

ПРЕОБРАЗОВАНИЕ БОЛЬШОГО ТЕРМИНОЛОГИЧЕСКОГО ТЕКСТА В БИНАРНУЮ ОДНОРОДНУЮ ОРИЕНТИРОВАННУЮ СЕТЬ ТЕРМИНОВ

Павловский И.С.

Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная д.65
pavlovskiy@ipu.ru

Аннотация: В статье рассмотрена проблема целостного представления терминологического текста. В качестве первого шага на пути решения этой проблемы предложен метод интеграции определений понятий в бинарную однородную ориентированную сеть терминов. Реализация метода основана на использовании модели определения понятия как системы терминов. Приведен пример реализации метода. Сделан вывод о необходимости автоматизации построения бинарной однородной ориентированной сети для больших терминологических текстов.

Ключевые слова: терминологический текст, бинарная однородная ориентированная сеть терминов, интеграция определений понятий, целостность системы, иерархическая упорядоченность системы.

Введение

Развитие знаний в различных областях человеческой деятельности связано с изменением понятийного аппарата научных теорий, в результате чего возникает потребность в его систематизации.

Как правило, понятийный аппарат раскрывается в виде текстов, содержащих определения понятий и термины. Будем называть такие тексты терминологическими. Примерами терминологических текстов являются терминологические стандарты, энциклопедии, терминологические словари, тезаурусы и др.

Одним из основных конструктивных признаков текста, отражающих его содержательную сущность, является целостность [1]. В условиях изменения терминологии важно обеспечить целостность терминологического текста как основную системообразующую характеристику.

Суть целостности системы раскрывают следующие замечания [2]:

1. Свойства системы (целого) Q_s не является простой суммой свойств составляющих ее элементов (частей) q_{ij} :

$$Q_s \neq \{q_{ij}\}, \quad i = \overline{1, n}, \quad j = \overline{1, n_i}$$

2. Свойства системы (целого) зависят от свойств составляющих ее элементов (частей):

$$Q_s = f(\{q_{ij}\}).$$

3. Объединенные в систему элементы (части), как правило, утрачивают часть своих свойств, присущих им вне системы, т.е. система как бы подавляет ряд свойств элементов. Кроме того, элементы (части), попав в систему, могут приобрести новые свойства.

$$q_{ij} \neq q_{ij}^s.$$

Таким образом, целостность проявляется в системе в возникновении у нее новых интегративных свойств, которые отсутствуют у ее элементов. При этом целостность всей системы обеспечивается существованием однозначных связей между элементами, сближение которых делает их совершенно необходимыми друг для друга в интересах системы [3].

Здесь необходимо отметить тесную связь между закономерностями целостности и иерархичности, согласно которым утверждается первичность системы как целого над ее элементами и принципиальная иерархическая организация любой системы [4].

В силу закономерности иерархичности существует возможность установления строгого порядка подчиненности между элементами системы. Это означает, что системообразующими являются иерархически упорядоченные связи. Связи, направление которых противоположно иерархически упорядоченному общему направлению связей, приводят к противоречиям в системе. Такие противоречивые связи не позволяют системе приобретать новые свойства и, тем самым, негативно влияют на степень целостности системы.

Таким образом, целостность проявляется в системе в иерархических связях между элементами, в которых отсутствуют противоречия.

Подводя итог, отметим, что закономерность иерархичности может служить своеобразным индикатором противоречивых связей между элементами в системе, количество которых позволяет сделать вывод о степени целостности системы [5, 6].

Применительно к терминологическому тексту его целостную смысловую картину может дать иерархическое упорядочение семантической сети терминов, но в силу разнородности ее элементов и связей между элементами это практически невыполнимо: возможно построить только локальные иерархические представления, которые объединяют однотипные элементы и связи семантической сети.

Поэтому проблема целостного представления терминологического текста в интересах его систематизации заключается в построении единой иерархической структуры, которая наиболее полно соответствует многообразию отношений между терминами.

