная некоммерческая образовательная организация высшего профессионального образования «Воронежский экономико-правовой институт»