Сформулированная проблема относится к области терминологического анализа. Содержание направлений проводимых в этой области исследований свидетельствует о том, что она ранее не решалась. Такими направлениями исследований являются:

- информационный поиск [7-9];
- категоризации (классификация) и кластеризации текстов [10-13];
- разработка тезаурусов (контролируемой терминологии) [9, 15-24];
- оценка семантического сходства (подобия) текстов [11, 14, 25-28];
- анализ семантической согласованности текста [18, 29-31];
- реферирование текста [32];
- анализ мнения в тексте [33];
- анализ настроения в тексте [11, 33];
- сопоставление нормативных текстов и организационных процессов [26].

1 Подход к решению проблемы

Первым шагом на пути решения обозначенной проблемы является преобразование терминологического текста в семантическую сеть, удовлетворяющую требованию, которое заключается в возможности ее иерархического упорядочения, т.е. приведения к иерархическому виду. Такая сеть относится к классу однородных бинарных ориентированных сетей.

Ключевым вопросом преобразования терминологического текста в однородную бинарную ориентированную семантическую сеть терминов является поиск системообразующего элемента терминологического текста.

В качестве такого системообразующего элемента выбрано определение понятия (дефиниция). Именно определение понятия характеризуется логико-семантической структурой, тождеством определяемой части и определяющей части, иерархической организацией этих частей [34]. Путем интеграции определений понятий возможно получить иерархическую смысловую структуру терминологического текста, необходимую для его целостного представления.

2 Модель определения понятия как система терминов

В соответствии с положениями теории понятия [35] определение понятия можно представить в виде логически взаимосвязанной совокупности терминов: одного определяемого термина и одного или нескольких определяющих терминов. Таким образом, связи между терминами упорядочиваются, и они приобретают направленный характер: «определяющий термин» → «определяемый термин».

Исходя из этого, определение можно представить в следующем виде:

$$d_k = \langle t_{def}^{d_k}, \{t_j^{d_k}\} \mid t_{def}^{d_k} \neq t_j^{d_k} \rangle, \quad (1)$$

d_k - элемент множества определений понятий $D = \{d_k \mid k=1, m\}$;

$t_{def}^{d_k}, t_j^{d_k}$ - определяемый и определяющие термины определения понятия d_k .

Очевидно, определяемый термин находится с каждым определяющим термином в отношении типа «определять»:

$$r_{d_k} = \{(t_{def}^{d_k}, t_j^{d_k})\}, \quad (2)$$

r_{d_k} - множество отношений между определяемым и определяющими терминами в определении d_k .

Тогда модель определения понятия d_k как система может быть представлена так:

$$s_{d_k} = \langle t_k^{def}, \{t_j\}, r_{d_k} \rangle \quad (3)$$

или в упрощенном виде

$$s_{d_k} = \langle d_k, r_{d_k} \rangle. \quad (4)$$

Элементами системы s_{d_k} являются определяемые и определяющие термины, которые составляют суть определения d_k . При этом устойчивые (системообразующие) связи между элементами

определения-системы s_{dk} соответствуют парным отношениям определяемого термина и каждого определяющего термина.

Модель определения понятия s_{dk} по сути своей является однородной моделью, так как её элементами являются термины, которые выражают смысловое значение описываемых в определении понятий. При этом для всех связанных терминов реализуется только один тип связи – «определять», которая является однонаправленной. Кроме того, модели определений понятий являются бинарными структурами – рассматриваются только парные отношения между терминами.

Терминологический текст может быть представлен как множество определений понятий:

$$S = \langle s_{d_1}, \dots, s_{d_m} \rangle. \quad (5)$$

Интегрируя определения понятий, осуществляется синтез системы терминов V , которая представляет собой бинарную однородную ориентированную сеть терминов.

3 Метод интеграция определений понятий в бинарную однородную ориентированную сеть терминов

Интеграция моделей «определение-система» в единую модель основана на проявление отдельными терминами свойства двойственности, которое выражается в том, что один и тот же термин может входить в состав двух и более моделей «определение-система». Наличие свойства двойственности у терминов является обязательным условием для проведения интеграции моделей «определение-система» [36].

В случае выполнения указанного условия смысл интеграции моделей «определение-система» сводится к объединению множества терминов d_k из каждой модели «определение-система» в одно множество терминов T терминологического текста:

$$\bigcup_{k=1}^l d_k \rightarrow T = \{t_i\}, \quad i = \overline{1, n}, \quad (6)$$

n – общее количество терминов в терминологическом тексте.

В процессе объединения связей терминов r_{dk} из моделей «определение-система» получают матрицу взаимосвязей (смежности) терминов R предметной области:

$$\bigcup_{k=1}^l r_{dk} \rightarrow R = \|r_{ij}\|, \quad i, j = \overline{1, n}. \quad (7)$$

В результате осуществляется трансформация моделей «определение-система» в семантическую сеть терминологического текста:

$$S = \langle s_{d_1}, \dots, s_{d_m} \rangle \rightarrow V = \langle T, R \rangle. \quad (8)$$

Таким образом, интеграция моделей «определение-система» заключается в совместном выполнении двух процессов:

- интеграция терминов моделей «определение-система» d_k ;
- интеграция связей между этими терминами.

Созданная таким образом семантическая сеть терминов может быть реструктурирована в иерархическое представление терминологического текста.

4 Алгоритм интеграции определений понятий в бинарную однородную ориентированную сеть терминов

Перед началом выполнения операций интеграции моделей «определение-система» вводятся два множества:

множество терминов терминологического текста $T = \{t_i\}, i = \overline{1, n}$;

множество, $X = \{x_w\}, w = \overline{1, m}$, которое выражает соответствие терминов отдельных моделей «определение-система» s_{dk} и терминов множества T . Каждый элемент x_w определяется как парное отношение (x_{S_w}, x_{T_w}) между каждым термином из множества моделей «определение-система» S и каждым элементом множества терминов T . Количество элементов множества X соответствует общему количеству связей между определяющими и определяемыми терминами в моделях «определение-система»:

$$m = \sum_{k=1}^l n_{dk}, \quad (9)$$

n_{dk} – количество терминов в определении понятия d_k .

В исходном состоянии n и m принимают нулевое значение:

$$n = m = 0. \quad (10)$$

В ходе объединения терминов моделей «определение-система» из каждой модели s_{dk} последовательно выбираются один ее термин (это относится как к определяющим t_f^{dk} , так и определяемым t_{def}^{dk} терминам); осуществляется проверка совпадения этого термина (например, t_f^{dk}) с одним из терминов t_i из множества T .

Если совпадение терминов t_f^{dk} и t_i не выявлено, во множество T добавляют новый элемент t_{n+1} , значение которого приравнивается значению термина t_f^{dk} . Во множество X так же добавляется новый элемент x_{m+1} , которому соответствует парное отношение (t_f^{dk}, t_{n+1}) :

$$T + t_{n+1}, t_f^{dk} \rightarrow t_{n+1}; X + x_{m+1}, (t_f^{dk}, t_{n+1}) \rightarrow x_{m+1}, \text{ если } t_f^{dk} \neq t_i. \quad (11)$$

В том случае, если термины t_f^{dk} и t_i совпадают, добавляется новый элемент только во множество X . При этом новому элементу x_{m+1} соответствует пара совпадающих элементов (t_f^{dk}, t_i) :

$$X + x_{m+1}, (t_f^{dk}, t_i) \rightarrow x_{m+1}, \text{ если } t_f^{dk} = t_i. \quad (12)$$

После выбора и проверки терминов из всех моделей «определение-система» формируется окончательно множество терминов T .

После завершения формирования множества T задается матрица смежности терминов R терминологического текста. Перед объединением связей между терминами моделей «определение-система», элементам матрицы R присваиваются нулевые значения:

$$\|r_{ij}\| = 0, \quad i, j = \overline{1, n}. \quad (13)$$

В ходе объединения связей между терминами моделей «определение-система» из каждого множества r^{dk} последовательно выбирается одна из связей (t_f^{dk}, t_{def}^{dk}) . Для определяющего термина t_f^{dk} во множестве X определяется соответствующий термин t_i . Тем же образом для определяемого термина t_{def}^{dk} определяется термин t_j . Далее элементу матрицы смежности R , индекс строки которого совпадает с индексом термина t_i , а индекс столбца – с индексом термина t_j , присваивается значение 1:

$$(t_f^{dk}, t_{def}^{dk}): t_f^{dk} \Rightarrow (x_w^T, x_w^S) \Rightarrow t_i; t_{def}^{dk} \Rightarrow (x_w^T, x_w^S) \Rightarrow t_j, \quad (14)$$

$$t_i, t_j \Rightarrow r_{ij} = 1. \quad (15)$$

В результате объединения связей между терминами моделей «определение-система» получают матрицу смежности терминов R , «ненулевые» элементы которой ($r_{ij}=1$) свидетельствуют о наличии отношения «определяет» между термином t_i и термином t_j .

После завершения формирования множества терминов T и установления связей между этими терминами в матрице смежности R построение однородной бинарной ориентированной семантической сети завершается.

5 Пример интеграции моделей «определение-система»

Для иллюстрации содержания приведенного выше алгоритма рассмотрим интеграцию двух моделей «определение-система» s_{d1} и s_{d2} (рис. 1).

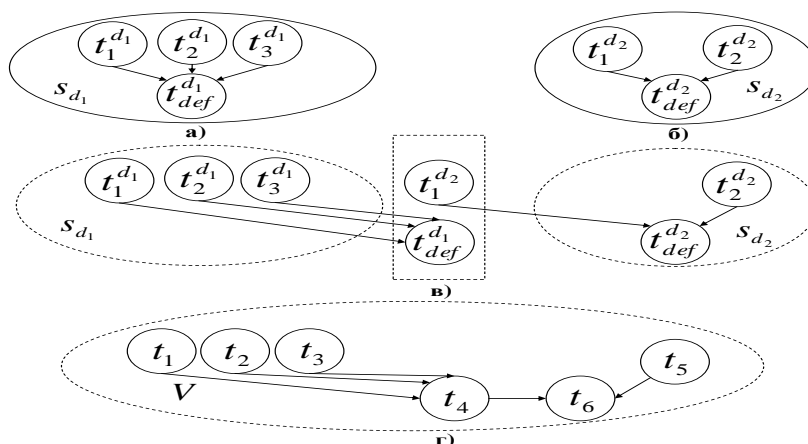


Рис. 1. Пример интеграции двух моделей «определение-система»:

Допустим, что в этих моделях термины t_{def}^{d1} и t_1^{d2} имеют равное значение (смысл), т.е. проявляют свойство двойственности: термин t_{def}^{d1} выступает в модели s_{d1} в роли определяемого термина (рис. 1, а); термин t_1^{d2} в модели s_{d2} играет роль определяющего термина (рис. 1, б).

В ходе интеграции (рис. 1, в) дублирование значений терминов t_{def}^{d1} и t_1^{d2} устраняется путем введения нового термина - t_4 . Кроме того, термин t_4 наследует связи терминов t_{def}^{d1} и t_1^{d2} . Другим терминам моделей s_{d1} и s_{d2} так же присваиваются новые имена (таблица 1).

Таблица 1. Соответствие терминов моделей «определение-система» и терминов семантической сети

Множество терминов модели «определение-система»	Элемент множества терминов модели «определение-система»	Элемент семантической сети V	Примечание
d_1	t_1^{d1}	t_1	
	t_2^{d1}	t_2	
	t_3^{d1}	t_3	
	t_{def}^{d1}	t_4	Двойственность терминов
d_2	t_1^{d2}	t_4	
	t_2^{d2}	t_5	
	t_{def}^{d2}	t_6	

- а) структура модели «определение-система» s_{d1} ;
- б) структура модели «определение система» s_{d2} ;
- в) процесс интеграции моделей s_{d1} и s_{d2} ;
- г) структура семантической сети V.

Результаты формирования матрицы смежности R для семантической сети приведены в таблице 2. Значения указанных в таблице 2 элементов r_{ij} равны 1, остальных элементов – 0.

В итоге формируется однородная бинарная ориентированная семантическая сеть терминов V (рис. 1, г), которая задана:

множеством терминов $T = \{t_1, t_2, t_3, t_4, t_5, t_6\}$;

матрицей смежности

$$R = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Таблица 2. Формирование элементов матрицы смежности семантической сети

Множество связей модели «определение-система»	Элемент множества связей модели «определение-система»	Элемент матрицы смежности
r_{d1}	(t_{def}^{d1}, t_1^{d1})	r_{14}
	(t_{def}^{d1}, t_2^{d1})	r_{24}
	(t_{def}^{d1}, t_3^{d1})	r_{34}
r_{d2}	(t_{def}^{d2}, t_1^{d2})	r_{46}
	(t_{def}^{d2}, t_2^{d2})	t_{56}

В дальнейшем сеть V преобразуется в иерархический вид. Такое представление используется для оценки целостности терминологического текста с целью выработки рекомендаций для систематизации его терминов.

6 Постановка задачи автоматизации

Критическим для реализации приведенного алгоритма является количество терминов в терминологическом тексте. По объему такие тексты могут включать сотни терминов. Часто термины относятся к разным отраслям знаний и требуют взаимодействия специалистов из разных научных и инженерных областей. Для анализа подобных текстов требуются существенные затраты ресурсов и времени. С учетом большого количества терминов и сложности связей между ними будем называть такие тексты большими терминологическими текстами (БТТ).

Экспериментальная проверка разработанного метода для БТТ объемом 297 терминов [37] выявила технологическую проблему, которая связана с подготовкой исходных данных для построения семантической сети. Данная проблема характеризуется значительными временными затратами,

появлением большого количества ошибок при выполнении рутинных операций построения моделей «определение-система», связанных с выявлением терминов и связей между терминами БТТ. Общее время этого этапа в ходе проведения эксперимента достигло 36 часов. Кроме времени, потраченного на выполнение собственно операций этапа, в результате многочисленных ошибок вынужденно тратилось дополнительно время на их устранение и повторное выполнение отдельных операций, что составило около 30% общего времени.

Выше изложенное делает актуальным решение задачи автоматизации преобразования терминологического текста в бинарную однородную ориентированную семантическую сеть терминов.

Заключение

В ходе проведенных исследований поставлена новая проблема целостного представления терминологического текста с целью систематизации его терминов.

Для решения поставленной задачи предложен подход, который основан на применении положений теории систем в части взаимоувязанных закономерностей целостности и иерархической упорядоченности.

В качестве системообразующего элемента терминологического текста, реализующего предложенный подход, выбрано определение понятия, которое рассматривается как система терминов и, вместе с тем, как элемент единой системы смыслового представления терминологического текста.

Разработанная на этой основе модель «определение-система» и метод интеграции таких моделей позволяют получить бинарную однородную ориентированную сеть терминов, которая может быть реструктуризирована в иерархическое представление терминологического текста. Такое представление терминологического текста является основой для оценки его целостности с целью выработки предложений для систематизации терминологии.

В ходе экспериментальной проверки разработанного методического аппарата возникла необходимость в автоматизации операций, связанных с извлечением элементов определений понятий из текста. В таком совершенствовании разработанного методического аппарата заключается дальнейшее направление исследований по рассматриваемой в статье тематике.

Литература

1. *Валгина Н.С.* Теория текста. М.: Логос, 2003. - 250 с.
2. Системный анализ и принятие решений: словарь-справочник / Под ред. *В.Н. Волковой, В.Н. Козлова.* — М.: Высш. шк., 2004 — 616 с: ил.
3. *Марков Ю.Г.* Функциональный подход в современном научном познании. Новосибирск: Наука, 1982. - 255 с.
4. *Садовский В.Н.* Основания общей теории систем. Логико-методологический анализ. – М.: Наука, 1974. – 280 с.
5. *Холл А.Д.* Опыт методологии для системотехники. – М.: Издательство «Советское радио», 1975. - 448 с.
6. *Денисов А.А.* Информационные основы управления. – Л.: Энергоатомиздат, 1983. – 72 с.
7. *Федорченко Л. А., Хайрова Н. Ф., Довнар А. И. и др.* Метод автоматизированного построения семантической сети терминов учебной дисциплины // *Радіоелектронні і комп'ютерні системи.* 2011. № 4. С. 115–120.
8. *J. Luo, B. Meng, C. Quan, X. Tu,* Exploiting salient semantic analysis for information retrieval, ENTERPRISE INFORMATION SYSTEMS, vol. 10, Dec 2016, iss. 9: Advanced Knowledge Discovery Techniques from Big Data and Cloud Computing, pp. 959-969.
9. *W. Lahbib, I. Bounhas, B. Elayeb,* Arabic-English Domain Terminology Extraction from Aligned Corpora, On the move to meaningful Internet system: OTM 2014 Conferences, Lecture Notes in Computer Science, vol. 8841, 2014, pp. 745-759.
10. *Новиков А.И.* Семантические расстояния в языке и тексте. - М.: Наука, 1990.
11. *Y. Gutierrez, S. Vazquez, A. Montoyo,* A semantic framework for textual data enrichment, Expert Systems with Applications, vol. 57, 15 September 2016, pp. 248–269.
12. *R. Palomares-Perraut, C. Gomez Camarero,* Strategies to improve the visibility and dissemination of the journal Fotocinema: creation of a multilingual controlled vocabulary, Scire-representacion y organizacion del conocimiento, vol. 22, iss. 2, Jul-Dec 2016, pp. 47-56.
13. *A. Vlachidis, D. Tudhope,* A knowledge-based approach to Information Extraction for semantic interoperability in the archaeology domain, Journal of the association for information science and technology, vol. 67, iss. 5, May 2016, pp. 1138-1152.
14. *Крюков К.В., Панкова Л.А., Пронина В.А. и др.* Меры семантической близости в онтологии // *Проблемы управления.* 2010. №5. С. 3-14.

15. Лукашевич Н.В., Четверкин И.И. Комбинирование тезаурусных и корпусных знаний для извлечения оценочных слов // Системы и средства информатики. 2015. Т. 25. Вып. 1. С.20-33.
16. Губанов Д. А., Макаренко А. В., Новигов Д. А. Методы анализа терминологической структуры предметной области (на примере методологии) // Управление большими системами. 2013. Вып. 43. С. 5-33.
17. C. Maia, P. Cesar; S.M. Vasconcellos, C. M. Teles, Terminology applied to scientific literature on environmental management: guidelines for a micro-thesaurus building, *Perspectivas em ciencia da informacao*, vol. 22, iss. 1, Jan-Mar 2017, pp. 80-99.
18. J. Misra, Terminological inconsistency analysis of natural language requirements, *Information and Software Technology*, vol. 74, November 2015, pp. 183–193.
19. J. Pathak, J. Wang, S. Kashyap, M. Basford, R. Li, D.R. Masys, C.G. Chute, Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience, *Journal of the American medical informatics association*, vol. 8, iss. 4, Jul 2011, pp. 376-386.
20. D.F. Podpolnyi; E.F. Skorokhodko, Interactive system of the network analysis of terminological vocabulary ASAT-2, “*Nauchno-tehnicheskaya informatsiya. Seriya 2. Informatsionnye protsessy i sistemy*”, iss. 2, 1988, pp. 2-5.
21. M.G. Mamedova, E.F. Skorokhodko, Automated-system of terminological dictionary analysis, “*Nauchno-tehnicheskaya informatsiya. Seriya 2. Informatsionnye protsessy i sistemy*”, iss. 1, 1981, pp. 14-18.
22. Добров Б.В., Лукашевич Н.В. Лингвистическая онтология по естественным наукам и технологиям для приложений в сфере информационного поиска // Ученые записки Казанского государственного университета. Серия «Физико-математические науки». 2007. Т. 149. Кн. 2. С.49-72.
23. Найханова Л. В. Анализ научного текста и формирование категорийно-понятийного аппарата в виде терминосистемы // Теоретические и прикладные вопросы современных информационных технологий : материалы Всероссийской научно-технической конференции. – Улан-Удэ, 2005. С. 130–139.
24. Langé J.-M. Discover and use real-world terminology with IBM Watson Content Analytics: Build sample domain dictionaries for data analysis, *IBM developerWorks*, 2014.
25. M. AL-Smadi, Z. Jaradat, M. AL-Ayyoub, Y. Jararweh. Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features, *Information Processing & Management*, vol. 53, iss. 3, May 2017, pp. 640–652.
26. K. Sapkota, A. Aldea, M. Younas, D.A. Duce, R. Banares-Alcantara, Automating the semantic mapping between regulatory guidelines and organizational processes, *Service Oriented Computing and Applications*, December 2016, vol. 10, iss. 4, pp 365–389.
27. J. Martinez-Gil, CoTO: A novel approach for fuzzy aggregation of semantic similarity measures, *Cognitive Systems Research*, vol. 40, December 2016, pp. 8–17.
28. R. Ferreira, R.D. Lins, S.J. Simske, F. Freitas, M. Riss, Assessing sentence similarity through lexical, syntactic and semantic analysis, *Computer Speech & Language*, vol 39, September 2016, pp. 1–28.
29. K. Zupanc, Z. Bosnić. Automated essay evaluation with semantic analysis, *Knowledge-Based Systems*, vol. 120, 15 March 2017, pp. 118–132.
30. J. Montero, H. Bustince, C. Franco, J.T. Rodríguez, D. Gómez, M. Pagola, J. Fernández, E. Barrenechea, Paired structures in knowledge representation, *Knowledge-Based Systems*, vol. 100, February 2016, pp. 50–58.
31. A. Lucas, L.M. Arruda Cafe, A.F. Godoy Viera, Business intelligence and competitive intelligence in brazilian information science: contributions to an analysis terminological, *Perspectivas em Ciencia da Informacao*, vol. 21, iss. 2, Apr-Jun 2016, pp. 168-187.
32. M. Gambhir, V. Gupta, Recent automatic text summarization techniques: a survey, *Artificial Intelligence Review*, January 2017, vol. 47, iss 1, pp 1–66.
33. R. Piryani, D. Madhav, V.K. Singh, Analytical mapping of opinion mining and sentiment analysis research during 2000-2015, *Information Processing & Management*, vol. 53, iss. 1, January 2017, pp. 122–150.
34. Зарва А.М. Дефиниция как типологическая разновидность научного текста / Автореферат. дис. ... канд. филол. наук. Нальчик, 2003.
35. Войшвилло Е.К. Понятие как форма мышления: логико-гносеологический анализ. Изд. 2-е. – М.: Издательство ЛКИ, 2007. – 240 с.
36. Павловский И.С. Автоматизация построения семантической сети на основе экспертной идентификации слов-терминов / Материалы 8-й Всероссийской мультikonференции по проблемам управления (МКПУ-2015, Ростов-на Дону). Ростов н/Д.: Издательство Южного федерального университета, 2015. Т. 1. С. 93-95.
37. Новигов А.М., Новигов Д.А. Методология: словарь системы основных понятий. – М.: Либроком, 2013. – 208 с.